Think: Cognitive Computing, Big Data, Cloud, Security and Privacy

Program Guide

Oct.29th - Oct.31st 2018

Sponsored By
IBM Advanced Studies
IBM Canada Software Lab
Table of Contents

Message From The Conference Chair 2
Message From The Program Co-Chairs 4
Keynote Speakers and Panelists 6
Conference Schedule
  Monday 10
  Tuesday 11
  Wednesday 12
CASCON Map 13
Full Paper Presentation 14
Schedule Full Papers 16
Position Paper Schedule Position Papers 24
CASCON Expo Map 26
  Monday 30
  Tuesday 32
  Wednesday 34
CASCON Expo 35
  Workshop Schedule At A Glance
    Monday 59
    Tuesday 60
    Wednesday 61
Workshop Info 63
Organizing Committee 80
About IBM Advanced Studies About CASCON 85 86
Message from the Conference Chair
CASCON 2018

Welcome to CASCON 2018!

For the past 28 years, IBM Centre for Advanced Studies (CAS has been hosting the Annual International Conference on Computer Science and Software Engineering. This is a testimony to our commitment to Academia and to the Applied Research in Canada. CASCON is well established yearly international meeting of minds that is unique in nature by being a purely academic conference sponsored by industry. The quality of papers, workshops, expo posters, and presentations showcased at CASCON is proof of the hard work of many academics and IBMers.

CASCON’s theme follows the latest trends in Computer Science and Software Engineering, and this year is Think: Cognitive Computing, Big Data, Cloud, Security and Privacy. We see these technological challenges being experienced in all the fields of modern society, with a tremendous impact on our lives. We are excited to have four thought-provoking keynote presentations, 23 technical and 10 position paper presentations, 29 workshops, and 81 exhibits. As with previous years, the CASCON proceedings, which include the technical papers, position papers, and detailed workshop abstracts, are also available online in the ACM Digital Library.

CASCON is an engaging 3 days conference offering a program of exciting sessions. Similar to previous years we have the regular and position paper tracks in the mornings, followed by keynotes around lunch time, workshops mid afternoons and Expo receptions to close off the days. Please plan to attend the Awards Ceremonies as well as the CAS picture happening right after lunch on the first day of the conference.

None of these would be possible without our dedicated community of Academics and IBMers. As conference chair, I am fortunate to be immersed in an exceptional team with professionals that go above and beyond their daily jobs and dedicate a considerable amount of time for the success of our conference. They are the role models I will always look up to and thank them all for their thought leadership, passion and commitment.

I would like to start by thanking the CASCON Steering Committee (Prof. Marin Litoiu, Prof. Kelly Lyons, Mr. Marcellus Mindel, Prof. Hausi Müller, Mrs. Tinny Ng, Mr. Joe Wigglesworth, Prof. Ken Wong and Prof. Ying Zou for all the planning and guidance that shaped this conference, and many more to come.

Special thanks go to our sponsor IBM Center for Advanced Studies and Mr. Marcellus Mindel, Head, IBM CAS for the continuous support, for always finding ways to give back to our Canadian computer science and software engineering community.

An academic conference is not possible without the dedication that the Program Chairs commit prior to the conference day and is always reflected in the quality of papers that are accepted. Mr. Wang Chen from IBM and particularly Professor Dorina Petriu from Carleton University have made a positive impact on the content this year, working tirelessly to orchestrate the paper.
submissions, revisions, and paper awards for our conference. Big thank you for the 71 Program Committee members who diligently peer-reviewed the papers and selected the top candidates; for the Best Paper Selection Committee and for the Most Influential Paper Award Committee. I would like to also thank Prof. Guy-Vincent Jourdan (Publication Chair) and Mr. Andrew Jaramillo (Conference Proceedings Editor) who were the force driving the proceedings and ensuring that all content was filtered, approved, and published in the ACM Library.

CASCON Workshops are always well received and well attended, many of which fill all their spots within a few days after the workshop registration is opened. Prof. Hausi Müller and Mrs. Tinny Ng acted this year as Workshop Chairs. I would like to thank them both for preparing a program consisting of an astonishing selection of top workshops. Additionally, I would also like to thank the Workshop Selection Committee members for reviewing the materials and making sure that the best workshops are accepted.

CASCON Technology Expo is the collaboration hub of the conference. With 81 exhibits and presentations, with new content daily and space designed to stimulate collaboration, it is one of the biggest orchestration challenges that was exceptionally handled by Prof. Kenneth Kent and Dr. Robert Enenkel, the Expo Chairs.

Special thanks to our IBM CAS Canada Team, Mr. Dennis Buttera, Mrs. Jennifer Collins, Mr. Bailey Duncan, Mr. Andrew Jaramillo, Mrs. Maria Gallaher, Ms. Clare Kim, Mr. Hanoor Manan, Mrs. Tinny Ng, and Mr. Temitayo Oyelowo for all the heavy lifting that goes behind the scenes and often is unnoticed but without which nothing is possible. Thank you to Mr. Rodney D’Silva and Mr. Alan Heighway for taking care of all the CASCON network related tasks.

I would like to thank all the CASCON volunteers and Prof. Marin Litoiu (Volunteer Chair) for all the help during the conference.

Finally, I would personally like to thank all the persons that submitted content to our conference and all our CAS Collaborators for promoting and contributing to this event, and last but not least, a big thank you to all CASCON participants for all the idea exchanges and brilliant discussions that happen during the conference.

I wish you all a wonderful and productive time at CASCON 2018!

Iosif-Viorel (Vio) Onut, Ph.D.,

Conference Chair | CASCON2018
Principal R&D Strategist | Centre for Advanced Studies | IBM Canada Lab
Adjunct Professor | University of Ottawa
Welcome to CASCON 2018, the 28th Annual International Conference on Computer Science and Software Engineering hosted by the IBM Advanced Studies (CAS)!

The theme of CASCON 2018 is Think: Cognitive Computing, Big Data, Cloud, Security and Privacy. This year we explore the research challenges as well as the economic and societal impacts of areas including cognitive computing, big data analytics, cloud computing, security and privacy through 4 thought-provoking keynote presentations, 23 technical and 10 position paper presentations, 29 workshops and 81 exhibits (posters, demos and short talks).

Our keynote presenters who will enlighten the audience on different topics are Steven Astorino, IBM, VP of Development, Hybrid Cloud, z Analytics and Canada Lab Director; Jessica Pointing, Quantum Computing Researcher at Stanford, Harvard, MIT; Walid Rjaibi, IBM, CTO, Data Security; and John Tsotsos, Distinguished Research Professor of Vision Science at York University and Director of the Centre for Innovation in Computing.

This year we received a total of 91 paper submissions, 68 full papers and 23 position papers from twenty different countries in North America, South America, Europe, South and East Asia and Africa. We accepted 23 full papers and 10 position papers. Acceptance rate for full papers was 33.8% and for position papers 43.5%. Each paper was rigorously reviewed by three members of the 71-member Program Committee and eight extra reviewers, resulting in a very high-quality program. The program is organized into the following sessions: Big data analytics, Cognitive computing, Model-Driven Engineering, Software Product Lines, Cloud Systems Mgmt, Machine Learning, Resource mgmt, Security and privacy, Adaptive systems, Compiler development, Microservices, Blockchain and Healthcare. As in previous years, the CASCON 2018 proceedings are archived for ease of access in the ACM Digital Library.

The Technology Expo provides an excellent opportunity to experience emerging research results and leading-edge products and developing product areas. For the third year in a row, CASCON 2018 features very short PechaKucha style presentations to showcase the posters and exhibits to the entire CASCON audience.

The 29 workshops at CASCON 2018 are wonderful forums for presenting, discussing, and debating issues, problems, ideas, technology gaps, works-in-progress, and gaining hands-on experience with new product directions.

One of the most gratifying aspects of the CASCON planning process is the selection of the Best Paper, Best Student Paper, and Most Influential Paper awards. This year, the
Best Paper Award goes to authors Kenny Wehling, David Wille, Christoph Seidl and Ina Schaefer for their paper, "Reducing Variability of Technically Related Software Systems in Large-Scale IT Landscapes." The Best Student Paper Award is given to Amine Barrak, Le An and Marc-André Laverdière for their paper, "Just-in-time Detection of Protection-Impacting Changes on WordPress and MediaWiki" co-authored with their supervisors Foutse Khomh and Ettore Merlo. The Most Influential Paper from CASCON 2008 is awarded to Giuliano Antoniol, Kamel Ayari, Massimiliano Di Penta, Foutse Khomh and Yann-Gaël Guéhéneuc for their paper, "Is it a Bug or an Enhancement? A Text-based Approach to Classify Change Requests".

We are very grateful to many people for their help and support in organizing CASCON 2018. First of all, we thank all the authors of technical papers, workshop proposals and technology expo submissions. We thank the hard-working members of the CASCON 2018 Program Committee for their dedication to excellence in completing the reviews and engaging in online discussion of the papers. We also recognize the workshops and technology showcase committees, as well as the awards committees, for their hard work. We thank the entire CASCON 2018 organizing team: Hausi Müller and Tinny Ng who coordinated the workshop selection and program, Kenneth Kent and Robert Enenkel who orchestrated the technology expo selection and program, Andrew Jaramillo and Guy-Vincent Jourdan who assembled the CASCON 2018 proceedings and Harnoor Manan who kept the CASCON 2018 website up-to-date.

We wish you a wonderful time at CASCON 2018 and hope you will find time to enjoy the opportunities for networking in the stimulating social events.

Dorina Petriu
Carleton University
CASCON 2018 Program Co-Chair

Wang Chen
IBM Canada Ltd.
CASCON 2018 Program Co-Chair
Keynote Speakers

Steven Astorino  
VP of Development, Hybrid Cloud, z Analytics and Canada Lab Director at IBM

Steven is a development executive with proven transformational leadership and expertise in leading large enterprise development teams and an experienced change agent with a proven track record to drive improvements and efficiency. Most recently he has been leading and driving the Machine Learning and Data Science strategy for IBM’s analytics group and has delivered key strategic solutions including IBM Machine Learning and Data Science Experience. Early in his career, Steven has spent several years working with network testing technologies for the telecom Industry and played a key role in providing VoIP testing solutions and has extensive experience with data technologies such as data movement and transformation from his days at DataMirror, to Database technologies such as DB2 LUW, DB2 for z/OS, IBM’s latest Data and Analytics platform enabling data science and machine learning. He has a Bachelor’s Degree in Computer Science with a strong business background.

Data Science for Everyone  
Monday, October 29th, 14:00 - 15:00

Although an empirical approach to data is as old as the scientific method, the data driven transformation of business and society has initiated a renaissance, with data science and analytics emerging as a key engine of wealth and prosperity for the 21st century. Will better tools and better training satisfy the exploding demand? Should data science take its place with physics, chemistry and biology as a basic science everyone needs to learn? What opportunities will be created, and how will we take advantage of them?

Jessica Pointing  
Quantum Computing Researcher at Stanford, Harvard, MIT

Jessica Pointing has conducted quantum computing research at Stanford, Harvard and MIT. She is currently a PhD student specializing in quantum computing at Stanford and has been awarded the Knight-Hennessy Fellowship. She completed her Bachelor’s degree in Physics and Computer Science, with high honors and as a member of the academic honors society Phi Beta Kappa, from Harvard University after spending her first two years of university studies at the Massachusetts Institute of Technology (MIT). She also founded the Harvard College Quantum Computing Association. Jessica was the winner (audience vote) for the Quantum Matters Science Communication Competition and has been invited to talk about quantum computing at conferences, including IBM’s conference in Las Vegas, the Chief Digital Officers conference in New York City and the Neo conference in Arizona. Jessica has interned as a software engineer at Google, management consultant at McKinsey & Company, investment
banker at Goldman Sachs, and strategist at Morgan Stanley. She is the founder and editor-in-chief of Optimize Guide, a blog about life optimization, and founded the Now Know Organization, which received funding from Google to teach technology to young people. She has been awarded the McKinsey Women's Impact Award, MIT Award for Distinguished Achievement in Leadership, and has been named a Google Anita Borg Scholar and John Harvard Scholar (top 5% of class). She is also a Microsoft Scholar, Palantir Scholar, Adobe Research Scholar, Morgan Stanley Scholar, Goldman Sachs Scholar, Society of Women Engineers Scholar and Society of Geophysicists Scholar.

The Quantum Computing Landscape
Tuesday, October 30th, 11:00 - 12:00

Quantum computers have the potential to advance many other fields, such as machine learning, medicine, and energy systems. We could solve problems on quantum computers that are not possible to solve on our current classical computers in our lifetime. What are the potential applications of quantum computing? How do quantum computers actually work? How do we build quantum computers? How much progress have we made? Jessica Pointing will unpack the topic of quantum computing with a fascinating primer.

Walid Rjaibi is Chief Technology Officer (CTO) for Data Security at IBM. As a Technical Executive and Distinguished Engineer, he drives the technical strategy and architecture for data security across products and cloud services (Guardium Database Activity Monitoring, Guardium Data Encryption, Security Key Lifecycle Manager, Guardium Multi-Cloud Data Encryption, and IBM Key Protect). Prior to his current role, he held several technical and management roles within IBM including Research Staff Member at the Zurich Research Lab, Security Architect for DB2 LUW, and Chief Security Architect for IBM Analytics. His Data Security work resulted in over 20 granted patents and several publications in leading scientific and academic conferences such as the international conference on Very Large Databases (VLDB), the International Conference on Data Engineering (ICDE), and the international conference on security and cryptography (SECRYPT). Walid is also a frequent speaker at industrial conferences such as IBM World of Watson, Interconnect, Security Congress, and the International DB2 User Group (IDUG). Specialties: Information Technology, Software Development, Data Security, Application Security, Cloud Security, Security Compliance, Secure Engineering Deployment, Database Technology, Leadership, Staff Management and Team Building, Client Partnering

Encryption, Key Management, and Quantum Computing
Tuesday, October 30th, 14:00 - 15:00

Encryption is a fundamental technology for data protection. While Transport Layer Security (TLS) is widely accepted
as the standard solution for encrypting data in transit, no single solution has achieved similar status for encrypting data at rest. We will review the most common strategies for encrypting data at rest and the use cases where each strategy is most appropriate. We will also discuss how exactly quantum computers will affect data encryption and how those concerns are being addressed.

John Tsotsos

Distinguished Research Professor of Vision Science at York University’s Lassonde School of Engineering, and Director of the Centre for Innovation in Computing at Lassonde

John Tsotsos is Distinguished Research Professor of Vision Science in the Dept. of Electrical Engineering and Computer Science and Director of the Centre for Innovation in Computing at Lassonde, at York University. He has Adjunct Professorships in Computer Science and in Ophthalmology and Vision Sciences at the University of Toronto. He received his doctorate in Computer Science from the University of Toronto developing the first computer system to interpret visual motion depicted in digital image sequences, with application to heart motion analysis. After a postdoctoral fellowship in Cardiology at Toronto General Hospital, he joined the University of Toronto on faculty in both Computer Science and in Medicine. In 1980 he founded the highly respected Computer Vision Group at the University of Toronto, which he led for 20 years. He moved to York University in 2000 as Director of the Centre for Vision Research. Under his directorship, the centre was ranked in the top six interdisciplinary vision research organizations in the world. He has been a Canadian Heart Foundation Research Scholar (1981-83), a Fellow of the Canadian Institute for Advanced Research (1985-95) and currently holds the Canada Research Chair in Computational Vision (2003-2024). He has received many awards and honours including several best paper awards, among them a 1987 inaugural Marr Prize citation, the 1997 CITO Innovation Award for Leadership in Product Development, the 2006 Canadian Image Processing and Pattern Recognition Society Award for Research Excellence and Service, the 1st President’s Research Excellence Award by York University on the occasion of the University’s 50th anniversary in 2009, the 2011 Geoffrey J. Burton Memorial Lectureship from the United Kingdom’s Applied Vision Association for significant contribution to vision science, has been an ACM Distinguished Speaker, and is an IEEE Fellow. He was elected as Fellow of the Royal Society of Canada in 2010, and was awarded its 2015 Sir John William Dawson Medal for sustained excellence in multidisciplinary research, the first computer scientist to be so honoured. He has co-founded or been a principal of several companies, and holds a number of patents and technology transfer licences. Visiting positions were held at: University of Hamburg, Germany; Polytechnical University of Crete, Greece; Center for Advanced Studies at IBM Canada; INRIA Sophia-Antipolis, France; and, the Massachusetts Institute of Technology, USA. His current research focuses on a comprehensive theory of visual attention in humans. A practical outlet for this theory forms a second focus, embodying
elements of the theory into the vision systems of mobile robots.

**It Only Took 60 Years to Solve Artificial Intelligence - That Wasn't so Hard, Was it?**

**Wednesday, October 31st, 11:00 - 12:00**

It has been 62 years since the founding Dartmouth AI workshop. In the 1960's, researchers at MIT believed that developing a computer program to simulate a significant part of human vision could be accomplished by employing several undergraduate students over a summer. They quickly found out they were mistaken. Fast forward to the present day. AI is everywhere, solving all manner of hard problems and triggering debates about its ethics and societal impact. In contrast, scientists studying human (and animal) vision and intelligence have made no corresponding claims for breakthroughs in their understanding of human intelligence. Although a great deal has been learned in the past 60 years to be sure, each major discovery seems to emphasize how much remains unknown. Understanding the brain and how intelligent behaviour is produced remains a major challenge. Noting that the current AI successes are often claimed to be due to mimicking brain processes, is there a disconnect here? This presentation will suggest that there has been some tacit, gradual moving of the goalposts taking place and that there is indeed a disconnect. An understanding of the brain and behavioural sciences reveals many directions for future AI research, and importantly, how far there is still to go. Dramatically, this will require the abandonment of some of the longest held computational standards. Examples from my lab will illustrate.

**Keynote Schedule**

**Monday, October 29th**
- 2:00 PM – 3:00 PM: Data Science for Everyone

**Tuesday, October 30th**
- 11:00 AM – 12:00 PM: The Quantum Computing Landscape
- 2:00 PM – 3:00 PM: Encryption, Key Management, and Quantum Computing

**Wednesday, October 31st**
- 11:00 AM – 12:00 PM: It Only Took 60 Years to Solve Artificial Intelligence - That Wasn't so Hard, Was it?
<table>
<thead>
<tr>
<th>Time</th>
<th>Event</th>
</tr>
</thead>
<tbody>
<tr>
<td>8:00</td>
<td>8:00 - 17:00 Registration</td>
</tr>
<tr>
<td>8:30</td>
<td>8:30 - 10:00 Regular Paper Track</td>
</tr>
<tr>
<td>10:00</td>
<td>Break</td>
</tr>
<tr>
<td>10:15</td>
<td>10:15 - 10:45 Position Papers</td>
</tr>
<tr>
<td>10:45</td>
<td>Nutrition Break</td>
</tr>
<tr>
<td>11:00</td>
<td>11:00 - 13:00 Expo</td>
</tr>
<tr>
<td>13:00</td>
<td>13:00 - 13:30 Opening Ceremony</td>
</tr>
<tr>
<td>13:30</td>
<td>13:30 - 13:45 CAS Pictures</td>
</tr>
<tr>
<td>13:45</td>
<td>13:45 - 14:00 CAS Awards</td>
</tr>
<tr>
<td>14:00</td>
<td>14:00 - 15:00 Keynote #1 Steven Astorino IBM</td>
</tr>
<tr>
<td>15:00</td>
<td>Nutrition Break</td>
</tr>
<tr>
<td>15:15</td>
<td>15:15 - 17:15 Workshops</td>
</tr>
<tr>
<td>17:15</td>
<td>Break</td>
</tr>
<tr>
<td>17:30</td>
<td>17:30 - 19:00 Expo Reception</td>
</tr>
<tr>
<td>19:00</td>
<td>Expo Closes at 19:00</td>
</tr>
</tbody>
</table>
# Schedule – Tuesday

<table>
<thead>
<tr>
<th>Time</th>
<th>8:00 - 11:00 Expo Setup</th>
<th>8:00 - 17:30 Registration</th>
</tr>
</thead>
<tbody>
<tr>
<td>8:00</td>
<td>8:30 - 10:45 Regular Paper Track</td>
<td>8:30 - 14:30 IBM-STEM 4 Girls Workshop</td>
</tr>
<tr>
<td>8:30</td>
<td>Break</td>
<td></td>
</tr>
<tr>
<td>10:00</td>
<td>10:15 - 10:45 Position Papers</td>
<td></td>
</tr>
<tr>
<td>10:45</td>
<td>Nutrition Break</td>
<td></td>
</tr>
<tr>
<td>11:00</td>
<td>11:00 - 12:00 Keynote #2 Jessica Pointing Harvard/MIT</td>
<td>12:00 - 13:30 Expo 12:00 - 13:00 Lunch</td>
</tr>
<tr>
<td>12:00</td>
<td>13:30 - 13:40 Paper Awards</td>
<td>13:30 - 14:00 Most Influential Paper (MIP) 13:45</td>
</tr>
<tr>
<td>13:30</td>
<td>14:00 - 15:00 Keynote #3 Walid Ihabib IBM</td>
<td></td>
</tr>
<tr>
<td>15:00</td>
<td>Nutrition Break</td>
<td></td>
</tr>
<tr>
<td>15:15</td>
<td>15:15 - 17:00 Workshops</td>
<td></td>
</tr>
<tr>
<td>17:15</td>
<td>Break</td>
<td></td>
</tr>
<tr>
<td>17:30</td>
<td>17:30 - 19:00 Expo Reception Posters / PechaKucha Talks / Short Talks</td>
<td>19:00 Expo Closes at 19:00</td>
</tr>
</tbody>
</table>
## Schedule – Wednesday

