Continuous Availability for FTM Immediate Payments and ACH: a Reference Approach

Introduction to Implementing IBM Financial Transaction Manager for Immediate Payment or ACH Services with High Availability and Minimizing Downtime

Introduction

Financial Transaction Manager (FTM) for Immediate Payment (IP) and ACH Services (ACH) are used by Financial Institutions to process payments. With the emerging requirement of immediate payments, financial institutions are looking for ways to keep the system processing 24x7x365.

What this paper is about

This paper discusses how FTM users can plan the product deployment to achieve high availability and in particular introduces the Blue-Green approach to minimize service interruption during product migration or product updates. After reading this paper, FTM users will have a better understanding of this approach that can be used as a guideline for their implementation. The Blue-Green deployment with Q-Replication is written based on a proof of concept on IP and ACH. Actual implementation of the Blue-Green deployment process will vary based on the user's requirements. Even though the proof of concept is based on the work on multi-platforms, most of the concepts discussed may apply to zOS.

On high availability

High availability (HA) typically refers to the ability that a solution can operate for a defined period of time and can quickly recover when either hardware or software failure happens. It is normally achieved in solution design by:

1. Avoiding single point of failure (SPOF) by leveraging redundancy in all tiers
2. Reducing failover time using high performing hardware or software settings when applicable
3. Having the capability of timely error detection or alert notification

FTM for Immediate Payment (IP) and FTM for ACH Services (ACH) use various middleware products such as IBM WebSphere Application Server (WAS), IBM Integration Bus (IIB/MB), and IBM MQ (MQ). Achieving high availability for FTM solutions requires configuring these middleware products for high availability. Although there are various ways to achieve high availability for middleware configuration, this paper limits the discussion to some specific configuration(s) for high availability.

Disaster Recovery (DR) consideration is out of this paper's scope.

High availability configuration

HA configuration can be done in several different ways. The following is a core list of middleware products used by an FTM solution

and basic HA configuration options, respectively.

For DB2, options are:

1. DB2 High Availability Disaster Recovery (HADR)
2. IBM PowerHA SystemMirror
3. DB2 pureScale
4. Variations of the above such as DB2 pureScale with HADR

For MQ/IIB, options are:

1. Single multi-instance MQ/IIB
2. Dual multi-instance MQ/IIB
3. Cluster
4. Variations

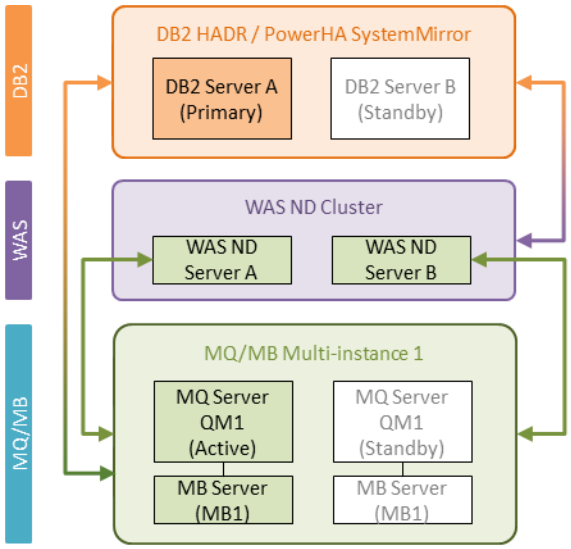
For WAS, options are:

