

**THE FILLMORE GROUP**  
Relational Database Solutions

# **IBM Replication – Data Delivery** **The DataView Show – June 2020**

**A Premier IBM Business Partner**





## Agenda

- ▶ Introduction
- ▶ Replication Review
  - ▶ Why do we replicate data?
  - ▶ Competitive Landscape
- ▶ IBM Replication Products
- ▶ IBM Replication Components



## Agenda (contd.)

- ▶ Sources
  - ▶ Classic mainframe: Db2 for z/OS, IMS, VSAM
    - ▶ Remote Captures
  - ▶ Heterogeneous
- ▶ Targets
  - ▶ Native targets
  - ▶ Federated targets
  - ▶ Message queues
    - ▶ Event Publishing
    - ▶ Kafka



## Agenda (contd.)

- ▶ Key features
- ▶ Upcoming enhancements
  - ▶ Remote Capture for VSAM
  - ▶ Integrated Synchronization
- ▶ Tooling
- ▶ Decision criteria
- ▶ Purchasing
- ▶ Resources



## Introduction

### The Fillmore Group, Inc.

- ▶ Founded in the US in Maryland, 1987
- ▶ IBM Business Partner since 1989
- ▶ Delivering IBM authorized education since 1994
- ▶ IBM Gold Consultant since 1998  
<https://www.ibm.com/analytics/ibm-gold-consultants>
- ▶ IBM Champions since 2009