<table>
<thead>
<tr>
<th>Time</th>
<th>8:00 - 17:00 Registration</th>
<th>8:00 - 8:30 Expo Setup</th>
</tr>
</thead>
<tbody>
<tr>
<td>8:00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8:30</td>
<td>8:30 - 10:45 Expo</td>
<td>8:30 - 10:45 Workshops</td>
</tr>
<tr>
<td>10:00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10:15</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10:45</td>
<td>Nutrition Break</td>
<td></td>
</tr>
<tr>
<td>11:00</td>
<td>11:00 - 12:00 Keynote #4</td>
<td></td>
</tr>
<tr>
<td></td>
<td>John K. Tsotsos</td>
<td></td>
</tr>
<tr>
<td></td>
<td>York University</td>
<td></td>
</tr>
<tr>
<td>12:00</td>
<td>12:00 - 13:30 Expo</td>
<td>12:00 - 13:00 Lunch</td>
</tr>
<tr>
<td></td>
<td>Continued</td>
<td></td>
</tr>
<tr>
<td>13:30</td>
<td>13:30 - 13:40 Expo Awards</td>
<td>13:45 - 14:00 CASCON 2019 Announcement</td>
</tr>
<tr>
<td>14:00</td>
<td>Break</td>
<td></td>
</tr>
<tr>
<td>15:15</td>
<td>14:15 - 17:30 Workshops</td>
<td></td>
</tr>
<tr>
<td>15:30</td>
<td>Nutrition Break</td>
<td></td>
</tr>
<tr>
<td>15:45</td>
<td>15:45 - 17:30 Workshops</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Continued</td>
<td></td>
</tr>
<tr>
<td>17:30</td>
<td>Expo Closes at 17:30</td>
<td></td>
</tr>
</tbody>
</table>

**Wednesday, October 31**
Map of CASCON

Registration Desk Hours:

October 29, 8:00 AM – 5:00 PM
October 30, 8:00 AM – 5:00 PM
October 31, 8:00 AM – 8:30 AM

CASCON Expo Hours:

October 29, 11:00 AM – 1:00 PM
5:30 PM – 7:00 PM
October 30, 12:00 PM – 1:30 PM
5:30 PM – 7:00 PM
October 31, 8:30 AM – 10:45 AM
12:00 PM – 1:30 PM
<table>
<thead>
<tr>
<th>ROOM</th>
<th>FULL PAPERS</th>
<th>SESSION NAME</th>
</tr>
</thead>
</table>
| Evergreen    | (1) The Impact of Design and UML Modeling on Codebase Quality and Sustainability  
(2) UML-Driven Automated Software Deployment  
(3) Automating the Detection of Third-Party Java Library Migration At The Function Level | Model-Driven Engineering           |
| Holly-Butternut | (1) Assuring the runtime behavior of self-adaptive cyber-physical systems using feature modeling  
(2) Reducing Variability of Technically Related Software Systems in Large-Scale IT Landscapes | Software Product Lines            |
| Jasmine      | (1) A Case Study of Spark Resource Configuration and Management for Image Processing Applications  
(2) Detecting Communities in Social Networks Using Concept Interestingness  
(3) Feature engineering in Big Data for detection of information systems misuse | Big Data Analytics                |
| Primrose     | (1) A Competitive Platform for Continuous Programming Skill Enhancement  
(2) Graphics Programming in Elm Develops Math Knowledge & Social Cohesion  
(3) Natural Language Question Answering in the Financial Domain | Cognitive Computing               |
<table>
<thead>
<tr>
<th>ROOM</th>
<th>FULL PAPERS</th>
<th>SESSION NAME</th>
</tr>
</thead>
<tbody>
<tr>
<td>Evergreen</td>
<td>(1) Feasibility of Internal Object Pools for Reduced Memory Management</td>
<td>Resource Mgmt.</td>
</tr>
<tr>
<td></td>
<td>(2) Hardware/Software CoDesign for Mathematical Function Acceleration</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(3) Persistent Memory Storage of Cold Regions in the OpenJ9 Java Virtual Machine</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(2) Just-in-time Detection of Protection-Impacting Changes on WordPress and MediaWiki</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(3) Scalable Practical Byzantine Fault Tolerance with Short-Lived Signature Schemes</td>
<td></td>
</tr>
<tr>
<td>Jasmine</td>
<td>(1) Design and Implementation of Loss Mitigation in Spot Instances</td>
<td>Cloud Systems Mgmt.</td>
</tr>
<tr>
<td></td>
<td>(2) Evaluating Efficiency, Effectiveness and Satisfaction of AWS and Azure from the Perspective of Cloud Beginners</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(3) Node.js Scalability Investigation in the Cloud</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(2) Evaluating Music Mastering Quality Using Machine Learning</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(3) (Semi)Automatic Construction of Access-Controlled Web Data Services</td>
<td></td>
</tr>
</tbody>
</table>
The Impact of Design and UML Modeling on Codebase Quality and Sustainability
Omar Badreddin, Khandoker Rahad

The general consensus of researchers and practitioners is that upfront and continuous software design using modeling languages such as UML improve code quality and reliability particularly as the software evolves over time. Software designs and models help in managing the underlying code complexities which are crucial for sustainability. Recently, there has been increasing evidence suggesting broader adoption of modeling languages such as UML. However, our understanding of the impact of using such modeling and design languages remains limited. This paper reports on a study that aims to characterize this impact on code quality and sustainability. We identify a sample of open source software repositories with extensive use of designs and modeling and compare their code qualities with similar code-centric repositories. Our evaluation focuses on various code quality attributes such as code smells and technical debt. We also conduct code evolution analysis over five-year period and collect additional data from questionnaires and interviews with active repository contributors. This study finds that repositories with significant use of models and design activities are associated with reduced critical code smells but are also associated with increase in non-critical code smells. The study also finds that modeling and design activities are associated with significant reduction in measures of technical debt. Analyzing code evolution over five year period reveals that UML repositories start with significantly lower technical debt density measures but tend to decline over time.

UML-Driven Automated Software Deployment
Luis Rivera, Norha Villegas, Gabriel Tamura, Miguel Jiménez, Hausi Müller

Software companies face the challenge of ensuring customer satisfaction through the continuous delivery of functionalities and rapid response to quality issues. However, achieving frequent software delivery is not a trivial task. It requires agile and continuous design, development and deployment of existing and new software features. Over time, managing these systems becomes increasingly complex. This complexity stems, in part, from the deployment pipelines and the myriad possible configurations of the software components. Furthermore, software deployment is a time-consuming and error-prone process, which, even when automated, can lead to configuration errors and cost overruns. In this paper, we address deployment challenges that developers face during continuous delivery and DevOps. Our proposal consists of Urano, a mechanism for automating the deployment process, which uses UML, an interoperable and de facto modeling standard, as a means of specifying a software architecture and its associated deployment. Our approach is based on the model-driven architecture principles to generate executable deployment specifications from user-defined UML deployment diagrams. We extend this kind of diagrams by defining and applying a UML profile that captures the semantics and requirements of the installation, configuration, and update of software components. Thus, enabling more expressive deployment specifications and their automatic realization. To evaluate Urano, we conducted three case studies that demonstrate its potential to effectively automate software deployment processes in industry.

Automating the Detection of Third-Party Java Library Migration At The Function Level
Hussein Alrubaye, Mohamed Mkaouer

The process of migrating between different third-party libraries is very complex. Typically, developers need to find functions in the new library that are most adequate in replacing the functions of the retired library. This process is subjective and time-consuming as developers need to fully understand the documentation of both libraries’ Application Programming Interfaces, and find the right match between their functions if it exists. In this context, several studies rely on mining existing library migrations to provide developers with by-example approaches for similar scenarios. In this paper, we propose a mining approach that extracts all the manually-performed function replacements for a given library migration. Our approach combines the mined function-change patterns with function-related lexical similarity to accurately detect mappings between replacing/replaced functions. Using our enhanced mining process, we perform a comparative study between state-of-art approaches for detecting migration traces at the function level. Our findings have shown its efficiency in accurately detecting migration fragments and it has enhanced the accuracy of state-of-art approaches in finding correct functions changes. We finally provide the community with a dataset of migrations between popular Java libraries, and their corresponding code changes at the function level.
Assuring the runtime behavior of self-adaptive cyber-physical systems using feature modeling

Nayreet Islam, Akramul Azim

A self-adaptive cyber-physical system (SACPS) can adjust its behavior and configurations at runtime in response to varying requirements obtained from the system and the environment. With the increasing use of the SACPS in different application domains, such variations are becoming more common. Users today expect the SACPS to guarantee its functional and timing behavior even in adverse environmental situations. However, uncertainties in the SACPS environment impose challenges on assuring the runtime behavior during system design. Software product line engineering (SPLE) is considered as a useful technique for handling varying requirements. In this paper, we present an approach for assuring the runtime behavior of the SACPS by applying an SPLE technique such as feature modeling. By representing the feature-based model at design time, we characterize the possible adaptation requirements to reusable configurations. The proposed approach aims to model two dynamic variability dimensions: 1) environment variability that describes the conditions under which the SACPS must adapt, and 2) structural variability, that defines the resulting architectural configurations. To validate our approach, the experimental analysis is performed using two case studies: 1) a traffic monitoring SACPS and 2) an automotive SACPS. We demonstrate that the proposed feature-based modeling approach can be used to achieve adaptivity which allows the SACPS to assure functional (defining execution of the correct set of adaptive tasks) and non-functional (defining execution of SACPS in the expected mode) correctness at runtime. The experimental results show that the feature-based SACPS demonstrates significant improvement in terms of self-configuration time, self-adaptation time and scalability with less probability of failure in different environmental situations.

Reducing Variability of Technically Related Software Systems in Large-Scale IT Landscapes

Kenny, David Wille, Christoph, Ina Schaefer

The number of software systems in a company typically grows with the business requirements. Therefore, IT landscapes in large companies can consist of hundreds or thousands of different software systems. As the evolution of such large-scale landscapes is often uncoordinated, they commonly comprise different groups of related software systems using a common core technology (e.g., Java Web-Application) implemented by a variety of architectural components (e.g., different application servers or databases). This leads to increased costs and higher effort for maintaining and evolving these software systems and the entire IT landscape. To alleviate these problems, the variability of such technically related software systems has to be reduced. For this purpose, experts have to assess and evaluate restructuring potentials in order to take appropriate restructuring decisions. As a manual analysis requires high effort and is not feasible for large-scale IT landscapes, experts face a major challenge. To overcome this challenge, we introduce a novel approach to automatically support experts in taking reasonable restructuring decisions. By providing automated methods for assessing, evaluating and simulating restructuring potentials, experts are capable of reducing the variability of related software systems in large-scale IT landscapes. We show suitability of our approach by expert interviews and an industrial case study with architectures of real-world software systems.
Big Data Analytics — Jasmine
Session chair: Mark Chignell
A Case Study of Spark Resource Configuration and Management for Image Processing Applications

Dwight Makaroff, Derek Eager, Winfried Grassmann, Habib Sabiu, Owolabi Adekoya

The world population is expected to reach an estimated 9.8 billion by 2050, necessitating substantial increases in food production. Achieving such increases will require large-scale application of computer informatics within the agricultural sector. In particular, application of informatics to crop breeding has the potential to greatly enhance our ability to develop new varieties quickly and economically. Achieving this potential, however, will require capabilities for analyzing huge volumes of data acquired from various field-deployed image acquisition technologies. Although numerous frameworks for big data processing have been developed, there are relatively few published case studies that describe user experiences with these frameworks in particular application science domains. In this paper, we describe our efforts to apply Apache Spark to three applications of initial interest within the Plant Phenotyping and Imaging Research Centre (P2IRC) at the University of Saskatchewan. We find that default Spark parameter settings do not work well for these applications. We carry out extensive performance experiments to investigate the impact of alternative Spark parameter settings, both for applications run individually and in scenarios with multiple concurrently executing applications. We find that optimizing Spark parameter settings is challenging, but can yield substantial performance improvements, particularly with concurrent applications, provided that the dataset characteristics are considered. This is a first step towards insights regarding Spark parameter tuning on these classes of applications that may be more generally applicable to broader ranges of applications.

Detecting Communities in Social Networks Using Concept Interestingness

Mohamed-Hamza Ibrahim, Rokia Missaoui, Abir Messaoudi

One key challenge in Social Network Analysis is to design an efficient and accurate community detection procedure as a means to discover intrinsic structures and extract relevant information. In this paper, we introduce a novel strategy called (COIN), which exploits COncept INterestingness measures to detect communities based on the concept lattice construction of the network. Thus, unlike off-the-shelf community detection algorithms, COIN leverages relevant conceptual characteristics inherited from Formal Concept Analysis to discover substantial local structures. On the first stage of COIN, we extract the formal concepts that capture all the cliques and bridges in the social network. On the second stage, we use the stability index to remove noisy bridges between communities and then percolate (merge) relevant adjacent cliques. Our experiments on several real-world social networks show that COIN can quickly detect communities more accurately than existing prominent algorithms such as Edge betweenness, Fast greedy modularity, and Infomap.

Feature engineering in Big Data for detection of information systems misuse

Eduardo Lopez, Kamran Sartipi

The increasing availability of very large volumes of digital data (i.e. Big Data) enables many interesting research streams on a wide variety of phenomena. However, there has been a paucity of Big Data sets in the area of cybersecurity in information systems, as organizations are reluctant to share data that may provide too much unrestricted visibility into their operations. In this study, we explore the use of a real-life, anonymized, very large dataset containing user behavior as captured in log files including both regular usage as well as misuse, typifying the dynamics found in a situation with compromised user credentials. Through the experiment, we validate that the existence of a large user behavior dataset in itself does not necessarily guarantee that abnormal behaviors can be found. It is essential that researchers apply deep domain knowledge, critical thinking and practical focus to ensure the data can produce the knowledge required for the ultimate objective of detecting an insider’s threat. In this paper we develop, formulate and calculate the features that best represent user behavior in the underlying information systems, maintaining a parsimonious balance between complexity, resource demands and detection effectiveness. We test the use of a classification model that proves the usefulness and aplicability of the features extracted.
Cognitive Computing - Primrose
Session Chair: Ulrike Stege
A Competitive Platform for Continuous Programming Skill Enhancement
Jen-Hao Kuo, Tsung-Han Wu, Hong-Bao Ye, Hewijin Christine Jiau
Enhancing programming skills is the key factor to keep up with current ever-changing technologies in IT industry. Implementing strategies on game-based platform is a common way for programmers to enhance programming skills. However, runtime simulation and game metrics provided by current game-based strategy platforms are ineffective to motivate continuous programming skill enhancement. We propose ELOP, a competitive game-based strategy platform to motivate programmers. ELOP will automatically schedule competitions for programmers, keep competition history, record changes in performance and provide needed personal information for further enhancement. To evaluate the effectiveness of ELOP, we conduct several studies. The result shows that ELOP does motivate programmers in continuous programming practice and enhance their programming skills.

Graphics Programming in Elm Develops Math Knowledge & Social Cohesion
John Zhang, Anirudh Verma, Chinmay Sheth, Christopher Schankula, Stephanie Koehl, Andrew Kelly, Yumna Irfan, Christopher K. Anand
At McMaster University, we have developed a framework for teaching computer science, including curricula and tools (iPad apps: Image 2 Bits and ElmJr; an open-source library GraphicSVG; and a web-based development environment). ElmJr is a projectional editor for Elm, with knowledge of our graphics library. Using ElmJr, children transform programs through contextual menus. As a result, they never see syntax or type errors. Children as young as 10 years old, who have just started learning about syntax in English, can be productive programmers. We will explain how ElmJr is designed to make programming simple for beginners, and how strong typing in Elm and our graphics library is key to making the list of program transformations manageable. We will explain the design and findings of a study of children in 14 classes (grade 4 and 5) in the Hamilton-Wentworth District School Board, who received 12 hours of instruction over eight weeks in ElmJr with the aim of improving their mathematics knowledge. In parallel, another eight grade 6 to 8 classes received at least two hours of instruction in Elm using our web-IDE, culminating in a Wordathon. The Wordathon is designed to reconcile the power of social computing with the need to protect children’s privacy. The intermediate classes were challenged to create animations in Elm of assigned words identified by teachers as K-4 core reading vocabulary. Joining the intermediate students in this activity, two high school classes were taught how to create interactive applications in Elm and were challenged to create a reading game using the word animations created by the intermediate children. In all, four games incorporating 408 animations were created, and some of the intermediate students presented the games to primary grades in their schools. Unlike typical network effects, we get a multiplier effect: more word animations make game development more attractive, and more games make animations more attractive. Having a critical mass then attracts the attention of other educators, including in this case the team implementing the board-wide reading strategy. All of this can be accomplished without any identifying information leaving the classroom.

Natural Language Question Answering in the Financial Domain
John Boyer
This paper describes a natural language question answering system focused on answering financial domain questions using a daily updated corpus of financial reports. Financial entity types of interest included company stocks, country bonds, currencies, industries, commodities, and diversified assets. Financial questions of interest included explanatory and factual questions about entities as well as financial outlook for entities. An important architectural divergence emerged between the approach required for answering financial outlook questions versus the approach for answering other financial information questions. The financial domain focus also introduced additional challenges to open domain natural language processing that were addressed in the areas of document ingestion, question classification accuracy, question analysis techniques, speed of machine learning, answer ranking by linguistic confidence versus temporality, and system accuracy assessment.
Konstantin Nasartschuk, Kenneth Kent, Stephen MacKay, Aleksander Micic

Object pools is a widely used software engineering pattern used to reuse object instances without the need of repeated allocation and instantiation. While the benefits of using object pool structures are still present when used in a garbage collected environment, it adds a memory management component to the development process. The paper investigates the feasibility of introducing automatically created and maintained object pools for predefined classes. Automatic object pools are implemented and discussed using the GenCon GC and Balanced GC policies.

Christopher Anand, Lucas Dutton, Adele Olejarz, Robert Enenkel, Wolfram Kahl

Many important workloads depend on the efficient computation of elementary functions like square root and logarithm. Accurate computation of these functions is time-consuming, and hard for compilers to schedule, because of conditional execution. These problems are exacerbated by SIMD computation, which does not mix well with conditional execution. Previously, we have outlined how performance can be improved by encapsulating the conditional execution in new instructions. In this paper, we refine this approach to take into account testability, the ability for code to be pipelined, and exploitation of processors with a gather-load instruction. In particular, we look at the decomposition of the previously described instruction pairs into three instructions. The instructions can incorporate table lookups, or complement existing load instructions. The variant which complements existing load instructions is expected to perform as well as the other variants, and is easier to test and to pipeline. This paper presents gate-level details for the instructions required to calculate various logarithm functions, including the circuit depth, count and approximate width. In addition, we highlight the relative complexity of verifying these instructions relative to other known instructions, and outline our strategy for light-weight verification. Finally, we show that this strategy would be expected to produce a doubling of performance on a wide class of processors, using an IBM POWER processor as an example.

Scott Young, Michael Flawn, Kenneth Kent, Gerhard Dueck, Charlie Gracie

In this paper an optimization technique for object-oriented language runtimes with automatic memory management is investigated. The technique involves segregating objects into different memory areas, backed by different memory devices, on a per object basis. This technique is compared to operating system paging mechanisms with swap partitions. Two different schemes for determining which objects should be segregated into slower memory are tested and their results are discussed. It has been observed that each technique can be the most or least optimal choice depending on the application.
Security and Privacy – Holly-Butternut
Session Chair: Jeremy Bradbury

Empirical Vulnerability Analysis of Automated Smart Contracts
Security Testing on Blockchains
Reza Parizi, Ali Dehghantanha, Kim Kwang Raymond Choo, Amritraj Singh

The emerging blockchain technology supports decentralized computing paradigm shift and is a rapidly approaching phenomenon. While blockchain is thought primarily as the basis of Bitcoin, its application has grown far beyond cryptocurrencies due to the introduction of smart contracts. Smart contracts are self-enforcing pieces of software, which reside and run over a hosting blockchain. Using blockchain-based smart contracts for secure and transparent management to govern interactions (authentication, connection, and transaction) in Internet-enabled environments, mostly IoT, is a niche area of research and practice. However, writing trustworthy and safe smart contracts can be tremendously challenging because of the complicated semantics of underlying domain-specific languages and its testability. There have been high-profile incidents that indicate blockchain smart contracts could contain various code-security vulnerabilities, instigating financial harms. When it involves security of smart contracts, developers embracing the ability to write the contracts should be capable of testing their code, for diagnosing security vulnerabilities, before deploying them to the immutable environments on blockchains. However, there are only a handful of security testing tools for smart contracts. This implies that the existing research on automatic smart contracts security testing is not adequate and remains in a very stage of infancy. With a specific goal to more readily realize the application of blockchain smart contracts in security and privacy, we should first understand their vulnerabilities before widespread implementation. Accordingly, the goal of this paper is to carry out a far-reaching experimental assessment of current static smart contracts security testing tools, for the most widely used blockchain, the Ethereum and its domain-specific programming language, Solidity, to provide the first body of knowledge for creating more secure blockchain-based software.