1. Horizontal cluster
2. Vertical cluster
3. Hybrid cluster
4. Variations

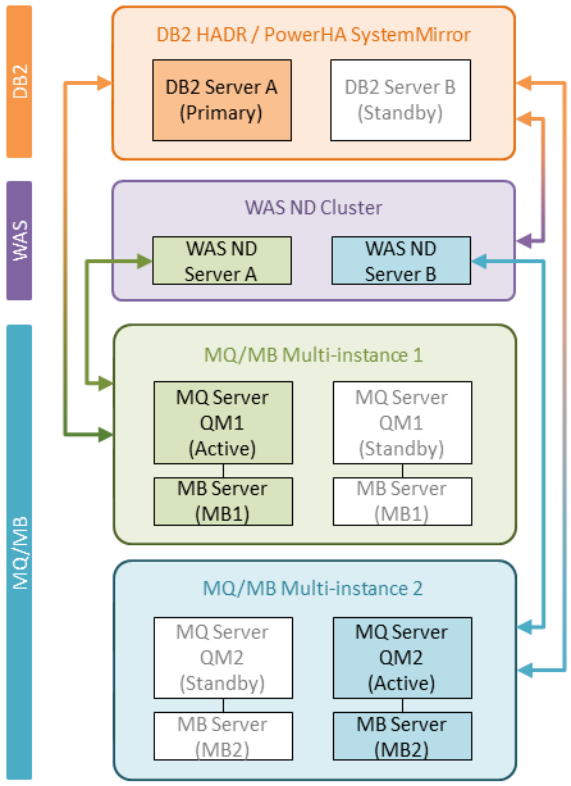
Sample HA configurations for IP are listed below

Options listed below focus on FTM internal components. HA of external systems that FTM interacts with should be considered in the actual deployment either using active/passive or active/active depending on the nature of the external systems. These configurations do not address disaster recovery which needs to be considered in actual production.

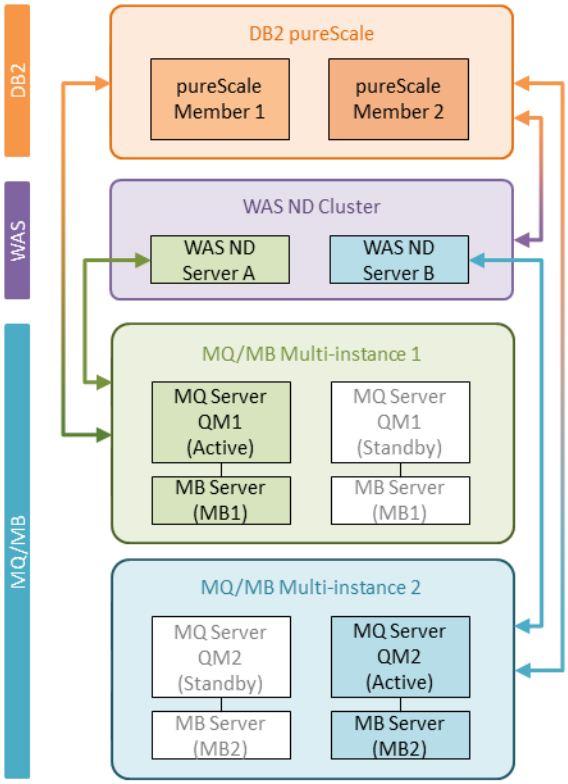
Option 1: In this configuration, the DB tier is configured active/passive and the MQ tier is configured active/passive using single multi-instance MQ. The following shows the configuration:



Option 2: In this configuration, the DB tier is configured active/passive and the MQ tier is configured with dual multi-instance MQ. The following shows the configuration:



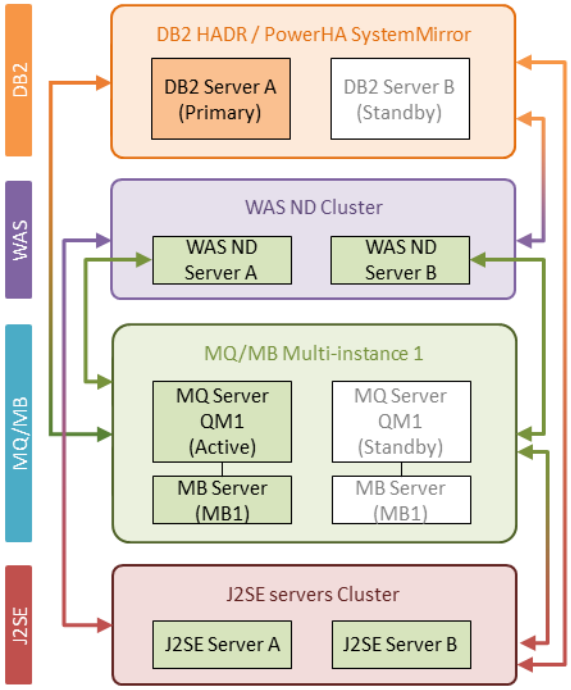
Option 3: In this configuration, the DB tier is configured active/active using DB2 pureScale. DB2 pureScale is not yet supported by FTM. The MQ tier is configured with dual multi-instance MQ. The following shows the hypothetical configuration once FTM supports DB2 pureScale.



