THE FILLMORE GROUP
Relational Database Solutions

IBM Replication Updates: 4+ in 45

The Fillmore Group – February 2019

A Premier IBM Business Partner

Premier Business Partner

Poll 1: Is your organization currently using replication in a production environment?

Agenda

- Replication Review
  - Why, What, How in under 10 minutes
- Replication: Key New Capabilities
  - Remote Db2 z/OS capture
  - Targeting Big Data (Kafka, Hadoop)
  - Targeting Db2 Warehouse
  - Sourcing PostgreSQL
  - Replication for Db2 Continuous Availability
- Replication Strategic Plans and updates: 2019 and beyond

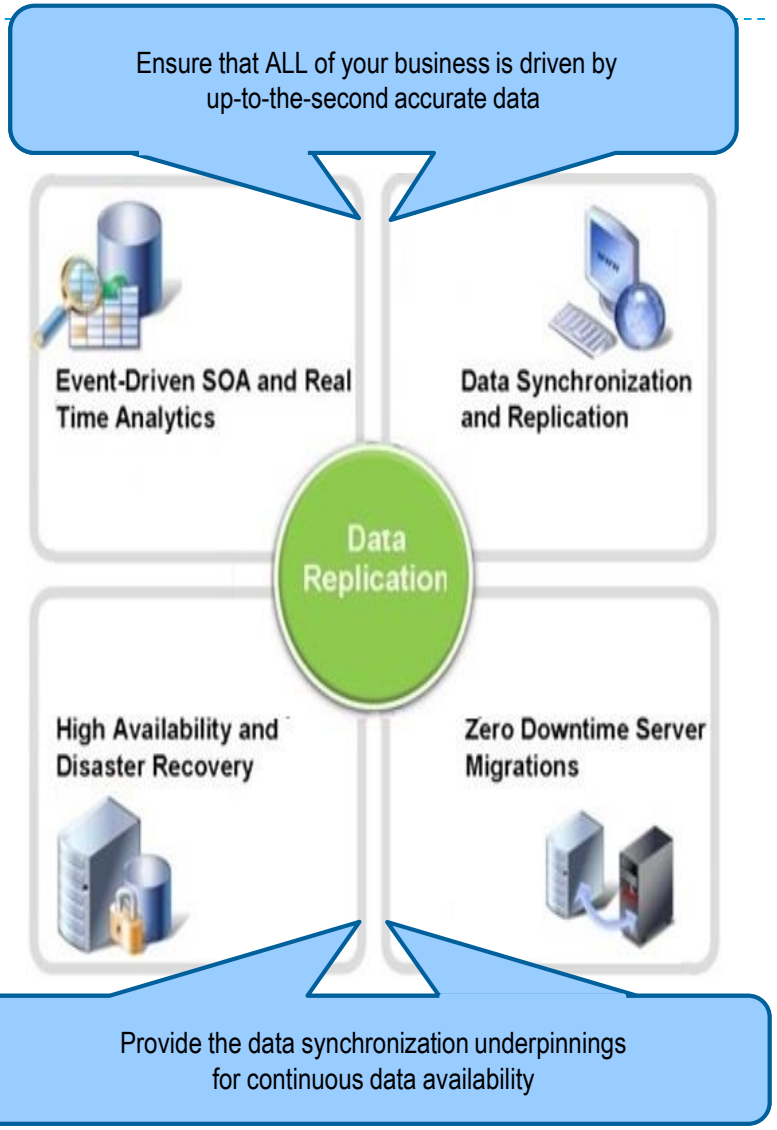
Poll 2: Which replication solutions are in use at your organization?

Replication Review:
Why, What , How… in under 10 min.

Replication in Action - *Why*

- ***Business in Real Time***
  - Detect and react to data events as they happen to drive the business
  - Optimize decision making with up to the second data, i.e. real time analytics

- ***Always On Information***
  - High availability with Active-Standby and Active-Active data deployments
  - Data synchronization for zero down time data migrations and upgrades



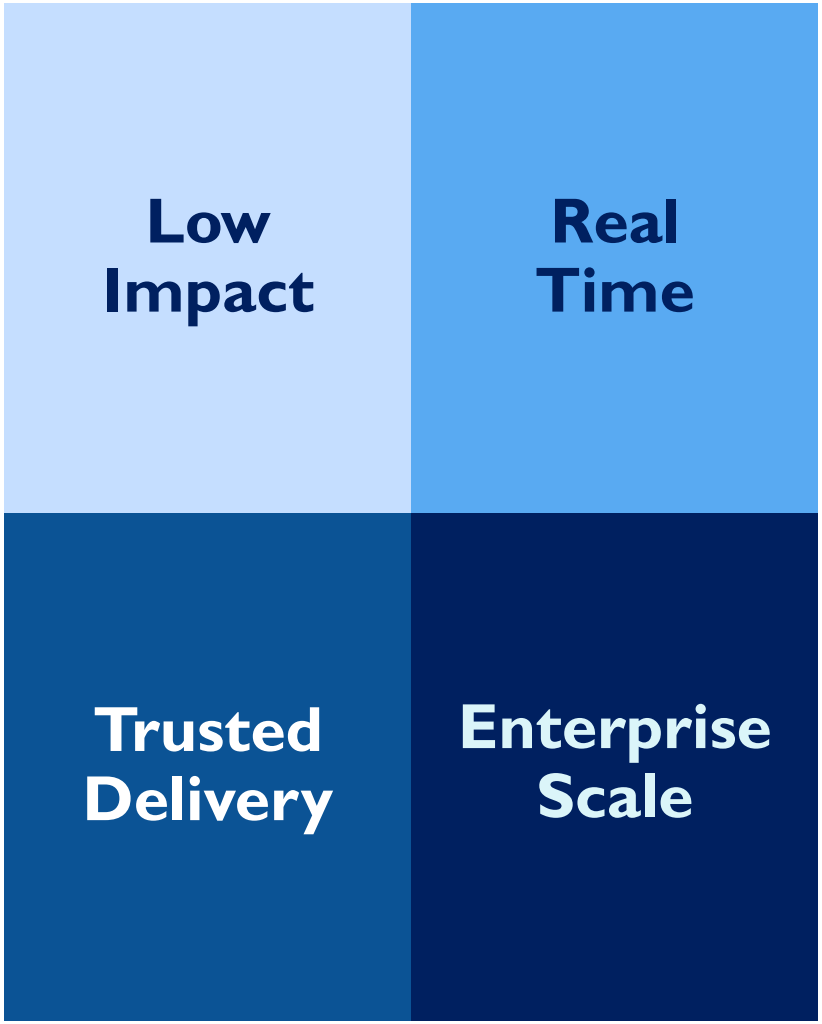
Replication - *Core Values* *(More Why)*