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





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

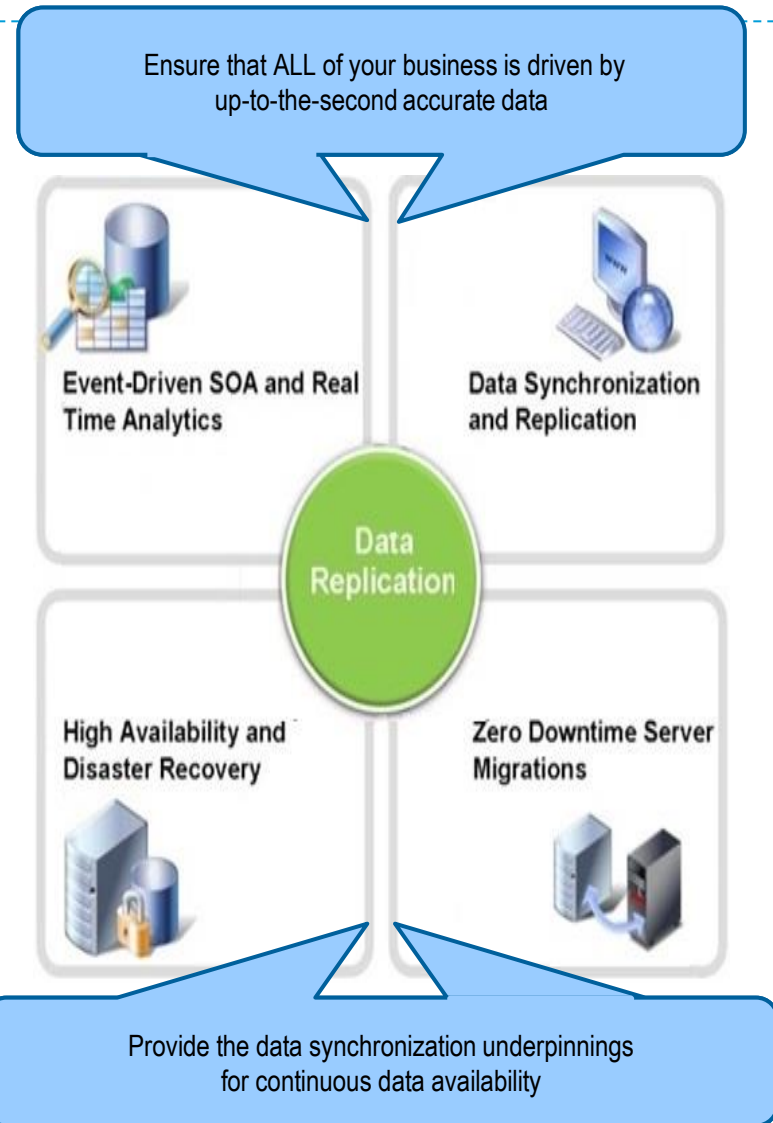
Premier  
Business  
Partner





## 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



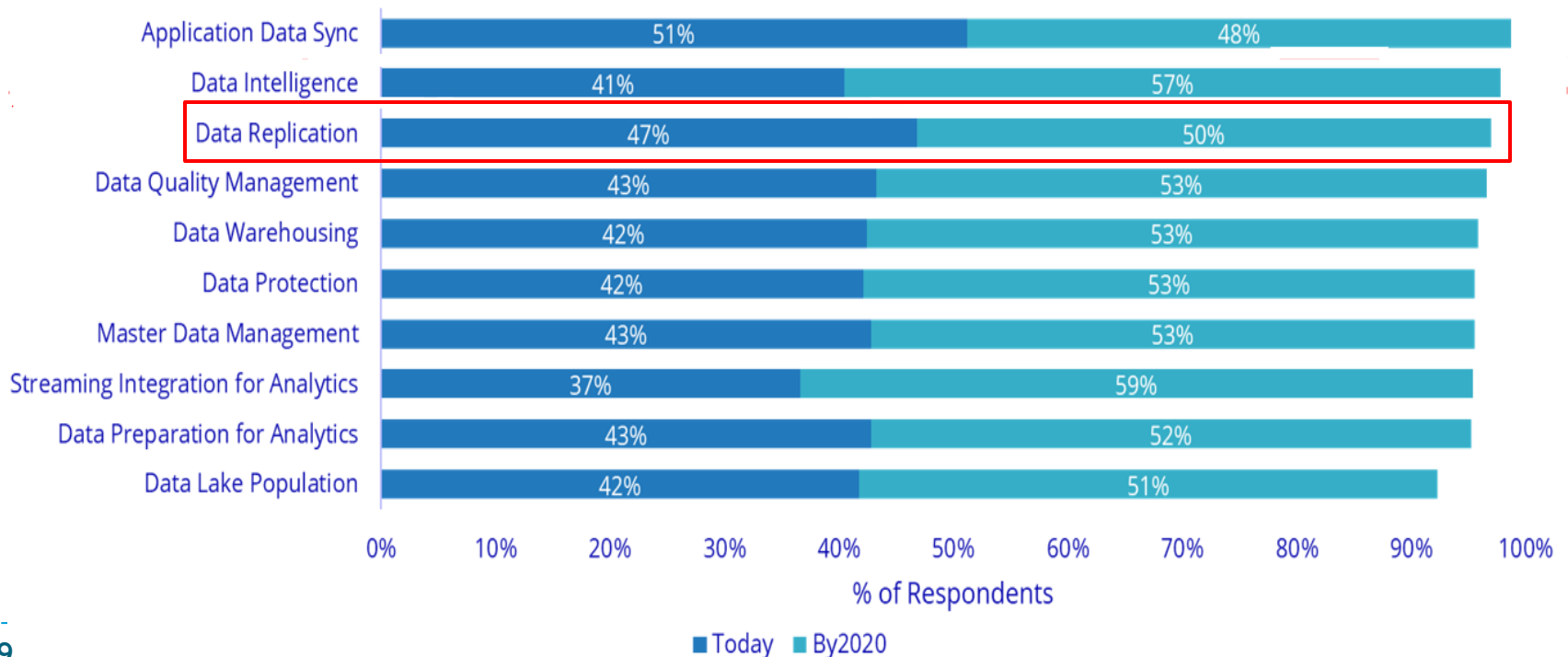




## Keeping data synchronized among applications is and will be the most prevalent use case for data integration

Please indicate what use cases of data integration and integrity solutions have been implemented, being implemented or are planned for in your environment.