Just-in-time Detection of Protection-Impacting Changes on WordPress and MediaWiki
Amine Barrak, Marc-André Laverdière, Foutse Khomh, Le An, Ettor Merlo

Access control mechanisms based on roles and privileges restrict the access of users to security sensitive resources in a multi-user software system. Unintentional privilege protection changes may occur during the evolution of a system, which may introduce security vulnerabilities; threatening user's confidential data, and causing other severe problems. In this paper, we use the Pattern Traversal Flow Analysis technique to identify definite protection differences in WordPress and MediaWiki systems. We analyse the evolution of privilege protections across 211 and 193 releases from respectively WordPress and MediaWiki, and observe that around 60% of commits affect privileges protections in both projects. We refer to these commits as protection-impacting change (PIC) commits. To help developers identify PIC commits just-in-time, we extract a series of metrics from commit logs and source code, and build statistical models. The evaluation of these models revealed that they can achieve a precision up to 73.8% and a recall up to 98.8% in WordPress and for MediaWiki, a precision up to 77.2% and recall up to 97.8%. Among the metrics examined, commit churn, bug fixing, author experiences and code complexity between two releases are the most important predictors in the models. We performed a qualitative analysis of false positives and false negatives and observe that PIC commits detectors should ignore documentation-only commits and process code changes without the comments.

Scalable Practical Byzantine Fault Tolerance with Short-Lived Signature Schemes
Xinxin Fan

The Practical Byzantine Fault Tolerance (PBFT) algorithm is a popular solution for establishing consensus in blockchain systems. The execution time of the PBFT consensus algorithm has an important effect on the blockchain throughput. Digital signatures are extensively used in PBFT to ensure the authenticity of messages during the different phases. Due to the round-based and broadcast natures of PBFT, nodes need to verify multiple signatures received from their peers, which incurs significant computational overhead and slows down the consensus process. To address this issue, we propose an efficient short-lived signature based PBFT variant, which utilizes short-length cryptographic keys to sign/verify messages in PBFT for a short period of time and blockchain-aided key distribution mechanisms to update those keys periodically. We also present efficient algorithms for accelerating the software implementation of the BLS threshold signature scheme. Our extensive experiments with three elliptic curves and two signature schemes demonstrate the efficacy of using short-lived signature schemes for improving the scalability of PBFT significantly.
Design and Implementation of Loss Mitigation in Spot Instances

Federico Sogaro, Eric Aubanel, Kenneth Kent, Marius Pirvu, Vijay Sundaresan and Peter Shipton

Spot instances (as provided in Amazon Elastic Compute Cloud, EC2) offer resources at a reduced cost, however, often provide less reliability. This happens as the resources assigned to spot instances can be withdrawn abruptly due to real-time variations in demand and price. A specialized mechanism to deal with this offered reduced cost while maintaining high reliability is desirable to users, as users generally try to minimize their loss maintaining a minimal cost. Such a mechanism is little focused in the literature till now. Therefore, in this paper, we propose a mechanism to mitigate the loss of these spot instances in road to providing high reliability without increasing the computational cost. To do so, first, we apply checkpointing at different stages of the computation. Further, we propose several algorithms to place checkpoints at proper points during a computation. Our proposed algorithms can mitigate the loss by up to 99.9% on an average. We confirm these findings by experimenting over real Amazon spot instance history.

Evaluating Efficiency, Effectiveness and Satisfaction of AWS and Azure from the Perspective of Cloud Beginners

Carlos Eduardo Carvalho Dantas and Marcelo de Almeida Maia

Quality has long been regarded as an important driver of cloud adoption. In particular, quality in use (QiU) of cloud platforms may drive cloud beginners to the cloud platform that offers the best cloud experience. Cloud beginners are critical to the cloud market because they currently represent nearly a third of cloud users. We carried out three experiments to measure the QiU (dependent variable) of public cloud platforms (independent variable) regarding efficiency, effectiveness and satisfaction. AWS EC2 and Azure Virtual Machines are the two cloud services used as representative proxies to evaluate cloud platforms (treatments). Eleven undergraduate students with limited cloud knowledge (participants) manually created 152 VMs (task) using the web interface of cloud platforms (instrument) following seven different configurations (trials) for each cloud platform. Whereas AWS performed significantly better than Azure for efficiency (p-value not exceeding 0.001, A-statistic = 0.68), we could not find a significant difference between platforms for effectiveness (p-value exceeding 0.05) although the effect size was found relevant (odds ratio = 0.41). Regarding satisfaction, most of our participants perceived the AWS as (i) having the best GUI to benefiting user interaction, (ii) the easiest platform to use, and (iii) the preferred cloud platform for creating VMs. Once confirmed by independent replications, our results suggest that AWS outperforms Azure regarding QiU. Therefore, cloud beginners might have a better cloud experience starting off their cloud projects by using AWS rather than Azure. In addition, our results may help to explain the AWS’s cloud leadership.

Node.js Scalability Investigation in the Cloud

Taees Eimouri, Kenneth Kent and Aleksandar Micic

Node.js has gained popularity in cloud development due to its asynchronous, non-blocking and event-driven nature. However, scalability issues can limit the number of concurrent requests while achieving an acceptable level of performance. To the best of our knowledge, no cloud-based benchmarks or metrics focusing on Node.js scalability exist. This paper presents the design and implementation of Ibenchjs, a scalability-oriented benchmarking framework, and a set of sample test applications. We deploy Ibenchjs in a local and isolated cloud to collect and report scalability-related measurements and issues of Node.js as well as performance bottlenecks. Our findings include: 1) the scaling performance of the tested Node.js test applications was sub-linear; 2) no improvements were measured when more CPUs were added without modifying the number of Node.js instances; and 3) leveraging cloud scaling solutions significantly outperformed Node.js-module-based scaling.
A Context-Aware Machine Learning-based Approach

Nathalia Nascimento, Carlos Lucena, Paulo Alencar, Donald Cowan

It is known that training a general and versatile Machine Learning (ML)-based model is more cost-effective than training several specialized ML-models for different operating contexts. However, as the volume of training information grows, the higher the probability of producing biased results. Learning bias is a critical problem for many applications, such as those related to healthcare scenarios, environmental monitoring and air traffic control. In this paper, we compare the use of a general model that was trained using all contexts against a system that is composed of a set of specialized models that was trained for each particular operating context. For this purpose, we propose a local learning approach based on context-awareness, which involves: (i) anticipating, analyzing and representing context changes; (ii) training and finding machine learning models to maximize a given scoring function for each operating context; (iii) storing trained ML-based models and associating them with corresponding operating contexts; and (iv) deploying a system that is able to select the best-fit ML-based model at runtime based on the context. To illustrate our proposed approach, we reproduce two experiments: one that uses a neural network regression-based model to perform predictions and another one that uses an evolutionary neural network-based approach to make decisions. For each application, we compare the results of the general model, which was trained based on all contexts, against the results of our proposed approach. We show that our context-aware approach can improve results by alleviating bias with different ML tasks.

Evaluating Music Mastering Quality Using Machine Learning

Mark Shtern, Pedro Casas, Vassilios Tzerpos

Machine learning has been applied in a vast array of applications in the recent years, including several qualitative problems in the arts. However, in the world of music production, including mixing and mastering, most tasks are still performed by music professionals with decades of experience. Aspiring mastering engineers typically have to apprentice with professionals to learn their craft. Access to professionals is a scarce resource though, as they are typically very busy. In this paper, we present a method to evaluate the mastering quality of a piece of music automatically. We delegate the task of determining what we deem to be a subjectively well mastered song to professional mastering engineers. Using professionally mastered music, we derive datasets with varying degrees of deviation from the original music and train models to recognize the changes that have been made. This allows us to provide novice mastering engineers with an automatic rating of their work based on the magnitude of the deviation from the gold standard. We present experiments that demonstrate the accuracy of our approach, as well as a user study that shows how the results of our approach correlate to assessments made by human evaluators.

(Semi)Automatic Construction of Access-Controlled Web Data Services

Kalvin Eng, Diego Serrano, Eleni Stroulia, Jacob Jaremko

The widespread adoption of the Internet of Things (IoT) is producing an ever-increasing stream of data that can be mined by multiple stakeholders, in support of different objectives and tasks. In fact, we are witnessing the emergence of data marketplaces that aim to share this data and harness economic value out of these transactions. The advent of data-as-a-service (DaaS) represents a key integrator opportunity that allows for the management of data collections, while providing specific privacy policies to delegated agents. To support Daas integrations, we develop a model-driven method for creating APIs to deliver DaaS. Our method supports data owners to: (1) automatically abstract the representation of relational database schemas into a visual model and map them to existing ontologies, (2) use the mappings in order to create different role-based access-control views of APIs, and (3) automatically generate API endpoints and their responses, based on these mappings. We develop a ‘plug-and-play’ prototype system for SQL databases to demonstrate this methodology and apply it to a use case of controlling data from a fitness monitoring application. Our aim is to enhance existing API creation methodologies that may be cumbersome by using semantics so that data can be easily shared and distributed.
# POSITION PAPER SCHEDULE

**MONDAY**  
**OCTOBER 29th, 2018**  
**10:15 AM - 10:45 AM**

<table>
<thead>
<tr>
<th>ROOM</th>
<th>POSITION PAPERS</th>
<th>SESSION NAME</th>
</tr>
</thead>
</table>
| Evergreen | (1) Adaptation as a Service  
(2) Uncertainty Quantification-as-a-Service                                  | Microservices         |
| Jasmine | (1) A DevOps Framework for Quality-Driven Self-Protection in Web Software Systems  
(2) Ontology Driven Temporal Event Annotator mHealth Application Framework     | Adaptive Systems      |
| Primrose | (1) A Survey of Ahead-of-Time Technologies in Dynamic Language Environments  
(2) All Timescale Window Co-occurrence                                         | Compiler Development  |
<table>
<thead>
<tr>
<th>ROOM</th>
<th>POSITION PAPERS</th>
<th>SESSION NAME</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jasmine</td>
<td>(1) Challenges and solutions on architecting Blockchain Systems</td>
<td>Blockchain</td>
</tr>
<tr>
<td></td>
<td>(2) Powering Software Sustainability with Blockchain</td>
<td></td>
</tr>
<tr>
<td>Primrose</td>
<td>(1) Ischemic Brain Stroke Detection using EEG signals.</td>
<td>Healthcare</td>
</tr>
<tr>
<td></td>
<td>(2) Predictive Analytics in Healthcare: Epileptic Seizure Recognition</td>
<td></td>
</tr>
</tbody>
</table>
Position Papers

Monday, October 29th, 2018
10:15AM – 10:45AM

Microservices – Evergreen

Session Chair: Hausi Müeller

Adaptation as a Service

Hamzeh Khazaei, Alireza Ghanbari, Marin Litoiu

Current and emerging complex systems of many types including but not limited to big data systems, web-based systems, data centers and cloud infrastructure, social networks and the Internet of Things (IoT) have increasingly distributed and dynamic architecture that provide unprecedented flexibility in creating and supporting applications. However, such highly distributed architecture also increases the complexity of end-to-end management of such systems. Due to the sheer complexity, uncertainty and at the same time programmability of cloud environments, microservices and finally big data analytics, it is now required, and possible, to enable autonomic management in distributed systems in a dependable manner. In this paper, we argue that building autonomic management systems is a challenging task and requires its own set of expertise and knowledge. Therefore, in the light of current challenges, available enablers and recent successful stories, we propose the idea of moving from self-adaptation to ADaptation-as-a-Service (ADaaS).

Uncertainty Quantification-as-a-Service

Malgorzata Zimon, Vadim Elisseev, Robert Sawko, Samuel Antão, Kirk Jordan

Uncertainty quantification (UQ), which enables non-destructive virtual testing, is the fast growing area of modern computational science. UQ methods are computationally intensive and require construction of complex workflows, which rely on a number of different software components often coming from different projects. Therefore, there is a need for developing a portable and scalable UQ pipeline that will enable efficient stochastic modeling. Our paper introduces a strategy for UQ as a Service using high performance computing and hybrid cloud infrastructures and presents its application to a heat transfer study in nuclear reactors simulation and modelling of tsunami events.

Adaptive Systems – Jasmine

Session Chair: Jin Li

A DevOps Framework for Quality-Driven Self-Protection in Web Software Systems

Nasim Beigi Mohammadi, Marin Litoiu, Mahsa Emami-Taba, Ladan Tahvildari, Marios Fokaefs, Ettore Merlo, Vio Onut

Modern software is developed, deployed and operates continuously. At the same time, cyberattacks are on the rise. The continuity of development and operations and the constant threat of attacks requires novel approaches to identify, analyze and address potential security vulnerabilities. In this continuous and volatile execution environment, factors like security, performance, cost and functionality may not be guaranteed in the same degree at the same time. In this work, we propose a DevOps framework for security adaptation that enables the development and operations teams to collaborate and address security vulnerabilities. The proposed framework spans across the different phases of software (development, operations, maintenance) and considers all other factors (performance, cost, functionality), when deciding for security adaptations. We demonstrate the approach on a prototype tool that shows how teams work together to tackle security concerns.
Ontology Driven Temporal Event Annotator mHealth Application Framework
Amente Bekele, Joe Samuel, Shermeen Nizami, Amna Basharat, Randy Giffen, James Green

We present an application (app) framework to facilitate the collection of gold standard temporal event annotations. These data will enable training and evaluation of machine learning algorithms for predicting events of clinical significance. Recording of such data using pen and paper can prove to be tedious and error-prone due to the variation in the types of events and the frequency of occurrence. To address this problem, we developed an mHealth application framework that presents an intuitive and configurable user interface for annotating a timeline with events. The presented Temporal Event Annotator (TEA) app framework supports dynamically building a customized application inclusive of events, event categories, and study attributes based on the design input of a specific study. This is accomplished by presenting a terminology schema for the hierarchical definition of event types and an additional user interface (UI) schema to support UI-specific attributes. We describe the framework architecture independent of specific technology implementations. We also describe specific instantiations of the framework that we used to develop and evaluate apps for three different use cases: 1) patient monitoring in the Neonatal Intensive Care Unit (NICU), 2) estimating patient stress levels during immersive rehabilitation therapy, and 3) quantifying the patient experience during emergency neonatal transport. The TEA framework provides a reliable and intuitive solution for temporal event annotation that accounts for the unique experimental requirements of each study.

Compiler Development – Primrose
Session Chair: Andrew Craik
A Survey of Ahead-of-Time Technologies in Dynamic Language Environments
Mark Thom, Gerhard Dueck, Kenneth Kent, Daryl Maier

Eclipse OMR is an open source collection of robust, reusable components for the construction of production-ready compilers. Great progress has been made on JITBuilder, OMR’s simplified interface to the compiler technology for building JIT compilers, but work on the planned interface for AOT compilation has only just begun. In this survey, we identify desirable characteristics for the design of OMR’s AOT by examining how several prominent open source compilers implement AOT. We conclude by discussing the advantages and disadvantages of the implementations seen, and how they might inform the final design of the OMR AOT component.

All Timescale Window Co-occurrence
Yumeng Liu, Daniel Busaba, Chen Ding, Daniel Gildea

Trace analysis is a common problem in system optimization and data analytics. This paper presents new efficient algorithms for window co-occurrence analysis, which is to find how likely two events will occur together in time windows of different lengths. The new solution requires a linear time preprocessing step, after which, it only takes logarithmic space and constant time to compute co-occurrence of a data pair in windows of any given length. One potential use of the new analysis is to reduce the asymptotic cost in affinity-based memory layout.
Tuesday, October 30th, 2018
10:15AM - 10:45AM
Blockchain – Jasmine
Session Chair: Biruk Habtemariam

Challenges and solutions on architecting Blockchain Systems
Gregory Fournier, Fabio Petrillo
Despite the fact that companies are gravitating more and more towards the use of blockchains in their systems, it is clear that the blockchains is no silver bullet. Many challenges such as scalability issues and frustrating trade-offs most notable in public decentralized blockchain systems are currently holding back blockchain's huge potential. In this paper we conduct a Systematic Literature Review in order to explore the current challenges of blockchain while presenting possible solutions to each of these challenges. We come to the conclusion that current challenges can be summarized in three categories: Scalability issues, security issues and a choice of consensus protocol. We also briefly discuss the use of blockchain in current systems, concluding that while blockchains current immaturity makes it hard to recommend for most projects, blockchains in their current state could be used in the Internet of Things.

Powering Software Sustainability with Blockchain
Omar Badreddin
Software sustainability is a systematic challenge that impacts broad segments of software systems. Software codebases must evolve overtime to address changing contexts and adapt to the flux in middlewares and platforms. In the process, it accumulates arbitrary complexities and its maintenance becomes progressively difficult. Current sustainability approaches focus on the symptoms and tend to be reactive in nature, and ignore the fundamental incentive structures that drive decision-making processes. Moreover, contemporary approaches are insensitive to the uniqueness of each software project context and operate on the assumption that sustainability measurements are universally applicable to the majority of software systems. This paper introduces a fundamentally novel peer-driven approach to managing software sustainability. The methodology ensures that software teams can define their own sustainability measures that adequately address the unique context of their project and its priorities. These measures are dynamically defined by the project peers to ensure their applicability as the project context evolves. Finally, the paper introduces Susereum, a blockchain platform that materializes the methodology and establishes novel incentive structures to systematically promote software sustainability throughout the project lifecycle.

Healthcare – Primrose
Session Chair: Fred Popowich
Ischemic Brain Stroke Detection using EEG signals
Arooj Ahmed Qureshi, Canxiu Zhang, Rong Zheng, Ahmed Elmeligi
Stroke is the second leading cause of death in the United States of America. 87% of all strokes are ischemic stroke, which is mainly caused by the blockage of small blood vessels around the brain. Magnetic resonance imaging (MRI) provides the gold standard for accurate diagnosis of ischemic strokes, but it is both time-consuming and unsuitable for 24/7 monitoring. In this paper, we propose an ischemic stroke detection method through the multi-domain analysis of EEG brain signal from wearable EEG devices and machine learning. Using 40 healthy and 40 patients' data, we find that Multi-Layered Perceptron (MLP) and Bootstrap models (Extra-Tree and Decision-Tree) can achieve test accuracy of 95% with an area under the ROC curve 0.85. auto-scalability and monitoring as-a-service for any type of cloud software system.
Predictive Analytics in Healthcare: Epileptic Seizure Recognition

Ashok Bhowmick, Tamer Abdou, Ayse Bener

Introduction Clinical applications of electroencephalography (EEG) span a very broad range of diagnostic conditions. Epileptic seizure is the fourth most common neurological disorder in that. Related Work There has been considerable progress in clinical understanding of epilepsy, however many aspects of seizure prevention are still a mystery. Predictive modeling of EEG can provide significant value addition to substantiate the diagnosis of epilepsy. Methodology Machine learning algorithms are applied to predict the probability of epileptic seizure using an open source multi-class dataset. Results and Discussion Comparing the F-score from different classifiers, it is found that XGBoost gives the best performance in binary classification and Random Forest provides the best performance in multinomial classification. Conclusion Our results show that it is possible to predict epileptic seizure with significant accuracy from non-epileptic parameters using a suitable machine learning algorithm. We also observe that binary classification methods have higher prediction accuracy.
Expo Map - Monday

Legend
- Joint R&D
- Technology Showcase
- Education and Skill
- Research

(B1) Document Signing on Blockchain
(B2) POWERAI - Democratizing AI for All Research Communities
(B3) Data Science 1 - Competency Model
(B4) Performance and Security Enhanced Containerized Analytics for IoT Systems
(B5) Randomized Offloading Algorithms for Mobile Cloud Computing
(B6) Exploring Challenges In adoption of Continuous Delivery
(B7) Container Technologies! Which One Best Suits Eclipse IoT Deployments?
(B8) Fall as an irregular pattern in IoT streaming data

(B9) DevOps Round-trip Engineering: Traceability from Dev to Ops and Back
(B10) Installing features to the Open Liberty kernel with Maven and Gradle
(B11) A Hierarchical Rule-based Security Manager System for Data-Intensive Applications
(B12) Continuous Delivery with Open Toolchains
(B13) Ahead-Of-Time Compilation of WebAssembly Using Eclipse OMR
(B14) Requirements-based Framework for Evaluating Trust in Service-oriented Systems
(B15) Tracking Unhandled Rejections in Node.js
(B16) A semi-automated Framework for Migrating Web applications from SQL to NoSQL database

(B17) Feature engineering in Big Data for the detection of information systems misuse
(B18) Composing Web Services Using a Multi-Agent Framework
(B19) Detection of Feature Interaction in Dynamic Scripting Languages
(B20) Enabling a highly-scalable, hybrid cloud-based microservices architecture
(B21) The Power of Accelerators for Cognitive Workloads
(B22) Predicting Cross-Sell Opportunities in financial Services Using Deep Learning
(B23) Attention models for gene promoter sequence characterization
(B24) Using AP-TED to Detect Phishing Attack Variations

