Les King

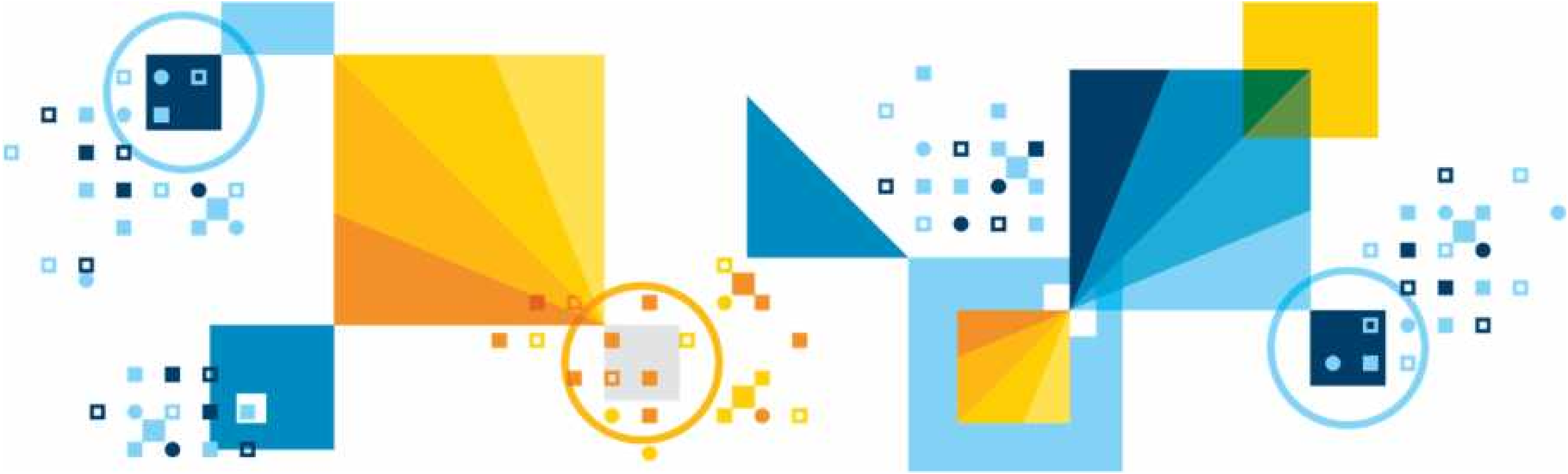
Director, Hybrid Data Management Solutions

May, 2018

lking@ca.ibm.com

ca.linkedin.com/pub/les-king/10/a68/426

Hybrid Data Management Strategy and New News !



Les King

Director, Hybrid Data Management Solutions
Professor, Big Data, Data Warehousing and Db2, Seneca College

lking@ca.ibm.com
ca.linkedin.com/pub/les-king/10/a68/426

Professional Highlights

- 27 years of Information Management, Database and Analytics
- Technical sales
- Technical customer support
- Software development
- Product / Offering management
- Product Marketing
- Product Sales
- Taught mathematics at University of Toronto
- Teaching data warehousing, big data and Db2 at Seneca College

Safe Harbor Statement

THE INFORMATION CONTAINED IN THIS PRESENTATION IS PROVIDED FOR INFORMATIONAL PURPOSES ONLY. WHILE EFFORTS WERE MADE TO VERIFY THE COMPLETENESS AND ACCURACY OF THE INFORMATION CONTAINED IN THIS PRESENTATION, IT IS PROVIDED "AS IS" WITHOUT WARRANTY OF ANY KIND, EXPRESS OR IMPLIED. IN ADDITION, THIS INFORMATION IS BASED ON CURRENT THINKING REGARDING TRENDS AND DIRECTIONS, WHICH ARE SUBJECT TO CHANGE BY IBM WITHOUT NOTICE. FUNCTION DESCRIBED HEREIN MY NEVER BE DELIVERED BY I BM. IBM SHALL NOT BE RESPONSIBLE FOR ANY DAMAGES ARISING OUT OF THE USE OF, OR OTHERWISE RELATED TO, THIS PRESENTATION OR ANY OTHER DOCUMENTATION. NOTHING CONTAINED IN THIS PRESENTATION IS INTENDED TO, NOR SHALL HAVE THE EFFECT OF, CREATING ANY WARRANTIES OR REPRESENTATIONS FROM IBM (OR ITS SUPPLIERS OR LICENSORS), OR ALTERING THE TERMS AND CONDITIONS OF ANY AGREEMENT OR LICENSE GOVERNING THE USE OF IBM PRODUCTS AND/OR SOFTWARE.

Topics for Today

- Strategy Overview
- Db2 V11.1.3.3 – Introduction !!
- Private Cloud – Introduction !!
- Flex Points and HDM Offering
- Appliance News
- Hadoop and Open Source
- Event Processing
- Next Generation Data Virtualization

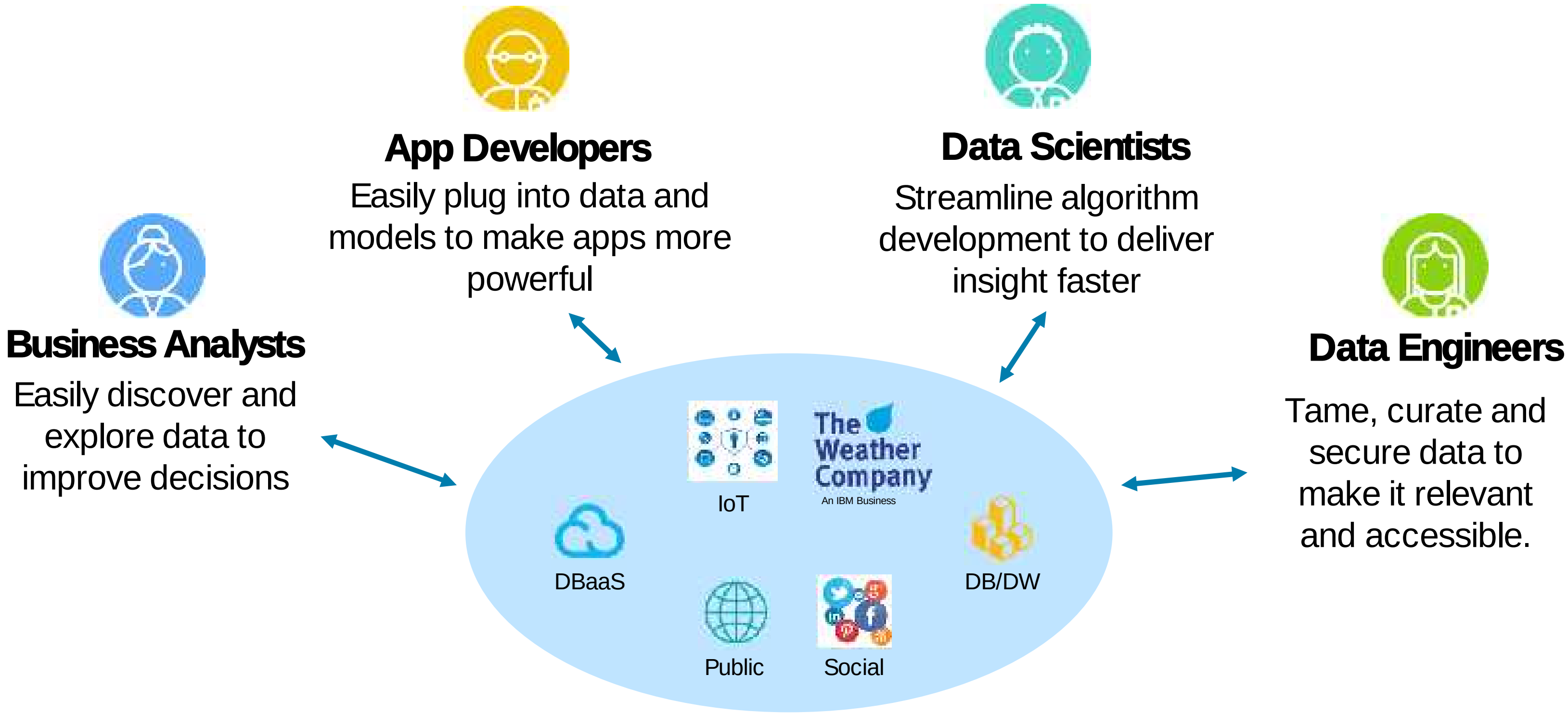
Topics for Today

- **Strategy Overview**
- **Db2 V11.1.3.3 – Introduction !!**
- **Private Cloud – Introduction !!**
- **Flex Points and HDM Offering**
- **Appliance News**
- **Hadoop and Open Source**
- **Event Processing**
- **Next Generation Data Virtualization**

All businesses have become data driven

Data Professionals – Evolving Roles

As data maturity increases, so does the number of data professionals who are hungry to put data to work



App Developers
Easily plug into data and models to make apps more powerful

Data Scientists
Streamline algorithm development to deliver insight faster

Business Analysts
Easily discover and explore data to improve decisions

Data Engineers
Tame, curate and secure data to make it relevant and accessible.

The Challenges of Fast Data

Data is arriving faster than ever before

- Billions of events processed every day
- Evident cross industry and driven by IoT
- Must land data quickly, or throw it away

Total data is large, and growing rapidly

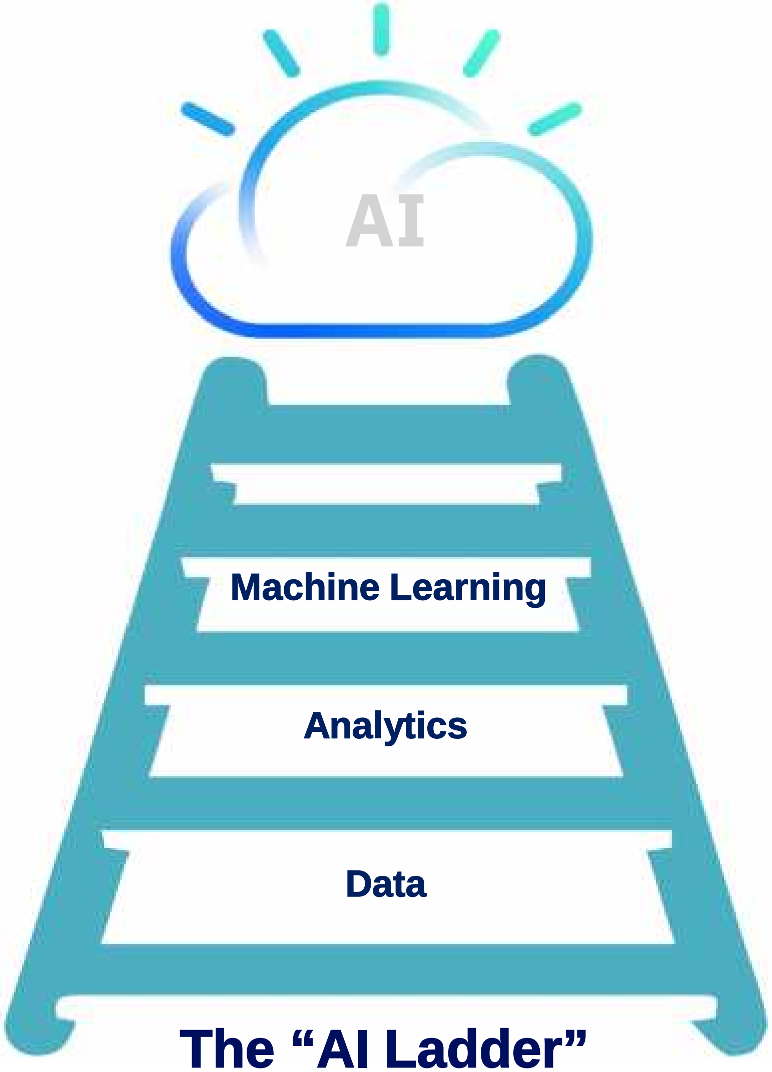
- Storing all events implies large data sets
- Storage costs are significant, and must be managed

Data is useless without fast insights

- Data value decays rapidly over time
- Insights must derived quickly, and use advanced analytics (ML)

Data availability without duplication

- Data must be available to the entire organization without requiring replication or duplication
- Maintain data in open format for future-proofing



The “AI Ladder”

Data Management Strategy is HYBRID

Its not about Cloud or On-Premises its about Cloud AND On-Premises

Its not about Traditional Relational or Open Source its about Traditional Relational AND Open Source

It’s About Hybrid

Its not about SQL or NoSQL its about SQL AND NoSQL

Its not about Structured or Unstructured Data its about Structured AND Unstructured Data

Common SQL Engine – Business Value

A **COMMON SQL ENGINE** **enabling true** **HYBRID** **data solutions for** **ALL WORKLOAD** **types**

Systems of Record

Systems of Engagement

Systems of Insight

Event Processing

Investment Protection

WRITE ONCE, RUN ANYWHERE

PUBLIC CLOUD

PRIVATE CLOUD

ON-PREMISES

Common SQL Engine – Consistent Technical Capabilities

A **COMMON SQL ENGINE enabling true HYBRID data solutions for ALL WORKLOAD types**

NEW			NEW	NEW		
Event Store	Managed Public Cloud Service	Hosted Public Cloud Service	On-premises Private Cloud	On-premises Appliance	On-premises Custom Software	Hadoop / Spark Environment
Db2 Event Store	Db2 [Warehouse] On Cloud	Db2 Hosted	Db2 Warehouse Db2 OLTP	IBM Integrated Analytics System	Db2	Big SQL

Foundation

- ✓ Full MPP scalability (GB-PB)
- ✓ High Concurrency
- ✓ Load and Go Simplicity
- ✓ Consistent Management and WLM
- ✓ HA, DR & Replication
- ✓ Integrated Security & Encryption