### DII Software Use Cases





## 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”

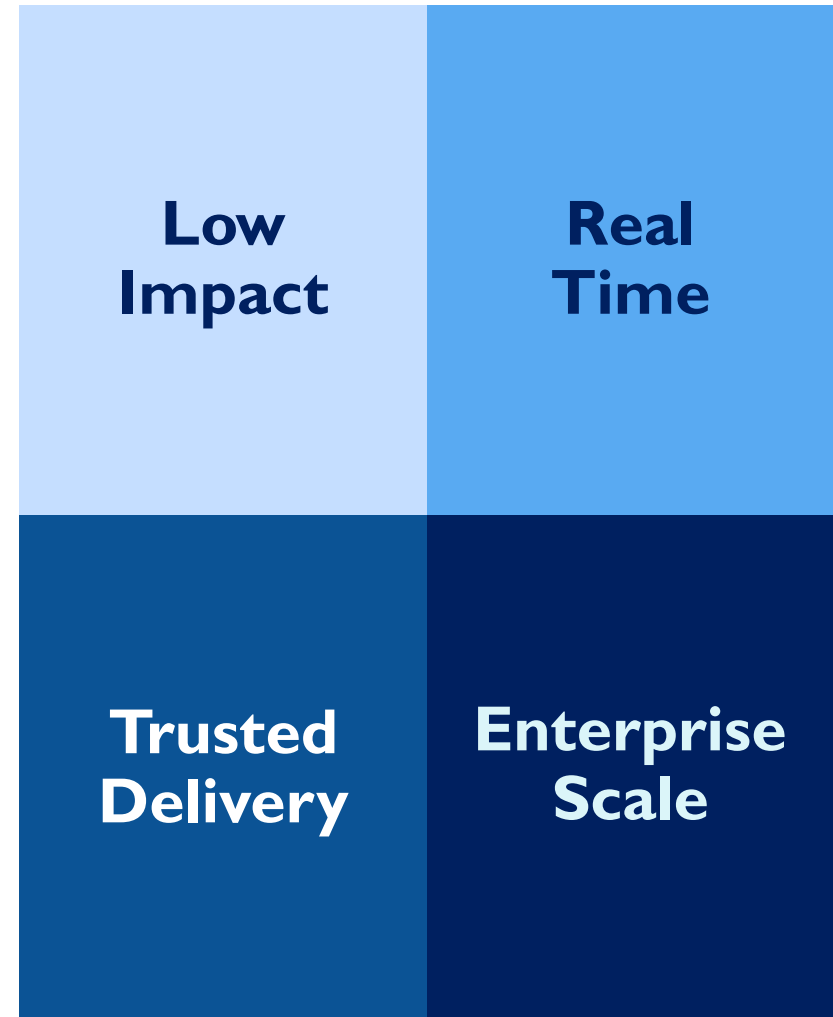
Fresh data improves business results

Reduced network traffic/costs

Audit capabilities for trusted data

Enterprise ready

Transactional integrity with very high volume throughput and low latency





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





# Gartner Magic Quadrant for Data Integration Tools

August 2019

ID G00369547

<https://www.gartner.com/doc/reprints?id=I-IQJ15E39&ct=190923&st=sb>





## IBM Strengths - Gartner

- ▶ **“Depth of integration offering.** Reference customers highlighted the completeness of IBM’s holistic data integration suite, including its rich functionality, variety of prebuilt functions and connectors, and its overall performance.”
- ▶ **“Diverse data integration delivery styles.** Reference customers use IBM’s products for traditional data delivery styles (data replication, batch processing), as well as more complex data delivery styles (including data synchronization and stream data integration). They praised IBM’s data integration tool portfolio for its ability to deliver complex data integration requirements that demand combinations of traditional and modern data integration styles, such as data replication, data virtualization and stream data integration for real-time analytics.”