(Please see back for the Talks Schedule)
PechaKucha Talks

17:30  B1: Document Signing on Blockchain
17:35  B15: Tracking Unhandled Rejections in Node.js
17:40  B16: A semi-automated Framework for Migrating Web applications from SQL to NoSQL database
17:45  B4: Performance and Security Enhanced Containerized Analytics for IoT Systems
17:50  B5: Randomized Offloading Algorithms for Mobile Cloud Computing
17:55  B6: Exploring Challenges In Adoption of Continuous Delivery
18:00  B7: Container Technologies! Which One Best Suits Eclipse IoT Deployments?
18:05  B8: Fall as an irregular pattern in IoT streaming data

Short Talks

18:10  S1: Db2 on Cloud Free Forever Cloud Database
18:20  S2: User Research: Mapping the end-to-end experience of a System Administrator’s deployment
18:30  S3: Assessing Databases for GDPR Compliance
18:40  S4: A Dynamic Knowledge Management System for Decision Support
18:50  S5: Distribution and Severity of Test Smells: An exploratory study

Legend

- Joint R&D
- Technology Showcase
- Education and Skill
- Research

(Please see back for the Expo Schedule)
Foodie: Conversational AI for the smart kitchen
Privacy-aware Efficient Visual Recognition Services for Smart Laboratories
Data Science 2 - Bootcamp Materials
In-Depth Evaluation of task Parallelism in Node.js
Supporting Software Variability in Eclipse OMR
Fear and Disgust during the 2015 Canadian Election: Automated Analysis with Watson
ROAD – Realtime Outlier and Anomaly Detection for IBM Cloud
Reconstruct the genealogy of phishing kits source code modifications

Machine Learning to Provide Intelligent Digital Health for Underserved Regions
Detecting Misuses of Crypto APIs
Pressure Sensitive Mats for Patient Monitoring in the NICU
Susereum - A Blockchain Platform for Sustainable Software Systems
Localizing Error Prone Code Using Software Analytics
Eclipse OMR on the Embedded AArch64 Platform
(Semi)Automatic Construction of Access-Controlled Web Data Services
Design Patterns for Machine Learning Solutions

Open Liberty and Eclipse MicroProfile: Foundation for Microservices
Node.js Scalability Investigation in the Cloud
Feasibility of Internal Object Pools to Reduce memory Management Activity
Cold Object Identification, Sequestrating and Revitalization
Persistent Memory Storage of Cold Regions in the OpenJ9 Java Virtual Machine
A novel approach to generational garbage collection
Math function performance acceleration via integrated HW and algorithm design
Automating the Detection of Third-Party java Library Migration At the Function Level

Legend
- Joint R&D
- Technology Showcase
- Education and Skill
- Research

(Please see back for the Talks Schedule)
Expo Map - Tuesday

PechaKucha Talks

17:30 B1: Foodie: Conversational AI for the smart kitchen
17:35 B2: Privacy-aware Efficient Visual Recognition Services for Smart Laboratories
17:40 B16: Design Patterns for Machine Learning Solutions
17:45 B4: In-Depth Evaluation of task Parallelism in Node.js
17:50 B5: Supporting Software Variability in Eclipse OMR
17:55 B6: Fear and Disgust during the 2015 Canadian Election: Automated Analysis with Watson
18:00 B7: ROAD - Realtime Outlier and Anomaly Detection for IBM Cloud
18:05 B8: Reconstruct the genealogy of phishing kits source code modifications
18:10 B9: Machine Learning to Provide Intelligent Digital Health for Underserved Regions
18:15 B10: Detecting Misuses of Crypto APIs

Short Talks

18:20 S1: Discovering business rules automatically from legacy applications
18:30 S2: REST APIs with LoopBack 4 and OpenAPI 3
18:40 S3: Dynamic Large-Scale Graph Processing over Data Streams
18:50 S4: A Multilevel Streaming Data Analytics Infrastructure for Predictive Analytics

Legend

- Joint R&D
- Technology Showcase
- Education and Skill
- Research

(Please see back for the Expo Schedule)
## Expo Map - Wednesday

### Gate 1
- B1: Model based Interference-aware Management System for Microservice Architectures
- B2: Modeling AUTOSAR Implementations in Simulink
- B3: Data Science 3 - Student Success Projects
- B4: Accessibility in the DevOps Era
- B5: Space Sentry Challenge: A game built with IoT devices and Microservices
- B6: Comparative Analysis of Cloud Console Systems
- B7: REST APIs with LoopBack 4 and OpenAPI 3
- B8: Interactive Visualization to Engage End-users in Prescriptive Analysis

### Gate 2
- B9: Finding Missing Labels for Large Industrial Datasets
- B10: Synthesizing Realistic Data for Machine Learning
- B11: Watch Phishing Evolution Over time
- B12: Easy Cloud-native Development with Microclimate
- B13: A Machine Learning Approach to Test Case Prioritization
- B14: Optimization Driven Inlining
- B15: Semantic-Aware Disambiguation for Entity Resolution
- B16: Create data-driven visual reports in Reactive Reports, a framework built for big data

### Gate 3
- B17: OptiMatch: Semantic-Web System for Automated Query Tuning
- B18: Third Party Container Tests
- B19: Runtime Automatic Parallelization of JVM Applications
- B20: Variability Aware analysis using Clang
- B21: Approximation Algorithms for Instruction Scheduling and Rapid Prototyping in Coconut
- B22: Static Dataflow Analysis for OpenMP Code Sanitization
- B23: Plasma Focused Ion Beam Curtaining Artefact Correction by Fourier-Based Linear Model
- B24: FLASH: Multithread Fault Localization

### Legend
- Orange: Joint R&D
- Blue: Technology Showcase
- Red: Education and Skill
- Green: Research

---

#CASCON2018
CASCON Expo
Monday, October 29th, 2018
11:00 to 13:00 and 17:30 to 19:00

Exhibit Sessions:

Document Signing on Blockchain
Asic Chen, Arno Jacobsen

Our project provides a fast, secure, and decentralized solution for document signing based on blockchain technology. Document endorsements are immutably recorded and reliably time-stamped on the blockchain, easily verifiable by third parties that neither need to authenticate the signatories, nor need to access the content of the document. User identities are verified without the need for a centralized certificate authority. Our prototype is built on Hyperledger Fabric with an Angular front end.

PowerAI - Democratizing AI for All Research Communities
Marija Mijalkovic

IBM's PowerAI platform provides a unique solution to support, enhance and progress the ground breaking research pursuits of diverse academic institutions and faculties. The solution is underpinned by the most powerful and innovative infrastructure for the development of artificial intelligence and layers on data science tools and utilities to enable both skilled computer science and mathematics experts as well as those with 'domain expertise' who are knowledgeable in their respective fields.

Data Science 1 - Competency Model
Marcellus Mindel, Dennis Buttera, Jennifer Collins, Ana Echeverri

In the first of three exhibits on Data Science Education, we will discuss applications of a Competency Model for Data Science originally developed by Ana Echeverri, IBM Data Science Growth Strategies Lead, to support the nontraditional education paths of new collar jobs. The model may also prove useful for postsecondary institutions that use competency matrices to manage the coverage of data and analytics skills in academic programs.

Performance and Security Enhanced Containerized Analytics for IoT Systems
Simon Nadeau, Nathaniel Holeksa, Marios Fokaefs

Our research delves into the pseudo dichotomy that is to be secure or to be performant, in a cloud based IoT architecture. We explore the two in the context of a sensor network in smart buildings, a novel technology that hopes to improve our environmental footprint. In our work, we deploy Spark streaming jobs in Docker containers and explore how to effectively, and automatically, scale them using Docker Compose pursuant to the CPU and memory usage of the standard and the security analytics.
Randomized Offloading Algorithms for Mobile Cloud Computing

Haleh Shahzad, Majid Risman Kar Zadeh

Computational offloading can improve the energy efficiency of mobile devices, by executing some tasks of a mobile application in the cloud. In my research, a set of new algorithms are presented to find the best decision for each task. These algorithms iteratively improve an offloading decision vector, trying to find the optimal one that can minimize the total energy consumption of the mobile device when the application is running while satisfying the delay requirement of the application.

Exploring Challenges In Adoption of Continuous Delivery

Trevor Rae

We look at the technical and cultural problems that exist in regards to adopting continuous delivery/integration into a company’s infrastructure.

Container Technologies! Which One Best Suits Eclipse IoT Deployments?

Mohab Aly, Foutse Khomh

The increasing apex in the IoT paradigm requires specific platforms to allow different components to communicate. Our research dives into assessing the performance implications of the Open-Sourced Eclipse IoT, including Hono, to provide guidelines while deploying them in production. In our work, we setup loads in container technologies, such as Kubernetes and OpenShift, and explore how to effectively scale them out using EnMasse while analyzing the CPU usage, memory consumption and network I/O.

Fall as an irregular pattern in IoT streaming data

Sazia Mahfuz

Detecting patterns in real time streaming data has been an interesting and challenging data analytics problem. In this work, we address the problem of learning an irregular human activity pattern, fall, from streaming IoT data using wearable sensors. We present a deep neural network model for detecting fall based on accelerometer data using an online dataset "MobiAct" by Vavoulas et al. The initial model was developed using IBM Watson studio and then later transferred and deployed on IBM Cloud.

DevOps Round-trip Engineering: Traceability from Dev to Ops and Back

Miguel Jimenez

DevOps engineers follow an iterative and incremental process to develop Deployment and Configuration specifications. Such process is likely to involve manual modifications to the running environment. Failing to appropriately update the specifications leads to technical debt. In this demo, we present a two-way Continuous Integration framework that automatically updates Terraform templates when the target OpenStack deployment changes, enabling bi-directional traceability of the modifications.
Installing features to the Open Liberty kernel with Maven and Gradle

Eric Lau

Open Liberty is a lightweight and modular server runtime for Java developers. It has features on Maven Central that can be installed to its kernel. We showcase the use of these features, the steps to add them into Maven and Gradle builds, and the benefits of doing so for developers. We explain how they provide relevant dependencies for compiling applications which simplifies the development process. We also provide high level details of the components used to make this solution possible.

A Hierarchical Rule-based Security Management System for Data-Intensive Applications

Marin Litoiu, Yar Rouf, Vio Onut

We propose a platform to support the SecDevOps framework, a hierarchical distributed architecture for security control that uses a Business Rules Engine. The Business Rules Engine simplifies security rules by allowing the teams to write them at an operational level rather than at the network level, which would require specialized knowledge.

Continuous Delivery with Open Toolchains

Chris Brealey, Curtis D'Entremont, Christophe ELEK, Joel Cayne

Continuous Delivery (n) - 1. The subtle science and exact art of releasing high quality, bite-sized changes to software at high velocity without sacrificing control or the ability to pivot with the needs of our clients. - 2. a feature of the IBM Cloud, built on the Open Toolchain platform, that delivers industry leading development tools and tool integrations as a service. Drop by to see toolchains in action, and to get the creative juices flowing on ways to evolve them!

Ahead-Of-Time Compilation of WebAssembly Using Eclipse OMR

Mark Thom, Gerhard Dueck, Daryl Maier, Kenneth Kent

WebAssembly is an emerging assembly language that runs on the web. It has nearly native performance, and is already the compilation target of several programming languages. There is an ongoing project to create a JIT compiler for WebAssembly using Eclipse OMR's JITBuilder. The poster will describe our attempted contribution of an AOT compiler implemented with OMR. We hope it will expose any disparity between OMR's current support for AOT and what its completed AOT module should be capable of.
Requirements-based Framework for Evaluating Trust in Service-oriented Systems

Konstantinos Kontogiannis, Kostas Tsiounis

We define trust as the degree of expectation that a service, resource, or system, will satisfy its predefined functional and non-functional requirements. Users define their requirement expectations sought by a service or a resource (i.e., their goals) as expressions. Fuzzy goal model reasoners, the TidalTrust algorithm, and information obtained from the system are used to compute the final trust a user associates that the service will be achieving the goals set in a given context.

Tracking Unhandled Rejections in Node.js

Maxim Uzun, Kenneth B. Kent, Michael Dawson

Node.js is an open source cross-platform runtime environment for executing server-side JavaScript. The framework is based on Google Chrome's V8 and is limited in regard to post-mortem diagnostics and debugging for Promises; a design pattern for managing asynchronous code. We propose and experimentally evaluate an implementation of Node.js that aims to improve troubleshooting for unhandled Promise rejections. The goal is to produce accurate core dumps and track unhandled rejections.

A semi-automated Framework for Migrating Web applications from SQL to NoSQL database

Rahma Al Mahruqi

We propose a semi-automated approach to migrate highly dynamic SQL-based web applications to ones that use NoSQL database. There are two parts to this approach: the migration of schema and data, and the migration of the actual application code with embedded queries. Our approach provides contributions to migrating and optimizing the embedded SQL queries to interact with the new database system and changing the application code to use the translated queries.

Feature engineering in Big Data for the detection of information systems misuse

Eduardo Lopez, Kamran Sartipi

This cybersecurity research project explores the Apache Spark-enabled feature engineering activities — conceptualization and generation — in machine learning detection of an insider's threat. Our experiment uses a very large, 1.6 billion records dataset of information systems logs, collected for more than 12,000 users over a period of 58 days. We test the feature set using classification models, validating feature relevance and significance using Receiver Operating Characteristics curves.
Composing Web Services Using a Multi-Agent Framework
Yu Zhao
Software agents are autonomous entities that can proactively activate web services, and autonomously react upon the results to determine the next steps. However, programming agents are complex due to the the proactive, autonomous and reactive characteristics of agents. To reduce the workload of developers when programming agents, we propose an easy-to-understand semi-natural language syntax to program agents for service composition.

Detection of Feature Interaction in Dynamic Scripting Languages
Omar Al Harthi
Our research indicates that duplicate names, from published plugins, can be a potential cause of conflict due to overlapping. Such feature interaction may create a challenge for smooth compatibility of JavaScript plugin. To detect duplicate names we need first to understand how the objects and properties are created, used and changed in JavaScript. This work aims to detect such conflict using static analysis approach.

Enabling a highly-scalable, hybrid cloud-based microservices architecture
Anthony Kwan, Jonathon Wong
Traditional microservice scaling methods perform either horizontal or vertical scaling exclusively. When used in combination, however, these methods offer complementary benefits and compensate for each other’s deficiencies. To leverage the high availability of horizontal scaling and the fine-grained resource control of vertical scaling, we developed two novel hybrid autoscaling algorithms and benchmarked them against Google’s popular Kubernetes horizontal autoscaling algorithm.

The Power of Accelerators for Cognitive Workloads
Mel Bakhshi, Theresa Xu, Neil Graham
Moore’s Law is no longer operative. Yet, cognitive workloads demand ever-more computational resources to process ever-increasing volumes of data. What options does he industry have to move forward? In this poster session, we’ll show how accelerators such as GPUs, FPGAs and high-bandwidth storage and network can help to address these challenges and show how POWER systems are uniquely designed to take advantage of these emerging technologies.
Predicting Cross-Sell Opportunities in Financial Services Using Deep Learning

Chux Ejieh, Diane Reynolds

In this exhibition, we will explore the design and implementation of the Product Propensity model (part of the IBM Client Insights for Wealth Management SaaS) and share how it impacts our clients. Cross-sell refers to the practice of selling an additional product to an existing client. Ideally, marketing campaigns in this domain are more targeted as they leverage information the business has about its customers.

Attention models for gene promoter sequence characterization

Raul Ivan Perez Martell, Alison Ziesel, Alex Thomo, Ulrike Stege

Augmented recurrent neural networks are well-suited to the analysis of eukaryotic promoter sequences. DNA sequences, like other biological sequences, are related to other sequential problems such as natural language processing (NLP). In this work we will investigate the efficacy of attention-augmented RNN networks on molecular biology data including DNA sequence, ChIP, and protein binding assays. We also investigate different biological data types and their relevance for a well-trained network.

Using AP-TED to Detect Phishing Attack Variations

Sophie Le Page

Detection of phishing attack variations is accordingly currently used as a prevention tool against phishing. We evaluate here the feasibility of applying Pawlik and Augsten's recent implementation of Tree Edit Distance (AP-TED) calculations as a way to compare DOMs and identify similar phishing attack instances. We also compare this tree method with an existing method that uses the distance between tag vectors to quantify similarity between phishing sites.
PechaKucha Sessions:

Document Signing on Blockchain
17:30 – 17:35

Asic Chen, Amo Jacobsen

Our project provides a fast, secure, and decentralized solution for document signing based on blockchain technology. Document endorsements are immutably recorded and reliably time-stamped on the blockchain, easily verifiable by third parties that neither need to authenticate the signatories, nor need to access the content of the document. User identities are verified without the need for a centralized certificate authority. Our prototype is built on Hyperledger Fabric with an Angular front end.

Tracking Unhandled Rejections in Node.js
17:35-17:40

Maxim Uzun, Kenneth B. Kent, Michael Dawson

Node.js is an open source cross-platform runtime environment for executing server-side JavaScript. The framework is based on Google Chrome’s V8 and is limited in regard to post-mortem diagnostics and debugging for Promises; a design pattern for managing asynchronous code. We propose and experimentally evaluate an implementation of Node.js that aims to improve troubleshooting for unhandled Promise rejections. The goal is to produce accurate core dumps and track unhandled rejections.

A semi-automated Framework for Migrating Web applications from SQL to NoSQL database
17:40-17:45

Rahma Al Mahruqi

We propose a semi-automated approach to migrate highly dynamic SQL-based web applications to ones that uses NoSQL database. There are two parts to this approach: the migration of schema and data, and the migration of the actual application code with embedded queries. Our approach provides contributions to migrating and optimizing the embedded SQL queries to interact with the new database system and changing the application code to use the translated queries.

Performance and Security Enhanced Containerized Analytics for IoT Systems
17:45 – 17:50

Simon Nadeau, Nathaniel Holeksa, Marios Fokaefs
Our research delves into the pseudo dichotomy that is to be secure or to be performant, in a cloud based IoT architecture. We explore the two in the context of a sensor network in smart buildings, a novel technology that hopes to improve our environmental footprint. In our work, we deploy Spark streaming jobs in Docker containers and explore how to effectively, and automatically, scale them using Docker Compose pursuant to the CPU and memory usage of the standard and the security analytics.

Randomized Offloading Algorithms for Mobile Cloud Computing

17:50-17:55

Haleh Shahzad, Majid Risman Kar Zadeh

Computational offloading can improve the energy efficiency of mobile devices, by executing some tasks of a mobile application in the cloud. In my research, a set of new algorithms are presented to find the best decision for each task. These algorithms iteratively improve an offloading decision vector, trying to find the optimal one that can minimize the total energy consumption of the mobile device when the application is running while satisfying the delay requirement of the application.

Exploring Challenges In Adoption of Continuous Delivery

17:55-18:00

Trevor Rae

We look at the technical and cultural problems that exist in regards to adopting continuous delivery/integration into a company's infrastructure.

Container Technologies! Which One Best Suits Eclipse IoT Deployments?

18:00-18:05

Mohab Aly, Foutse Khomh

The increasing apex in the IoT paradigm requires specific platforms to allow different components to communicate. Our research dives into assessing the performance implications of the Open-Sourced Eclipse IoT, including Hono, to provide guidelines while deploying them in production. In our work, we setup loads in container technologies, such as Kubernetes and OpenShift, and explore how to effectively scale them out using EnMasse while analyzing the CPU usage, memory consumption and network I/O.

Fall as an irregular pattern in IoT streaming data

18:05-18:10

Sazia Mahfuz

Detecting patterns in real time streaming data has been an interesting and challenging data analytics problem. In this work, we address the problem of learning an irregular human activity pattern, fall, from streaming IoT data using wearable sensors. We present a deep neural network model for detecting fall based on accelerometer data using an online dataset "MobiAct" by Vavoulas et al. The initial model was developed using IBM Watson studio and then later transferred and deployed on IBM Cloud.
Short Talk Sessions:

Db2 on Cloud Free Forever Cloud Database
18:10 – 18:20

Saygin Yag, Simon Lightstone

In 2018, Db2 on Cloud introduced a lite plan that is completely free for as long as you need it. This plan provides users with 100MB of cloud database space on a multi-tenant system. This presents a perfect starting point for students, university classes, idea-stage startups and anyone who wants to have a world-class cloud database, for free!

User Research: Mapping the end-to-end experience of a System Administrator’s deployment
18:20-18:30

Arun Martin, Steve Murphy

In this talk, we share how design research helped the product stakeholders identify issues in the end-to-end experiences and various support channels accessed while deploying complex enterprise applications. During this design research journey, the product stakeholders understood the ecosystem of primary & secondary users, dependencies and handoffs and how the current customer journey impacted the organizational teams and processes. The findings from the study informed the creation of a MVP.

Assessing Databases for GDPR Compliance
18:30-18:40

Devan Shah, Walid Rjaibi, Larry Lindsay, Josue Diaz

The General Data Protection Regulation (GDPR) requires organization to implement adequate controls to protect Personally Identifiable Information (PII). A critical step in that journey is impact assessment. That is, understanding where data subject to GDPR is located and how vulnerable it is. In this session, we will introduce IBM Security Guardium Analyzer, a new SaaS offering which intends to help organizations easily complete the impact assessment for their databases.
A Dynamic Knowledge Management System for Decision Support

18:40 – 18:50

Ftoon Kedwan, Farhana Zulkernine

We focus on; a) studying the state-of-the-art KMS to select the right system to store hybrid medical knowledge, b) define a multi-level graph structure with schema and metadata to dynamically and efficiently link knowledge sources and also to contain frequently accessed knowledge, and c) develop machine learning models to learn query patterns for efficient knowledge organization and retrieval. We are still at the very early phase of the research. Fig. 1 shows the conceptual architecture.