Application

- ✓ Built-in analytics (OLAP)
- ✓ Data Virtualization
- ✓ Application portability
- ✓ Hybrid by design
- ✓ Oracle Compatibility
- ✓ Netezza Compatibility

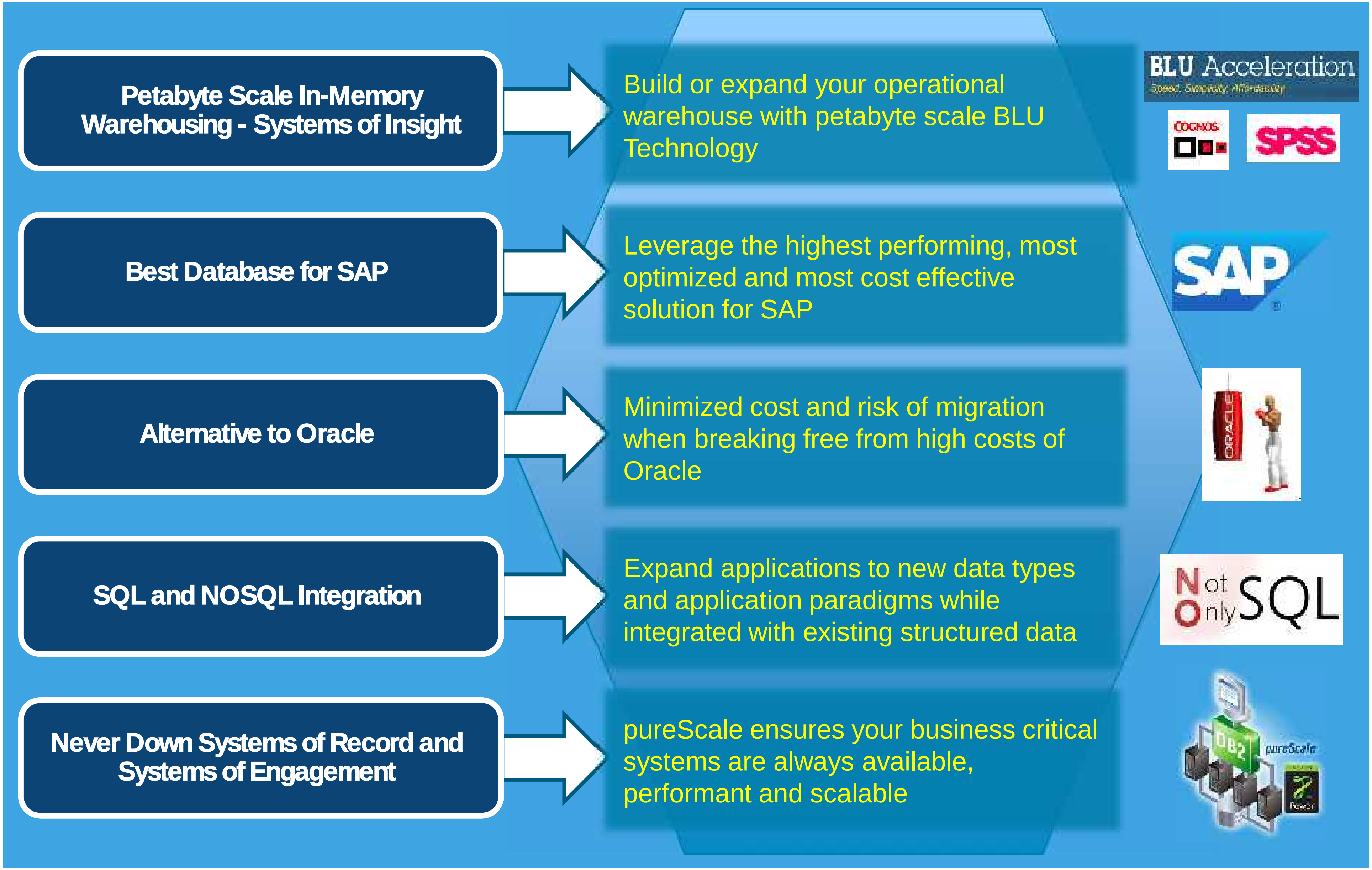
New Growth Trends

- ✓ Spark Integration
- ✓ HTAP Support
- ✓ SQL & NOSQL Capabilities
- ✓ Native JSON Support
- ✓ R Language Support
- ✓ Structured & Unstructured Data

Topics for Today

- Strategy Overview
- Db2 V11.1.3.3 – Introduction !!
- Private Cloud – Introduction !!
- Flex Points and HDM Offering
- Appliance News
- Hadoop and Open Source
- Event Processing
- Next Generation Data Virtualization

DB2 - Highlights and Strategic Investment Areas



Petabyte Scale In-Memory Warehousing - Systems of Insight

Build or expand your operational warehouse with petabyte scale BLU Technology

Best Database for SAP

Leverage the highest performing, most optimized and most cost effective solution for SAP

Alternative to Oracle

Minimized cost and risk of migration when breaking free from high costs of Oracle

SQL and NOSQL Integration

Expand applications to new data types and application paradigms while integrated with existing structured data

Never Down Systems of Record and Systems of Engagement

pureScale ensures your business critical systems are always available, performant and scalable

Check George Baklarz’s Presentation this afternoon

Db2 Version 11.1.2.2 Highlights

Higher Availability and Core Capabilities

Near-zero outage recovery

- Online crash recovery
- pureScale REBUILD restore

NoSQL Support

Native JSON support

- JSON SQL support Part 1
- Built-in UDFs for enhanced JSON capabilities

Not Only SQL

Db2 Tooling Capabilities

- Data Server Manager
- DB2 Connect
- DS Driver
- DS Gateway
- Advanced Recovery Tools

Column-Organized (BLU) Tables

Deeper BLU Optimizations for Operational Workloads

- Performance enhancements
- Builds on 4Q ‘16 advances
- Enables use of BLU beyond strictly analytic workloads

Additional Operating System Support

Solaris Support – by exception

MacOS Support – by exception

Packaging Changes

- Developer Community Edition
- Introduction of non-production licenses
- Data Management Bundle V1

Check George Baklarz’s Presentation this afternoon

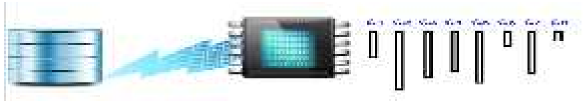
Db2 Version 11.1.3.3 Highlights

Higher Availability and Core Capabilities

- Faster Rollback of very large transactions
- WLM – Improve deadlock detection
- HADR Resilience and SSL Encryption
- Db2iupdt – ADD/DROP CFs on-line
- pureScale – on-line CREATE INDEX w/R/W access to table
- pureScale – faster member crash recovery

Column-Organized (BLU) Tables

UDF Cacheing for BLU
BLU Memory Usage enhancements
Temporal Query Support
Index Support



Additional Operating System Support

Solaris Support – 11.3+

Data Virtualization

MariaDB Connectivity Support
Db2 iSeries 7.2&7.3 Connectivity Support
Teradata 16 Connectivity Support
JSON over RESTful Service (MongoDB)
Boolean, Binary/Varbinary Data Type Mapping Enhancement
Pushdown Improvement for Hadoop Datasource
Function Mapping Pushdown Enhancement

Packaging Changes

- Hybrid Data Management Packaging

Topics for Today

- Strategy Overview
- Db2 V11.1.3.3 – Introduction !!
- Private Cloud – Introduction !!
- Flex Points and HDM Offering
- Appliance News
- Hadoop and Open Source
- Event Processing
- Next Generation Data Virtualization

Db2 and the Cloud

		Provisioning & Db2 Setup	Maintenance	Management
"Bring Your Own License"	• Custom-deployable software on your own infrastructure or private cloud or public cloud • Fully customizable for any type of workload • Complete flexibility including DPF and pureScale * • Customer managed	👥	👥	👥
Db2 Hosted	• Hosted database-as-a-service • Pre-defined hardware configurations • Fully customizable for any type of workload • Available on SoftLayer and AWS • Customer managed	IBM	👥	👥
Db2 on Cloud	• Fully managed database-as-a-service • Pre-defined and flexible hardware configurations optimized for transactional and general purpose workloads • Available on Bluemix public cloud	IBM	IBM	IBM
Db2 Warehouse on Cloud	• Fully managed database-as-a-service • Pre-defined hardware configurations optimized for analytics workloads • In-database analytics • Available on Bluemix and AWS public cloud	IBM	IBM	IBM
Db2 Warehouse	• Deploy on your own infrastructure or private cloud • Docker container technology for fast and simple deployment • Optimized for analytic workloads • Scalable, elastic • Customer managed	IBM	👥	👥
Db2 OLTP	• Deploy on your own infrastructure or private cloud • Docker container technology for fast and simple deployment • Optimized for operational and OLTP workloads • Scalable, elastic • Customer managed	IBM	👥	👥

Db2 and the Cloud

		Provisioning & Db2 Setup	Maintenance	Management
"Bring Your Own License"	• Custom-deployable software on your own infrastructure or private cloud or public cloud • Fully customizable for any type of workload • Complete flexibility including DPF and pureScale * • Customer managed	👥	👥	👥
Db2 Hosted	• Hosted database-as-a-service • Pre-defined hardware configurations • Fully customizable for any type of workload • Available on SoftLayer and AWS • Customer managed	IBM	👥	👥
Db2 on Cloud	• Fully managed database-as-a-service • Pre-defined and flexible hardware configurations optimized for transactional and general purpose workloads • Available on Bluemix public cloud	IBM	IBM	IBM
Db2 Warehouse on Cloud	• Fully managed database-as-a-service • Pre-defined hardware configurations optimized for analytics workloads • In-database analytics • Available on Bluemix and AWS public cloud	IBM	IBM	IBM
Db2 Warehouse	• Deploy on your own infrastructure or private cloud • Docker container technology for fast and simple deployment • Optimized for analytic workloads • Scalable, elastic • Customer managed	IBM	👥	👥
Db2 OLTP	• Deploy on your own infrastructure or private cloud • Docker container technology for fast and simple deployment • Optimized for operational and OLTP workloads • Scalable, elastic • Customer managed	IBM	👥	👥

Introducing IBM Cloud Private

Innovation

Kubernetes-based container platform

Cloud Foundry for prescribed container-based application development and deployment and life cycle management

Integrated DevOps toolchain

Integration

Catalog of integration services

API availability and management to integrate applications and data across environments

Investment Protection

Prescriptive guidance on where to run and how to architect your critical workloads

Next generation versions of industry leading IBM Middleware and Analytics
(MQ, Db2, Data Science, Cognos, Blockchain, IIB)

Management and Compliance

Core operational services, including monitoring, log mgmt, and security

Integration with existing systems and operations management solutions

Analytics Roadmap : Offerings / Capabilities on ICp

(as of Nov 2017)

Preliminary & Subject to changes
* To be confirmed

2017 Q4

- Db2 OLTP
- Db2 Warehouse
- Data Server Manager
- Data Science Experience

2018 1H

1. Hybrid Data Management

- Db2 OLTP
- Db2 Warehouse MPP
- Data Server Manager
- Big SQL *

2. Unified Governance

- Data Stage
- IGC

3. Data Science & BA

- Data Science Experience

Common / Foundational

✓	Metering	✓	Logging
✓	Monitoring	✓	IAAM & SSO
✓	Catalog		

2018 2H

1. Hybrid Data Management

- Db2 OLTP MPP
- Db2 Event Store

2. Unified Governance

- WEX *

3. Data Science & BA

- SPSS Modeler *
- SPSS Statistics *
- Cognos *

Common / Foundational

✓	Metering	✓	Logging
✓	Monitoring	✓	IAAM & SSO
✓	Catalog		

Why Analytics on IBM Cloud Private

True Hybrid Solution - consistency between public cloud and private cloud

No vendor lock-in. **Open Platform as a Service** (PaaS) for maximum integration ability

Container-based platform with very fast time to value (hours instead of weeks)

Extensive service-oriented analytic and machine learning capabilities ready for **Data Scientists** and **Business Analysts**

Optimized and secure **Data Management Services** for SQL, NoSQL, structured, semi-structured and unstructured data

Secure, governed and **compliant** platform for integration with any data source

IBM Cloud Private – More Information

Catch Kelly Schlamb's session this afternoon !!