Optimize resource utilization
- Send only the changes with continuous "feeds". Read Source only for "refresh"

Extend Application Availability
- Shorten batch windows by streaming changes in real time as they occur

Improve the "Bottom Line"
- Fresh data improves business results
- Reduced network traffic/costs
- Audit capabilities for trusted data

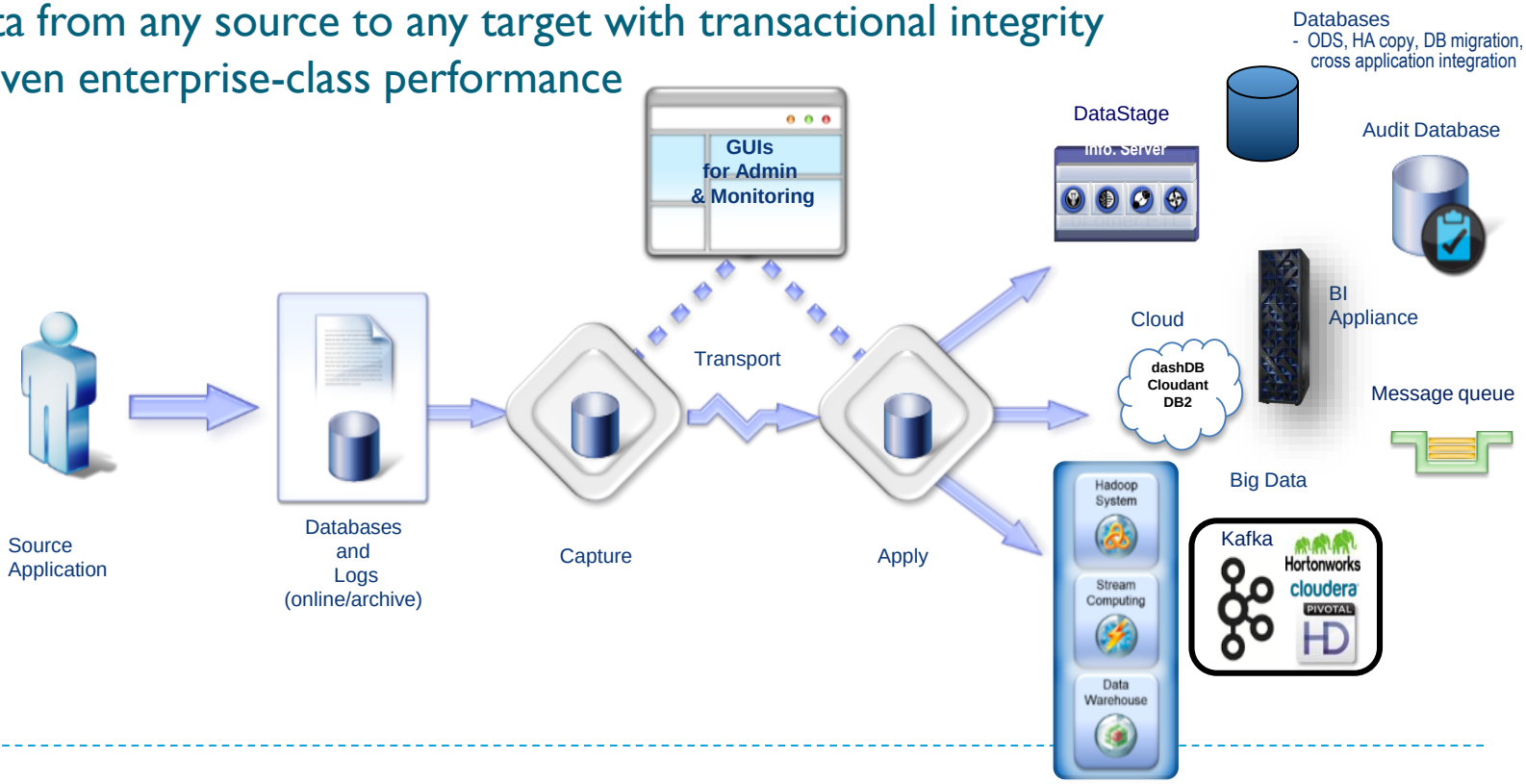
Enterprise ready
- Transactional integrity with very high volume throughput and low latency