## IBM Strengths – Gartner (contd.)

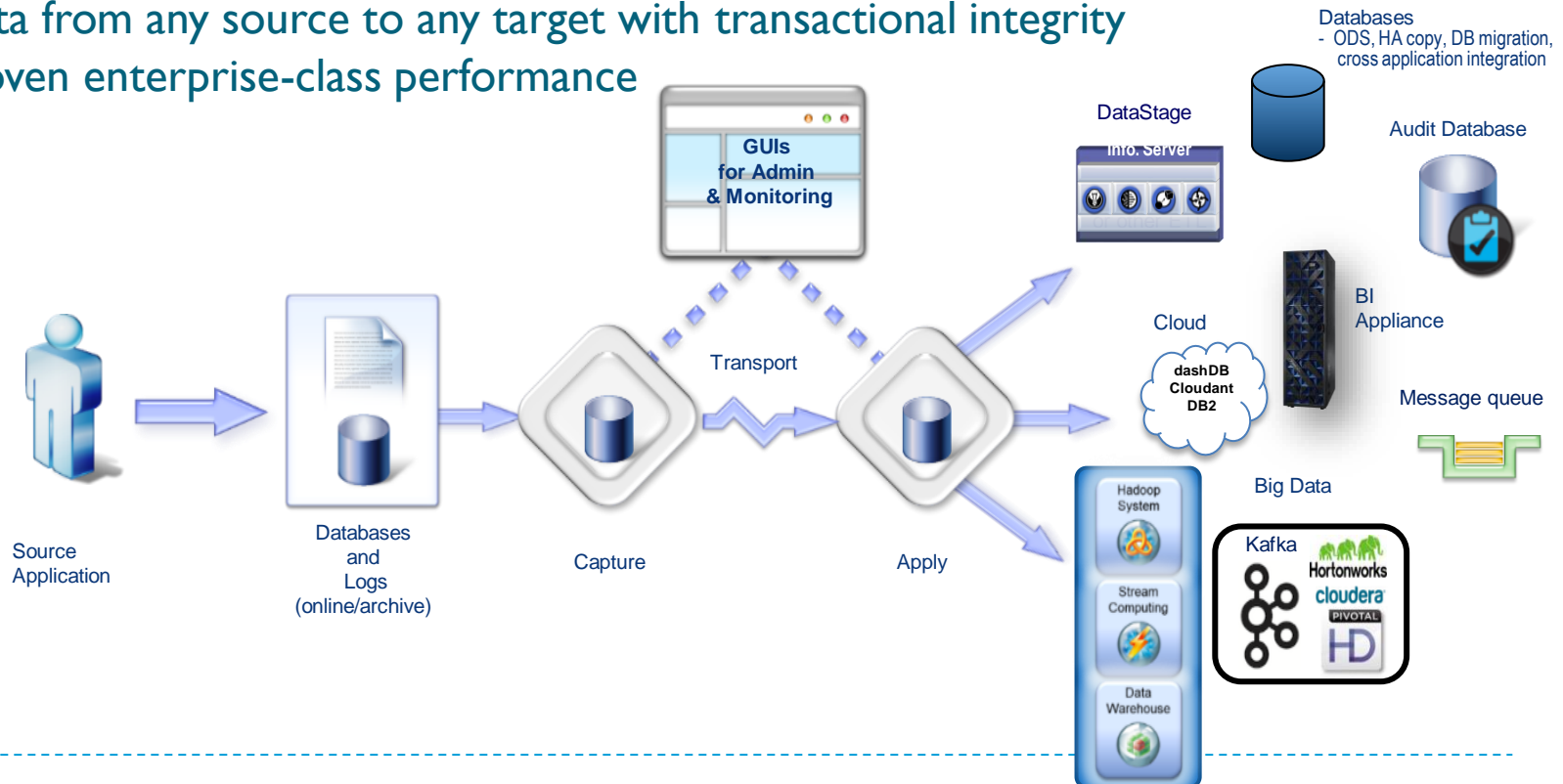
- ▶ **“Brand awareness and market presence.** IBM’s size and the global coverage of its business systems, infrastructure platforms and analytics solutions enable it to draw on a huge customer base and a wide product distribution model for positioning its data integration tools. Broad usage of IBM technologies within its customer base has driven the wide availability of implementation service providers and approaches to solving complex integration challenges.”



# IBM's Replication 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





## **IBM Replication Products**

- ▶ IBM Data Replication (IDR)
- ▶ IBM InfoSphere Data Replication (IIDR)
- ▶ Db2 for z/OS Remote Capture
- ▶ IBM InfoSphere Classic Replication (for VSAM, IMS, and with Federation to others)
- ▶ IBM Data Replication for Db2 Continuous Availability





## IBM Replication Components

- ▶ SQL Replication
  - ▶ Staging Tables
  - ▶ Used in broadcast topology
- ▶ Q Replication
  - ▶ IBM MQ
  - ▶ High volume, low latency
- ▶ Change Data Capture (CDC)
  - ▶ TCP/IP
  - ▶ Heterogeneous