Learn More

- IBM Cloud private home: https://ibm.biz/Bdj4Bz
- White paper: https://ibm.biz/Bdj4UJ

See it In Action

- Offering demo: https://youtu.be/yzXA3qhfaq0
- Try It: https://ibm.biz/Bdj4UC
- Free Community Edition: https://hub.docker.com/r/ibmcom/cfc-installer/

Topics for Today

- Strategy Overview
- Db2 V11.1.3.3 – Introduction !!
- Private Cloud – Introduction !!
- Flex Points and HDM Offering
- Appliance News
- Hadoop and Open Source
- Event Processing
- Next Generation Data Virtualization

Portfolio Simplification:

Three new bundles

Hybrid Data Management

- Db2
- Db2 Warehouse
- Db2 Event Store
- Db2 Big SQL

Unified Governance & Integration

- Information Server
- Entity Analytics
- Master Data Management
- Info Governance Catalog
- Data Replication
- Test Data Fabrication
- Info Lifecycle Governance
- Industry Models
- BigIntegrate, BigQuality BigMatch

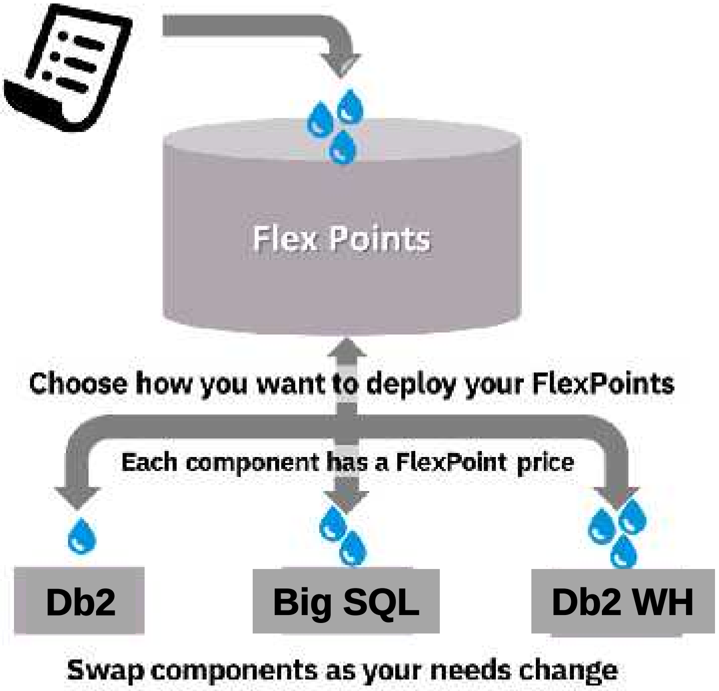
Data Science & Business Analytics

- DSx Local
- SPSS Modeler
- Decision Optimization
- Watson Explorer (v12 +)
- Cognos Analytics
- Planning Analytics

We will now focus on Hybrid Data Management

FlexPoints: How It Works

Buy FlexPoint licenses for the "Platform of your Choice"


Flex Points
Choose how you want to deploy your FlexPoints
Each component has a FlexPoint price
Db2
Big SQL
Db2 WH
Swap components as your needs change

Platform Offerings deliver integrated capabilities – now offered as flex bundles to simplify planning for adoption and growth at the lowest cost

Available for Our 3 Platform Offerings:

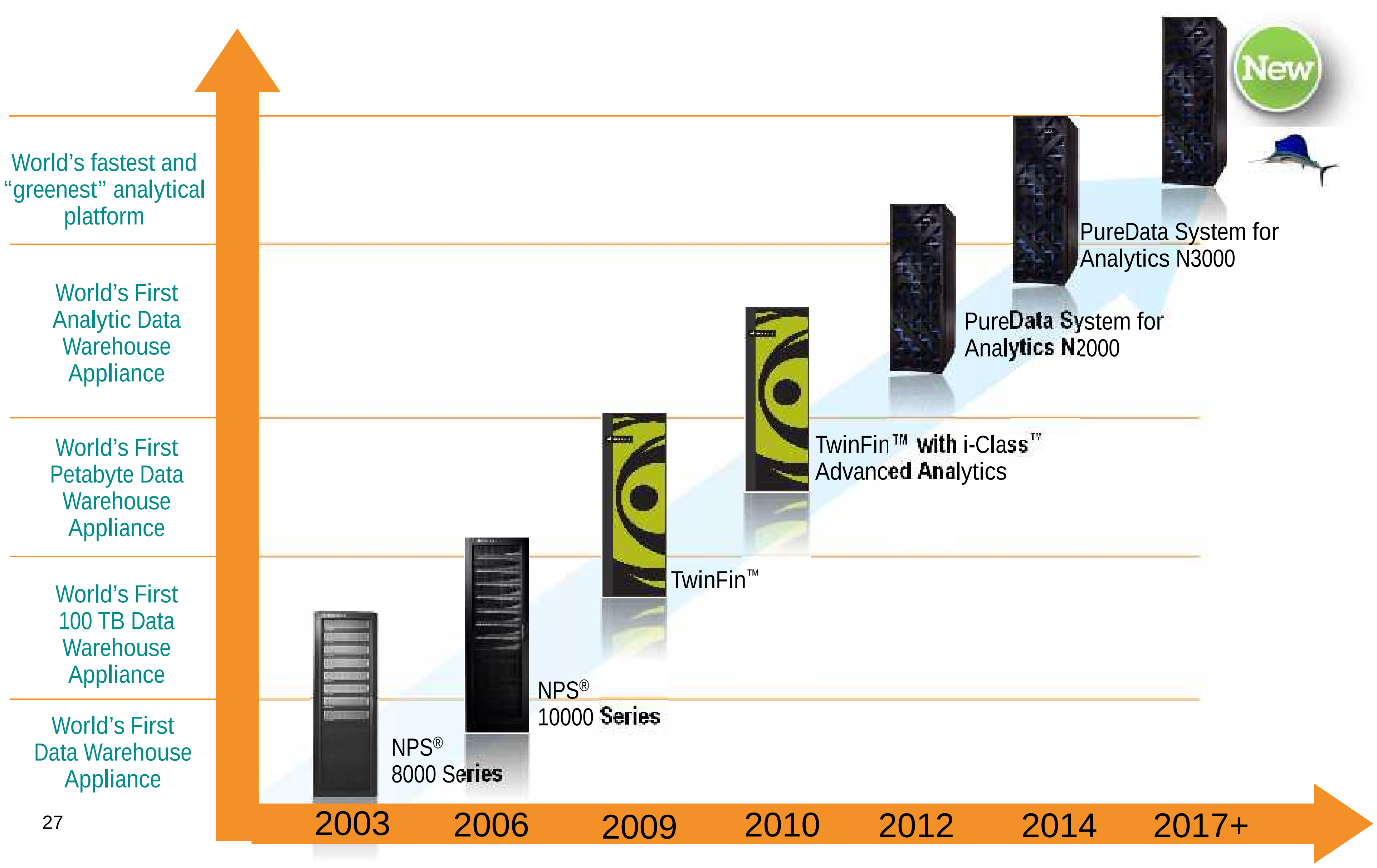
- Hybrid Data Management
  - Db2
  - Db2 Warehouse
  - Db2 Event Store
  - Db2 Big SQL
- Unified Governance & Integration
- Data Science & Business Analytics

FlexPoints CANNOT be used across PLATFORMS
As an example, Data Science and Business Analytics FlexPoints are NOT valid for Hybrid Data Management

Topics for Today

- Strategy Overview
- Db2 V11.1.3.3 – Introduction !!
- Private Cloud – Introduction !!
- Flex Points and HDM Offering
- Appliance News
- Hadoop and Open Source
- Event Processing
- Next Generation Data Virtualization

Next Generation Analytics Appliance



World's fastest and "greenest" analytical platform

World's First Analytic Data Warehouse Appliance

World's First Petabyte Data Warehouse Appliance

World's First 100 TB Data Warehouse Appliance

World's First Data Warehouse Appliance

NPS® 8000 Series

NPS® 10000 Series

TwinFin™

TwinFin™ with i-Class™ Advanced Analytics

PureData System for Analytics N2000

PureData System for Analytics N3000

2003 2006 2009 2010 2012 2014 2017+

Next Generation Analytics Appliance – Names

4000-series
“Sailfish”
IBM Integrated Analytics System
IIAS

World’s fastest and “greenest” analytical platform

World’s First Analytic Data Warehouse Appliance

World’s First Petabyte Data Warehouse Appliance

World’s First 100 TB Data Warehouse Appliance

World’s First Data Warehouse Appliance

PureData System for Analytics N3000

Our Next-Generation Hybrid DWH vision:

__Sailfish__ is IBM’s industry-leading hybrid - Private Cloud Data Warehouse and Analytics Platform that will __integrate__ seamlessly with other ground and cloud data warehouse services, delivering ultra __fast & scalable__ performance, cloud __elasticity__ together with end to end security - and the ultimate in __simplicity__ across all dimensions of the client’s experience.

...Integrated, Insightful, Simple, Agile and Secure!

2003 2006 2009 2010 2012 2014 2017+

IBM Integrated Analytics System

Next Generation Hybrid Data Warehouse

Optimized for **high performance** to support the broadest array of workload options for structured and unstructured data in your **hybrid data management** infrastructures

Real time analytics with **machine learning** that accelerates decision making, bringing new opportunities to the business – ready for **business analysts** and **data scientists**

Cloud-ready to support multiple workload deployment options

Reliable, elastic and flexible system that reduces and **simplifies management** resources

Leverages a **Common SQL Engine** for workload portability and skill sharing across public and private cloud

Built-in **IBM Data Science Experience** to collaboratively analyze data

Addressing Top Customer Requirements

Broader set of workloads

- Combination of reporting, analytics, operational analytics and data stores

Higher Concurrency

- Expand number of business analytics and machine learning activities within a single system

In-Place Expansion

- Independently scale both compute and storage as needed while protecting existing investments

Richer Availability Solutions

- High Availability, Disaster Recovery and replication solutions

Less admin & more analytics

Simplicity

Load and Go

Low TCO

One Touch Support

Accelerate Time to Insight

Easy to Deploy and Easy to Operate
Faster Time to Value - Load and Go…it's an appliance!
Lower Total Cost of Ownership
Built-in Tools for data migration and data movement

BI Developers & DBAs – faster delivery times

No configuration
No storage administrations
No physical modeling
No indexes and tuning
Data model agnostic
Self Service Management dashboard

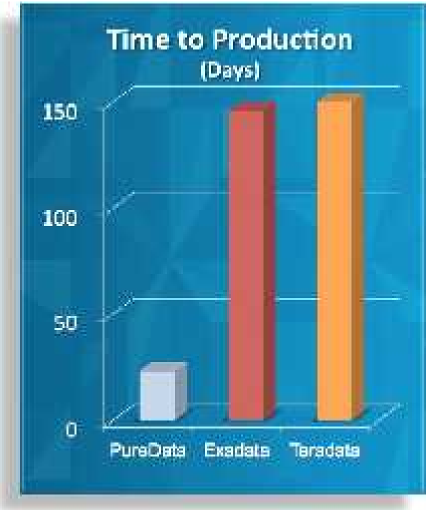
ETL Developers

No aggregate tables needed – simpler ETL logic
Faster load and transformation times

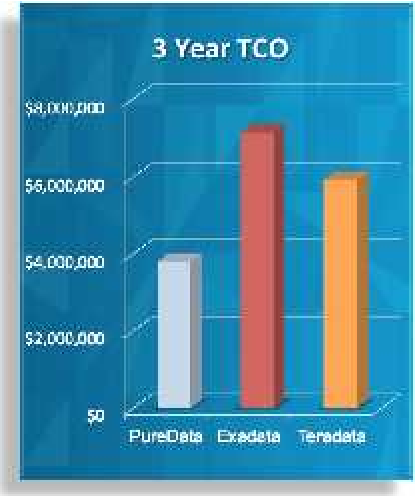
Business Analysts

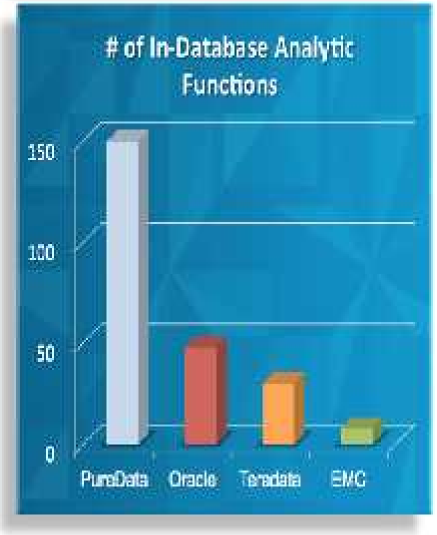
True ad hoc queries – no tuning, no indexes
Ask complex queries against large datasets
Load & query simultaneously

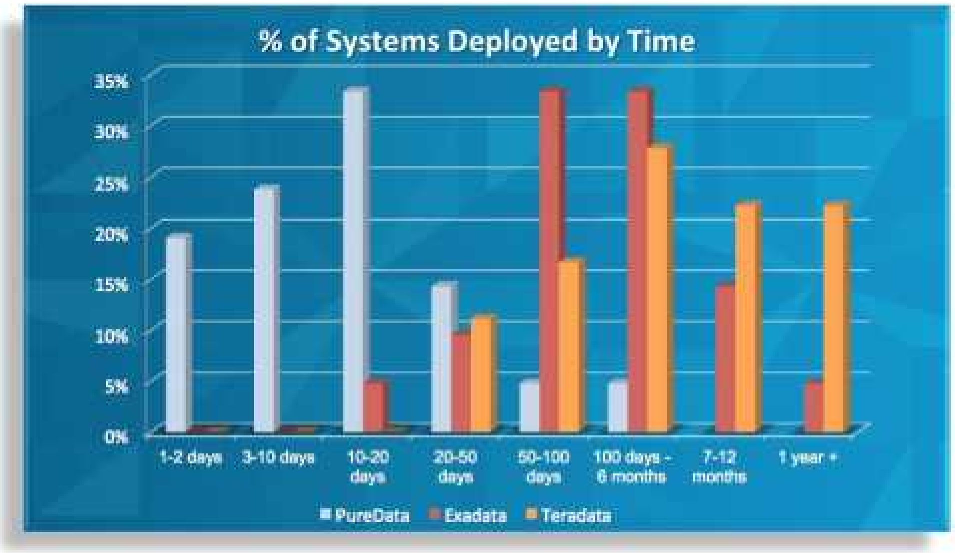
Maintain Core Values

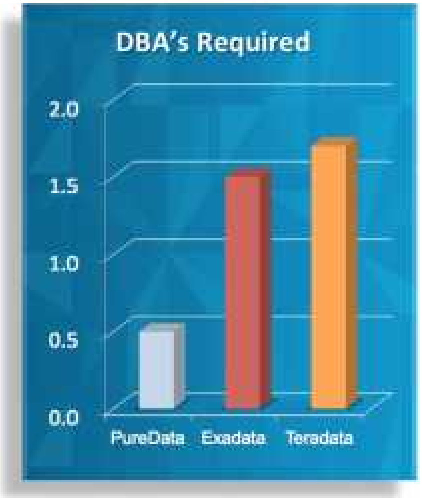


-Reduced administration
-Performance portal
-Lower end starting point
-More scale-out
-Fast time to deployment
-Low TCO