Distribution and Severity of Test Smells: An exploratory study

18:50-19:00

Mohamed Wiem Mkaouer

Similar to production code, unit tests are also susceptible to bad programming practices, which can, in turn, have a negative impact not only on the quality of the software system but also on maintenance activities. In this talk, I showcase various test smell types, demonstrate their existence in Java projects, analyze their impact on software maintenance in terms of code vulnerability to changes and bugs.
Tuesday, October 30th, 2018

12:00 to 13:30 and 17:30 to 19:00

Exhibit Sessions:

Foodie: Conversational AI for the smart kitchen

Prashanti Priya Angara, Miguel Jimenez, Hausi Muller, Ulrike Stege

Foodie is a cognitive text-and-voice based conversational agent that augments the capabilities of home cooks by incorporating health-related information to aid one’s eating habits. This exhibit demonstrates the capabilities and limitations of existing conversational frameworks via a hands-on demo of Foodie which is built with IBM’s Natural Language Understanding and voice services. On the back-end, Foodie is connected to food databases such as Spoonacular (for recipes).

Privacy-aware Efficient Visual Recognition Services for Smart Laboratories

Christianne Huber, Mahima Chaudhary, Joydeep Mukherjee, Marin Litoiu

With concerns about personal privacy on the rise, individuals may prefer that their images captured by visual recognition services have controlled privacy levels. IBM Watson Visual Recognition Service ensures efficiency but individuals may prefer not to send images to such a public service. We propose an intelligent privacy-aware framework that maintains efficiency while delivering a desired privacy level to its end users by dynamically directing images to a localized privacy guaranteed server.

Data Science 2 - Bootcamp Materials

Marcellus Mindel, Dennis Buttera, Jennifer Collins

In the second of three exhibits on Data Science Education, we will discuss uses for 12 hours of course materials that cover the data science workflow using IBM Watson Studio. These materials are suitable for inclusion as part of an introductory Data Science course. They could also be delivered as an intensive hands-on bootcamp, for example, to complement a business program, or to enable graduate students from different backgrounds to begin with a consistent practical foundation for data science.

In-Depth Evaluation of Task Parallelism in Node.js

Maria Patrou

Node.js is a framework for asynchronous I/O, event-driven, server-side Javascript. Computationally intensive tasks are bound to the performance of a single core. We evaluate the performance of some representative Node.js multi-processed and multi-threaded techniques used for scaling, running a CPU-intensive task. Computation, memory and garbage collection metrics, the effect of the execution environment and similarities between these modules are investigated and recommendations are provided.
Supporting Software Variability in Eclipse OMR

Samer Al Masri

Eclipse OMR is a C++ library of language runtime components that supports multiple architectures and various languages. Aiming to support variability in OMR, we created OMRStatistics, a static analysis tool that helps developers collect statistics about the project and make development decisions about variability. I will be explaining how we use OMRStatistics to support variability in OMR and our experience switching from static polymorphism to dynamic polymorphism in OMR’s implementation.

Fear and Disgust during the 2015 Canadian Election: Automated Analysis with Watson

Trevor Deley

There is a tremendous opportunity for automated content analysis solutions in communications research as well as the media industry at large. This poster presents a proof of concept for automated content analysis in a media studies context. Currently there is a gap between the long standing field of content analysis in media studies and the field of sentiment analysis that arose as a subset of natural language processing. This exhibit hopes to begin to bring the two fields closer together.

ROAD - Realtime Outlier and Anomaly Detection for IBM Cloud

William Pourmajidi, Lei Zhang, Andriy Miranskyy

We refer to logs when there are errors or anomalies. Hence, anomaly detection is a crucial part of the log analysis, and if anomalies are detected at an early stage, many issues can be prevented. In this poster, we discuss practical issues and potential solutions needed to create an anomaly detection solution tailored for a large-scale Cloud platform consisting of multiple software and hardware components.

Reconstruct the genealogy of phishing kits source code modifications

Stéphane Heudron

In this project, we will study the similarity between phishing kits belonging to a reference database and develop an approach to extract a plausible genealogy to explain source code modifications. Our initial approach is based on static analysis of PHP scripts. To investigate the similarity, we will explore several distances and algorithms responding to our constraints (incrementality, efficiency). We will present preliminary results of similarity between kits and some candidate genealogy trees.

Machine Learning to Provide Intelligent Digital Health for Underserved Regions

Kamran Sartipi, Jing Chong

This project provides advanced information intelligence for physicians in underserved regions to utilize mined knowledge of medical specialties. A pilot project is ongoing with collaboration of Gastroenterologists and Family Physicians at McMaster University. Resources are: anonymized specialty patient data; research datasets; expert decisions of specialists. Concept lattice, machine learning will produce RDF knowledgebase. Intelligent consultant service navigates RDF and interacts with doctors.
Detecting Misuses of Crypto APIs

Kristen Newbury

This exhibit will present the integration of cryptography API misuse detection into a Just-in-Time context. My goal for a novel contribution to this research will be to enable a JIT compiler to perform security checks that have been generated by a pre-existing API misuse detection framework. Such an integration is expected to benefit the precision and performance of the security analyses, since runtime values can be exploited in this context.

Pressure Sensitive Mats for Patient Monitoring in the NICU

Shermeen Nizami, James R. Green

A research project that is evaluating the use of the pressure sensitive mat (PSM) technology as well RGB-D streaming video as continuous patient monitoring modalities. This research is conducted at the Neonatal Intensive Care Unit (NICU) at CHEO, Ottawa. We have developed and deployed two software apps at the patient bedside to collect physiologic and clinical event data from patients in real-time. We have acquired data from fifteen critically ill newborn babies up to date.

Susereum - A Blockchain Platform for Sustainable Software Systems

Omar Badreddin

Software sustainability is a systematic challenge, and not an individual or team failings. This systematic challenge stems because engineers often take shortcuts to maximize immediate gains that become detrimental to software longevity. We demo Susereum, a blockchain platform that creates a balance between long term longevity and short term necessities. Susereum creates a permanent public ledger to monitor the sustainability of software code modifications and credit their authors.

Localizing Error Prone Code Using Software Analytics

Konstantinos Kontogiannis, Sanjay Ghanathey, Marios Grigoriou

Evaluation study into how software analytics and machine learning can be used to predict whether a source code file contains one or more errors that may contribute to a major system failure. - The approach is utilizing information extracted both from the system's source code, such as code metrics, and from a series of DevOps tools. - Results indicate that data obtained from repositories and used as training features performed equal to or better than source code metrics.

Eclipse OMR on the Embedded AArch64 Platform

Jean-Philippe Legault, Aaron Graham

The AArch64 architecture—a subset of the ARM architectural model—has the characteristics to be driving the next generation of data centers. With low power and cost overhead, AArch64 helps move computation closer to the originating data; following the “fog computing” model. We develop support for AArch64 in the Eclipse OMR Toolkit to evaluate modern runtimes with new computing paradigms.
(Semi)Automatic Construction of Access-Controlled Web Data Services

Kalvin Eng

A demo of a prototype system implementing a novel fitness application use case. This system uses model-driven methodology to (1) automatically abstract the representation of relational database schemas into a visual model and map them to existing ontologies, (2) use the mappings in order to create different role-based access-control views of APIs, and (3) automatically generate API endpoints and their responses, based on these mappings.

Design Patterns for Machine Learning Solutions

Soroosh Nalchigar

dvanced analytics and machine learning algorithms have a wide and rapidly growing range of applications in business contexts. In this project, we apply a design patterns approach to reduce the level of expertise and efforts needed to develop machine learning solutions to be incorporated into cognitive business operations. The approach is based on design knowledge catalogs encoded according to a meta-model. Implementation of the approach is being investigated in the context of a BPM suite.

Open Liberty and Eclipse MicroProfile: Foundation for Microservices

Eveline Cai, Gilbert Kwan, YK Chang, Panagiotis Roubatsis

We will introduce Open Liberty, which is the largest open source contributions ever done by IBM, what it is, and why Open Liberty. We invite you to come see what Open Liberty is, collaborate with us, and contribute to this fast, dynamic, and simply amazing application server! Eclipse MicroProfile is the next step in the evolution of enterprise Java technologies for the microservices and cloud native world.

Node.js Scalability Investigation in the Cloud

Jiapeng Zhu, Maria Patrou

Node.js has gained popularity in cloud development. However, scalability issues can limit the number of concurrent requests while achieving an acceptable level of performance. We design and implement Ibencjjs, a scalability-oriented benchmarking framework, and a set of sample test applications. We deploy Ibencjjs in a local and isolated cloud to report scalability-related measurements and issues of Node.js as well as performance bottlenecks to investigate Node.js scalability issues in the cloud.

Feasibility of Internal Object Pools to Reduce Memory Management Activity

Konstantin Nasartschuk

Object pools is a software engineering pattern used to reuse object instances without the need of repeated allocation and instantiation. While the benefits of using object pool structures are still present when used in a garbage collected environment, it adds a memory management component to the development process. This work investigates automatically created object pools for predefined classes. Automatic object pools are implemented and discussed using the GenCon GC and Balanced GC policies.
Cold Object Identification, Sequestrating and Revitalization

**Abhijit Taware**

Cold objects are alive and infrequently accessed. Such objects are an overhead for garbage collection. Identifying and moving them to a separate memory space would benefit the real memory footprint and cache coherence. Intel's NVRam technology will be used as a cold heap for this research. Application will be profiled to gather statistics about object temperature. This will be utilized for cold object sequestration and revitalization.

Persistent Memory Storage of Cold Regions in the OpenJ9 Java Virtual Machine

**Scott Young**

In this poster an optimization technique for language runtimes with automatic memory management is presented. The technique involves segregating objects into different memory areas, backed by different memory devices, on a per object basis. This technique is compared against operating system paging mechanisms with swap partitions. Two different schemes for determining which objects should be segregated into slower memory are tested and their results are discussed.

A novel approach to generational garbage collection

**Kim Briggs**

Presenting a new algorithm for evacuating live objects from the nursery during generational collection cycles. This algorithm uses a stack to follow object reference chains and selectively copies small objects close to referring pointers, thereby improving collector throughput and object locality within the generational heap. This algorithm is being developed within IBM's open source OMR framework, which provides language-agnostic components to provide common components for language runtimes.

Math function performance acceleration via integrated HW and algorithm design

**Christopher Anand, Lucas Dutton, Adele Olejarz, Robert Enenkel**

We present results from our CAS project to design HW instructions together with algorithms to work together to produce fast, accurate mathematical functions.

Automating the Detection of Third-Party Java Library Migration At The Function Level

**Mohamed Wiem Mkaouer**

This poster complements our accepted paper entitled "Automating the Detection of Third-Party Java Library Migration At The Function Level". In this poster, we show the design of algorithms related to mining all the manually-performed function replacements for a given library migration. We further show the detailed results of the comparative study between state-of-art approaches for detecting migration traces at the function level.
PechaKucha Sessions:

Foodie: Conversational AI for the smart kitchen
17:30 – 17:35
Prashanti Priya Angara, Miguel Jimenez, Hausi Muller, Ulrike Stege
Foodie is a cognitive text-and-voice based conversational agent that augments the capabilities of home cooks by incorporating health-related information to aid one’s eating habits. This exhibit demonstrates the capabilities and limitations of existing conversational frameworks via a hands-on demo of Foodie which is built with IBM’s Natural Language Understanding and voice services. On the back-end, Foodie is connected to food databases such as Spoonacular (for recipes).

Privacy-aware Efficient Visual Recognition Services for Smart Laboratories
17:35-17:40
Christanne Huber, Mahima Chaudhary, Joydeep Mukherjee, Marin Litoiu
With concerns about personal privacy on the rise, individuals may prefer that their images captured by visual recognition services have controlled privacy levels. IBM Watson Visual Recognition Service ensures efficiency but individuals may prefer not to send images to such a public service. We propose an intelligent privacy-aware framework that maintains efficiency while delivering a desired privacy level to its end users by dynamically directing images to a localized privacy guaranteed server.

Design Patterns for Machine Learning Solutions
17:40-17:45
Soroosh Nalchigar
advanced analytics and machine learning algorithms have a wide and rapidly growing range of applications in business contexts. In this project, we apply a design patterns approach to reduce the level of expertise and efforts needed to develop machine learning solutions to be incorporated into cognitive business operations. The approach is based on design knowledge catalogs encoded according to a meta-model. Implementation of the approach is being investigated in the context of a BPM suite.

In-Depth Evaluation of Task Parallelism in Node.js
17:45-17:50
Maria Patrou
Node.js is a framework for asynchronous I/O, event-driven, server-side Javascript. Computationally intensive tasks are bound to the performance of a single core. We evaluate the performance of some representative Node.js multi-processed and multi-threaded techniques used for scaling, running a CPU-intensive task. Computation, memory and garbage collection metrics, the effect of the execution environment and similarities between these modules are investigated and recommendations are provided.
Supporting Software Variability in Eclipse OMR
17:50 – 17:55
Prashanti Priya Angara, Miguel Jimenez, Hausi Muller, Ulrike Stege
Eclipse OMR is a C++ library of language run-time components that supports multiple architectures and various languages. Aiming to support variability in OMR, we created OMRStatistics, a static analysis tool that helps developers collect statistics about the project and make development decisions about variability. I will be explaining how we use OMRStatistics to support variability in OMR and our experience switching from static polymorphism to dynamic polymorphism in OMR's implementation.

Fear and Disgust during the 2015 Canadian Election: Automated Analysis with Watson
17:55 – 18:00
Trevor Deley
There is a tremendous opportunity for automated content analysis solutions in communications research as well as the media industry at large. This poster presents a proof of concept for automated content analysis in a media studies context. Currently there is a gap between the long standing field of content analysis in media studies and the field of sentiment analysis that arose as a subset of natural language processing. This exhibit hopes to begin to bring the two fields closer together.

ROAD - Realtime Outlier and Anomaly Detection for IBM Cloud
18:00 – 18:05
William Pourmajidi, Lei Zhang, Andriy Miranskyy
We refer to logs when there are errors or anomalies. Hence, anomaly detection is a crucial part of the log analysis, and if anomalies are detected at an early stage, many issues can be prevented. In this poster, we discuss practical issues and potential solutions needed to create an anomaly detection solution tailored for a large-scale Cloud platform consisting of multiple software and hardware components.

Reconstruct the genealogy of phishing kits source code modifications
18:05 – 18:10
Stéphane Heudron
In this project, we will study the similarity between phishing kits belonging to a reference database and develop an approach to extract a plausible genealogy to explain source code modifications. Our initial approach is based on static analysis of PHP scripts. To investigate the similarity, we will explore several distances and algorithms responding to our constraints (incrementality, efficiency). We will present preliminary results of similarity between kits and some candidate genealogy trees.
Machine Learning to Provide Intelligent Digital Health for Underserved Regions

18:10 – 18:15

Kamran Sartipi, Jing Chong

This project provides advanced information intelligence for physicians in underserved regions to utilize mined knowledge of medical specialties. A pilot project is ongoing with collaboration of Gastroenterologists and Family Physicians at McMaster University. Resources are: anonymized specialty patient data; research datasets; expert decisions of specialists. Concept lattice, machine learning will produce RDF knowledgebase. Intelligent consultant service navigates RDF and interacts with doctors.

Detecting Misuses of Crypto APIs

18:15 – 18:20

Kristen Newbury

This exhibit will present the integration of cryptography API misuse detection into a Just-in-Time context. My goal for a novel contribution to this research will be to enable a JIT compiler to perform security checks that have been generated by a pre-existing API misuse detection framework. Such an integration is expected to benefit the precision and performance of the security analyses, since runtime values can be exploited in this context.
Short Talk Sessions:

Discovering business rules automatically from legacy applications
18:20 – 18:30
Arun Martin, David Budreau, Tala El Hallak
The BRM research and development team will demonstrate how our revised architecture focusing on machine learning and user research has transformed last year’s BRM preview in Application Discovery and Delivery Intelligence (ADDI) to enable automated discovery of terms and to attach them to relevant code snippets through a collaborative workflow between business and information technology teams resulting in increased knowledge, reach, and ability for members to make faster business decisions.

REST APIs with LoopBack 4 and OpenAPI 3
18:30-18:40
Taranveer Virk, Diana Lau, Biniam Admikew, Janny Hou
REST APIs power web applications, mobile applications, IoT devices & more. LoopBack 4 is an API creation framework that has been rewritten from the ground up and features a new programming interface, OpenAPI support, Dependency Injection, a new CLI and so much more. LoopBack makes it easy to create powerful OpenAPI 3.x (an industry standard) compliant REST APIs in minutes. It supports connecting to multiple SQL and NoSQL databases. Poster will show how to get started in minutes with

Dynamic Large-Scale Graph Processing over Data Streams
18:40-18:50
Tariq Abughofa
Processing time-evolving graphs is still a challenging problem. In this work, we studied approaches to dynamic graph processing with Spark. The study helped us design and develop Sprouter, a dynamic graph processing framework that enables storing enormous graphs, real-time updates, and efficient OLTP and analytics. We selected community detection as a case study of incremental analytics with Sprouter and proposed IDWCC, a novel community detection algorithm for dynamic graphs.

A Multilevel Streaming Data Analytics Infrastructure for Predictive Analytics
18:50-19:00
Haruna Isah, Farhana Zulkernine
We are developing a real-time multilevel streaming data analytics infrastructure using cutting-edge streaming data processing engines, in-memory data structure and analytics tools which will preprocess, filter and store necessary data in the memory for higher level processing. This talk is aimed at showcasing our preliminary framework and is targeted at individuals and organizations that are planning to optimize their current infrastructure to embrace multilevel streaming data analytics.
Wednesday, October 31th, 2018

8:30 to 10:45 and 12:00 to 13:30

Exhibit Sessions:

Model based Interference-aware Management System for Microservice Architectures

Joydeep Mukherjee, Marin Litoiu, Cornel Barna

The performance of microservices deployed in containers on a cloud platform can often be impacted due to interference. Interference is caused by other applications that run concurrently with the microservice containers on the same host and compete for host level shared resources. We build a model based management system that takes into account the effect of interference and works in parallel with existing platforms such as IBM Cloud Automation Manager for efficient deployment of microservices.

Modeling AUTOSAR Implementations in Simulink

Jian Chen

AUTOSAR defines the automotive three-layered software architecture. Application layer specifies functional behaviours and are realized as a set of tasks. Simulink is employed to address AUTOSAR’s lack of support for modelling behaviours of tasks. Simulink simulations assume tasks are completed in zero execution time, while real executions require a finite execution time. We present a Simulink block that can schedule tasks with a non-zero simulation time to enable a more realistic simulation.

Data Science 3 - Student Success Projects

Marcellus Mindel, Jennifer Collins, Dennis Buttera

In the third of three exhibits on Data Science Education, we will discuss how students at three different institutions are using Watson Studio to analyze institutional data to improve student retention and satisfaction. This approach is intended to create a pattern for supporting a wide variety of student capstone projects involving IBM tools and practices. Drop by to learn more about tools like Watson Studio and practices like Enterprise Design Thinking and Outcome Delivery.

Accessibility in the DevOps Era

Ali Asghar, Larry Lindsay, Thomas Brunet, Babz Kukoyi

As enterprises adopt continuous deployment practices, the DevOps testing practice must adapt accessibility testing tools and processes or user experience will suffer. These tools must work hand in hand with your application DevOps methodology - both required by & enabling those practices. In this hands-on demo, the audience will learn about IBM’s cutting-edge accessibility DevOps tooling and see how simple it is to integrate the tools into an existing continuous integration pipeline (CI).
Space Sentry Challenge: A game built with IoT devices and Microservices

Frank Ji, Prashanth Gunapalasingam

Space Sentry Challenge is a fun arcade game that allows players to take control of a 3D printed spaceship to shoot down pop-up targets. This game is built for the purpose of showcasing the application of Microservices in an IoT context. It serves as a good sample project for anyone who wants to build IoT projects with Microservices. It uses popular IoT kits like Arduino, ESP wifi modules and Raspberry Pi with Microservices Java applications running on Open Liberty.

Comparative Analysis of Cloud Console Systems

Julia Rubin, Satish Sarraf, John Steinbacher, Anthony W Erwin

Cloud Console systems feature a customizable presentation layer that allows users to manage their cloud resources, such as VMs, storage, etc. Designing such systems is challenging due to the demand to (a) handle massive amounts of data, (b) integrate with different resource providers, and (c) deliver data from these providers in near real-time. In this work, we analyze several Cloud Consoles, including that of IBM Cloud, and discuss design decisions and trade-offs made to satisfy these demands.

REST APIs with LoopBack 4 and OpenAPI 3

Taranveer Virk, Diana Lau, Biniam Admikew, Janny Hou

REST APIs power web applications, mobile applications, IoT devices & more. LoopBack 4 is an API creation framework that has been rewritten from the ground up and features a new programming interface, OpenAPI support, Dependency Injection, a new CLI and so much more. LoopBack makes it easy to create powerful OpenAPI 3.x (an industry standard) compliant REST APIs in minutes. It supports connecting to multiple SQL and NoSQL databases. Poster will show how to get started in minutes with LoopBack 4.