## Supported source databases

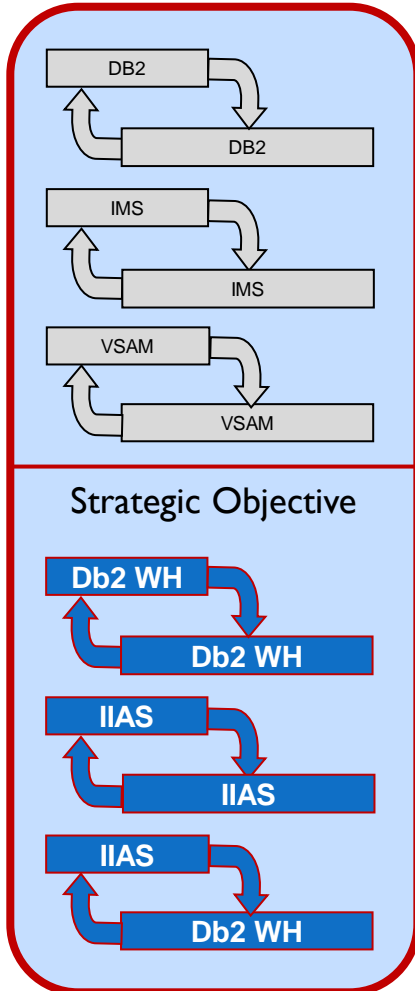
IBM Db2 for Linux, UNIX and Windows (LUW)  
IBM Db2 for i  
IBM Db2 for z/OS  
IMS  
Microsoft SQL Server  
Oracle  
Sybase  
Informix  
Db2 on Cloud  
VSAM  
PostgreSQL  
Db2 Warehouse on Cloud  
Db2 Warehouse  
MySQL

## Supported target databases and middleware applications

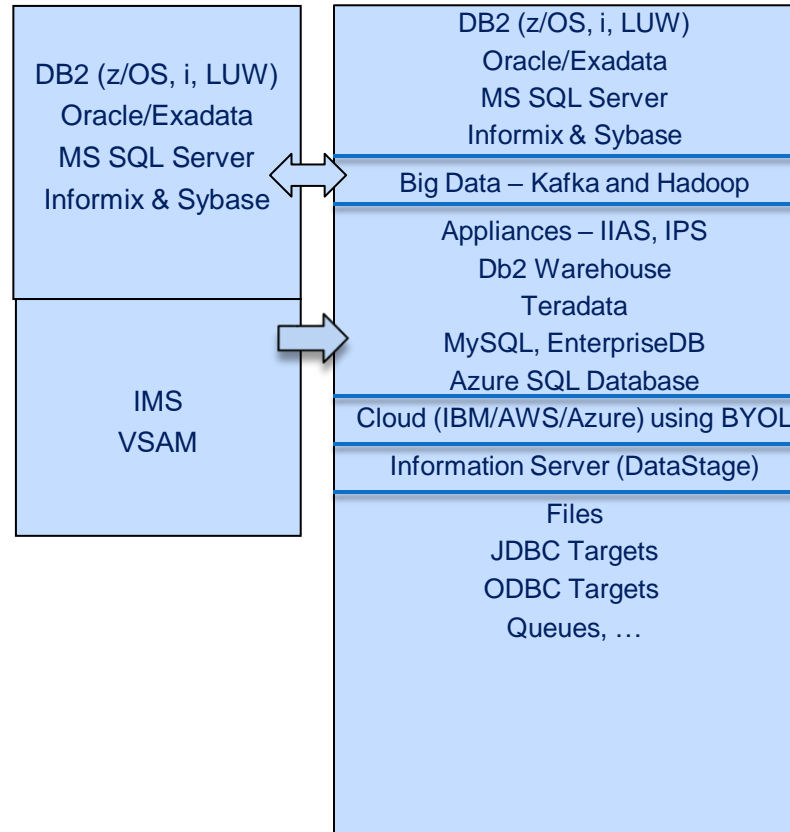
IBM Db2 for Linux, UNIX and Windows (LUW)  
IBM Db2 for i  
IBM Db2 for z/OS  
IBM InfoSphere DataStage  
IBM Netezza  
IBM Informix  
Microsoft SQL Server  
Microsoft Azure SQL Database  
Microsoft Azure SQL Database Managed Instance  
Oracle  
Sybase  
CDC Replication Engine for Event Server  
CDC Replication Engine for FlexRep (JDBC)  
IBM Cloudant  
IBM Integrated Analytics System  
Apache Hadoop  
Apache Kafka  
Db2 Warehouse on Cloud  
Db2 on Cloud  
Db2 Warehouse  
IBM MQ for z/OS (using Classic CDC for z/OS)  
Teradata