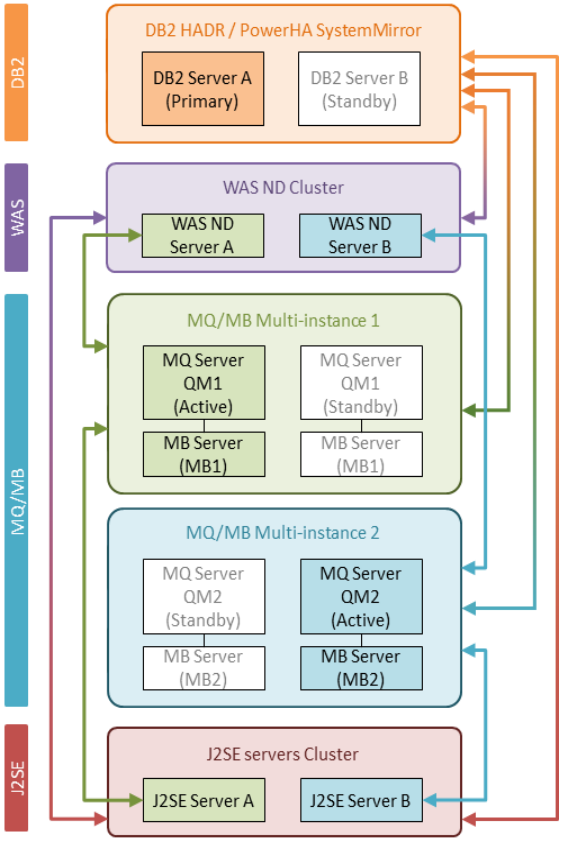
Sample HA configurations for ACH are listed below

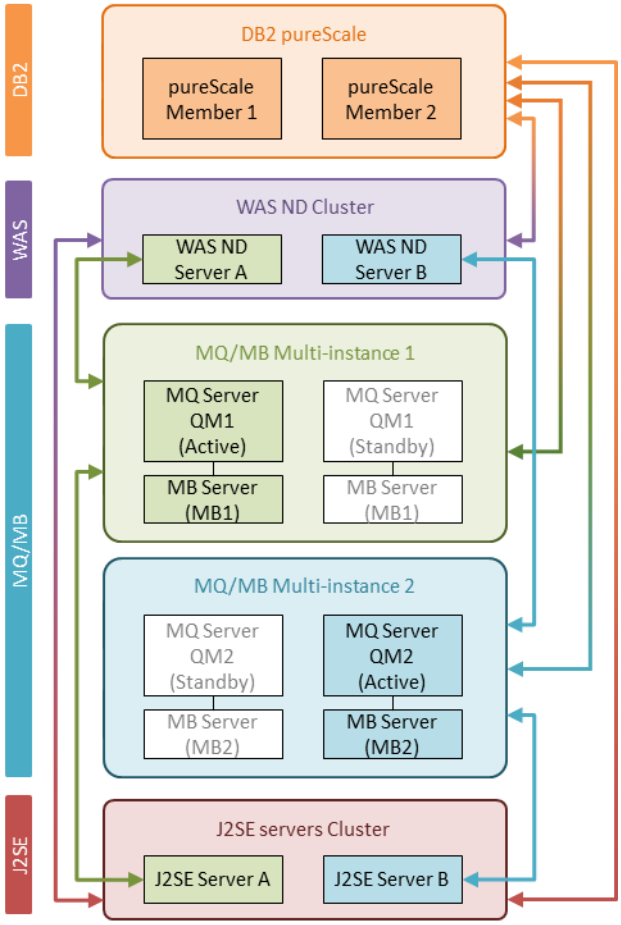
Options listed below focus on FTM internal components. HA of external systems that FTM interacts with should be considered in the actual deployment either using active/passive or active/active depending on the nature of the external systems. These configurations do not address disaster recovery which needs to be considered in actual production.

Option 1: In this configuration the DB tier is configured active/passive and the MQ tier is configured active/passive using single multi-instance MQ. The following shows the configuration:



Option 2: In this configuration, the DB tier is configured active/passive and the MQ tier is configured with dual multi-instance MQ. The following shows the configuration:





Option 3: In this configuration, the DB tier is configured active/active using DB2 pureScale. DB2 pureScale is not yet supported by FTM. The MQ tier is configured with dual multi-instance MQ. The following shows the hypothetical configuration once FTM supports DB2 pureScale.