IBM's Data Replication (IDR) Portfolio - ***What***

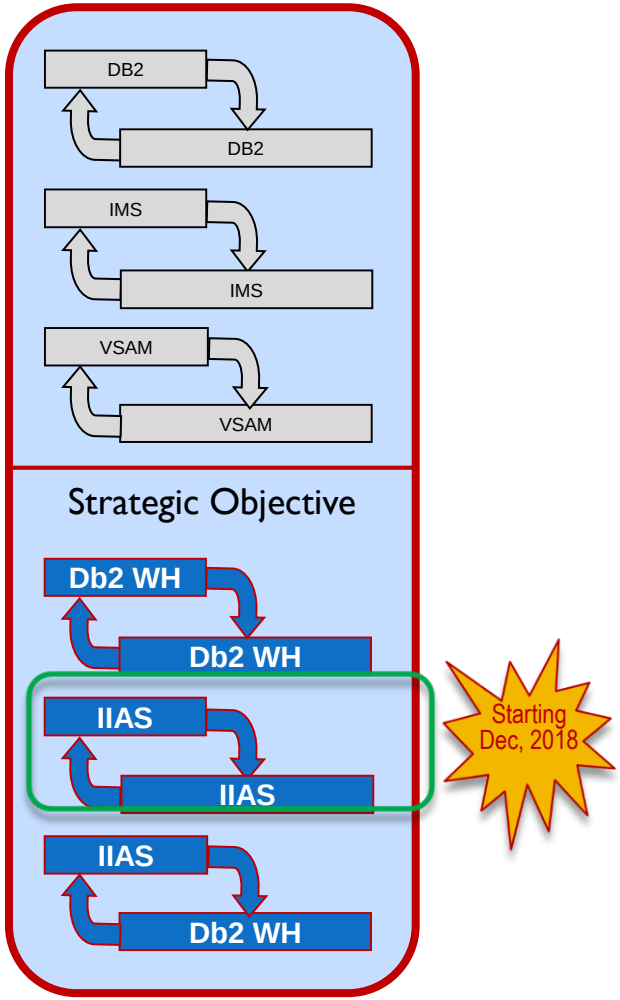
Dynamically synchronize changing structured data with diverse targets, apps, platforms

- Real time updating using continuous delivery with full refresh
- Low impact on source applications using log files for continuous, not source data
- Data from any source to any target with transactional integrity
- Proven enterprise-class performance

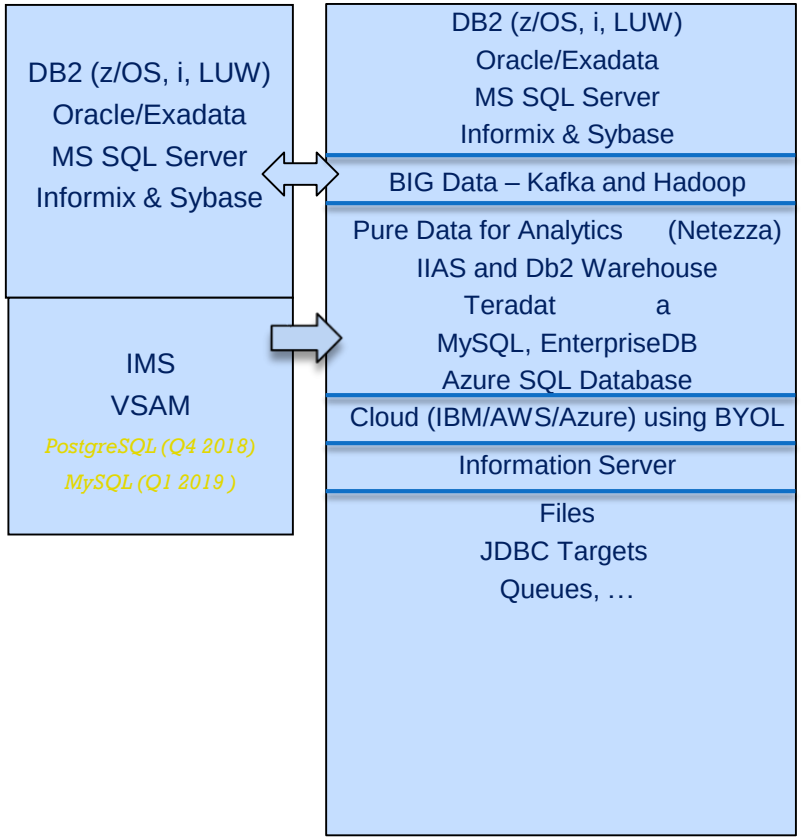


Poll 3: What are the purposes of replication in your organization?

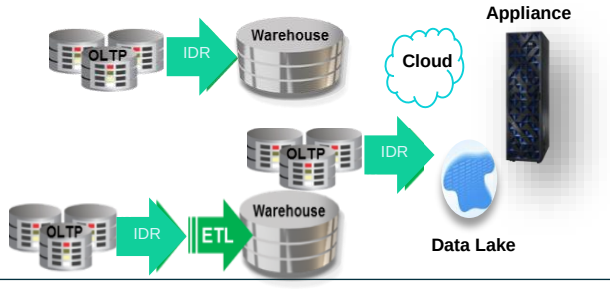
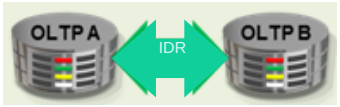
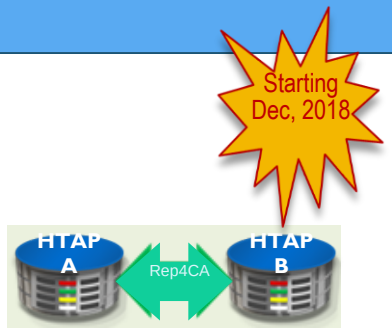
Always On Information