Speed of Thought Analytics

Performance

2X – 5X Performance Gain

POWER

Powered by RedHat® Linux on Power

Optimized for Analytics with 4X Threads per core, 4X Memory bandwidth and 4X more cache at lower latency compared to x86

ALL Flash Storage

Hardware Accelerated architecture enabling faster insights with extreme performance, 99.999% reliability and operational efficiency

MPP Scale out

Memory Optimized

In-memory BLU columnar processing with dynamic movement of

data from storage

Data Skipping

Skips unnecessary processing of irrelevant data

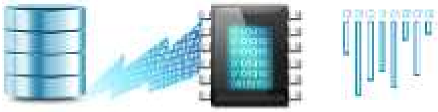
Actionable Compression

Patented compression technique that preserves order so data can be used without decompressing

Optimized Analytics Performance

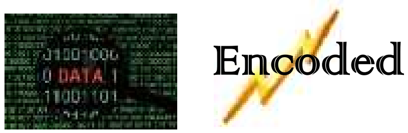
Next Generation In-Memory

In-memory columnar processing with dynamic movement of data from storage



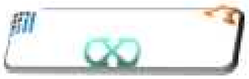
Analyze Compressed Data

Patented compression technique that preserves order so data can be used without decompressing



Embedded Spark

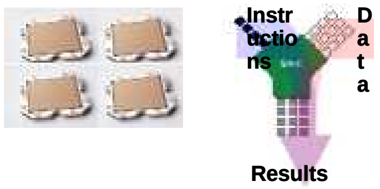
Spark As an Analytics Engine



Spark/R, Spark/ML, Rest API, Object Store ETL, Complex Transformations (ELT), Streaming

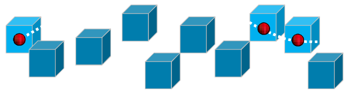
CPU Acceleration

Multi-core and SIMD parallelism (Single Instruction Multiple Data)



Data Skipping

Skips unnecessary processing of irrelevant data



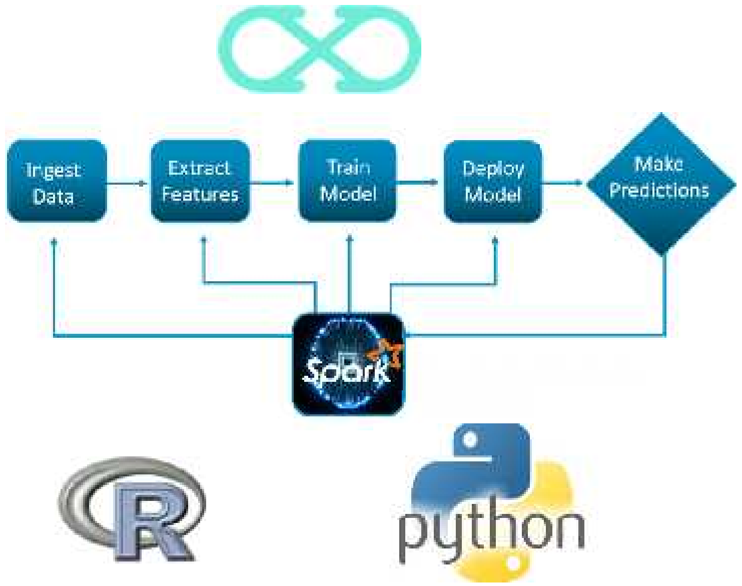
Powered by Hardware

Designed for Deep Complex Analytics

4X Threads per core
4X Memory Bandwidth
4X More cache at Lower Latency

Ready for Data Scientists and Business Analysts

Machine Learning



Integrated Cognitive Assist for Machine Learning

DSX for Interactive & Collaborative Data Science

Scalable ML Model Training, Deployment and Scoring with Spark embed Predictive / Prescriptive In place Analytics

Embedded

Data mining, prediction, transformations, statistics, geospatial, data preparation

Full integration with tools for BI & visualization

IBM Cognos, Tableau, Microstrategy, Business Objects, SAS, MS Excel, SSRS, Kognitio, Qlikview

Full integration with tools for model building and scoring

IBM SPSS, SAS, Open Source R, Fuzzy Logix

Full integration for custom analytics

Open Source R, Java, C, C++, Python, LUA

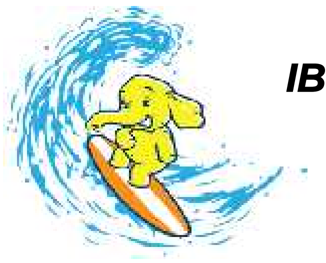
DSX Local on IIAS Benefits

- **The inclusion of DSX Local widens the audience for IIAS**
  - DSX Local is a on-prem platform which manages and provides access to the data, tools and packages that data scientist need
    - Jupyter, Zeppelin*, and RStudio
    - Anaconda for Python 2 and 3* support
    - Support for Python, Scala, and R languages
- **DSX Local extends IIAS federation support**
  - Livy included for connecting to and running jobs on external Spark clusters
  - GUI for connecting to external data sources and data sets
    - DB2, DB2 Z, Netezza, Informix, Oracle, dashDB, HDFS*, Hive*, and more to come
  - Easily combines data from multiple sources to create new data sets
- **DSX Local provides full model management for IIAS**
  - Create models with the built-in model builder GUI or programmatically from a notebook

Write Once, Run Anywhere

Hybrid

IBM Data Lift



IBM Fluid Query

Application Agility

Common SQL Engine with comprehensive tools and capabilities across all deployment models: Public/Private Cloud, On-premise Appliance.

One ISV certification for all deployments .

Operational Compatibility

Single consistent interface powered by IBM Data Server Manager for Management and Maintenance

Make Data Simple and Accessible to All

Data Virtualization capabilities enabled by Fluid across deployment models

Querable Archive Query historical data on Hadoop or other content stores

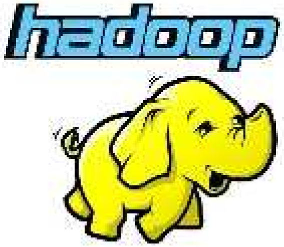
Discovery & Exploration Implement the Logical Data Warehouse; Land data in Hadoop for discovery, exploration & “day 0” archive

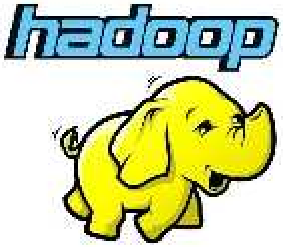
Build Bridges to RDBMS Islands Combine data from different enterprise divisions currently trapped in silos ; Federate to other data sources such as Oracle, SQL Server, PostgreSQL, Teradata, etc.,

Ground to Cloud Blazing-fast Data Transfer

Integrated high speed IBM Data Lift using IBM Aspera for secure ground to cloud data movement

Hybrid – Common SQL Engine





Hybrid – Common SQL Engine

Db2 Warehouse On Cloud

Db2 Big SQL

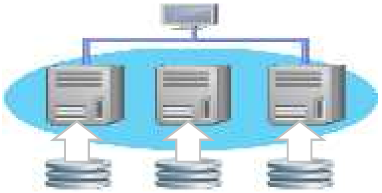
Db2 Warehouse

IBM Integrated Analytics System

Db2 Big SQL

Unmatched multi-dimensional Flexibility

Flexible



Scalable

Versatile Workloads

In-Place Incremental Expansion
Easily and incrementally scale out your environment by adding Compute and Storage capacity to meet your growth needs

In-place Tiered Storage Expansion
Independently scale storage for cost effective capacity growth

HTAP with IBM Db2 Analytics Accelerator
Seamlessly integrate with IBM z Systems infrastructure to enable real-time analytics combining transactional data, historical data and predictive analytics

Truly a Mixed Workload Appliance
Whether it be high scan performance needed to answer your business's strategic questions, high concurrency, low-latency requirements to support your operational systems, or even use as an operational data store. Perform all your enterprise Analytics needs on a single platform with mission critical availability.

Flexible Licensing
Flexible entitlements for business agility & cost-optimization

Expansion capabilities

Non-disruptive in-place incremental expansion

- Reduce disruptions to your analytics systems as you scale out

Cloud-ready

- Tools to move workloads seamlessly to the cloud based on your requirements

Non-disruptive in-place tiered storage expansion

- Independently scale storage for cost effective capacity growth

Cost efficient multi-temperature storage

- Most frequently accessed data ("hot") on faster flash storage
- Less frequently accessed data ("colder") on cost efficient enterprise storage systems

IBM Db2 Analytics Accelerator

High performance for complex queries

- Unprecedented response times to enable ’train of thought’ analyzes frequently blocked by poor query performance

Seamless integration with z Applications

- Brings high performance queries to existing z systems while protecting the core OLTP workloads

Self-managed workloads

- Queries are executed in the most efficient location

Transparent application access

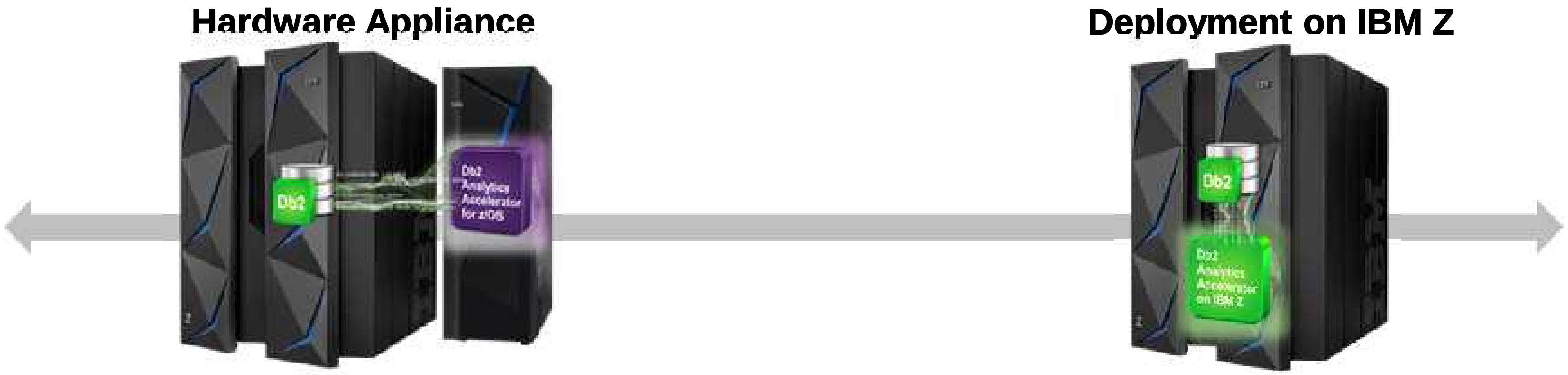
- Brings the value of the Common SQL Engine to the z environment
- Applications connected to Db2 are entirely unaware of the Accelerator, all security is handled by Db2 z/OS

Fast deployment and time to value

- Non-disruptive installation. Plug it in, load data and go in 1-2 days
- Db2 for z/OS query router automatically sends analytic queries to source which will provide optimal performance

A high performance appliance that integrates the IBM Integrated Analytics System with zEnterprise technology to deliver dramatically faster business analysis

One API – One implementation – Two deployment options



Hardware Appliance

Deployment on IBM Z

Db2
Db2 Analytics Accelerator for z/OS
Db2
Db2 Analytics Accelerator on IBM Z

Uniform experience, simultaneous use, and easy transition between different implementations

Common analytics engine across all the platforms: Db2 Warehouse

IBM Integrated Analytics System configurations

IBM Power 8 S822L 24 core server 3.02GHz
IBM FlashSystem 900

In-place Expansion Tiered storage

Mellanox 10G Ethernet switches
Brocade SAN switches

	M4001-003 1/3 Rack	M4001-006 2/3 Rack	M4001-010 Full Rack	M4001-020 2 Racks	M4001-040 **4 Racks**
Servers	3	5	7	14	28
Cores	72	120	168	336	672
Memory	1.5 TB	2.5 TB	3.5 TB	7 TB	14 TB
User capacity (Assumes 4x compression)	64 TB	128 TB	192 TB	384	768
Tiered storage (Optional)			TBD—GA 1H 2018		

≥ 2 Racks + Tiered Storage targeted for 1H 2018; In place expansion targeted for 2H 2018

Hardware architecture overview

7 Compute Nodes in 1 rack containing

- IBM Power 8 S822L 24 core server 3.02GHz
- 512 GB of RAM (each node)
- 2x 600GB SAS HDD
- Red Hat® Linux OS

Up to 3 Flash Arrays in 1 rack containing

- IBM FlashSystem 900
- Dual Flash controllers
- Micro Latency Flash modules
- 2-Dimensional RAID5 and hot swappable spares for high availability

2x Mellanox 10G Ethernet switches

- 48x10G ports
- 12x40/50G ports
- Dual switches form resilient network

IBM SAN64B 32G Fibre Channel SAN

- 16Gb FC Switch
- 48x 32Gb/s SFP+ ports

User Data Capacity:
192 TB*
(Assumes 4x compression)

Power Requirements:
9.4 kW

Cooling Requirements:
32,000 BTU/hr

Scales from:
1/3rd Rack to 8 Racks
(initial GA is 1/3rd to 1 Rack)