Interactive Visualization to Engage End-users in Prescriptive Analysis

Aindrila Basak

The research proposes a Visual Analytics System that can help end-users of Data Programming to visualize and understand patterns in input datasets and to write better labeling functions for generating labels. The proposed approach aims at enhancing the end-users’ trust in predictive models and helping with the conversion of the predictions into actions. The design requirements of this work were identified by performing semi-structured interviews with end-users, data analysts, and domain experts.

Finding Missing Labels for Large Industrial Datasets

Mona Nashaat

Obtaining labelled training datasets is a common challenge in the business domain. It costs a considerable amount of time and domain experience. We present a new approach to generate labelled datasets, in which we employ active learning in the data programming process. The proposed method aims at improving the accuracy of the training data. It optimizes the user engagement with the labelling process along with the annotation cost. It helps the users to trust the predictions of the final model.
Synthesizing Realistic Data for Machine Learning

DAVID D.COSTA

This exhibit is to show several techniques to generate large volumes of synthetic data required by machine learning algorithms. The aspects explored will show the need and how to build proper relationships in attribute features using copulas to assist in testing machine learning algorithms. We will show ways to efficiently build and output large datasets in the order of 5 Million clients. In addition, a way to scale dataset creation in a parallel process across several virtual machines.

Watch Phishing Evolution Over time

Qian Cui, Sophie Le Page, Emad Badawi, Ali Almokhtar

We proposed a model to monitor and analyze the phishing evolution. By using our model, we could observe the evolution pattern over different phishing groups.

Easy Cloud-Native Development with Microclimate

Elson Yuen, John Collier, Sakib Hasan, Stephanie Cao

Microclimate is a brand-new development environment that offers a end-to-end development experience for Microservices from local or IBM Cloud private cluster. With Microclimate, you can create or import Java, Node.js, or Swift applications, and using any editor of your choosing to develop applications. We will demonstrate Microclimate’s features on Iterative Development, DevOps pipeline and Intelligent Feedback, and will show how the tools will help developers to accelerate their development.

A Machine Learning Approach to Test Case Prioritization

Francis Palma

We propose to use the defect-proneness value of source files as the essential measure for test case prioritization. Our proposed test case prioritization model comprises two stages: (1) Metric extraction and defect prediction and (2) Test case rescheduling.

Optimization Driven Inlining

Erick Ochoa

Inlining is a compiler optimization that should be selectively applied in JIT compilation. Inlining too much increases compilation time and inlining too little decreases throughput. Different measures (like invocation count and method size) have been used to decide which methods to inline. Our proposal is to use abstract interpretation to determine which optimizations are feasible and use this calculation to drive inlining.

Semantic-Aware Disambiguation for Entity Resolution

Yu Huang

As data is integrated, duplicates inevitably occur in the integrated instance. Existing similarity functions often rely on string similarity to identify duplicates, which is insufficient in capturing the semantic meaning of terms. We propose a semantic-aware deduplication framework which can identify the duplicates by extracting the semantic information to measure ‘closeness’ of entities.
Create data-driven visual reports in Reactive Reports, a framework built for big data

Tereza Nedelescu, Mihai Nicolae, Samar Sajnani

This expo will showcase Reactive Reports, an open source framework for generating data-driven reports. The framework can use Scala and Java code to take input from mixed sources, such as database records, objects, or streams. The generated output is one or more PDF reports that can be used to summarize or visualize data. Reactive Reports scales well for large amounts of input data due to its memory efficient design, the use of Akka streaming, and its internal data storage architecture.

OptImatch: Semantic-Web System for Automated Query Tuning

Guilherme Damasio, Alexandar Mihaylov

OptImatch is a system which discovers problem patterns automatically by matching them to templates of plans from a previously populated knowledge base. The system extends upon our previous work, but now alleviating all need for manual expert input by introducing a customizable knowledge base capable of learning problem patterns from previously ran execution plans. The system output is a query with the embedded optimization guideline which enforces the optimal execution plan to be applied.

Third Party Container Tests

Mesbah Alam

This poster session depicts the '3rd party container tests' that we curate and run at the AdoptOpenJDK project. The IBM Runtime technologies team that produces the 'Eclipse OpenJ9' product uses many test materials to verify the SDK binaries. In case of 3rd party application tests, we run a variety of Java applications inside Docker containers. This poster session will describe the test automation techniques we use to execute, triage and debug application tests inside Docker.

Runtime Automatic Parallelization of JVM Applications

Akihiro Hayashi, Gita Koblents

In this IBM-Rice University collaboration, we push the bounds on integrating JVM applications with GPU accelerators. This work uses advanced code generation and performance prediction techniques to effectively and automatically offload JVM applications including Apache Spark to hardware accelerators.

Variability Aware analysis using Clang

Jacob Reckhard

Tool-based analysis is important to catch errors before the codebase becomes too dependent on the erroneous parts. One well known project for compilation and analysis of code is Clang. Here, we describe the changes required to coerce Clang into conducting variability-aware analysis of C++ code. Variability-aware analysis is a technique in which the codebase is analysed for all possible configurations of the code.
Approximation Algorithms for Instruction Scheduling and Rapid Prototyping in Coconut

Curtis D’Alves, Bill O’Farrell

Instruction scheduling is an NP-complete problem which allows for altering the execution order of instructions in a function without altering the function’s semantics. We present a continuous optimization based model for solving near-optimal schedules and a prototype implemented for the IBM z13 and z14 architectures and used to generate fast schedules for common cryptographic functions used in Blockchain operations implemented in GOLang assembly.

Static Dataflow Analysis for OpenMP Code Sanitization

Prithayan Barua, Jun Shirako, Whitney Tsang, Vivek Sarkar

OpenMP provides high-level abstractions and rich functionality for GPU programming. However, understanding the complex semantics of various combinations of pragmas is not trivial and leads to their wrong usage. We present a static analysis technique to detect incorrect usage of OpenMP target data mapping. Assuming the correctness of underlying sequential programs, our approach detects the OpenMP errors by comparing the data flow information of the original OpenMP version with the sequential.

Plasma Focused Ion Beam Curtaining Artefact Correction by Fourier-Based Linear Model

Christopher Schankula, Christopher K. Anand

This exhibit describes one example of a multidisciplinary partnership between the McMaster Computing & Software and Materials Science & Engineering departments to better solve modern Materials Science problems computationally. Specifically, we describe a software package for reducing artefacts in focused ion beam scanning electron microscope (FIB-SEM) images. We will describe and provide a live demonstration of the linear optimization model used to reduce the curtaining artefacts.

FLASH: Multithreaded Fault Localization

Luisa Rojas Garcia, Jeremy Bradbury

Multithreaded programs can have considerable performance benefits over sequential programs. However, this often comes at a cost with respect to program understandability as well as testing and debugging. To address these challenges we have developed an automatic fault localization tool for multithreaded Java programs (FLASH). Our tool utilizes a combination of code injection and a heuristic search algorithm to identify high probability code blocks.
<table>
<thead>
<tr>
<th>ROOM</th>
<th>WORKSHOP</th>
<th>FORMAT</th>
<th>CHAIRS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Holly-Butternut</td>
<td>Build a cognitive serverless Slack app with IBM Cloud Functions &amp; IBM Watson API</td>
<td>HandsOn</td>
<td>Serjik Dikaleh, Eric Charpentier, John Liu, Neil DeLima, Vince Yuen</td>
</tr>
<tr>
<td>Orchid</td>
<td>Build better APIs with the next generation of API testing and monitoring</td>
<td>HandsOn</td>
<td>Ivy Ho, JJ Tang, Jisoo Lee, Amirali Jafarian, Peter El-koreh</td>
</tr>
<tr>
<td>Violet</td>
<td>Come learn how to deploy Open Liberty applications using Docker, Kubernetes, Helm and MicroProfile!</td>
<td>HandsOn</td>
<td>Arthur De Magalhaes, Leo Christy Jesuraj</td>
</tr>
<tr>
<td>Conf Center 1</td>
<td>13th Workshop on Challenges For Parallel Computing</td>
<td>Speakers</td>
<td>Jeeva Paudel, Wael Yehia, Jeremy Bradbury</td>
</tr>
<tr>
<td>Conf Center 2</td>
<td>Building AI models using IBM Watson Studio</td>
<td>HandsOn</td>
<td>Sarah Packowski, Wendy Switzer</td>
</tr>
<tr>
<td>Jasmine</td>
<td>IBM Academic Skills Academy - Syllabi and Other Things</td>
<td>Speakers</td>
<td>Stephen Perelgut, Dennis Buttera, Colette Lacroix, Lila Adamec</td>
</tr>
<tr>
<td>Evergreen</td>
<td>Best Practices and Lessons Learned in Microservices</td>
<td>Panel</td>
<td>Julia Rubin, Yingying Wang, Harshavardhan Kadiyala, John Steinbacher, Tony Erwin</td>
</tr>
<tr>
<td>Primrose</td>
<td>Data-driven medicine: promise and challenges</td>
<td>Panel</td>
<td>Tomas Tokar, Igor Jurisica</td>
</tr>
<tr>
<td>Elm 2</td>
<td>The 3rd International Workshop on Dew Computing</td>
<td>Speakers</td>
<td>Yingwei Wang, Karolj Skala</td>
</tr>
<tr>
<td>ROOM</td>
<td>WORKSHOP</td>
<td>FORMAT</td>
<td>CHAIRS</td>
</tr>
<tr>
<td>--------------</td>
<td>---------------------------------------------------------------------------</td>
<td>---------</td>
<td>------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Holly-Butternut</td>
<td>Modernize digital applications with Microservices management using the Istio service mesh</td>
<td>HandsOn</td>
<td>Ozair Sheikh, Serjik Dikaleh, Dharmesh Mistry, Darren Pape, Chris Felix</td>
</tr>
<tr>
<td>Orchid</td>
<td>Hands-On: EasyMicroservices Application Development with Microclimate</td>
<td>HandsOn</td>
<td>Elson Yuen, Eric Peters, Rajiv Senthilnathan, Maysun Jamil Faisal, Steven Hung</td>
</tr>
<tr>
<td>Violet</td>
<td>IBM Voice Agent with Watson</td>
<td>HandsOn</td>
<td>Alice Yeung, Rick Chen, Philip Kurowski, Trevor Crawford, Meswan Bhaugerutty</td>
</tr>
<tr>
<td>Conf Center 1</td>
<td>The 10th CASCON Workshop on Cloud Computing</td>
<td>Speakers</td>
<td>Marin Litoiu, Joe Wigglesworth</td>
</tr>
<tr>
<td>Conf Center 2</td>
<td>Blockchain Fundamentals and Development Platforms</td>
<td>HandsOn</td>
<td>Omar Badreddin</td>
</tr>
<tr>
<td>Jasmine</td>
<td>Third Annual Workshop on Data-Driven Knowledge Mobilization</td>
<td>Panel</td>
<td>Kelly Lyons, Eleni Stroulia, Marcellus Mindel</td>
</tr>
<tr>
<td>Evergreen</td>
<td>The Best of IBM Innovation: Advancements through Overcoming Technological Uncertainties</td>
<td>Speakers</td>
<td>Jerrold Landau, Perry Fuller</td>
</tr>
<tr>
<td>Primrose</td>
<td>Large-Scale Multilevel Streaming Data Analytics</td>
<td>Speakers</td>
<td>Farhana Zulkernine, Haruna Isah</td>
</tr>
<tr>
<td>Elm 2</td>
<td>CASCON Workshop on Developing Big Data Applications and Services - BDAS 2018</td>
<td>Speakers</td>
<td>Darlan Arruda, Nazim H. Madhvji, Colin Taylor</td>
</tr>
<tr>
<td>ROOM</td>
<td>WORKSHOP</td>
<td>FORMAT</td>
<td>CHAIRS</td>
</tr>
<tr>
<td>--------------</td>
<td>-----------------------------------------------------------------------------------------------</td>
<td>--------</td>
<td>-----------------------------------------------</td>
</tr>
<tr>
<td>Holly-Butternut</td>
<td>Introduction to the IBM Q experience and Quantum Computing</td>
<td>HandsOn</td>
<td>Mehdi Bozzo-Rey, Robert Loredo</td>
</tr>
<tr>
<td>Orchid</td>
<td>Building Microservices in a Cloud-Native World using Eclipse MicroProfile and Open Liberty</td>
<td>HandsOn</td>
<td>Eveline Cai, Gilbert Kwan, YK Chang, Panagiotis Roubatsis</td>
</tr>
<tr>
<td>Violet</td>
<td>Refine, restructure and make sense of data visually, using IBM Watson Studio</td>
<td>HandsOn</td>
<td>Serjik Dikaleh, Darren Pape, Dharmesh Mistry, Chris Felix, Ozair Sheikh</td>
</tr>
<tr>
<td>Conf Center 1</td>
<td>Compiler-Driven Performance Workshop</td>
<td>Speakers</td>
<td>Gennady Pekhimenko, Ettore Tiotto</td>
</tr>
<tr>
<td>Conf Center 2</td>
<td>Practical Machine Learning with Python on DSX</td>
<td>HandsOn</td>
<td>Shaikh Quader, Mark Ryan, Eric Dong</td>
</tr>
<tr>
<td>Jasmine</td>
<td>2nd Workshop on DevOps and Software Analytics for Continuous Engineering and Improvement</td>
<td>Speakers</td>
<td>Konstantinos Kontogiannis, Chris Brealey, Alberto Giammaria, Brian Countryman, Marios-Stavros Grigoriou</td>
</tr>
<tr>
<td>Evergreen</td>
<td>Distributed Ledgers and Blockchain: Concepts and Applications</td>
<td>Speakers</td>
<td>Asic Chen, Arno Jacobsen</td>
</tr>
<tr>
<td>Primrose</td>
<td>2nd Workshop on Advances in Open Runtime Technology for Cloud Environments</td>
<td>Speakers</td>
<td>Daryl Maier, Kenneth Kent</td>
</tr>
<tr>
<td>ROOM</td>
<td>WORKSHOP</td>
<td>FORMAT</td>
<td>CHAIRS</td>
</tr>
<tr>
<td>----------------</td>
<td>-----------------------------------------------------------------</td>
<td>---------</td>
<td>-------------------------------------</td>
</tr>
<tr>
<td>Holly-Butternut</td>
<td>Introduction to the IBM Q experience and Quantum Computing</td>
<td>HandsOn</td>
<td>Mehdi Bozzo-Rey, Robert Loredo</td>
</tr>
<tr>
<td>Orchid</td>
<td>IBM Security Guardium Analyzer Bootcamp</td>
<td>HandsOn</td>
<td>Devan Shah, Larry Lindsay, Josue Diaz, Sagi Shechter, Andy Becher</td>
</tr>
<tr>
<td>Violet</td>
<td>Deriving Client Insights in the Financial Sector</td>
<td>HandsOn</td>
<td>Diane Reynolds, David Dcosta, David Xie, Seacy Zhen</td>
</tr>
<tr>
<td>Conf Center 1</td>
<td>Compiler-Driven Performance Workshop</td>
<td>Speakers</td>
<td>Gennady Pekhimenko, Ettore Tiotto</td>
</tr>
<tr>
<td>Conf Center 2</td>
<td>Practical Machine Learning with Python on DSX</td>
<td>HandsOn</td>
<td>Shaikh Quader, Mark Ryan, Eric Dong</td>
</tr>
<tr>
<td>Jasmine</td>
<td>iCity - Big Data and Visualization Urban Transportation Strategies</td>
<td>Speakers</td>
<td>Sara Diamond</td>
</tr>
<tr>
<td>Evergreen</td>
<td>Distributed Ledgers and Blockchain: Concepts and Applications</td>
<td>Speakers</td>
<td>Asic Chen, Arno Jacobsen</td>
</tr>
<tr>
<td>Primrose</td>
<td>2nd Workshop on Advances in Open Runtime Technology for Cloud Environments</td>
<td>Speakers</td>
<td>Daryl Maier, Kenneth Kent</td>
</tr>
</tbody>
</table>
Workshops
Monday, October 29th, 2018

Afternoon 03:15PM – 05:15PM
Build a cognitive serverless Slack app with IBM Cloud Functions & IBM Watson API
Location: Holly-Butternut
Theme: Cloud Computing
Prereq:
- Some JavaScript and Node.js knowledge
- Sign up for an IBM Cloud and a Slack account
Format: HandsOn
Level: Intermediate
Slack is an easy use collaboration tool to serve as a digital communication hub in many companies and teams. Based on Apache OpenWhisk, IBM Cloud Functions is a functions-as-a-service (FaaS) programming platform for developing lightweight code that scalably executes on demand.

This workshop will teach the audience to build a Slack app by implementing several serverless IBM cloud functions and integrating them into slack channels through Slack Events API. The application will also leverage IBM Watson APIs to have a Slack chatbot to chat with users to provide cognitive services within a demonstrated knowledge domain.

Build better APIs with the next generation of API testing and monitoring
Location: Orchid
Theme: Data and Analytics
Chairs: Ivy Ho, JJ Tang, Jisoo Lee, Amirali Jafarian, Peter El-koreh
Prereq: none
Format: HandsOn
Level: Beginner
May need own laptop with https://www.ibm.com/cloud/api-connect/api-test signed up before joining the workshop. Proliferation of APIs across all facets of life continued to explode and grow. The quality of the APIs and the data retrieval become a critical factor. In this workshop, we will walk you through a new no-code way of validating the API quality. How you can validate the API payload accuracy. How you can get new insights into API data from real business use-cases in different sectors. You will witness the innovation of the API test and monitor approach in this workshop.

Come learn how to deploy Open Liberty applications using Docker, Kubernetes, Helm and MicroProfile!
Location: Violet
Theme: Cloud Computing
Chairs: Arthur De Magalhaes, Leo Christy Jesuraj
Prereq: none
Format: HandsOn
Level: Beginner
Application modernization is in everyone’s mind - but what environment do you migrate your legacy application into? Can that environment also host your new cloud-native applications?

In this hands-on lab you’ll learn about how to leverage Open Liberty’s Docker container from Open Source to package your
applications (legacy or cloud-native) into a container and store them in a private, on-premises Docker registry.

You will then use IBM's Open Source Helm charts to deploy these applications (plus a database!) into Kubernetes using production-grade settings such as auto-scaling and health monitoring.

Lastly, you will see how MicroProfile OpenAPI can help your containerized microservices expose their REST APIs and enable an internal API economy between legacy and cloud-native applications.

The environment for this lab will be IBM Cloud Private - a production-ready kubernetes platform.

## 13th Workshop on Challenges For Parallel Computing

**Location:** Conf Center 1  
**Theme:** Systems  
**Chairs:** Julia Rubin, Yingying Wang, Harshavardhan Kadiyala, John Steinbacher, Tony Erwin  
**Prereq:** none  
**Format:** Speakers  
**Level:** beginner  

Parallel computing has expanded significantly over the past decade and now includes the development of applications for multi-core systems, distributed systems and heterogeneous systems. The goals of this workshop are to bring together different groups from the parallel community (application developers, language developers, compiler and tools developers, system architects and academic researchers) to explore the current challenges that parallel computing faces and present ideas on how to deal with these challenges.

## Building AI models using IBM Watson Studio

**Location:** Conf Center 2  
**Theme:** Cognitive Computing  
**Chairs:** Sarah Packowski, Wendy Switzer  
**Prereq:**  
1. You MUST bring your own laptops for the hands-on exercise  
2. Set up Watson Studio and Watson Knowledge Catalog on IBM Cloud: https://dataplatform.cloud.ibm.com/registration/stepone  
**Format:** HandsOn  
**Level:** Beginner  

In this workshop you will train AI models that process different types of sample input data, including: structured data (tabular) and unstructured data (images and sound.) Then you will see how to integrate those models into fun, sample apps that solve everyday challenges. With our help, you will create models in different ways using a variety of tools in IBM Watson Studio.

About Watson Studio: IBM Watson Studio provides a range of tools to help you train AI models - from graphical tools that guide you, step-by-step, in choosing machine learning algorithms to notebooks where you can construct complex neural networks by hand. Watson Studio simplifies AI development, whether you are looking for tools to make learning AI easier, tools to rapidly prototype your AI inventions, or a cost-effective, powerful platform for AI research or enterprise AI solutions. See more: https://medium.com/ibm-watson/introducing-ibm-watson-studio-e93638f0bb47

## IBM Academic Skills Academy - Syllabi and Other Things

**Location:** Jasmine  
**Theme:** Other  
**Chairs:** Stephen Perelgut, Dennis Buttera, Colette Lacroix, Lila Adamec  
**Prereq:** This workshop is intended primarily for academic faculty and administrators  
**Format:** Speakers  
**Level:** Beginner  

IBM has initiated a program to "Teach the Teacher" how to use the latest technologies. This session details 4 of the most requested topics, giving the complete syllabus for the 40hr course as well as highlights from the available badges for: Explorer, Mastery and Instructor.