On zero downtime

Zero downtime is a further requirement on top of high availability. High availability is focused on normal operation while zero downtime is focused on minimizing downtime typically caused by maintenance and migration.

Requirements on zero downtime are typically translated into Service Level Agreements (SLA), Service Level Objectives (SLO), or uptime such as 99.999% (five 9s).

While HA configuration mentioned above will help the solution from a high availability point of view, customers still need to consider what happens to the solution during maintenance activity that may include but are not limited to: ifix, fixpack, release update, etc. Further consideration is needed when the activity involves DB schema updates that require a database update to be performed offline.

While certain updates to the FTM applications may be done online or with rolling update features supported by the middleware products (i.e. components in different nodes of an application cluster can be updated independently without application totally down), updates involving DB schema changes may require applications to be stopped throughout the update process. In such cases, to minimize the downtime, this paper discusses an approach called **Blue-Green deployment** with unidirectional Q-Replication that allows customers to conduct maintenance activity while keeping the production system running with minimal interruption.

Blue-Green deployment with Q-Replication explained

Blue-Green deployment may have different meanings to different people. To be clear, in this paper Blue-Green deployment refers to a process to migrate an FTM environment in current production called Green to a newer version of the FTM solution on another environment called Blue. Each environment consists typically of separate hardware, operating system, and middleware components as DB2, MQ, WAS. The Blue environment is prepared to run production workload. The approach discussed below is targeted to minimize the service interruption needed to perform an update on FTM solution that would require database to be stopped such as to perform certain database schema changes.

Blue-Green deployment for Immediate Payment and Blue-Green deployment for ACH are discussed separately below.

Blue-Green deployment for FTM for Immediate Payment (IP) with Q-Replication

In the case where migrating IP from older version to newer version requires the application to be down for an extended period of time, Blue-Green deployment with Q-Replication will minimize the impact. For description purposes, we will refer to the FTM IP current production solution as IP Green and the new FTM IP solution to replace the current production solution as IP Blue. The following highlights the steps that can be taken for migration, and key points will follow:

1. Create Blue DB as a copy of Green DB
2. Set up Q-Replication from Green DB to Blue DB and start the Q-Replication

process (see reference in section "*For More Information*")

3. Stop the Q-replication, then upgrade the Blue DB and install new application.
4. Start the Q-replication, monitor the Q-replication process, and make sure the Blue DB catches up to the Green DB. This step is important in determining if and when the switchover can be done. It is important to ensure the IP Green DB and IP Blue DB are in sync or only contain a small amount of new transactions on IP Green DB, but not on IP Blue DB such that new transactions can be replicated in seconds before switchover.
5. Determine the best point of time (BPOT) (either no traffic time frame, lowest traffic time frame, no internal scheduled job running, or other criteria best for business flows) for switchover from the IP solution on IP Green to the IP solution on IP Blue
6. During the switchover from IP Green to IP Blue

    A. Stop accepting new traffic to IP solution on Green

    B. Drain the transactions already processing on IP Green

    C. Stop the Q-Replication process once transactions leftover on IP Green are fully processed including those that might come in based on the scheduler, if any

    D. Start accepting new traffic to the IP solution on Blue

    E. Repurpose the IP Green for backup or other usage. From this point on, IP Blue is now production

    Note: Strictly following steps A through D may result in a brief service interruption if SLA, SLO, and uptime allow it. Depending on the solution and requirement, steps A through D may be adjusted to achieve no service interruption while keeping both Blue and Green running for brief period of time. However, thorough testing needs to be conducted to account for transactions in various stages of the transaction life cycle during the switchover if transactions are not drained first to avoid inconsistency.