Business in Real Time



IBM's Replication Portfolio – ***Technologies with Use Cases***

	Business in real time	*Always on Information*	
Architecture			
Usage	CDC direct apply for real time analytics, application integration: - Hadoop (WebHDFS) - Cloud & On-Premise • Cloudant • Db2 Warehouse • IIAS • BigInsights on Cloud - Kafka - DataStage - General Relational DB Targets • All Db2, MS SQL, Oracle/Exadata, … CDC and Classic CDC	Stand-alone replication and integration with GDPS Active/Active: - Any Db2 to/from Any Db2 - IMS to IMS - VSAM to VSAM Q-Rep / IMS Rep / VSAM Rep	Containerized replication tightly integrated with Db2 Warehouse and IIAS: 1. IIAS to/from IIAS (Q4 2018) 2. IIAS to/from Db2 Warehouse 3. Db2 Warehouse to/from Db2 Warehouse Rep. for Continuous Availability

Poll 4: Do you replicate data from IBM z Systems?

Replication: Key New Capabilities

1) Replication from Db2 for z/OS using a new Remote Capture source engine

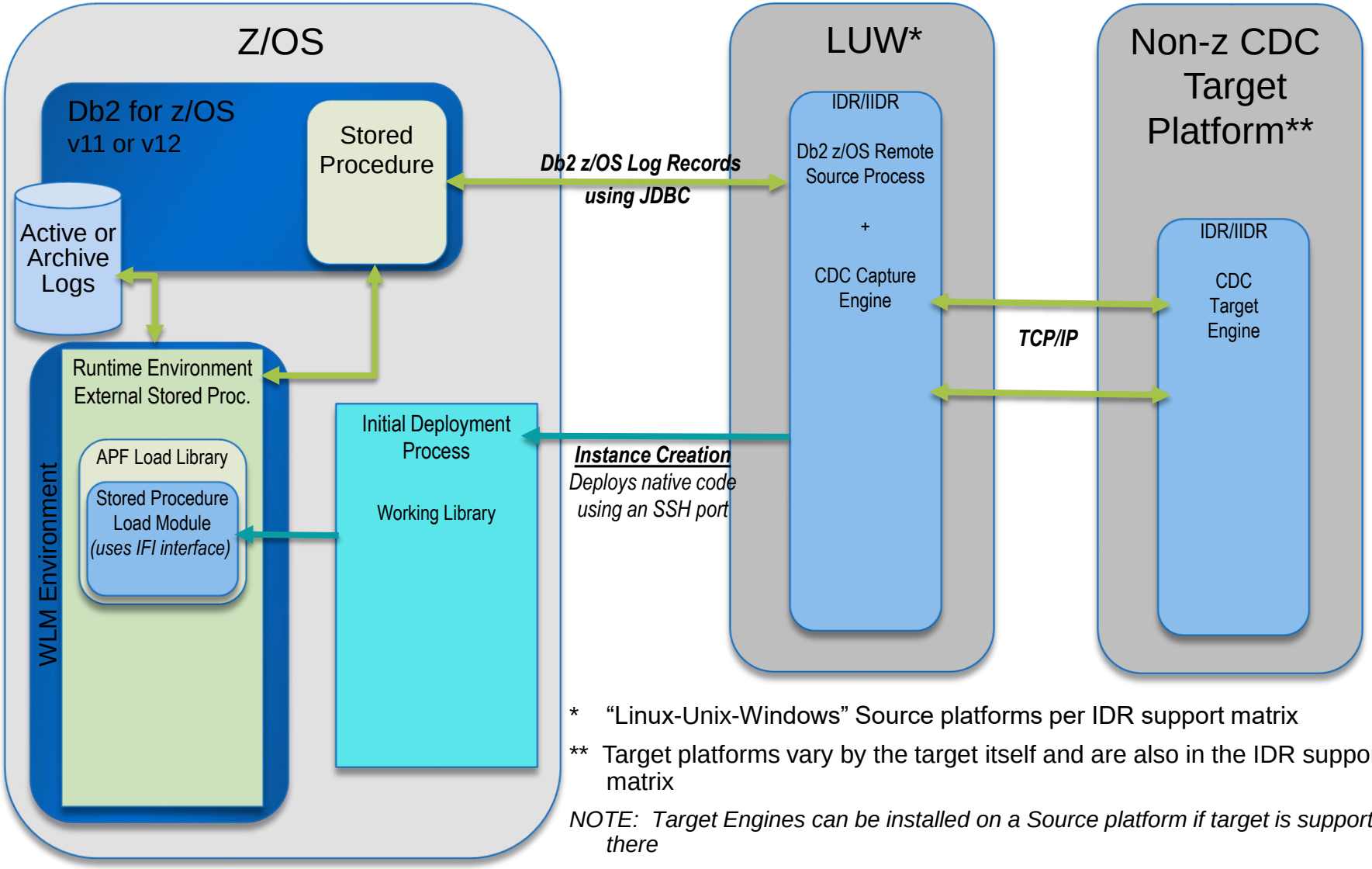
Available today!