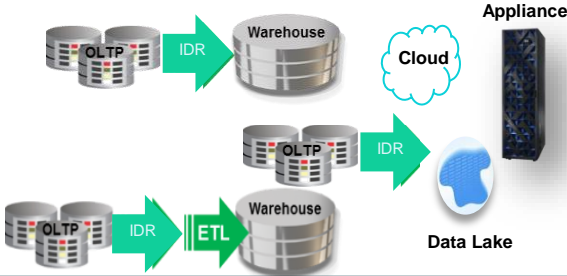


### Always On Information



### Business in Real Time



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

	<i>Business in real time</i>	<i>Always on Information</i>	
<b>Architecture</b>			
<b>Usage</b>	<p>CDC direct apply for real time analytics, application integration:</p> <ul style="list-style-type: none"> <li>- Hadoop (WebHDFS)</li> <li>- Cloud &amp; On-Premise               <ul style="list-style-type: none"> <li>• Cloudant</li> <li>• Db2 Warehouse</li> <li>• IIAS</li> </ul> </li> <li>- Kafka</li> <li>- DataStage</li> <li>- General Relational DB Targets               <ul style="list-style-type: none"> <li>• All Db2, MS SQL, Oracle/Exadata, ...</li> </ul> </li> </ul> <p style="text-align: center;"><b>CDC and Classic CDC</b></p>	<p>Stand-alone replication and integration with GDPS Active/Active:</p> <ul style="list-style-type: none"> <li>- Any Db2 to/from Any Db2</li> <li>- IMS to IMS</li> <li>- VSAM to VSAM</li> </ul> <p style="text-align: center;"><b>Q-Rep / IMS Rep / VSAM Rep</b></p>	<p>Containerized replication tightly integrated with Db2 Warehouse and IIAS:</p> <ol style="list-style-type: none"> <li>1. IIAS to/from IIAS</li> <li>2. IIAS to/from Db2 Warehouse</li> <li>3. Db2 Warehouse to/from Db2 Warehouse</li> </ol> <p style="text-align: center;"><b>Replication for Continuous Availability</b></p>