Q-Replication setup consideration from IP Green DB to IP Blue DB

For the IP solution, you need to prepare the following:

List of tables for replication: analysis needs to be done for your specific solution to determine what tables to replicate. The following is the minimum set of tables for replication:

- OBJ_BASE
- H_OBJ_BASE
- BATCH_BASE
- TRANSMISSION_BASE
- TRANSACTION_BASE
- TXN_PAYMENT_BASE
- SVC_PARTICIPANT_BASE
- SCHEDULER_TASK_BASE
- OBJ_OBJ_REL
- OBJ_VALUE

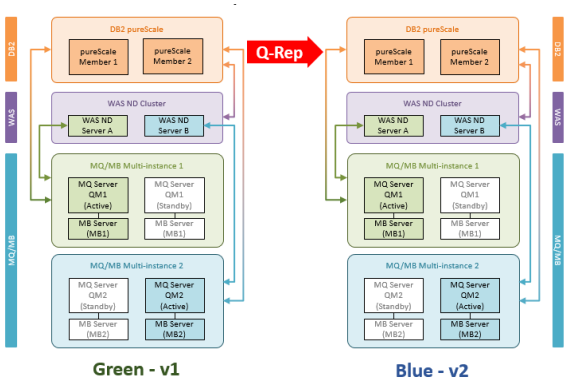
- COUNTER
- ERROR
- EVENT
- OUT_REQ_CORREL

Note: If tables such as UI_AUDIT_LOG are replicated, identity column contained in such table may need to be redefined to allow replication to proceed.

Database sequence consideration: the IP solution uses sequences for ID generation. During the switch from IP Green to IP Blue, if the ID sequence is not properly handled, new transactions to IP Blue will encounter exceptions due to duplicate ID generation. To solve this, it is recommended that **IP Green** be setup using ID management discussed in the white paper: "*Financial Transaction Manager: Increase Your Transaction Throughput with FTM ID Management*" using suitable values for PRIMARY_ID_GRP_NUM and PRIMARY_ID_GRP_SIZE for IP Blue and IP Green to control the ranges of ID values used. For example: PRIMARY_ID_GRP_NUM 0 and PRIMARY_ID_GRP_SIZE 2 on **IP Green**, and PRIMARY_ID_GRP_NUM 1 and PRIMARY_ID_GRP_SIZE 2 on **IP Blue**. However, if ID management is already used for performance improvement, the exact settings used for Blue-Green Deployment need to be reviewed and tested.

IP migration with HA

The following diagram shows Blue-Green deployment for migration when HA is configured. DB2 tier could be configured with PowerHA SystemMirror or HADR. DB2 pureScale is not yet supported by FTM.



Blue-Green deployment for FTM for ACH Services (ACH) with Q-Replication

In the case where migrating ACH from an older version to a newer version requires the application to be stopped, Blue-Green deployment with Q-Replication will minimize the impact. For description purposes, we will refer to the FTM ACH solution in current production as ACH Green and the new FTM ACH solution to replace current production as ACH Blue. The following highlights the steps that can be used for migration, and key points will follow:

1. Create a copy of Green DB
2. Set up Q-Replication from Green DB to Blue DB and start the Q-Replication process (see reference in section "*For More Information*")
3. Stop the Q-replication, then upgrade the Green DB copy in step 1 to the Blue DB and install the new application
4. Start the Q-replication, monitor the Q-replication process, and make sure Blue DB catches up to Green DB
5. Determine the best point of time (BPOT) (either no traffic time frame, lowest traffic time frame, no internal scheduled job running, or other criteria best for business flows) for switchover

from the ACH solution on ACH Green to the ACH solution on ACH Blue

6. During the switchover from ACH Green to ACH Blue

    A. Stop accepting inbound traffic to ACH Green

    B. Drain the transactions already in on ACH Green

    C. Stop the Q-Replication process once transactions leftover on ACH Green are fully processed including those that might come in based on the scheduler, if any

    D. Start accepting inbound traffic to ACH Blue

    E. Repurpose the ACH Green for future migration, backup or other usage. From this point on ACH Blue is now production

    Note: Strictly following steps A through D may result in a brief service interruption if SLA, SLO, and uptime allow it. Depending on the solution and requirement, steps A through D may be adjusted. However, thorough testing needs to be conducted to account for transactions in various stages of the transaction life cycle during the switchover if transactions are not drained first to avoid inconsistency and/or duplicates. Additional steps may need to be introduced if transactions are not drained first before the switchover. BPOT of switchover needs to be determined through testing based on implementation and business requirements.