Remote Capture from Db2 z/OS – ***Business Value***

IBM Data Replication for Db2 z/OS Remote Source

- Share mission critical Db2 z/OS data with new environments for Analytics, integration and cloud projects including targeting:
  - RDBMS data warehouse
  - Kafka hub
  - Information Server (ETL solutions)
  - OLAP appliances (IIAS) and databases
  - HDFS clusters (Hadoop)
  - Cloud targets
- Deploy the replication capture remotely from the mainframe
  - **Reduce z/OS MIPS needed to replicate Db2 for z/OS data by up to 90%**
  - **Reduce dependency on specialized System z skills for:**
    - Deploying: installation from LUW platform
    - Configuring: via Management Console UI or scripting
    - Monitoring: via Management Console UI or scripting

Remote Capture from Db2 z/OS – ***Deployment and Operations***



* "Linux-Unix-Windows" Source platforms per IDR support matrix

** Target platforms vary by the target itself and are also in the IDR support matrix

NOTE: Target Engines can be installed on a Source platform if target is supported there

Remote Capture from Db2 z/OS – ***Deployment and Operations***

Deployment

No SMP/E Install, No agent on System z

- **Stored Procedure** deployed via SSH during install process into Advanced Program Facility (APF) authorized library
- User ID on Db2 for z/OS that invokes the Stored Proc needs authority to execute IFI306 calls

Operations (Why Reduced z/OS MIPS?)

- Only the Stored Procedure (which invokes the IFI360 interface) runs on System z
- LUW remote capture engine in turn takes care of log record parsing, transaction processing, filtering, etc.
- SP only responsible for continuously providing byte buffers of log data to the CDC Capture engine

Operations (Connections, Memory..)

- JDBC sessions/connections needed for log reading and initial refresh
- Minimal memory needed on System z: no above the bar memory needed
- Memory allocated to the LUW remote capture engine is configurable

Premier Business Partner IBM

2) Replication from Db2 z, i, LUW, Oracle, SQL Server, … into Kafka to feed an Operational Data Store (ODS) or Data Lake

Available today!

Premier Business Partner IBM

Hadoop HDFS and Kafka

Hadoop HDFS

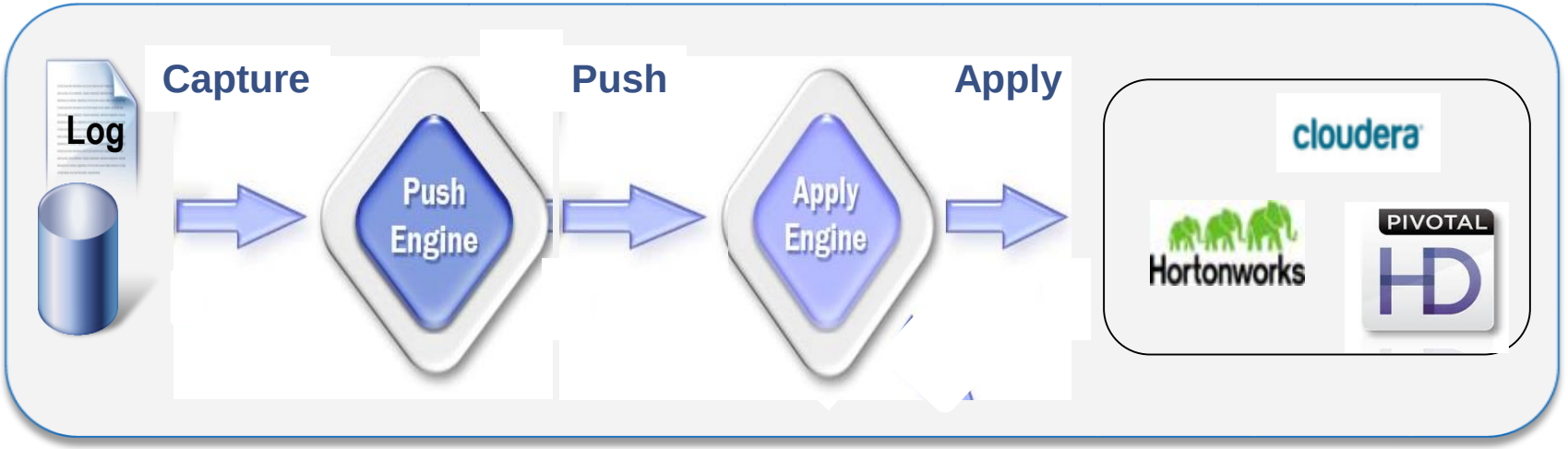
- Designed to process *large relatively static data sets*
- Bulk *append only*, no update resulting in many files, more storage
- HDFS file system designed to distribute copies of data across *commodity nodes* to provide availability & scalability at low cost
- *No set format* to the data, Source/Consumer agree on record layouts in advance

Kafka

- Designed from the outset to deal with *constantly changing events/data*
- Built in Insert plus log compaction (Delete) to *emulate an update*
- Also exploits *commodity nodes* for scalability, availability and cost but with *self managed storage* dramatically reduced footprint
- Storage is *self described JSON (Java Script Object Notation)* document wrapped in Apache Avro binary format

Replication's Real Time Creation of Hadoop Files

- Leverage webHDFS to replicate data directly to any Hadoop distribution
  - Kerberos authentication for secure data movement to Hadoop
  - In-memory buffering with HDFS.Append API to write predetermined amounts of data (configurable)
  - Fully featured, graphical and programmatic configuration and monitoring