Attendees will leave with a clear idea of what they can expect from the Skills Academy program and how they can learn materials to pass on to others.
Topics will include:
- Business Intelligence Analytics
- Mobile Application Development and IoT
- Blockchain and Design Thinking for Blockchain
- Quantum Computing

Best Practices and Lessons Learned in Microservices

**Location:** Evergreen  
**Theme:** Cloud Computing  
**Chairs:** Julia Rubin, Yingying Wang, Harshavardhan Kadiyala, John Steinbacher, Tony Erwin  
**Prereq:** none  
**Format:** Panel  
**Level:** Beginner

Microservice-based architecture is an approach to developing a single application as a suite of independent services. The services run in separate processes and communicate with each other via lightweight language-agnostic protocols, such as HTTP REST. The services are split following business capabilities; each service has a fully automated pipeline and is independently deployable.

Microservices aim at shortening the development lifecycle while improving the quality, availability, and scalability of applications at runtime. From the development perspective, cutting one big application into small independent pieces reinforces the component abstraction, and makes it easier for the system to maintain clear boundaries between components. At runtime, microservices can be individually scaled by adding more instances of those microservices that experience increasing traffic.

Due to these advantages, microservice-based architectures are now becoming increasingly popular in industry. Examples of companies that have been using microservices include Amazon, Netflix, IBM, Uber, LinkedIn, Groupon, and eBay. Yet, adopting microservice-based architectures and implementing it “right” is no a trivial endeavour. Just “jumping on the microservices trend” and expecting that the transition itself, together with the adoption of advanced technology, such as Docker and Kubernetes, will allow companies to achieve significant improvements is a false belief.

In this workshop, we intend to explore best practices, lessons learned, and technical challenges practitioners face when adopting and implementing microservices. These include considerations for identifying the right service granularity and topology, issues related to synchronization and constancy, security of microservices, performance debugging, efficient monitoring and troubleshooting, and more.

Our goal is to gather researchers and practitioners interested in exchanging ideas on the topic. For practitioners, the workshop will provide venue to learn from each other, borrow successful ideas, and avoid common mistakes. For researchers, a description of current practices and challenges practitioners face can inspire novel software engineering methods and techniques.

The workshop will be structured as a series panel discussions and invited talks by participants from industry and academia.

Data-driven medicine: promise and challenges

**Location:** Primrose  
**Theme:** Data and Analytics  
**Chairs:** Tomas Tokar, Igor Jurisica  
**Prereq:** none  
**Format:** Panel  
**Level:** Beginner

In the last few years, use of machine learning (ML) and artificial intelligence (AI) in the sector of healthcare started to gain broader acceptance. These technologies may revolutionize medicine by improving diagnostic accuracy and increasing therapeutic efficiency. The successful application of these technologies requires a constant circulation of a large amounts of data from patients, through healthcare professionals, to scientists and software developers. The data are not only required for the initial training and testing of the algorithms, they are also essential for monitoring algorithms’ performance once deployed in a clinical practice. Nowadays medical data may include range of modalities. These include genetic profiles, medical imaging records, data from wearable devices, clinical findings and various socioeconomic characteristics of patients. This poses several technical and ethical challenges, addressing which requires introduction of a novel technologies and development of a new healthcare policies. This is only possible through broad communication between the medical professionals, computer science experts and information privacy specialists.
The 3rd International Workshop on Dew Computing

**Location:** Elm2  
**Theme:** Cloud Computing  
**Chairs:** Yingwei Wang, Karolj Skala  
**Prereq:** none  
**Format:** Speakers  
**Level:** Beginner

DEWCOM is an annual international workshop on dew computing. The first one, DEWCOM 2016, was held in Charlottetown, Canada. The second one, DEWCOM 2017, was held in Opatija, Croatia. The third one, DEWCOM 2018, is generously sponsored by IBM Centre for Advanced Studies and CASCON 2018 and will be held together with CASCON 2018. The details of DEWCOM 2018 can be found in [http://www.dewcomputing.org/index.php/dewcom-2018/](http://www.dewcomputing.org/index.php/dewcom-2018/).

Dew computing is a new post-cloud computing model appeared in 2015. While cloud computing uses centralized servers to provide various services, dew computing uses on-premises computers to provide decentralized, cloud-friendly, and collaborative micro services to end-users.

Dew computing is an on-premises computer software-hardware organization paradigm in the cloud computing environment, which does not contradict with cloud computing, does not replace cloud computing, but it is complementary to cloud computing. The key features of dew computing are that on-premises computers provide functionality independent of cloud services and they also collaborate with cloud services. Briefly speaking, dew computing is a better way of using local computers in the age of cloud computing.

In this workshop, research progress in dew computing will be presented. Ideas and future directions will be discussed.

This workshop will have 5 sessions. The schedule of this workshop is: Session 1: Oct. 29, 3:15 – 5:15pm. Session 2: Oct. 30, 8:30 – 10:00am. Session 3: Oct. 30, 10:15am. – 12:00pm. Session 4: Oct.30, 1:00 – 3:00pm. Session 5: 3:15 – 5:15pm. Here we introduce the content of the first session.

This session will include “Dew Computing Tutorial” and a presentation: “Dewblock: A Blockchain System Based on Dew Computing.”

In the tutorial, we will focus on the following aspects: What is dew computing? What are the features of dew computing? Where can dew computing be applied to?

The presentation is about an application of dew computing to the blockchain technology. Blockchain is great and has huge potentials, but the size of a blockchain is always increasing. This will eventually cause problems for the use of blockchains. In this presentation, a new kind of blockchain system based on dew computing, Dewblock, will be introduced. The major feature of this new blockchain is that the data size of a client is very small and the features of a full node are still kept. This approach opens the door for the blockchain technology to be widely used in personal computers and mobile devices.
Tuesday, October 30th, 2018
Afternoon 03:15PM – 05:15PM

Modernize digital applications with Microservices management using the
Istio service mesh

Location: Holly-Butternut
Theme: Cloud Computing
Chairs: Ozair Sheikh, Serjik Dikaleh, Dharmesh Mistry, Darren Pape, Chris Felix

Prereq:
- Intermediate Kubernetes knowledge
- A free trial IBM Cloud (Bluemix) account
- Minimal familiarity with Linux command line

Format: HandsOn
Level: Intermediate

Digital solutions are being built on modernized enterprise platforms deployed on cloud infrastructure and managed using container platforms. Foundational infrastructure capabilities such as load balancing and routing, previously available as software are now being provided as part of the underlying cloud platform. When designing your next generation architecture, its integral to understand the capabilities available from the cloud platform versus acquiring / developing it with software. For example, load balancing and automatic scaling are features that are built-into container orchestration platforms such as Kubernetes; therefore, you should not expect your applications to develop these capabilities, rather write your applications in a manner that allows you to embrace the container platform.

These key application design principals are based on API / Microservices architecture, where business functions are packaged and deployed within containers and communicate with each other using API interfaces. As the number of microservices grow, the need to manage the interactions and provide key runtime capabilities becomes a critical requirement for success ... lets explore why the service mesh is the right architecture for microservices-based applications.

A service mesh is an infrastructure layer for controlling container (ie microservice-to-microservice) traffic in microservices-based applications. Each container (ie microservice) is deployed together with a separate “sidecar” proxy, which interacts with a “control plane”, enforcing access between microservices. The service mesh provides a clear boundary between runtime operations and microservices functionality. It standardize the runtime operations using a declarative approach, so you can write policies to enforce runtime behaviour without developing any code. For example, lets explore the circuit breaker pattern. This pattern helps prevent failure for your entire application when a single service / component is unresponsive. Netflix Hystrix is a popular library used within Java applications to provide circuit breaker functionality. The challenge with using a shared library is that it gets embedded in code and becomes difficult to manage when code changes need to be made; instead, using an out-of-process proxy (ie sidecar) allows your microservice to add circuit breaker capability without modifying your application.

Istio is an open community project that implements the service mesh architecture. It is built on top of Kubernetes and provides an additional runtime layer that adds scalability, routing, A/B testing and more. It allows you to inject a “side proxy” into an existing kubernetes pod without changing your application, reducing the friction for adoption. You automatically gain the benefits of telemetry, security and circuit breaking without writing code or modifying configuration.

In this workshop, you will learn how to deploy an single-page application (SPA) built with API/Microservices design principals into the Istio service mesh. You will get hands-on experience in configuring Istio-based policies to manage microservices interaction (ie service routing) and improve infrastructure resiliency (ie circuit breaker) without writing application code. The overall application resiliency is enhanced when you deploy your application within the Istio service mesh.
Hands-On: Easy Microservices Application Development with Microclimate

Location: Orchid
Theme: Cloud Computing
Chairs: Elson Yuen, Eric Peters, Rajiv Senthilnathan, Maysun Jamil Faisal, Steven Hung
Prereq: Basic knowledge on Java and JavaScript is recommended.
Format: HandsOn
Level: Beginner

Microclimate is a brand-new, cloud native development environment that offers a complete, end-to-end development experience for Microservices. Since Microclimate has been designed with a focus on containerization, it can run anywhere from your local laptop, to an IBM Cloud private cluster.

With Microclimate, you can create or import Java, Node.js, or Swift applications into the development environment, and using any editor of your choosing, you can quickly start development on your application in a containerized environment. Through a process called Rapid Iteration, Microclimate will quickly detect any changes that occur in your project and determine the minimal and best course of action to update your application. From there, using our integrated DevOps pipeline, you can deploy your application with Jenkins to a live ICP cluster. With these features, Microclimate offers a fully featured development experience that many other environments don’t offer today.

During the hands-on workshop, we will give you an introduction to Microclimate, starting from product installation to write Microservices applications to run on Microclimate in a Docker environment. You will get hands on experiences to create new applications and import existing applications into Microclimate. For developers, a crucial part of the development cycle is the ability to quickly develop and test applications changes on a running application. The develop-deploy-test-repeat cycle must be as short as possible in order to prevent lost developer productivity due to deployment downtime. You will be given the opportunity to experience this rapid iterative development support by developing Java and JavaScript applications in this workshop.

Finally, during the workshop we will introduce the integrated DevOps pipeline functions provided that allows you get into production fast with a preconfigured DevOps pipeline and deploy application to IBM Cloud Private (ICP). We will also show you the diagnostic services that helps you to do problem determination in production.

IBM Voice Agent with Watson

Location: Violet
Theme: Cognitive Computing
Chairs: Alice Yeung, Rick Chen, Philip Kurowski, Trevor Crawford, Meswan Bhaugeerutty
Prereq: Attendees need to create trial Twilio and IBM Cloud accounts during the workshop. The accounts are free and no credit card information is needed.
Format: HandsOn
Level: Beginner

Cognitive chatbots are changing the way businesses are interacting with their customers. Whether embedded in a web page, or talking to users via a mobile application, Watson powered cognitive bots can resolve queries quickly and efficiently. To better leverage this powerful technology, businesses can use Voice Agent with Watson on IBM Cloud to quickly build Watson powered chat bots (voice agents) and connect them to the telephone network.

Through connecting Speech to Text, Text to Speech, and Watson Assistant, voice agents can identify what a caller is saying and to respond back in real time. As Watson Assistant can process natural language, it is able to converse with callers using complete sentences, helping to improve the experience for callers. Voice agents can act as a self-service solution to solve the most common calls entering call centers. By answering the main volume of calls with voice agents, businesses can resolve calls more effectively. This results in a reduction to the volume of calls sent to human agents.
The 10th CASCON Workshop on Cloud Computing

**Location:** Conf Center 1  
**Theme:** Cloud Computing  
**Chairs:** Marin Litoiu, Joe Wigglesworth  
**Prereq:** none  
**Format:** Speakers  
**Level:** Intermediate

The goal of the workshop was to bring together researchers and practitioners from government, industry and academia to present and share the best practices and research agendas at the intersection of Cloud Computing and Internet of Things: development, deployment, runtime management, quality of services and runtime models. We particularly focused on several main topics: cloud requirements for Internet of Things, deployment and adaptive runtime management, cognitive capabilities, security and privacy. DevOps plays an important role in the IoT and cloud ecosystems, providing the mechanisms that enable agile development and operations and it was a topic of the workshop. Application domains such as smart buildings and smart cities were illustrated.

This half-day workshop consisted in presentations and a panel. The presentations were structured along the main themes of the workshop. To encourage discussion and provide a more open discussion and perspective, we included a panel where industry and academic experts presented their visions and answered questions from the audience.

Blockchain Fundamentals and Development Platforms

**Location:** Conf Center 2  
**Theme:** Security  
**Chairs:** Omar Badreddin  
**Prereq:** none  
**Format:** HandsOn  
**Level:** Beginner

Blockchain is an emerging computing and development platform. It is a new paradigm that aims at empowering peers and eliminate the need for central authenticating authority. In essence, blockchain has introduced a novel level of distributed sovereignty. Crypto currency is one prominent outcome of this new computing paradigm that has gained broad attention. However, Blockchain paradigm has demonstrated broader potential impacts in many disciplines including secure software engineering, supply chain, banking, and peer to peer commerce.

This half day workshop will give participants a brief background on the technology along with hands on practice on development using some prominent open source blockchain platforms. The hands-on exercises will be led by experienced Blockchain developer.

The first talk in the workshop will introduce the fundamental Blockchain concepts. The next talk will provide an overview of current and emerging impacts of Blockchain technologies covering many disciplines and industries, including financial industry, supply chains, authentication and security, as well as the recent emergence of blockchain based distributed social networks.

After the first two talks, participants will engage with guided hands-on exercises to develop a basic blockchain application. Participants will be given the required software and will also be made available online.

By the end of the workshop, participants are expected to have gained in-depth understanding of the emerging Blockchain technology and its applications. Participants will also gain knowledge and skills on existing blockchain development platforms. Therefore, this workshop is ideal for researchers and practitioners alike who are considering using Blockchain in their research or work, as well as middle technical managers who want to gain understanding on how Blockchains and the emergence of distributed sovereignty may impact their current line of businesses. The workshop is useful for educators who wish to introduce Blockchain in their undergraduate and graduate courses.
Third Annual Workshop on Data-Driven Knowledge Mobilization

**Location:** Jasmine  
**Theme:** Data and Analytics  
**Chairs:** Kelly Lyons, Eleni Stroulia, Marcellus Mindel  
**Prereq:** none  
**Format:** Panel  
**Level:** Beginner

Knowledge mobilization and translation describes the process of moving knowledge from research and development (R&D) labs into environments where it can be put to use. There is increasing interest in understanding mechanisms for knowledge mobilization, specifically with respect to academia and industry collaborations. At the same time, the number of available datasets and accessible analysis tools is growing.

Building on the discussions and results presented at previous workshops, the third annual workshop on data-driven knowledge mobilization will bring together researchers, students, and industry partners to present results and discuss challenges associated with the analysis of datasets associated with knowledge mobilization. In order to understand the processes of knowledge mobilization, we need access to certain datasets and specific analyses techniques. We will present details of curated datasets and analysis techniques that support analysis of individuals and resources, expertise and work activities, and work outputs and innovations.

The goals of the workshop are to bring participants together to share results and outcomes and to discuss challenges and future directions. In this workshop, we will report on research underway as part of a Strategic Partnership Project on Data-Driven Knowledge Mobilization, Translation, and Innovation. The Strategic Partnership Project is in its final year. Since the start of the project, several people have participated in the research including six investigators, two postdoctoral fellows, eight Ph.D. students, five masters students, and fourteen industrial and partner scientists. One of the goals of the project is to produce a repository of datasets and analysis tools. The theme of this 3rd annual workshop focuses on tools to enable the analysis of datasets that will help in understanding the processes of knowledge mobilization.

There will be five presentations by researchers and students involved in the project followed by a panel discussion.

The Best of IBM Innovation: Advancements through Overcoming Technological Uncertainties

**Location:** Evergreen  
**Theme:** Data and Analytics  
**Chairs:** Jerrold Landau, Perry Fuller  
**Prereq:** none  
**Format:** Speaker  
**Level:** Beginner

There are many factors that spur innovation in the field of technology. One such factor is governmental support. The Canadian government provides such support through the SR&ED (Scientific Research and Experimental Development) tax credit program. Not surprisingly, IBM has availed itself of this tax credit for many years. The fundamentals of SR&ED are based on three pillars: a) identifying a technological uncertainty, b) overcoming the uncertainty through a scientific experimental process, c) leading to an advancement in technology. The advancement is often defined as the acquisition of new knowledge in the domain. It should be noted that the advancement need not be incorporated into a product, and indeed need not necessarily be successful in the classical business definition of the term. One advances technology just as well by proving that something is not feasible than by proving something is feasible. This definition of advancements in technology, while nuanced and subject to governmental SR&ED program regulations, can be applied in a most general sense to many fields of scientific innovation. It has been noted that both the IBM CAS organization and IBM’s participation in the SR&ED program promote innovation through an exploration of the proverbial ‘bleeding edge’ of technology. In this workshop, we will provide an overview of the technological criteria for participation in the SR&ED program, and present SR&ED as an indicator of technical vitality. This will be followed by presentations from several IBM Lab teams highlighting their SR&ED claims over the past several years. It is expected that the audience will come away with a new perspective on scientific innovation as applied to the field of software development.
Large-Scale Multilevel Streaming Data Analytics

Location: Primrose  
Theme: Data and Analytics  
Chairs: Farhana Zulkernine, Haruna Isah  
Format: Speakers  
Prereq: none  
Level: Beginner  

Motivation and Justification:

There is a monumental shift happening in how data powers organizational and business operations. This shift is about moving away from traditional batch data analytics to real-time and hybrid data analytics involving both static and continuous data to avoid delay in generating insights and storing massive amount of streaming data. A good number of analytics systems currently utilize stream processing without storing the data to quickly ingest, analyze and to correlate information as it arrives from thousands of real-time sources (devices, sensors, and applications). Such systems often provide real time dashboards and critical alerts, and therefore, are required to be fast, efficient, effective, scalable, and reliable.

In most cases stream processing is followed by batch processing for deeper analytical processing. Modern streaming analytic systems, therefore, try to unify batch and streaming analytics into a seamless data processing pipeline. A general architecture of a large-scale multilevel analytics system consists of (i) an ingestion mechanism at the front-end, (ii) streaming and batch data processing engines for data transformation, scoring, modelling of historical data, and real-time prediction, (iii) data storage units for persisting, indexing, searching, and knowledge management, (iv) resource management unit for the coordination of distributed compute and storage resources, and (v) visualization units to present results and knowledge for decision support.

Some of the deeper analytics of streaming data requires longer execution time and can choke the data processing pipeline. The stream plus batch analytics solves that problem. However, in our progression towards the Internet of Things (IoT), we will face serious computational and storage challenges in such an approach. Innovative solutions are needed to selectively store streaming data, enable near real time micro batch processing, and perform multi-level in-memory analytics. Large-scale multilevel analytics on a unified platform is increasingly gaining attention in the industry as it can potentially enhance business and operational decision making. However, it faces the following challenges, a) implementing an efficient front-end for ingestion and integration of massive data streams across the globe, b) combining streaming and in-memory data analytics, c) developing a knowledge management strategy to store, manage and link big data and distributed knowledge, and d) other challenges including cluster management, knowledge representation, and visualization. The above challenges make the development of methods, algorithms, and infrastructures for multilevel streaming analytics a challenging but interesting research problem.

Goals and Outcomes:

This workshop aims to provide a forum for researchers and industry practitioners to discuss new ideas and share their experiences in the areas of streaming data analytics. Participants will present their work on topics including methods, models, algorithms, infrastructures, quality issues, applications, and open problems for large-scale streaming data analytics. The workshop can serve as a guide for organizations and individuals planning to implement a real-time data stream processing and multilevel data analytics framework.

Workshop Structure:

The half-day workshop will feature invited talks by experts, practitioners, researchers, and industry partners working on massive streaming analytics research. There will be a time for discussion after each presentation to instigate the audiences to share their comments and views and ask questions to the speaker.
CASCON Workshop on Developing Big Data Applications and Services - BDAS 2018

Location: Elm2
Theme: Other
Chairs: Darlan Arruda, Nazim H. Madhavji, Colin Taylor
Prereq: none
Format: Speakers
Level: Intermediate

1. Background
Research from Gartner (2015) indicates that, in 2017, 60% of Big Data projects failed or did not provide the expected benefits. However, in November 2017, Nick Heudecker, a Gartner analyst, posted in his twitter account that they were too conservative. The Big Data project failure rate is now close to 85%. The reasons are not only related to technology itself. It is a mix of environmental, technological and managerial problems. Some of the reasons for Big Data projects failure are: At the project level: missing link to business objectives, lacking big data skills, relying too much on the data, failing to convince executives, and poor planning; At the technical level: Rapid technology changes, difficulty in selecting Big Data technologies to address the systems and project requirements, complex integration between new and old systems, computation of intensive analytics, and the necessity of high scalability, availability and reliability, to name a few. Further, our previous study has shown that there is approximately a 80:20 split in the industry focus in favour of “algorithms for analytics” and “infrastructure”, thereby shortchanging the aspects of creating and evolving “applications” and “services” concerned with Big Data.

2. Importance
The emerging data on project challenges or failure should be of immense concern to the Big Data software community. It calls for meeting of the minds to deliberate about, and share experiences concerning the development of Big Data applications and services. Both industry and academia representation are needed to cut through the barriers facing the community today. Working in isolation may prolong the pain and agony of challenges faced in the Big Data software community. In turn, the society at large is deprived of the potential benefits of Big Data applications and services. The proposed workshop comes at a critical juncture in the fast-emerging field of Big Data applications and services development. The workshop aims to be a catalyst in the movement on Big Data applications building and services creation. It will form a platform for participants, from both practice and research, to deliberate on, and achieve a deeper understanding of, the different activities, methods, tools, processes, system artefacts, constraints, conditions, etc., involved in Big Data projects.