Q-Replication setup consideration from ACH Green DB to ACH Blue DB

List of tables for replication: analysis needs to be done for your specific solution to determine what tables to replicate. Due to the sheer number of tables in ACH, it may be acceptable to replicate all tables except the following:

- IIA_RULES
- LOCKEVMON
- TIFF_IQA_RULES

Database identity column consideration: the ACH solution uses identity columns for ID generation. During the switchover from ACH Green to ACH Blue, if an identity column is not handled properly, new transactions to ACH Blue will potentially encounter exceptions due to duplicates in ID generation. To solve this, it is recommended that identity columns be setup initially using odd/even configuration as below:

On ACH Green DB

ALTER TABLE <table> ALTER COLUMN <identity column> SET INCREMENT BY 2;

ALTER TABLE <table> ALTER COLUMN <identity column> RESTART WITH <max id + 1 >;

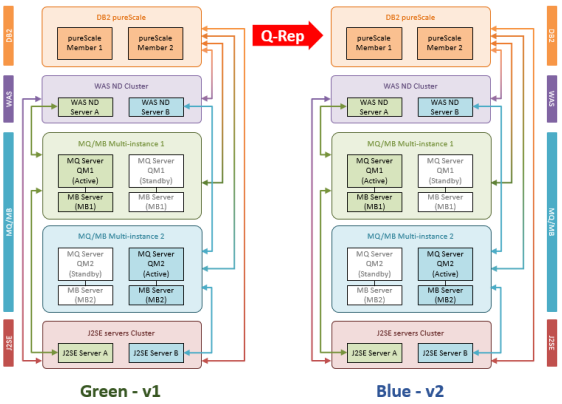
On ACH Blue DB

ALTER TABLE <table> ALTER COLUMN <identity column> SET INCREMENT BY 2;

ALTER TABLE <table> ALTER COLUMN <identity column> RESTART WITH <max id + 2 >;

ACH migration with HA

The following diagram shows ACH Blue-Green deployment for migration when HA is configured. DB2 tier could be configured with PowerHA SystemMirror or HADR. DB2 pureScale is not yet supported by FTM.



Migration considerations

Migration is a complex process. It requires comprehensive planning and thorough testing. It is **strongly** recommended to

1. Thoroughly test the Blue-Green deployment migration process in a non-production environment using production data
2. Select BPOT for switchover from Green to Blue
3. Plan retry logic from the application interface in case a query or update on an existing transaction has not been replicated over to Blue if switchover is done before replication is completed
4. Know your solution transaction rate and plan well for handling transactions that may be still on Green if you choose to allow Blue to become live before replication is completed
5. Take caution if configuration tables are replicated. Any configuration change on Green after the DB copy is taken will be replicated to Blue if the change is in the table being replicated. During the actual migration, a new copy of the Green DB should be taken to minimize the Blue DB catchup time. From this point on, configuration changes should be locked until migration is complete.
6. Migration should be done while the system is in a clean state where user does not have to worry about any intermediate files or messages resulting from unresolved issues

Conclusion

Selection of configuration for high availability and minimizing downtime depends on SLA, SLO, or uptime requirements along with cost analysis such as software licensing and hardware needed. The selection of hardware is also critical in achieving SLA, SLO, or uptime, and disaster recovery. The Blue-Green deployment approach discussed in this paper allows customers to migrate an FTM solution from older versions to newer versions with minimum interruption in production. This will reduce the typical downtime from hours to just a small window: the switchover time of inbound traffic + processing time of transactions leftover on Green if any + replication time. Selecting BPOT will further minimize the potential service interruption.

As a rule of thumb, all configurations and solutions should be thoroughly tested to ensure they meet SLA, SLO, or uptime set by the business before production goes live.

For More Information

To learn more about IBM Financial Transaction Manager for Multiplatform, visit the IBM Knowledge Center:

https://www.ibm.com/support/knowledgecenter/SSRH46/welcomeSSRH46.html

To learn more about ID Management, visit:

https://www.ibm.com/support/knowledgecenter/en/SSRH46_3.0.0/fxhplanidmanagement.html

To learn how to set up Q-replication, visit:
https://www.ibm.com/support/knowledgecenter/SSEPGG_10.1.0/com.ibm.swg.im.iis.repl.qtutorial.doc/topics/iiyrqtutintro2.html

To learn more on ID Management, visit:

http://www.ibm.com/support/docview.wss?uid=swg27049594