Application and Operational Compatibility

… compared to Netezza

Perspective	September, 2017 (First release of Sailfish)	December, 2018 (Completion of Sailfish)
Applications	95% SQL compatibility nz* commands not available yet Manual conversion of stored procedures Performance degradation of INZA functions	100% Application Compatibility Equal or better performance for all applications
Operations & Management	nz* commands not available yet Workload Management tools change Replication solutions changed (NRS) Multi-tenancy (single database)	Some areas of operational management will continue to be different with Sailfish in order to provide a richer set of capabilities (WLM, HA/DR)

… compared to PDOA and ISAS (Db2)

Perspective	September, 2017 (First release of Sailfish)	December, 2018 (Completion of Sailfish)
Applications	100% compatibility	100% compatibility
Operations & Management	Some limitations such as multi-tenancy	100% compatibility

… compared to Oracle

Perspective	September, 2017 (First release of Sailfish)	December, 2018 (Completion of Sailfish)
Applications	95%-98% compatibility Leverages Oracle Application Compatibility Layer	95%-98% compatibility Leverages Oracle Application Compatibility Layer

Topics for Today

- Strategy Overview
- Db2 V11.1.3.3 – Introduction !!
- Private Cloud – Introduction !!
- Flex Points and HDM Offering
- Appliance News
- Hadoop and Open Source
- Event Processing
- Next Generation Data Virtualization

IBM and Hortonworks Deliver Data Science at Scale

Focus on extending data science and machine learning to analyze the data in Apache Hadoop systems

Consumers get the best in class open technology

IBM + Hortonworks

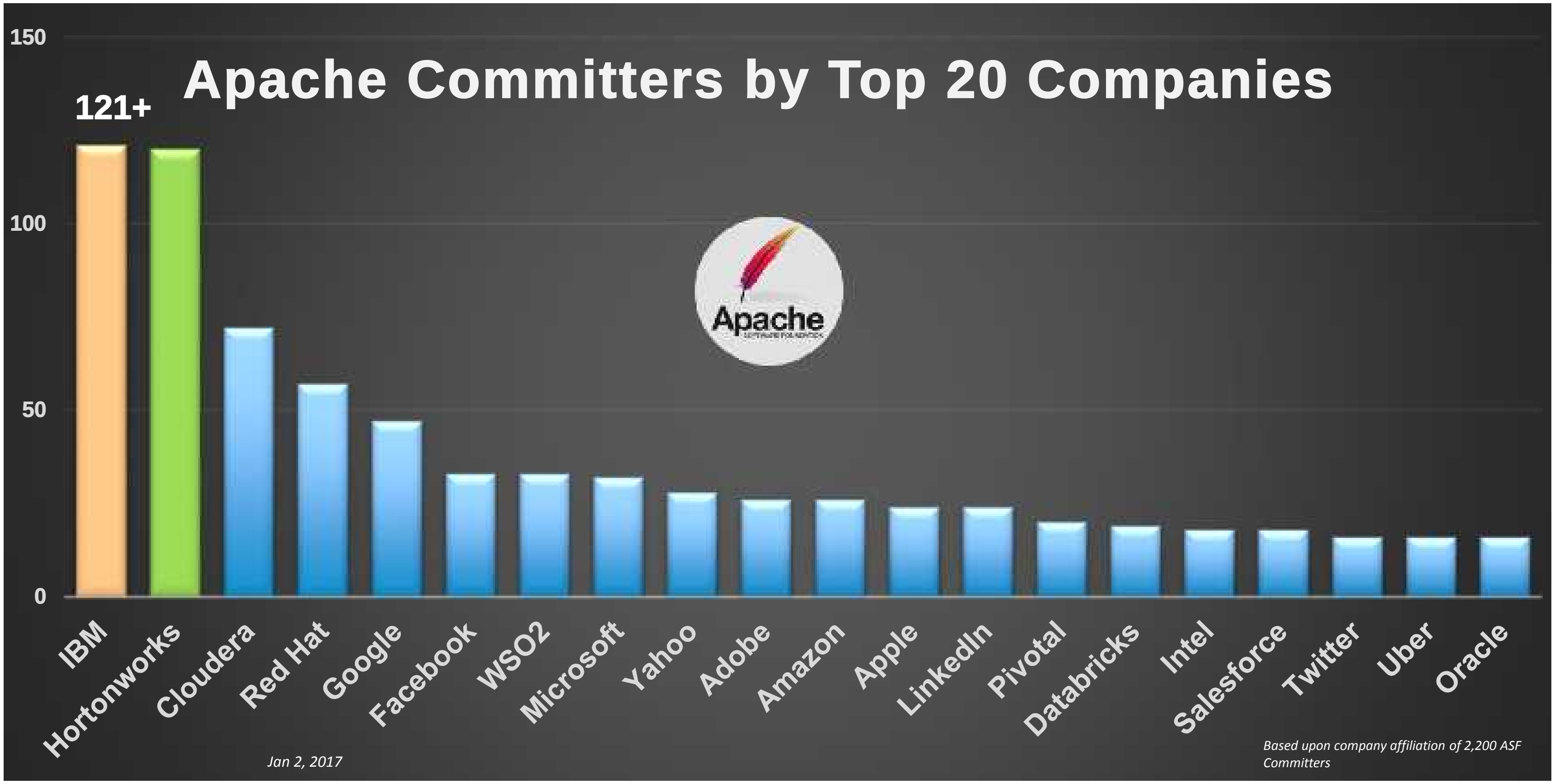
- **#1 Rank by Gartner** 2017 Data Science Magic Quadrant
- **Leader in SQL technology** for Hadoop (www.tpc.org)
- **Leader in data and analytics** solutions for hybrid cloud
- Provides Data Science **& Machine Learning**

- **Leader in Hadoop** Open Source Distribution
- **1000+ customers** and 2100+ ecosystem partners
- **Hadoop original architects, developers** employed by Hortonworks
- **Provides Open Hadoop** Data Platform

Commitment to progressing advanced analytics through open source

IBM and Hortonworks - Open Source Commitment

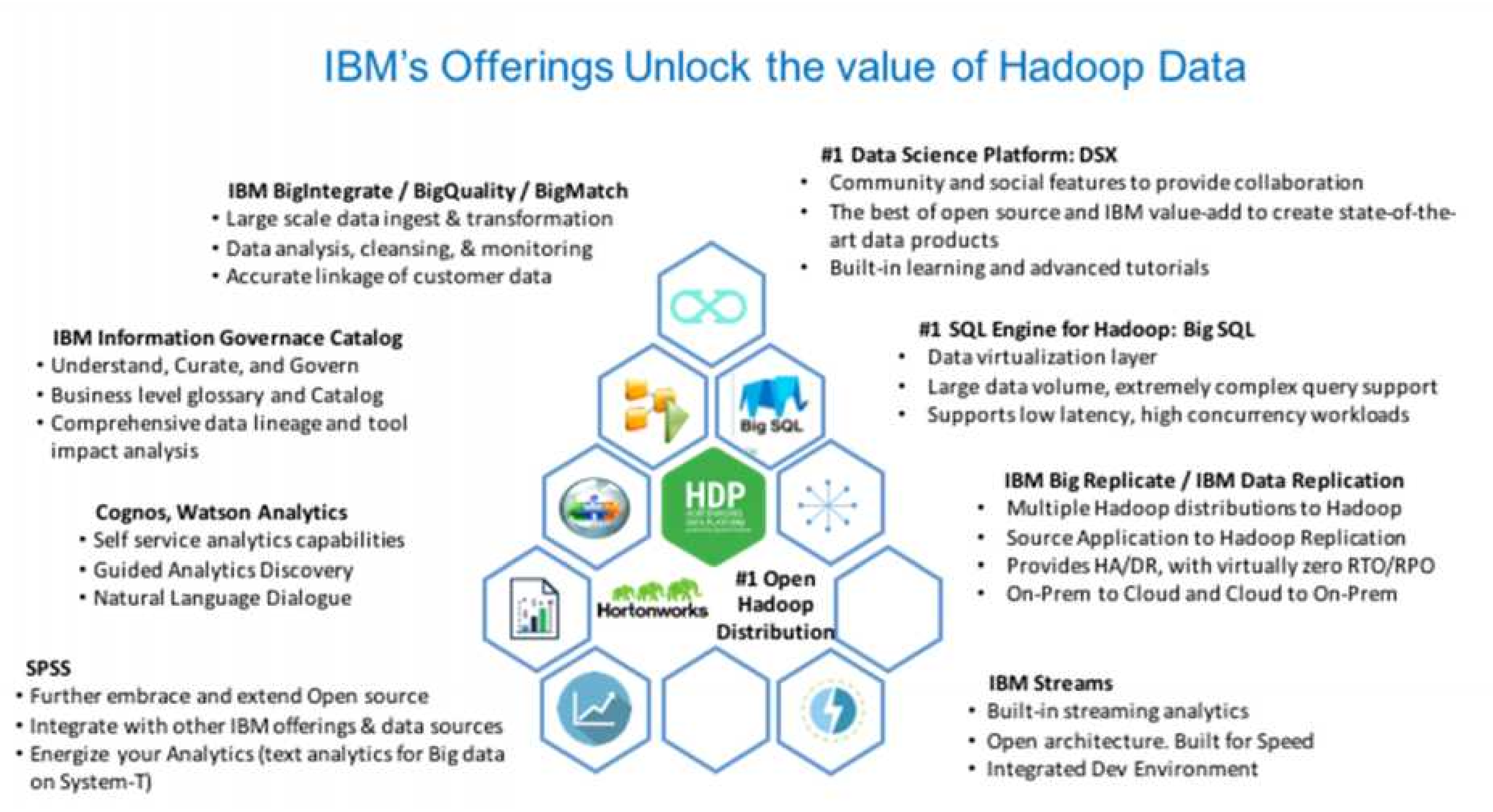
...and our combined commitment to Open Standards is Unmatched.



Jan 2, 2017

Based upon company affiliation of 2,200 ASF Committers

IBM Big Data High Value with Hortonworks



IBM's Offerings Unlock the value of Hadoop Data

IBM BigIntegrate / BigQuality / BigMatch
- Large scale data ingest & transformation
- Data analysis, cleansing, & monitoring
- Accurate linkage of customer data

IBM Information Governace Catalog
- Understand, Curate, and Govern
- Business level glossary and Catalog
- Comprehensive data lineage and tool impact analysis

Cognos, Watson Analytics
- Self service analytics capabilities
- Guided Analytics Discovery
- Natural Language Dialogue

SPSS
- Further embrace and extend Open source
- Integrate with other IBM offerings & data sources
- Energize your Analytics (text analytics for Big data on System-T)

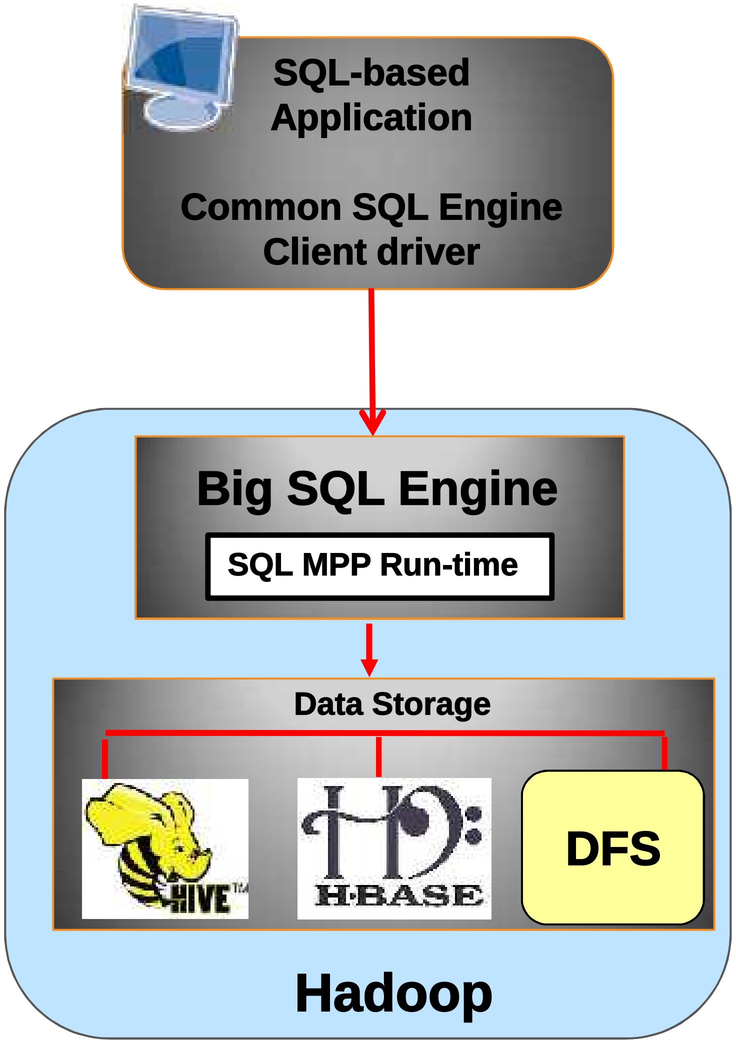
#1 Data Science Platform: DSX
- Community and social features to provide collaboration
- The best of open source and IBM value-add to create state-of-the-art data products
- Built-in learning and advanced tutorials