3. Purpose and Interest
Given the described importance of the proposed workshop, the purpose of the workshop includes: (1) sharing Big Data project experiences among the participants, and identifying challenges in the design, implementation, deployment, and evolution of Big Data applications and services; (2) fostering a Big Data community of researchers and practitioners focused on applications and services, and (3) compiling an agenda for future research. All application domains are of interest in this workshop. In the workshop, we shall identify, debate about, and discuss solutions to, the barriers challenging the development, deployment, evolution, and success of Big Data Applications and Services.
Introduction to the IBM Q experience and Quantum Computing

**Location:** Holly-Butternut  
**Theme:** Other  
**Chairs:** Mehdi Bozzo-Rey, Robert Loredo  
**Prereq:**  
- bring their own laptops  
- have the anaconda python distribution installed (www.anaconda.com)  
- have an IBM Q experience account (quantumexperience.ng.bluemix.net)

**Format:** HandsOn  
**Level:** Beginner

IBM’s work in quantum computing started in the 1970s with the birth of quantum information theory and the first conference on Physics of Computation was co-hosted by IBM and MIT in 1981. The quantum foundations time, where quantum computing was the exclusive domain of scientists and theoreticians is past history. We are now in a quantum readiness phase where education, algorithm development and use case identification that may lead to demonstrations of quantum advantage are key.

In 2016 IBM made quantum computing capable devices available for the public at no cost in the cloud and then launched the IBM Q network in 2017. Programming a quantum computer is quite different from “classical programming” and is done with a circuit made of quantum gates that execute the quantum algorithm. For devices with a small number of qubits, a GUI driven interface can be used to build the circuit, making quantum programming available and viable for even high schoolers. For devices with more qubits, or more complex circuits a programming framework is needed.

In this hands-on workshop, we will review the basics of quantum computing, go through account creation on the IBM Q experience and basic use of the Composer (GUI driven interface), installation of the Open Source Qiskit and Qiskit AQUA frameworks, and how to execute simple quantum circuits on either a local quantum simulator or on a real quantum device that operates at a temperature colder than outer space.

Building Microservices in a Cloud-Native World using Eclipse MicroProfile and Open Liberty

**Location:** Orchid  
**Theme:** Cloud Computing  
**Chairs:** Eveline Cai, Gilbert Kwan, YK Chang, Panagiotis Roubatsis  
**Prereq:** none  
**Format:** HandsOn  
**Level:** Beginner

Eclipse MicroProfile is a set of open technologies to optimize enterprise Java for a microservices architecture. Open Liberty is the open source foundation of WebSphere Liberty, IBM’s strategic Java application server for a cloud-native world. Come and learn how you can easily build microservices with Eclipse MicroProfile and Open Liberty. Get your hands dirty with various aspects of building cloud-native applications, from foundation for RESTful services to what you need to manage many microservices and what you need for reliable operations.
Refine, restructure and make sense of data visually, using IBM Watson Studio

Location: Violet
Theme: Cognitive Computing
Chairs: Serijk Dikaleh, Darren Pape, Dharmesh Mistry, Chris Felix, Ozair Sheikh
Prereq: Must sign up for IBM Cloud account prior to the workshop (trial account/free tier is fine) - https://www.ibm.com/cloud/
Optional basic coding knowledge
Format: HandsOn
Level: Beginner

More than ever before, larger and more comprehensive data sets are being made publicly available on the internet. You can find data on all sorts of topic such as housing prices, sports data, wine reviews, weather, movies, TV shows, gun violence and anything you can think of. How does one make sense of all this data, and can you combine different data sets to get new insights for your needs?

In this workshop, we will begin by talking about where you can find open data sets and show some examples of how they have been used to gain insights. We will then take some sample data sets and explore it through IBM Watson Studio. We will further the workshop by creating visualizations of the data using both open source programming concepts and using tools available in IBM Watson Studio.

Compiler-Driven Performance Workshop Strategies

Location: Conf Center 1
Theme: Systems
Chairs: Gennady Pekhimenko, Ettore Tiotto
Prereq: none
Format: Speakers
Level: Intermediate

The workshop has a particular focus on (but not limited to):
Innovative compiler analysis, transformation, and optimization techniques
Languages, compilers, and optimization techniques for multicore processors and other parallel architectures
Compiling for streaming or heterogeneous hardware
Dynamic compilation for high-performance and real-time environments
Compilation, optimization, and analysis for dynamic languages
Compilation techniques for reducing power
Program safety
Whole system optimization and analysis
Tools and infrastructure for compiler research

Practical Machine Learning with Python on DSX

Location: Conf Center 2
Theme: Data and Analytics
Chairs: Shaikh Quader, Mark Ryan, Eric Dong
Prereq:
* Coding experience in any programming language
* Users MUST bring their own laptops for the hands-on exercise
* IBM Cloud id: https://console.bluemix.net/registration/
Format: HandsOn
Level: Beginner

In this FULL-DAY workshop, after a brief introduction to Machine Learning, we’ll take the students through the hands-on exercise of building a Machine Learning model from the scratch. They will learn and code in different phases of Machine Learning pipeline, including acquiring, cleaning, and exploring data; building and evaluating ML model. Finally, we’ll tell them how to build a discipline for continuous learning of ML and applying this to solving real problems.
2nd Workshop on DevOps and Software Analytics for Continuous Engineering and Improvement

**Location:** Jasmine  
**Theme:** Systems  
**Chairs:** Konstantinos Kontogiannis, Chris Brealey, Alberto Giammaria, Brian Countryman, Marios-Stavros Grigoriou  
**Prereq:** Knowledge on software development process and software engineering principles  
Knowledge on DevOps tools and frameworks  
Experience in software development and its life-cycle  
**Format:** Speaker  
**Level:** Intermediate

A key issue that emerges in the software engineering community is how to provide efficient DevOps tools and processes that facilitate continuous delivery and improvement, particularly in and for Cloud based environments where continuous delivery at speed with high quality can be crucial to business success.

This workshop aims to bring experts from industry and academia to discuss and debate the latest trends related to the design of frameworks that support DevOps practices of complex systems which are developed and evolved within a “Measure-Analyze-Assess-Act” loop. Such frameworks utilize software repositories, software analytics, process analytics, the quantification of technical debt as a failure risk predictor, and the system’s run-time behavior to dynamically assess deploy/no-deploy choices and achieve continuous deployment.

Distributed Ledgers and Blockchain: Concepts and Applications

**Location:** Evergreen  
**Theme:** Systems  
**Chairs:** Asic Chen, Arno Jacobsen  
**Prereq:** General computer science and computer engineering (information systems background as taught in Bachelor’s curricula)  
**Format:** Speakers  
**Level:** Beginner

Blockchain has been, without doubt, one of the hottest topics in technology in recent years. As is in the case of most “buzzwords”, most people, even software professionals, have no more than a surface comprehension of the technology. We hope that through this workshop, we can demystify blockchain and distributed ledgers, giving the attendees a working understanding while expanding on use cases far beyond the most popular: cryptocurrency. During this workshop, we will provide a tutorial-style introduction to various distributed ledgers and blockchain technologies. We will be focusing on first principles and algorithms while identifying emerging blockchain use cases.

2nd Workshop on Advances in Open Runtime Technology for Cloud Environments

**Location:** Primrose  
**Theme:** Cloud Computing  
**Chairs:** Daryl Maier, Kenneth Kent  
**Prereq:** None  
**Format:** Speakers  
**Level:** Intermediate

Modern language runtimes are complex, dynamic environments that involve a myriad of components that must work cooperatively to achieve the functional and performance requirements of a given language. Typical core runtime technologies include dynamic just-in-time compilers for performance, garbage collection for heap management, platform abstraction for ease of portability to different hardware and operating system environments, developer tooling for diagnosis and tuning of the various components, and interoperability between different language environments.

Cloud services such as IBM Cloud, Microsoft Azure, or Amazon Web Services (AWS) are increasingly becoming the environments where applications are developed and deployed, data is stored, and businesses are run. Many of the features that define a cloud (e.g., resiliency, elasticity, consistency, security) are realized through runtime technologies. Clouds are polyglot environments, and therefore advances in cloud development are directly driven by innovation in
runtime technologies. However, cloud environments pose unique, often conflicting demands on runtime systems that are often less of a concern in isolated systems. Throughput performance (how fast is my application?), density (how many instances of my application can I run simultaneously in my provisioned environment?), startup performance (how quickly can I launch a new instance of my application?), and language interoperability (how can my JavaScript code efficiently call a function in a Python module?) are all important considerations that require innovation to solve effectively.

The goal of this workshop is to bring together research, industry, and development communities to share and discuss innovations, challenges, and research across a broad set of open-source runtime technologies (such as Eclipse OMR, LLVM, Eclipse OpenJ9, Node.js) for cloud environments. The focus on open technology solutions rather than proprietary is key as it allows for greater collaboration amongst individuals, communities, researchers, and companies through shared learning on common technology.
Wednesday, October 31st, 2018
Afternoon 2:15PM – 5:30PM

Introduction to the IBM Q experience and Quantum Computing

Location: Holly-Butternut
Theme: Other
Chairs: Mehdi Bozzo-Rey, Robert Loredo
Prereq:
- It is assumed that participants will:
  - bring their own laptops
  - have the anaconda python distribution installed (www.anaconda.com)
  - have an IBM Q experience account (quantumexperience.ng.bluemix.net)
Format: HandsOn
Level: Beginner

IBM’s work in quantum computing started in the 1970s with the birth of quantum information theory and the first conference on Physics of Computation was co-hosted by IBM and MIT in 1981. The quantum foundations time, where quantum computing was the exclusive domain of scientists and theoreticians is past history. We are now in a quantum readiness phase where education, algorithm development and use case identification that may lead to demonstrations of quantum advantage are key.

In 2016 IBM made quantum computing capable devices available for the public at no cost in the cloud and then launched the IBM Q network in 2017. Programming a quantum computer is quite different from “classical programming” and is done with a circuit made of quantum gates that execute the quantum algorithm. For devices with a small number of qubits, a GUI driven interface can be used to build the circuit, making quantum programming available and viable for even high schoolers. For devices with more qubits, or more complex circuits a programming framework is needed.

In this hands-on workshop, we will review the basics of quantum computing, go through account creation on the IBM Q experience and basic use of the Composer (GUI driven interface), installation of the Open Source Qiskit and Qiskit AQUA frameworks, and how to execute simple quantum circuits on either a local quantum simulator or on a real quantum device that operates at a temperature colder than outer space.

IBM Security Guardium Analyzer Bootcamp

Location: Orchid
Theme: Security
Chairs: Devan Shah, Larry Lindsay, Josue Diaz, Sagi Shechter, Andy Becher
Prereq: IBM ID account already created
Format: HandsOn
Level: Beginner

General Data Protection Regulation (GDPR) requires organizations to implement adequate controls to protect personal and sensitive personal. A critical step in that journey is impact assessment; understanding where data (subject to GDPR) is located and how vulnerable it is. In this session, we will introduce IBM Security Guardium Analyzer, a new SaaS offering, which intends to help organizations easily complete the impact assessment for their databases. You will receive hands-on experience on quickly utilizing the Guardium Analyzer solution to locate GDPR data and determining risk on existing databases.
Deriving Client Insights in the Financial Sector

Location: Violet
Theme: Data and Analytics
Chairs: Diane Reynolds, DAVID DCOSTA, David Xie, Seacy Zhen
Prereq:
- solid knowledge of coding in Python
- exposure to wealth management / financial sector
- willingness to participate in team/group activities
Format: HandsOn
Level: Intermediate

Join us in this hands-on workshop to create your own reports and dashboards to support a financial advisor in completing key elements of his/her day-to-day activities. We'll look at data requirements, experiment hands-on with the data, clean it, load it to a data-science-friendly environment, run some standard models and then extend those models in different ways. Finally, we'll bring together the results in a user-friendly way.

If you've wondered about how to operationalize your machine learning algorithms, want to get deeper into data science as a financial-sector participant or are interested in IBM's ecosystem for machine learning innovation, this is the right workshop for you!

Compiler-Driven Performance Workshop Strategies

Location: Conf Center 1
Theme: Systems
Chairs: Gennady Pekhimenko, Ettore Tiotto
Prereq: none
Format: Speakers
Level: Intermediate

The workshop has a particular focus on (but not limited to):
- Innovative compiler analysis, transformation, and optimization techniques
- Languages, compilers, and optimization techniques for multicore processors and other parallel architectures
- Compiling for streaming or heterogeneous hardware
- Dynamic compilation for high-performance and real-time environments
- Compilation, optimization, and analysis for dynamic languages
- Compilation techniques for reducing power
- Program safety
- Whole system optimization and analysis
- Tools and infrastructure for compiler research

Practical Machine Learning with Python on DSX

Location: Conf Center 2
Theme: Data and Analytics
Chairs: Shaikh Quader, Mark Ryan, Eric Dong
Prereq:
- Coding experience in any programming language
- Users MUST bring their own laptops for the hands-on exercise
- IBM Cloud id: https://console.bluemix.net/registration/
Format: HandsOn
Level: Beginner

In this FULL-DAY workshop, after a brief introduction to Machine Learning, we'll take the students through the hands-on exercise of building a Machine Learning model from the scratch. They will learn and code in different phases of Machine Learning pipeline, including acquiring, cleaning, and exploring data; building and evaluating ML model. Finally, we'll tell them how to build a discipline for continuous learning of ML and applying this to solving real problems.
iCity - Big Data and Visualization Urban Transportation Strategies

- **Location**: Jasmine
- **Theme**: Data and Analytics
- **Chairs**: Sara Diamond
- **Prereq**: none
- **Format**: Speakers
- **Level**: Intermediate

Providing efficient, cost-effective, sustainable transportation networks and services is a major challenge for cities around the world – not only for individual cities, but for connectivity between cities. High quality transportation services, notably well-designed transit hubs within comprehensive networks are fundamental prerequisites for effective cities and spur economic, social and cultural inclusion, development and growth. Transportation strategies must be at the heart of smart city strategies. The melding of machine learning, simulations, predictive analytics and design create capacity and connectivity that will help policy and makers gain insight into complex decision-making processes and support evidence-based decision making. Solving transportation and transit challenges requires integrating transdisciplinary knowledge, including computer science, engineering into city planning.

Distributed Ledgers and Blockchain: Concepts and Applications

- **Location**: Evergreen
- **Theme**: Systems
- **Chairs**: Asic Chen, Arno Jacobsen
- **Prereq**: General computer science and computer engineering (information systems background as taught in Bachelor’s curricula)
- **Format**: Speakers
- **Level**: Beginner

Blockchain has been, without doubt, one of the hottest topics in technology in recent years. As is in the case of most “buzzwords”, most people, even software professionals, have no more than a surface comprehension of the technology. We hope that through this workshop, we can demystify blockchain and distributed ledgers, giving the attendees a working understanding while expanding on use cases far beyond the most popular: cryptocurrency. During this workshop, we will provide a tutorial-style introduction to various distributed ledgers and blockchain technologies. We will be focusing on first principles and algorithms while identifying emerging blockchain use cases.

2nd Workshop on Advances in Open Runtime Technology for Cloud Environments

- **Location**: Primrose
- **Theme**: Cloud Computing
- **Chairs**: Daryl Maier, Kenneth Kent
- **Prereq**: None
- **Format**: Speakers
- **Level**: Intermediate

Modern language runtimes are complex, dynamic environments that involve a myriad of components that must work cooperatively to achieve the functional and performance requirements of a given language. Typical core runtime technologies include dynamic just-in-time compilers for performance, garbage collection for heap management, platform abstraction for ease of portability to different hardware and operating system environments, developer tooling for diagnosis and tuning of the various components, and interoperability between different language environments.

Cloud services such as IBM Cloud, Microsoft Azure, or Amazon Web Services (AWS) are increasingly becoming the environments where applications are developed and deployed, data is stored, and businesses are run. Many of the features that define a cloud (e.g., resiliency, elasticity, consistency, security) are realized through runtime technologies. Clouds are polyglot environments, and therefore advances in cloud development are directly driven by innovation in runtime technologies. However, cloud environments pose unique, often conflicting demands on runtime systems that are often less of a concern in isolated systems. Throughput performance (how fast is my application?), density (how many instances of my application
Organizing Committee

Conference Chair
Iosif Viorel Onut

Conference Program Co-Chair
Dorina Petriu
Wang Chen

Workshop Co-Chair
Hausi Müller
Tinny Ng

Expo Co-Chair
Kenneth Kent
Robert Enenkel

Finance and Registration Chair
Marcellus Mindel

Website Chair
Harnoor Manan

Local Arrangements Co-Chair
Marin Litoiu
Iosif Viorel Onut

Volunteer Chair
Marin Litoiu

Publication Chair
Guy-Vincent Jourdan

Conference Proceedings Editor
Andrew Jaramillo

Network Chair
Rodney D'Silva
Alan Heighway
Steering Committee

Marin Litoiu
Kelly Lyons
Marcellus Mindel
Hausi Müller
Tinny Ng
Iosif Viorel Onut
Joe Wigglesworth
Ken Wong
Ying Zou

York University
University of Toronto
IBM Canada Ltd.
University of Victoria
IBM Canada Ltd.
IBM Canada Ltd.
University of Alberta
Queen’s University

Program Committee

Bram Adams
Jose Nelson Amaral
Daniel Amyot
Christopher Anand
Ebrahim Bagheri
Ayse Bener
Jeremy Bradbury
David Bremner
Tse-Hsun Peter Chen
Mark Chignell
Andrew Craik
Frank Dehne
Chen Ding
Juergen Dingel
Herhard Dueck
Gbizlane El Boussaidi
Michalis Famelis
Marios Fokaefs
Sudhakar Ganti

École Polytechnique de Montréal
University of Alberta
University of Ottawa
McMaster University
Ryerson University
Ryerson University
University of Ontario Institute of Technology
University of New Brunswick
Concordia University
University of Toronto
IBM Canada Ltd.
Carleton University
University of Rochester
Queen’s University
University of New Brunswick
École de Technologie Supérieure
University of Montreal
École Polytechnique de Montréal
University of Victoria
Best Paper Awards Selection Committee

Dorina Petriu
Wang Chen
Ebrahim Bagheri
Andrew Craik
Juergen Dingel

Most Influential Paper Award Committee

Dorina Petriu
Wang Chen
Marsha Chechik
Ken Wong
Workshop Selection Committee

Hausi Muller
Tinny Ng
Frank Li
Marin Litoiu
Kishor Patil
Maria Patrou
Ladan Tahvildari
Derek Truong
Ying Zou
Calisto Zuzarte

University of Victoria
IBM Canada Ltd.
IB...
IBM Advanced Studies (CAS) is a system of engagement for academic collaboration with IBM. We network IBMers with a community of students, educators, researchers and industry professionals to connect ideas, people and organizations around research, technology and design.

**Programs**

**RESEARCH** — Engage students and faculty in joint projects with IBM product teams to shape industry-leading offerings, and advance academic research and publication.

**EDUCATION** — Create learning opportunities and proof-of-concept solutions based on IBM product offerings and expertise. Embrace diverse experiences and perspectives, reframe technical and social challenges, and facilitate meaningful engagement and transformation.

**Projects**

For 2018, we currently have 49 joint academic-industry research projects associated with 102 IBM products, and involving 107 IBMers, 43 faculty and 94 graduate students. In advancing the thesis research of each student, these projects generate publications, patents and IBM product contributions.

**Conference**

Hosted by IBM Advanced Studies, the Community for Advanced Studies Conference (CASCON) is the largest software engineering conference in Canada, with annual attendance averaging around 1000. CASCON is a full academic conference with 18 years of annual proceedings published in the Association of Computing Machinery digital library. Through keynotes, paper presentations, workshops and exhibits, CASCON brings top researchers, students and industry professionals together to promote joint research in computer science and software engineering.

Marcellus Mindel
Head, IBM Advanced Studies
About CASCON

CASCON is a premier industrial and academic conference in computer science and software engineering. Hosted by IBM Advanced Studies and IBM Canada Laboratory, CASCON attracts developers, researchers, innovators, technologists and decision makers from a variety of sectors including academia, industry, and government. We all come together to discover and discuss technology trends, share knowledge, form collaborations, write proposals, present papers, exhibit prototypes and showcase products.

Conference content and format include:

- Speakers
- Expo
- Papers
- Workshops

Papers
Submit your paper and showcase your research findings to a variety of attendees from different backgrounds and specialties. If your paper is accepted, it will be showcased in the conference proceedings published by CASCON and included in the ACM Digital Library. Visit the Papers section to find out more about the submission steps and requirements.

Best Paper Awards
CASCON recognizes the best technical contributions of the event in terms of originality, clarity, and potential impact with two awards: Best Paper and Best Student Paper. To be eligible for the Best Student Paper award, a student must have primarily authored the paper, and student(s) at the time must have done the work described.

Workshops
You will have the opportunity to present, discuss, and debate ideas, challenges, emerging technologies and project directions in an interdisciplinary environment.

CASCON Expo
Our famous technology showcase is interactive, innovative, and includes diverse exhibits of research interests, technologies, joint projects, and collaboration opportunities.