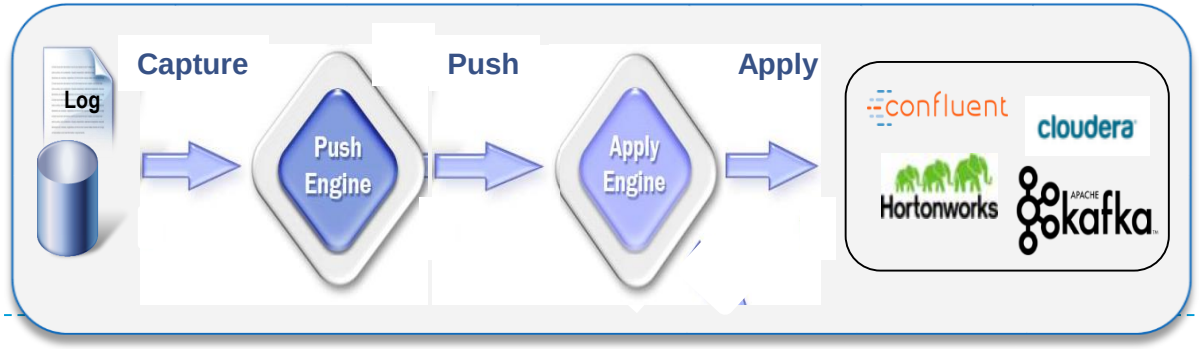
Replication's Real Time Data Feeds to Kafka

- Two Write methods:
  - **REST API** … well suited to targeting managed environments, restricted access
  - **Native java API** … higher performing for inside the firewall or with "tunnels"

- Some Details:
  - Targeted Kafka must be at 0.10 standard level (a.k.a. Kafka 10) or higher
  - We recommend a schema registry service and corresponding deserializer that supports the Confluent open APIs

OR

  - Use provided Kafka Custom Operator Process (KCOP) that eliminates the use of a schema registry



How are IBM CDC Kafka target customers using this?

What are CDC Kafka target customers doing?

- *Architectural Data Hub / Landing Zone*
  - The central point in the architecture for all data to be landed and consumed by various apps
- *Staging for Data Lake and other Analytics platforms*
  - Handles OLTP transactions, staging results for consumption by non-OLTP targets
    - Hadoop/HIVE, Column Store Databases, Key Value Databases (Aggregation, Net Change)
- *Staging for the Cloud*
  - Data is replicated to Kafka, in some cases manipulated (Compaction, Streams Jobs), Results / original data then copied to the Cloud
- *Delivering OLTP data directly to Cloud*
  - Data written to Cloud-hosted Kafka for enterprise use

THE FILLMORE GROUP
Relational Database Solutions

Apache Hadoop/Hive Data Lake: Parts and Repair Services Firm

Challenge

- Moving to a Hadoop/Hive based data lake for a consolidated view of the data available from multiple subsidiaries
- Need for a standard/normalized pattern of feeding the data lake from disparate transactional databases
- Insulate downstream applications from changes e.g. schema changes in their transactional RDBMs

Solution

- Use Data Replication to replicate Oracle, iSeries, Db2 transactional data in geographically separate data centers in real time to a Hadoop/Hive data lake via a Kafka landing zone
- Schema evolution support in Replication's Kafka apply means downstream applications are agnostic to schema changes on the original transactional systems.
- Data from all sources represented in a common AVRO format in Kafka enabling standardized consumption/processing of data.

Business Benefits

- Better coordination in the IT infrastructure for division analytics enablement.
- Consistent message format to enable use of a standard Kafka to Hive connector.
- Option to develop more real time processing applications now enabled.

3) Replication from any supported source to feed IBM Integrated Analytics System (IIAS) and Db2 Warehouse

Available today!

IDR's CDC Native Apply for IIAS & Db2 Warehouse

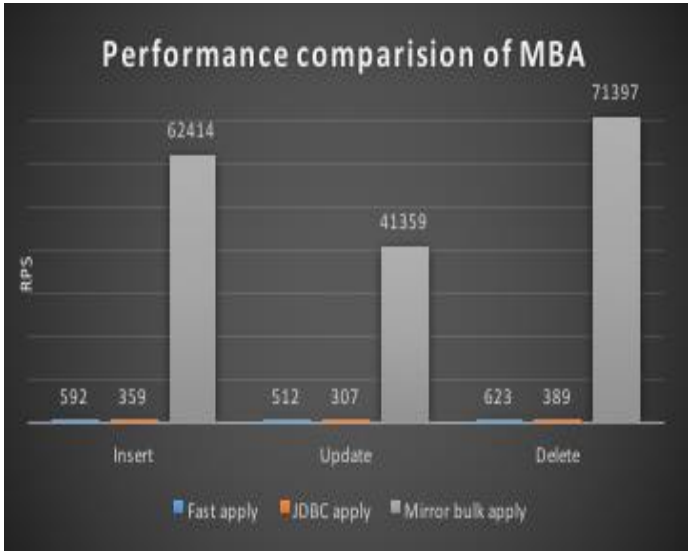
Features:

- ✓ Feeds changed data from ALL supported IIDR (CDC sources) into IIAS and Db2 Warehouse **directly** in one hop
- ✓ **CDC** Fast Apply has a NEW Mirror Bulk Apply option that supports Db2's External Table as the apply mechanism for fast ingest into IIAS and Db2 Warehouse BLU Tables
- ✓ Selectable Type 4 JDBC connection recommended for Apply to remote Db2 targets like IIAS and Db2 Warehouse
- ✓ Vanilla CDC Fast Apply usually sufficient for row based Db2 tables on target
- ✓ Mirror Bulk Apply Capability is embedded in IBM Db2 Analytics Accelerator (IDAA) for **pre-configured** incremental feeds from **Db2 for z/OS**
- ✓ Standard CDC UI and APIs/scripting for configuration, operation control, and monitoring
- ✓ Remote CDC apply to Db2 Warehouse and IIAS does not require Db2 catalog when JDBC apply is used for refresh