# Key Features





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



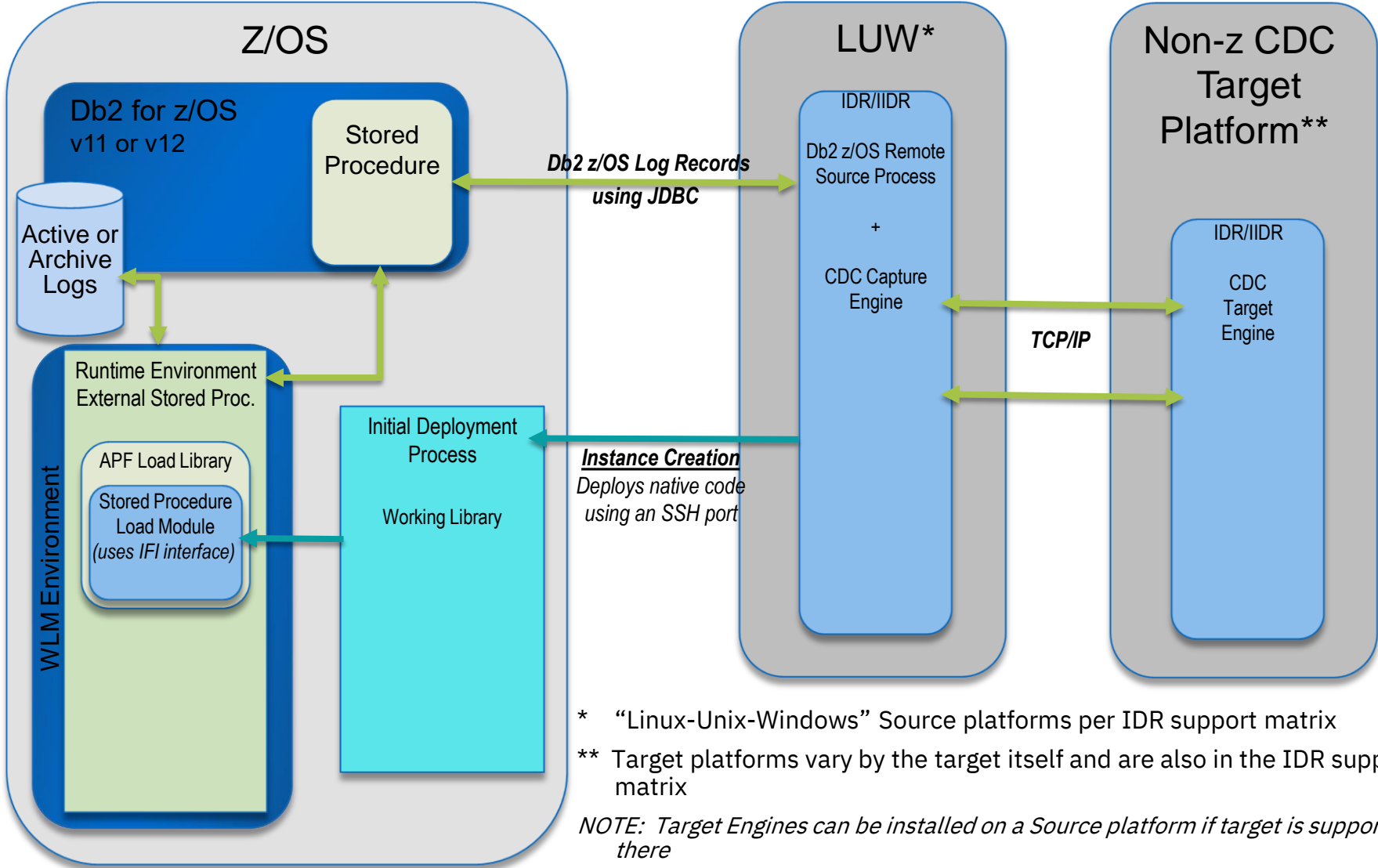


## Remote Capture from Db2 z/OS – *Business Value*

### *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
  - OLAP appliances (IIAS) and databases
  - Kafka hub
  - HDFS clusters (Hadoop)
  - Information Server (ETL solutions)
  - 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 50%**
  - **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*







# 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

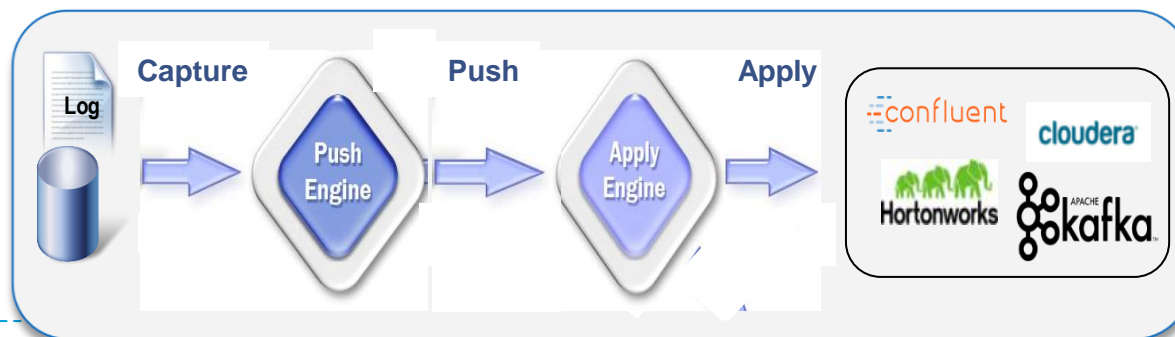


## Replication's Real Time Data Feeds to Kafka

- ▶ 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





## What are CDC Kafka 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



## PostgreSQL Capture Engine

- ▶ 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





## Replication for Continuous Availability

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





## Upcoming Enhancements\*

- ▶ Remote Capture for VSAM – June 2020
  - ▶ MIPS offload
  - ▶ Requires VSAM logging
  - ▶ Early availability/beta
- ▶ Integrated Synchronization