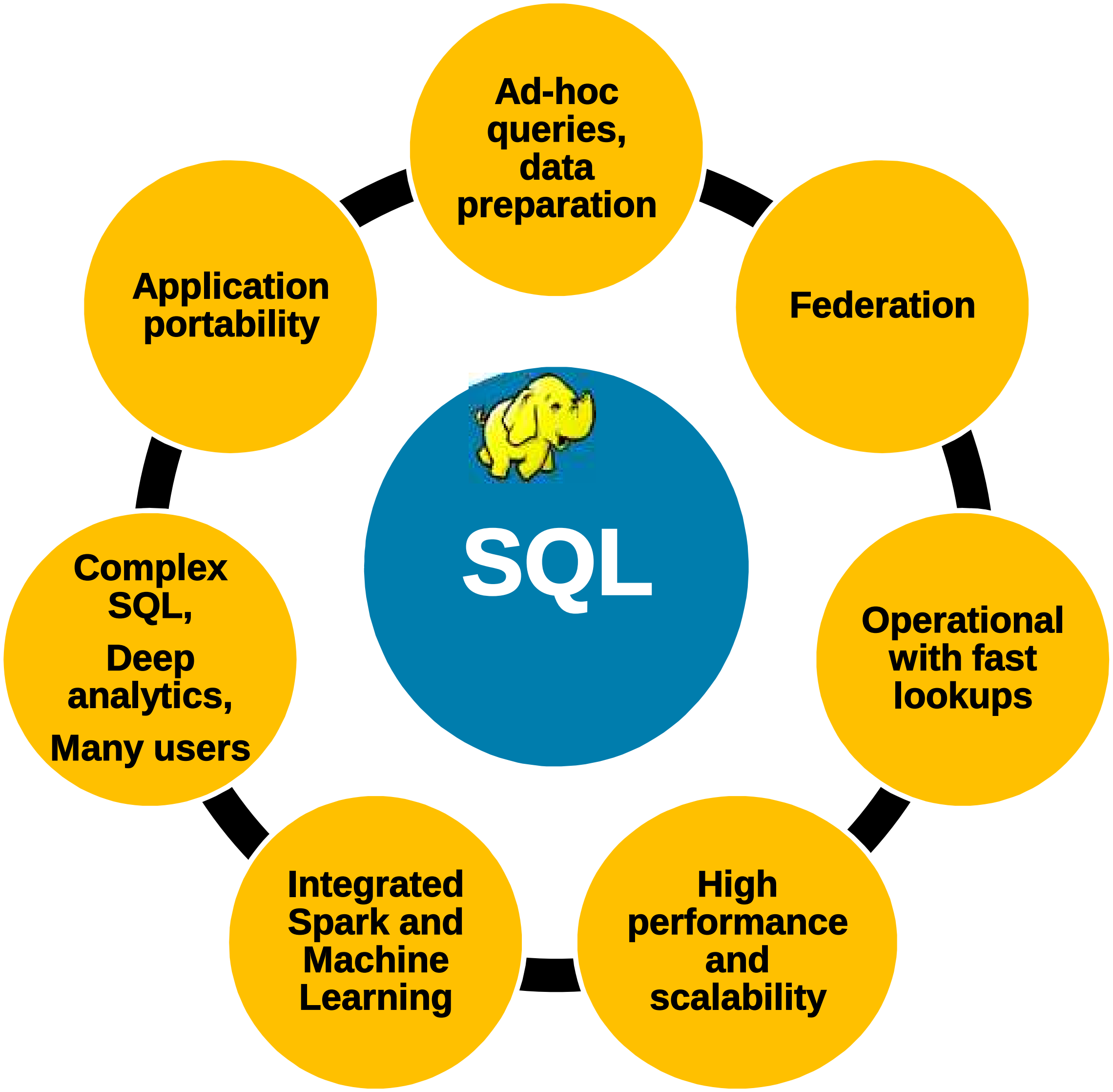
#1 SQL Engine for Hadoop: Big SQL
- Data virtualization layer
- Large data volume, extremely complex query support
- Supports low latency, high concurrency workloads

IBM Big Replicate / IBM Data Replication
- Multiple Hadoop distributions to Hadoop
- Source Application to Hadoop Replication
- Provides HA/DR, with virtually zero RTO/RPO
- On-Prem to Cloud and Cloud to On-Prem

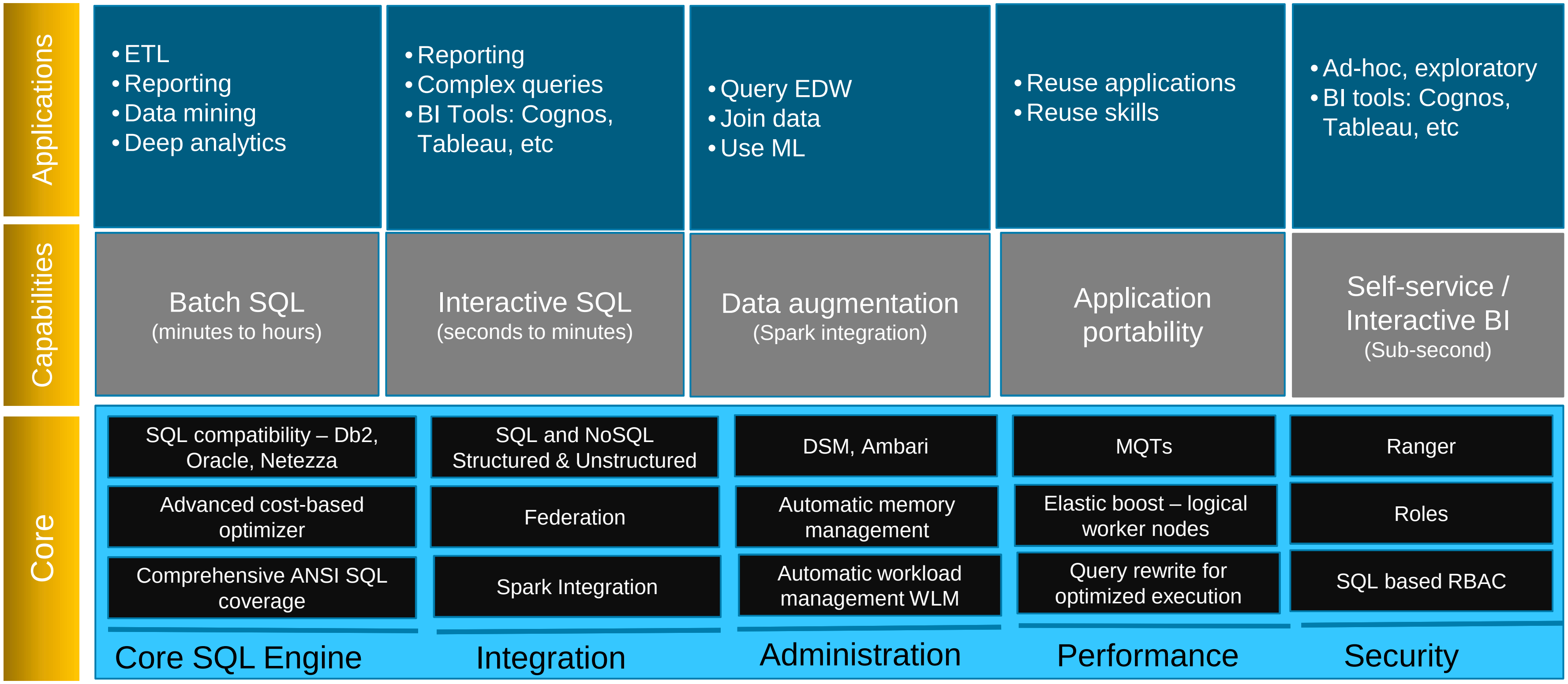
IBM Streams
- Built-in streaming analytics
- Open architecture. Built for Speed
- Integrated Dev Environment

Db2 Big SQL – For all WH Needs in Hadoop




SQL
Ad-hoc queries, data preparation
Federation
Operational with fast lookups
High performance and scalability
Integrated Spark and Machine Learning
Complex SQL, Deep analytics, Many users
Application portability

Db2 Big SQL V5.0

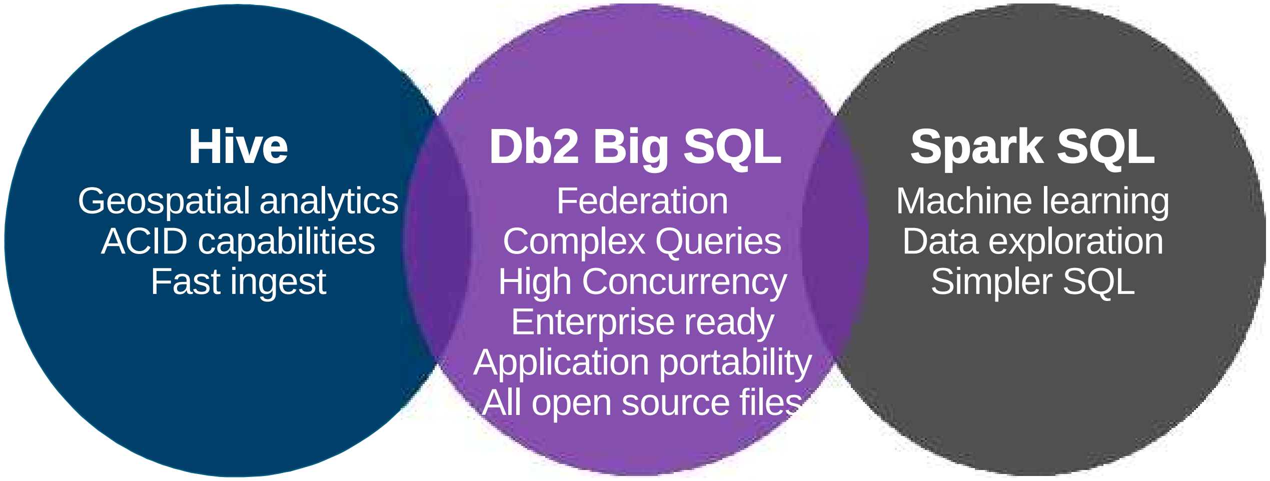


Applications	• ETL • Reporting • Data mining • Deep analytics	• Reporting • Complex queries • BI Tools: Cognos, Tableau, etc	• Query EDW • Join data • Use ML	• Reuse applications • Reuse skills	• Ad-hoc, exploratory • BI tools: Cognos, Tableau, etc
Capabilities	Batch SQL (minutes to hours)	Interactive SQL (seconds to minutes)	Data augmentation (Spark integration)	Application portability	Self-service / Interactive BI (Sub-second)
Core	SQL compatibility – Db2, Oracle, Netezza Advanced cost-based optimizer Comprehensive ANSI SQL coverage	SQL and NoSQL Structured & Unstructured Federation Spark Integration	DSM, Ambari Automatic memory management Automatic workload management WLM	MQTs Elastic boost – logical worker nodes Query rewrite for optimized execution	Ranger Roles SQL based RBAC
	Core SQL Engine	Integration	Administration	Performance	Security

www.tpc.org – check out TPC-H and TPC-DS – Big SQL vs Impala vs Hive
Db2 Big SQL 5.0 is **2X** faster than Hive LLAP with Tez – and much more functional
Db2 Big SQL 5.0 is **3X** faster than Spark SQL 2.1

Combining Hadoop Technologies

Not Mutually Exclusive. Hive, Db2 Big SQL & Spark SQL can co-exist and **complement** and **leverage** each other in a cluster



Hive

Geospatial analytics
ACID capabilities
Fast ingest

Db2 Big SQL

Federation
Complex Queries
High Concurrency
Enterprise ready
Application portability
All open source files

Spark SQL

Machine learning
Data exploration
Simpler SQL

Leveraged for metastore and UDFs

IBM's proprietary SQL engine for Hadoop

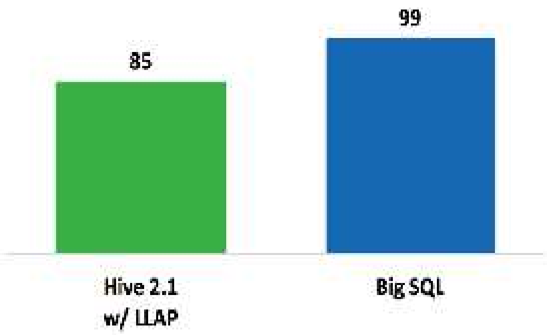
IBM's open source SQL engine for Hadoop

Query Performance at a Glance – vs Hive LLAP with Tez

HADOOP-DS @ 10TB

85 COMMON QUERIES

WORKING COMPLIANT QUERIES: 6-streams

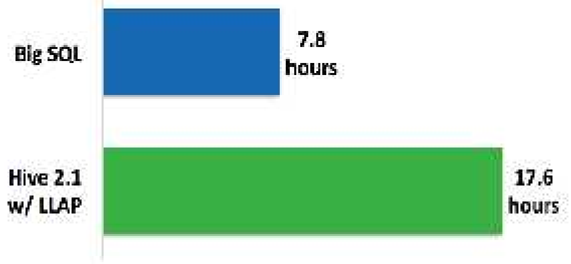


RESOURCE UTILIZATION: 6-STREAMS

1.5x FEWER CPU CYCLES USED

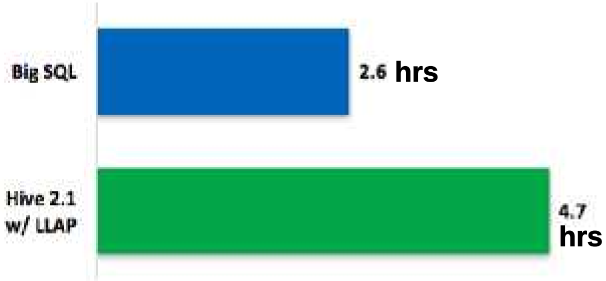
PERFORMANCE: 6-streams

Db2 Big SQL **2.3x** FASTER



PERFORMANCE: 1-stream

Db2 Big SQL **1.8x** FASTER



WORKLOAD

SCALE FACTOR: **10 TB**
FILE FORMAT: **ORC** (ZLIB)
CONCURRENCY: **6 STREAMS**
QUERY SUBSET: **85 QUERIES**

INTERESTING FACTS

FASTEST QUERY

5.4x FASTER (Db2 Big SQL: 1.5 SEC, HIVE: 8.1 SEC)

