Big Data Announcement
Vivisimo, Inc. Acquisition & Cloudera Partnership

The purpose of this document is to address questions and provide additional insight about IBM's big data platform, IBM's acquisition of Vivisimo and its partnership with Cloudera. To view the full press release, go to IBM Press Release.

BIG DATA PLATFORM

1. **What is big data and why is it important?**
    The explosion in the volume, variety and velocity of information is creating new opportunities to optimize business outcomes in an information-centric, insight-driven world. This big data is creating a new era of computing, much like the emergence of the Web did 15 years ago. IBM has the opportunity to lead in this industry transition.

2. **What is IBM's big data platform strategy?**
    To help clients take advantage of the big data opportunity, IBM is creating an enterprise-class big data platform that integrates, manages and applies analytics to the full volume, variety, and velocity of data. This big data platform leverages and extends open source technologies, adding capabilities such as stream computing, warehousing, enterprise integration, visualization, systems management and governance, required for the enterprise. IBM's big data platform includes:
    *   Hadoop systems - cost-effectively analyze petabytes of structured and unstructured information
    *   Stream Computing - analyze streaming data and large data bursts for real-time insights
    *   Data Warehousing - deliver deep insight with advanced in-database analytics and operational analytics
    *   Information Integration & Governance – integrate diverse sources of information, govern data quality and manage information lifecycle
    *   Accelerators - speed time to value with analytic and application accelerators
    *   Visualization - gather, extract and explore data
    *   Application Development – develop a new class of analytic applications across multiple big data engines
    *   Systems Management – monitor and manage big data systems for secure and optimized performance.

3. **What is IBM's Hadoop system offering?**
    IBM's Hadoop system offering is InfoSphere BigInsights. InfoSphere BigInsights includes the basic edition (open source distribution of Hadoop) and the enterprise edition (the IBM value-add on top of open source distributions). InfoSphere BigInsights is available as a software package and is also available on the cloud.

4. **How is IBM's big data approach different than alternatives?**
    IBM's big data strategy is different than competitive offerings such as those from Oracle, Teradata, EMC and others in several key aspects:

    1. **Comprehensive Platform** – IBM is the only vendor who defines a broad big data platform that includes structured data warehousing, Hadoop systems, and stream computing.
    2. **Stream Computing** – IBM is the only vendor who identifies Stream Computing as a key element in the big data platform. IBM InfoSphere Streams is a unique big data offering that was built by IBM research and development.

3. **Analytic Accelerators** – IBM delivers analytic accelerators that are designed to work on a variety of data types – to analyze data in its native format (be it textual content, video, audio, among others).
4. **Adding Value on top of Open Source** – IBM's strategy is to build on top of open source Hadoop and open source distributions from third parties. Many alternatives have simply packaged or partnered with open source distributions, or have added a very small amount of intellectual property on top of open source Hadoop. IBM is unique in adding advanced analytic accelerators, visualization, development tooling, performance enhancements, and security features on top of open source.
5. **Treating a Hadoop System as much more than a pre-processor for data warehousing** – many data warehouse vendors position Hadoop as a pre-processor of unstructured data designed to load structured data into the data warehouse, where more sophisticated analytics are applied. IBM is unique in positioning analytics designed for variety; there are many use cases in which Hadoop is a logical destination for data and analytics, not simply a pre-processor.
6. **Advancing cost-efficient alternatives to simply putting 'all' data into Hadoop for analysis** – Hadoop is a very cost-efficient system for processing large volumes of unstructured data. Many vendors position it as the logical place to put ALL information to analyze and determine its value. For many requirements this is a valid approach. But IBM recognizes that it is possible to analyze data before storing all of it in a Hadoop system. One example is stream computing, which offers the ability to analyze data in motion and to persist only the relevant data and/or insights in Hadoop or a data warehouse. Another alternative is federation and discovery – the ability to understand big data content in other repositories before importing all of it into a Hadoop system. IBM is unique in offering both of these capabilities that allow clients to understand the value of information before deciding to persist it in Hadoop system or a data warehouse. This further reduces the time and cost of analyzing big data by identifying the relevant data before integrating and persisting it.

VIVISIMO ANNOUNCEMENT

GENERAL

1. **What are you announcing today?**
On May 24, 2012, IBM completed the acquisition of Vivisimo a Pittsburgh, Pennsylvania-based provider of federated discovery and navigation software for big data and Vivisimo is now an IBM Company. The acquisition aims to accelerate IBM's efforts to help clients integrate, manage, and analyze the full variety, velocity and volume of big data. Financial terms were not disclosed.