<http://ibm.biz/Namik-AccessDb2Fillmore>





## Tooling

- ▶ **Out**
  - ▶ Replication Center
  - ▶ Q Replication Dashboard
- ▶ **In**
  - ▶ Watson Knowledge Catalog
  - ▶ Watson Studio
  - ▶ Enterprise Db2 Unified Console

Premier  
Business  
Partner



# User Interface – integrated into IAS console

IBM Integrated Analytics System | Status: Ready | Storage: 40% used | Discover

## Replication\_Set\_Name

Target: BLUDB@9.11.221.22 | Active

[Edit Replication Set](#) [Stop Replication](#)

### CURRENT STATUS

Last Consistency Point: 12:33 PM, August 13, 2018

- 1 Inactive
- 1 Loading
- 2 Errors
- 1 Warnings
- 3 Active

### AVERAGE REPLICATION LATENCY

Last 6 Hours: 51 seconds

Time (sec) vs. 01:00, 03:00, 06:00

- Capture Latency
- Apply Latency
- Transport Latency
- Point in Time Latency

### AVERAGE REPLICATION THROUGHPUT

Last 6 Hours: 1925 rows/second

Rows (/sec) vs. 06:45, 07:45, 08:45, 09:45, 10:45, 11:45, 12:45

TABLE NAME	ALERT	STATUS	SCHEMA
Table A	ASN0009E - Error	Inactive	Schema A12
Table B		Active	Schema D6
Table C	ASN0009E - Error	Active	Schema B3
Table D		Loading	Schema A11
Table E		Inactive	Schema F34
Table F		Inactive	Schema B4

1-25 of 67 Applications | < 1 2 3 4 5 > | Applications per page 25







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





## Decision criteria

- ▶ **Cost**
  - ▶ Software: licensing, maintenance
  - ▶ Implementation
  - ▶ BAU processing cycles
- ▶ Data volume
- ▶ Tolerance for data loss
- ▶ Latency
- ▶ Skills
- ▶ Simplicity vs. complexity
- ▶ Risk



## How to Buy

- ▶ **Step 1:** Identify all sources and targets
- ▶ **Step 2:** Determine whether replication will be unidirectional or bidirectional
- ▶ **Step 3:** Determine the best IBM replication solution
- ▶ **Step 4:** Figure out how to get YOUR best deal



## IBM Products and Metrics

- ▶ Processor Value Units (PVU) – used for all non-mainframe product sizing
- ▶ Resource Value Units (RVU's) – for z/OS only
- ▶ Virtual Processor Core License – for Db2 WH only
- ▶ “Per Install” target only licenses to add targets to existing replication installations



# Resources and Contacts





## 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
  - ▶ InfoSphere Data Replication v11.3.0
- ▶ IBM Authorized Training
  - ▶ Change Data Capture
  - ▶ Q Replication

**Kim May**

- [kim.may@thefillmoregroup.com](mailto:kim.may@thefillmoregroup.com)
- 410-465-6335

EoS: 2018-04-30

EoS: 2019-09-30



## What's coming: Detailed Roadmaps

- ▶ CDC Roadmap

<https://bigblue.aha.io/published/0e98b1d896d4f9d46038ca7a7bd7ecab?page=5>

- ▶ Q Replication Roadmap

<https://bigblue.aha.io/published/0e98b1d896d4f9d46038ca7a7bd7ecab?page=6>

- ▶ IMS/VSAM Roadmap

<https://bigblue.aha.io/published/0e98b1d896d4f9d46038ca7a7bd7ecab?page=7>



## Resources

- ▶ Submit and vote on Requests for Enhancements (RFEs)

<https://ibm.biz/IBM-Data-and-AI-Ideas>

- ▶ Data Replication Community

<https://community.ibm.com/community/user/dataops/communities/community-home?CommunityKey=013638d8-5a9c-4470-904a-6c9e4fdabc97>

- ▶ IBM Data Replication Community Wiki

<https://www.ibm.com/support/pages/node/1104465>





# Thank you!

Kim May, Vice President Business Development

[kim.may@thefillmoregroup.com](mailto:kim.may@thefillmoregroup.com)

[www.thefillmoregroup.com](http://www.thefillmoregroup.com)