SLOWEST QUERY (QUERY 67)

1.7x FASTER (Db2 Big SQL: 6827 SEC, HIVE: 11830 SEC)

Db2 Big SQL FASTER FOR **80%** OF QUERIES RUN

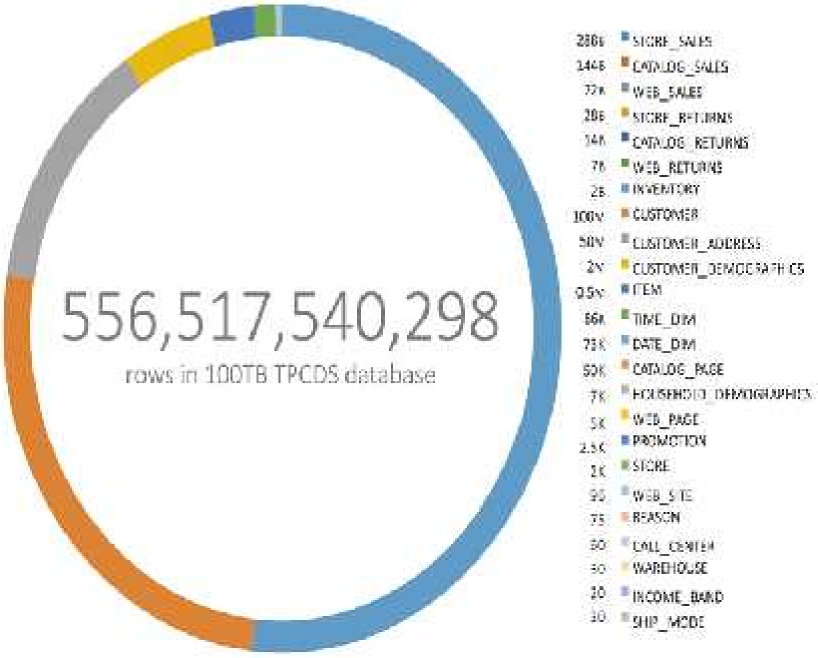
STACK

HDP 2.6.1
Db2 Big SQL 5.0.1
HIVE 2.1 LLAP ON TEZ

Query Performance at a Glance – Db2 Big SQL & Spark SQL

Leads performance metrics on high volumes of data and concurrent streams

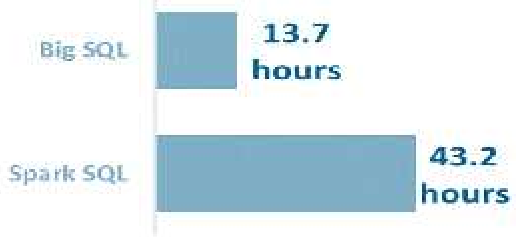
SNAPSHOT OF 100TB HADOOP-DS



PERFORMANCE

Db2 Big SQL 5.0 is **3.2x** faster than Spark SQL 2.1

(4 Concurrent Streams)



COMPRESSION

60%

SPACE SAVED WITH PARQUET

AVERAGE CPU USAGE

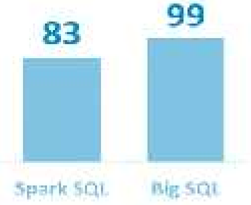
76.4%

MAX I/O THROUGHPUT

READ 4.4 GB/SEC

WRITE 2.8 GB/SEC

WORKING QUERIES



I/O (vs Spark)

Db2 Big SQL reads **12x** less data

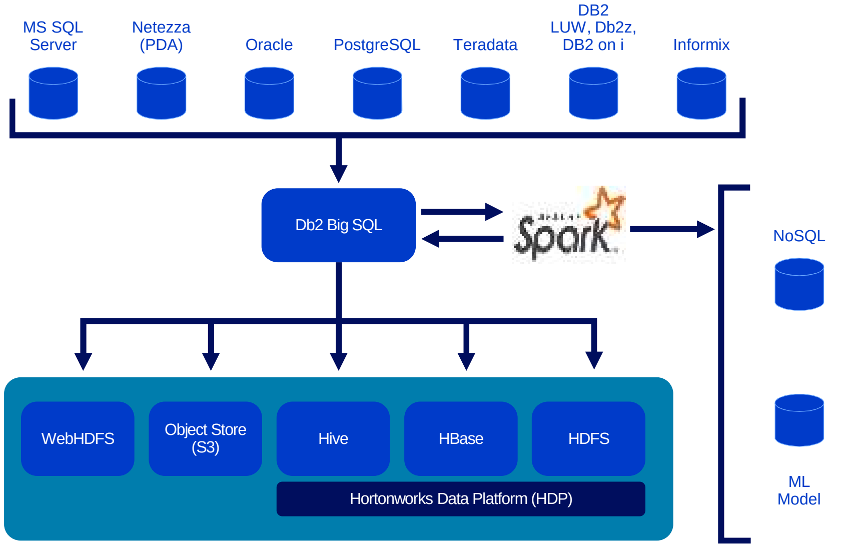
Db2 Big SQL writes **30x** less data

Blog on benchmark: https://developer.ibm.com/hadoop/2017/02/07/experiences-comparing-big-sql-and-spark-sql-at-100tb/

Federation – Query one connection and Virtualize Heterogeneous Data

Db2 Big SQL queries heterogeneous systems in a single query

Only SQL-on-Hadoop that virtualizes more than 10 different data sources: RDBMS, NoSQL, HDFS or Object Store



Transparent

- **Appears to be one source**
- **Programmers don't need to know how / where data is stored**

High Function

- **Full query support against all data**
- **Capabilities of sources as well**

Autonomous

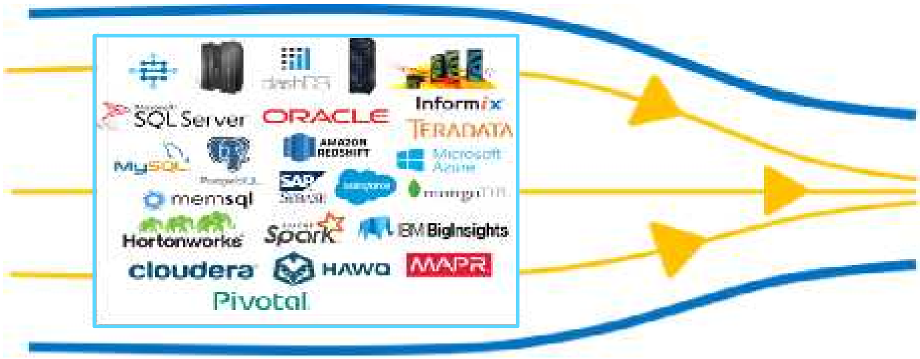
- **Non-disruptive to data sources, existing applications, systems.**

High Performance

- **Optimization of distributed queries**

Federation: Rich Capabilities that Brings Data Together

- ✓ *Easily access information on demand*
- ✓ *Combine data in Hadoop with disparate sources to form a data lake*
- ✓ *Quickly extend your data warehouse by enriching it*



Connect

- Quick access to Data value
- Common Framework
- ODBC/JDBC
- Spark integration enables new data sources
- Connect all data sources in single query

Query

- Intelligent Query Routing
- Cost-based optimizer
- SQL pushdown
- Local data caching
- ANSI-compliant SQL

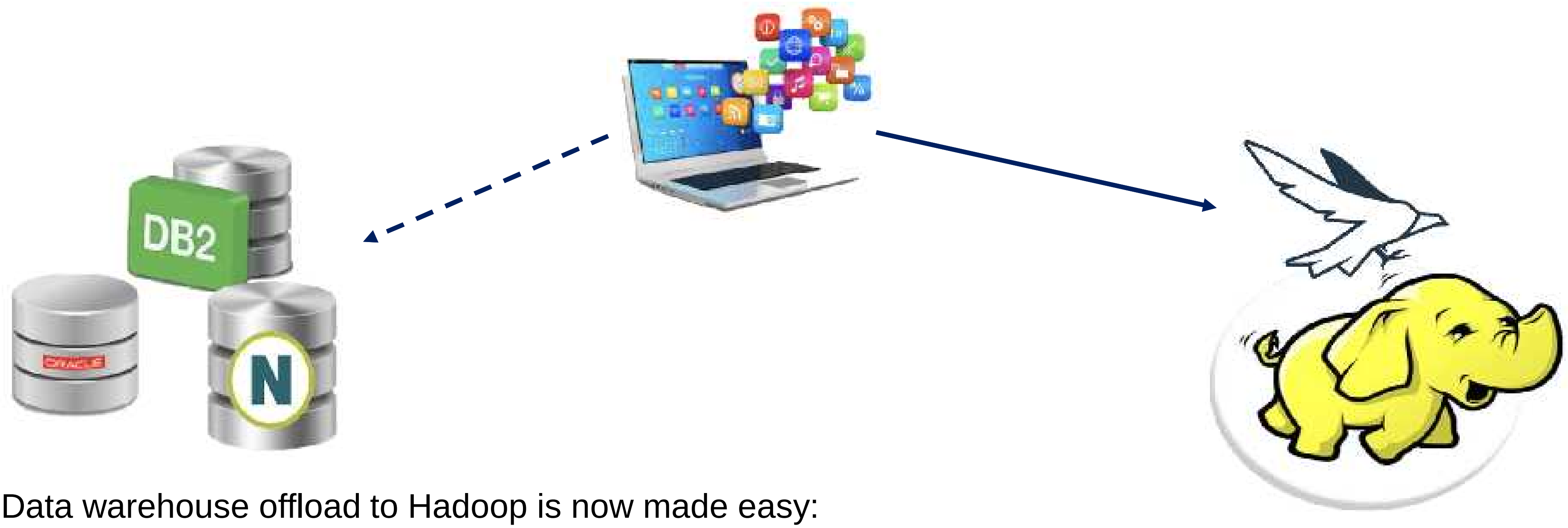
Monitor

- Easily define & manage through a common UI
- Simple point & click to discover and query
- Monitor and visualize active queries

Data Placement

- Schema conversion when moving data
- Bulk data copy to Hadoop
- Filtered subsets of data

Application Portability: Move Applications without Re-tooling

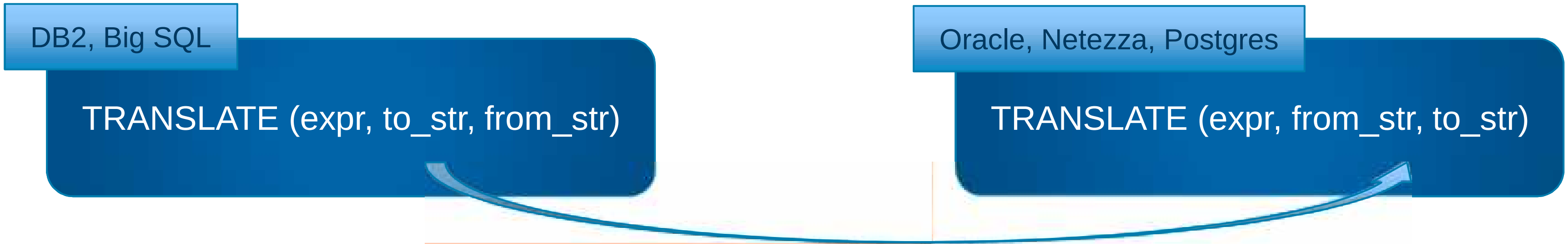


Data warehouse offload to Hadoop is now made easy:

- Write one, run anywhere…
- Easy porting of applications
- Reuse skills of DBAs/ developers who know ANSI SQL

Db2 Big SQL is the best platform for offloading Oracle Data Marts and Warehouses to Hadoop

Oracle Compatibility - SET sql_compat='ORA'



DB2, Big SQL

TRANSLATE (expr, to_str, from_str)

Oracle, Netezza, Postgres

TRANSLATE (expr, from_str, to_str)

Same function, parameters reversed!

- **SQL_COMPAT global variable enables support for both parameter orderings (and other syntax/behavioral conflicts when offloading SQL to Hadoop)**

- **Excellent Oracle PL/SQL support! (New for V5.0)**
  - SQL data-access-level enforcement
  - Enforce data access levels at run time rather than at compile time.
  - Oracle database link syntax (@ symbol)

- **Note:**
  - Setting of the DB2_COMPATIBILITY_VECTOR registry variable (inherited from DB2) is **not recommended** in Big SQL. Custom compatibility features should be enabled **only** by using the SQL_COMPAT global variable.

Oracle PL/SQL Support

```
set sql_compat=’ORA’

create or replace procedure plsql_proc (fetchval out integer)
as

cursor cur1 is
select count(*) from syscat.tables ;

-- begin

open cur1 ;

fetch cur1 into fetchval ;
close cur1 ;

end
```

Easy session variable to switch modes!