THE FILLMORE GROUP
Relational Database Solutions

IDR's CDC Technology Native Apply for IIAS and Db2 Warehouse: Performance

- Mirror Bulk Apply**: 60K**+ ops/sec seen in IBM test environments
- **Much better** throughput compared to other apply methods in test scenarios
- Similar results seen with remote and local CDC applies/External Tables



Premier Business Partner IBM

4) Replication from PostgreSQL

Available 1Q19 - Early support program

Premier
Business
Partner
IBM

PostgreSQL Capture Engine

New Source engine added to existing IDR and IIDR offerings:

- PostgreSQL is the #1 growth database in today's market
- Share PostgreSQL data into new environments for Analytics and integration
- Delivered via continuous delivery (fixpack) into IDR/IIDR v11.4
  - *Users with sufficient entitlements (PVUs) can immediately use this new engine!*

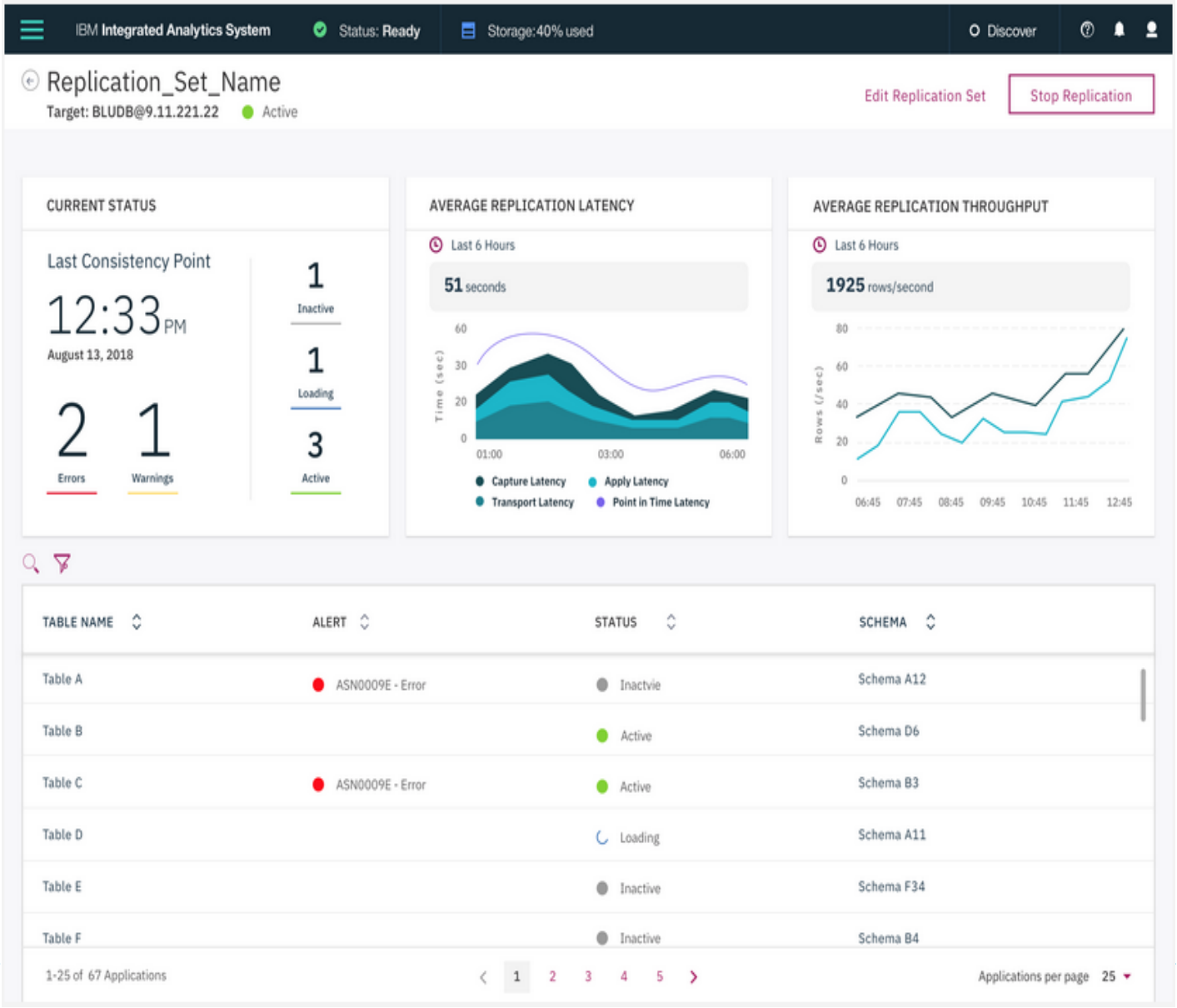
5) Data Replication for Db2 Continuous Availability

Available today!

Replication for Continuous Availability

- IIAS <-> IIAS Now
- Db2WH <-> Db2WH and Db2WH <-> IIAS 2019
- Provides asynchronous transactionally consistent Db2 replication
- Near real-time mirror of primary platform data on secondary
- Both primary and secondary are active

User Interface – integrated into IIAS console



Poll 5: Which of the IBM replication enhancements presented sounds *most* interesting based on your organization's needs?