2. **Who is Vivisimo? What do they do?**
Vivisimo is a software company whose federated discovery and navigation capabilities (of a variety of information sources) help organizations access, understand and navigate through vast amounts of structured and unstructured information to get more value from it. Vivisimo provides a single point of access to disparate data sources, extracting relevant information and providing ad hoc analysis without moving the data.
   - Founded in 2000, Vivisimo is headquartered in Pittsburgh, Pennsylvania
     - 120 employees worldwide
     - Track record of success and experience in information optimizations through enterprise search solutions
   - >140 clients; with enterprise customers and large presence in government
     - Airbus, U.S. Air Force, Social Security Administration, Defense Intelligence Agency, U.S. Navy, Procter & Gamble, Bupa and LexisNexis

- Partnered with key technology partners, systems integrators and resellers, both global and regional, in markets worldwide.

3. **Why did IBM acquire Vivisimo?**
   Vivísimo provides the ability to leverage and increase the value of traditional approaches to data (structured, warehouses etc.) while also adding the ability to gain insight via new approaches and new sources.

   Vivisimo's key capabilities in conjunction with the IBM big data platform include:
   - Providing a single point of access to multiple internal and external sources held within different engines – Hadoop, warehouse or external sources such as Lotus Notes and other content repositories like MS SharePoint
   - Enabling ad hoc exploration and analysis across federated data
   - Discovering and profiling sources to select which data to bring into a Hadoop system for deeper analysis
   - Indexing big data held within BigInsights for broad consumption by business users
   - Enhancing structure and classification for unstructured information so that it can be more effectively used for insight
   - Faceted navigation, to explore a large corpus of data
   - Ability to ingest industry and domain-specific taxonomies and ontologies
   - Rapid deployments from POCs to production or fast time to value.

4. **How will Vivisimo fit within the IBM portfolio?**
   Vivisimo products will become part of the big data portfolio within Information Management. Vivisimo will be integrated within the InfoSphere BigInsights product offering and continue to be offered separately as well.

5. **How will Vivisimo clients benefit?**
   Vivísimo clients will benefit from the increased investment, global reach, industry expertise and support available from IBM. IBM remains committed to the values – such as ease of deployment, simplicity of management and high performance that have made Vivisimo successful to date and intends to continue to support and enhance the Vivisimo products. Vivisimo clients will benefit from the breadth of IBM's big data platform.

6. **How will the Vivisimo team fit organizationally within IBM?**
   Vivisimo is now part of the big data portfolio within Information Management led by Arvind Krishna, General Manager, IBM Information Management.

BUSINESS PARTNERS

1. **How will this affect current Vivisimo Business Partners?**
   Vivisimo Business Partners will continue to operate as Business Partners of Vivisimo, now an IBM Company. As before this announcement, Vivisimo Business Partners are also welcome to enroll in IBM's Business Partner program – IBM PartnerWorld – to access a vast array of benefits and resources, including marketing, selling, technical and training/certification resources.

   Visit IBM PartnerWorld for additional information at www.ibm.com/partnerworld.

CLOUDERA ANNOUNCEMENT

1. **What did IBM announce?**
   On April 25th, 2012, IBM announced its support for third party Apache Hadoop distributions; in addition to IBMs own distribution. The first partner distribution supported is from Cloudera.

2. **Who is Cloudera, why has IBM selected Cloudera as its Apache Hadoop partnership?**
   Cloudera is one of the early providers of Hadoop systems and has its own distribution of open source Apache Hadoop. Cloudera has numerous companies who have downloaded and/or are clients of its distribution of Hadoop. IBM is partnering with Cloudera as it helps clients protect their investment in Cloudera, while benefiting from IBM's enterprise platform capabilities. Supporting multiple Hadoop distributions builds on IBM's ongoing commitment to fuel innovation through the open source community.

3. **Will IBM continue to work with its own distribution?**
   Yes, IBM will continue to support its own distribution while testing and supporting a growing number of Apache Hadoop distributions.

4. **How will IBM clients benefit?**
   With IBM's support of multiple Hadoop distributions (Cloudera being the first); customers have flexibility in the selection of their Hadoop distribution. IBM's strategy is squarely focused on adding value on top of open source distributions of Hadoop, a component of IBM's enterprise big data platform.

5. **Will IBM sell InfoSphere BigInsights to clients with Cloudera?**
   Yes. IBM will sell InfoSphere BigInsights to Cloudera's client base.

6. **How will Cloudera clients benefit?**
   Clients who have invested in Cloudera can benefit from IBM's enterprise big data capabilities, including analytic accelerators, enterprise integration capabilities, performance improvements to the Hadoop system (workload management and improvements to parallel processing), visualization, security, and development tooling, and also Vivisimo's federated discovery and navigation capabilities while supporting their investment in Cloudera's Hadoop distribution.

DISCLAIMER