Replication Strategic Plans and Initiatives 2018 and Beyond ("Business in Real Time" focus)

What's New with Data Replication

Recent or Q4 2018-Q1 2019 Deliverables

- GDPR Guidance across the board (Q-Rep, CDC, and Classic)
- Q-Replication
  - Usability: e.g. Commit by row count, oldest commit time
  - Monitoring & Resiliency: e.g. Latency exceed warnings
  - Improved handling of: multi-row inserts, LOB data, Changing column sizes…
- IMS & VSAM:
  - Resiliency: Force Apply VSAM
  - Usability: API driven configuration for IMS-IMS and VSAM-VSAM
  - Currency: IMS 15 support

NEW Models

Replication premium add-on for ICP4D

- *Containerized Delivery*
- *Remote capture from Db2 z/OS into Db2WH on ICP4D (Q1 2019)*

Replication for Continuous Availability

- CDC
  - Sources & Targets: Oracle performance (S), Db2 LUW Binary/Varbinary (S&T), JDBC Type 4 Driver (S&T), ***Oracle 18C, Oracle extended datatypes, Oracle TDE support***
  - ***Kafka: Improved resiliency,*** Transactionally consistent consumers, KCOP
  - Remote capture from SQL Server
  - IIAS/Db2 Warehouse Performance Optimization
  - ***Remote Db2 z/OS Source Engine***
  - ***PostgreSQL Source Engine***
  - ***MySQL Source Engine***
  - ***Oracle: Source both Little and Big Endian remotely from Linux***
  - ***MongoDb Target: support via user exit***

What's New with Data Replication: Detailed Roadmap

https://bigblue.aha.io/published/19ec783a1fa457b43dd486612257f845?page=1

User demand and trends driving Replication Strategies

2019 and Beyond

- Additional DB sources and more targets
  SOURCES: Oracle Capture from standby
  TARGETS: Additional Cloud targets on AWS, Azure, IBM Cloud
- Enhanced Deployment capabilities for QREP and CDC

  SaaS deployment model for hybrid, private and public cloud integration
- Improved Usability (QREP #1, CDC #2, Classics #3)
  Configuration, setup, management, monitoring via REST APIs and integration with UIs
- Automated and "Learned" Resilience – more long term...
  More self-correcting capabilities baked in
- Integrated Security – subject to customer feedback
  Multi-factor authentication, tighter integration with external security, ...

Replication in Summary

Replication - *Wrap-up*

Proven in thousands of customer deployments around the world!

- Efficient delivery of only the changes, not the entire source
  - Low impact incremental update to reduce risk and cost
  - Near zero latency ensuring up-to-date information to the business
  - Built in source to target transformations reduce or eliminate hand coding
- Non-intrusive to applications and databases
  - Native DB logs, NO use of database triggers
  - Readily integrated into existing IT operations reducing operational risk
  - Data event aware to enable broader business process integration when needed
- Enterprise Ready
  - Transaction consistency maintained to preserve unit of work referential integrity
  - Granularity with before and after image of transactional changes
  - Fault tolerance ensuring recovery to last committed transaction

IBM Oracle Attack *Special Offer*

Oracle GoldenGate Licenses Qualify

- **IBM offers current Oracle customers software license discounts > *50%***
- **Discounts are calculated on current Oracle *support* costs for GoldenGate**

The math:

If Oracle GG support = 20% of new license cost, you can purchase an equal quantity of IBM InfoSphere Data Replication for ~ *10% of the cost of new licenses!!*

The fine print:

1) ***This requires an IBM Special Bid***
2) ***You must share your current OGG support costs with IBM***
3) ***You need to factor in the cost and effort to migrate to IIDR***

Still, it's an amazing deal!

Replication @ Think San Francisco 2/12 - 15

Established Bank Modernizes Legacy Data with a Data Lake and Creates New Business Opportunities

- **Thursday, 2:30 PM - 3:10 PM | Session ID: 1583A**
- **Moscone South, Level 2 | San Francisco Ballroom 212**

A major establishment bank seeks to modernize its legacy mainframe VSAM data together with other relational data stores. The goal is to provide a 360-degree view of its customers by delivering data into a data lake in near real-time. IBM InfoSphere data integration technologies are a core part of this multi-vendor environment. See how IBM's InfoSphere data replication technologies provide the capture and delivery mechanisms that not only enable this near real-time data lake to create new business capabilities, but also reduce the cost and burden of current legacy data methods.

https://myibm.ibm.com/events/think/all-sessions/session/1583A

Resources and Contacts

THE FILLMORE GROUP
Relational Database Solutions

Replication Professional Services and Training

- Architecture and Implementation
  - Resilient, robust, scalable deployments
- Healthchecks
  - Monitoring and automation
  - Scripting
  - Schema evolution
- Version Upgrades
  - InfoSphere Data Replication v10.2.1 — EoS: 2018-04-30
  - InfoSphere Data Replication v11.3.0 — EoS: 2019-09-30
- IBM Authorized Training
  - Change Data Capture
  - Q Replication

Kim May

- kim.may@thefillmoregroup.com
- 410-465-6335

Replication - Get Involved!

- Submit and vote on Requests for Enhancements (RFEs)
- Check roadmaps

https://bigblue.aha.io/published/3bcb05bd498d8253b81becf647a73350?page=1

Replication Resource Links

- Solution Brief

What's New

- InfoSphere Change Data Capture (ICDC)
- Q Replication
- ICDC Classic Data Replication and Federation
- Big Data Solution Brief
- Developer Works (IBM Data Replication Community)

THE FILLMORE GROUP
Relational Database Solutions

Thank you!

Kim May, Vice President Business Development

kim.may@thefillmoregroup.com

www.thefillmoregroup.com

Premier Business Partner IBM