Big SQL is the best platform for offloading Oracle Data Marts and Warehouses to Hadoop

- Big SQL V4.2 already supports some Oracle SQL compatibility (but not PL/SQL)
- Big SQL V5.0 adds support for Oracle PL/SQL procedural language

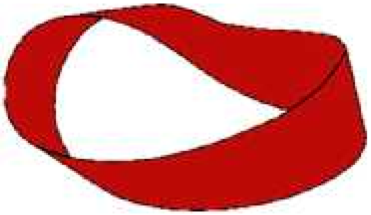
Query Execution

Here's why Db2 Big SQL can get you the best execution for complex queries and many concurrent users with high performance

Performance

Materialized Query Tables

Advanced Statistics



Elastic Boost

World Class Cost Based Optimizer

Concurrent users

Self Tuning Memory Manager

Advanced Workload manager

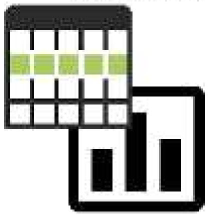
Complex query

Hardened runtime

SQL Compatibility

Query rewrite

Native Row & Columnar stores

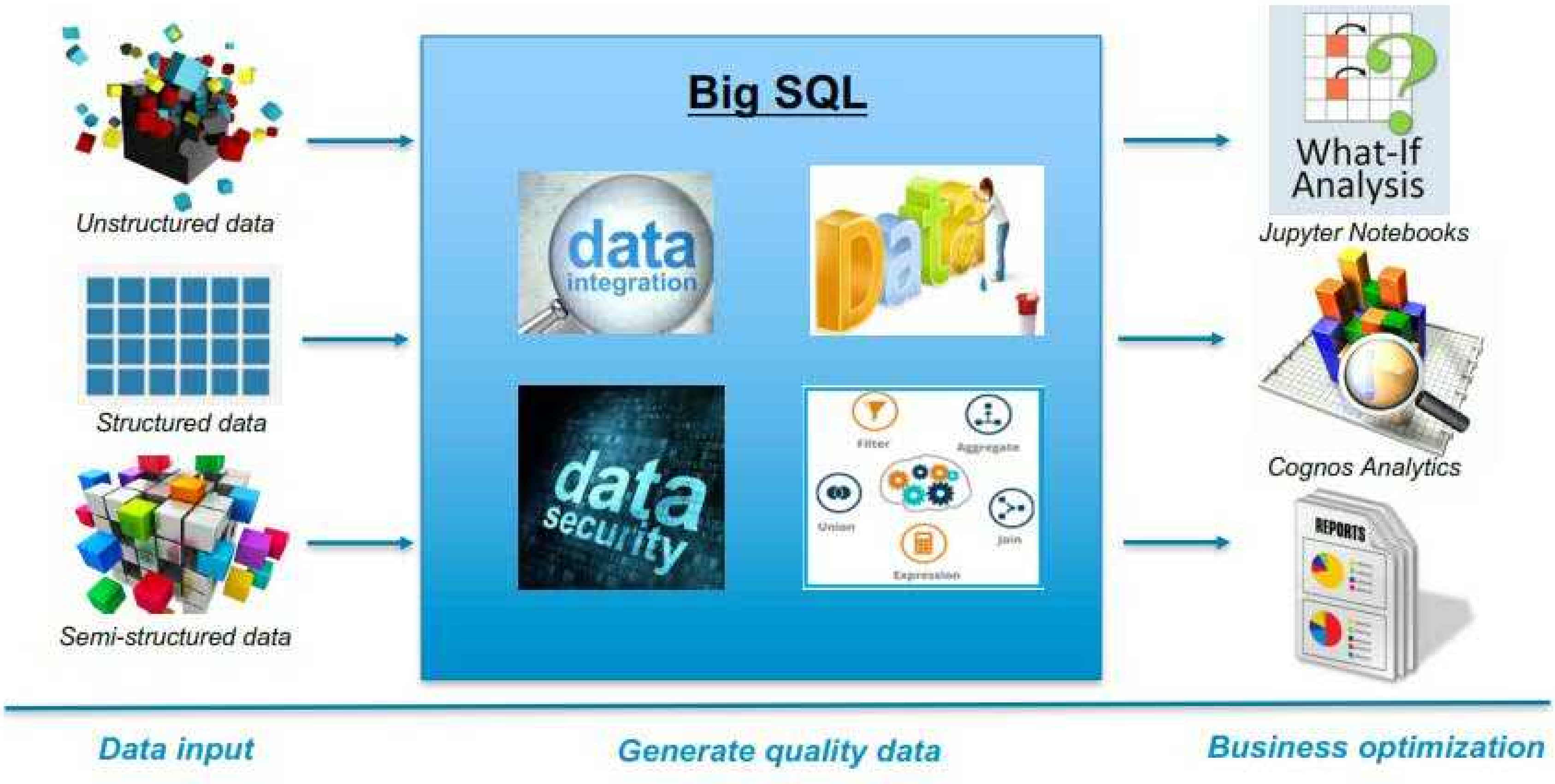


Big SQL – Rich Analytics

Big SQL is a powerful hybrid analytical engine

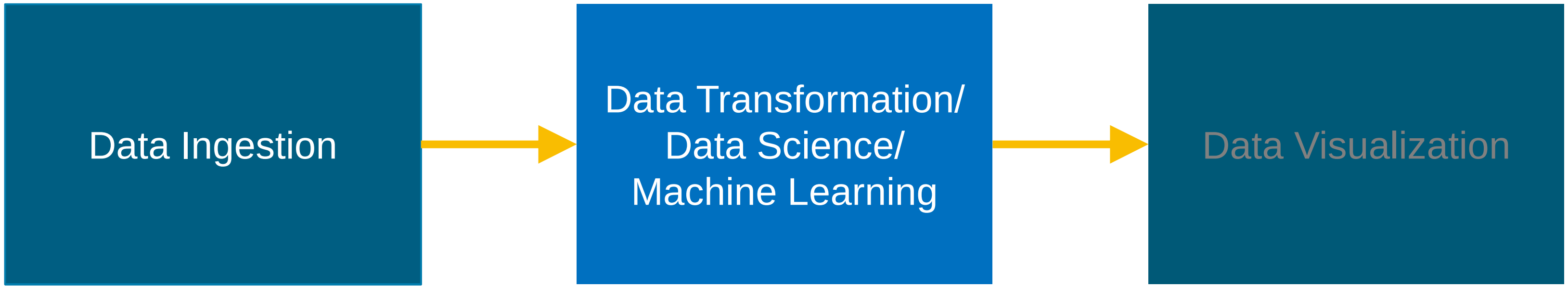
Data engineer

Offers leading performance metrics on high volumes of data to combine, transform, cleanse data in a secure environment, to generate a dataset to derive insights on data



Self-service Analytics: Democratize Data Science and ML

Leverage ***Db2 Big SQL*** *throughout your journey*



Data Ingestion

Data Transformation/ Data Science/ Machine Learning

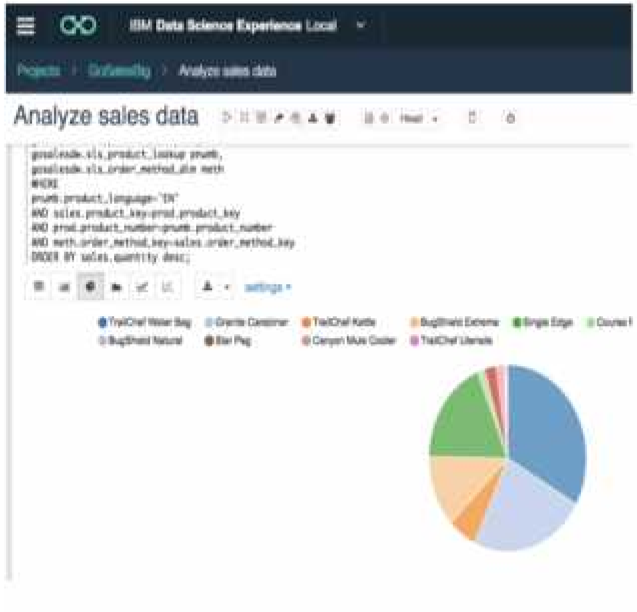
Data Visualization

Virtualize disparate data sources like Hadoop, RDBMS, and Object Stores (S3) to join data in a single query

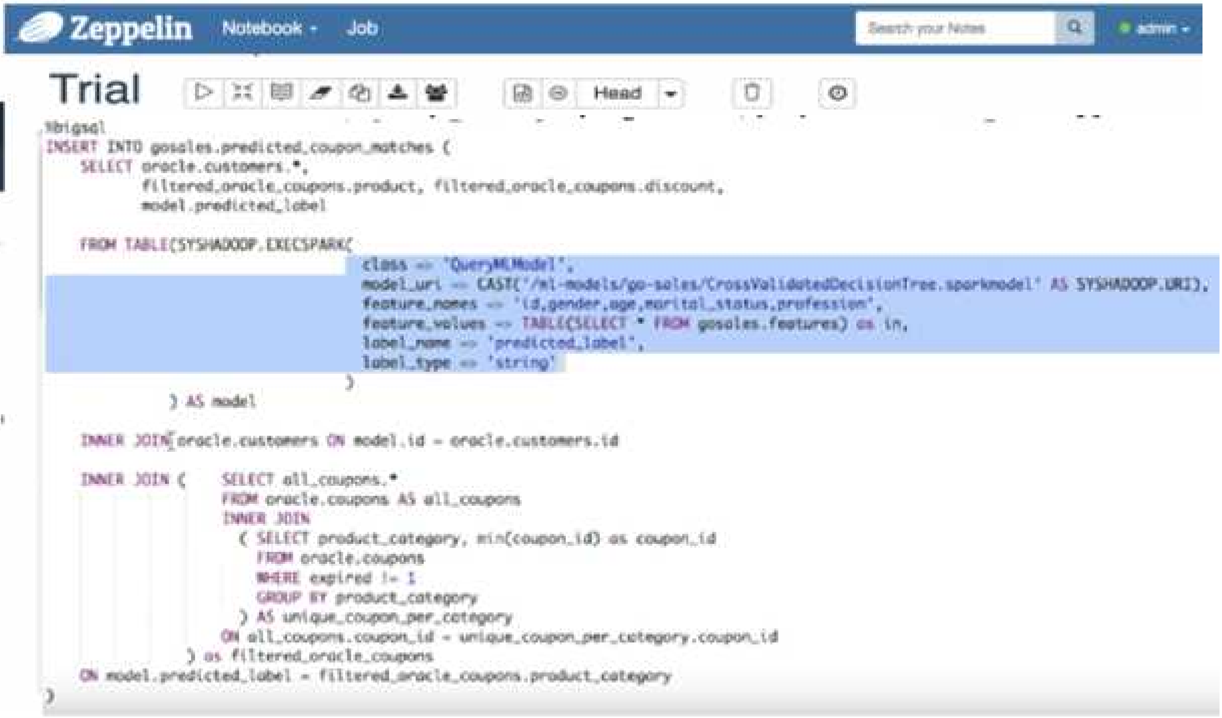
Manipulate data and operationalize data science models written in various languages

Perform data discovery, analyze, and visualize business results in notebooks or other BI tools

Operationalize Machine Learning Models using SQL



```
gosalesdw.sls_product_lookup pnumb,
gosalesdw.sls_order_method_dim meth
WHERE
pnumb.product_language='EN'
AND sales.product_key=prod.product_key
AND prod.product_number=pnumb.product_number
AND meth.order_method_key=sales.order_method_key
ORDER BY sales.quantity desc;
```



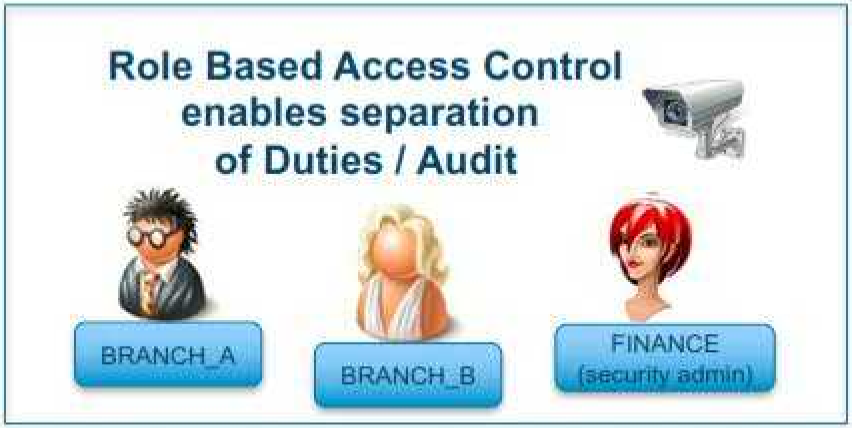
```
%bigsql
INSERT INTO gosales.predicted_coupon_matches (
    SELECT oracle.customers.*,
           filtered_oracle_coupons.product, filtered_oracle_coupons.discount,
           model.predicted_label

    FROM TABLE(SYSHADOOP.EXECSPARK(
                                class => 'QueryMLModel',
                                model_uri => CAST('/ml-models/go-sales/CrossValidatedDecisionTree.sparkmodel' AS SYSHADOOP.URI),
                                feature_names => 'id,gender,age,marital_status,profession',
                                feature_values => TABLE(SELECT * FROM gosales.features) as in,
                                label_name => 'predicted_label',
                                label_type => 'string'
                              )
              ) AS model

    INNER JOIN oracle.customers ON model.id = oracle.customers.id

    INNER JOIN (     SELECT all_coupons.*
                     FROM oracle.coupons AS all_coupons
                     INNER JOIN
                       ( SELECT product_category, min(coupon_id) as coupon_id
                         FROM oracle.coupons
                         WHERE expired != 1
                         GROUP BY product_category
                       ) AS unique_coupon_per_category
                     ON all_coupons.coupon_id = unique_coupon_per_category.coupon_id
               ) as filtered_oracle_coupons
    ON model.predicted_label = filtered_oracle_coupons.product_category
)
```

For more details check the blog: https://developer.ibm.com/hadoop/2017/11/07/ibm-big-sql-machine-learning-demo/

Big SQL - Security

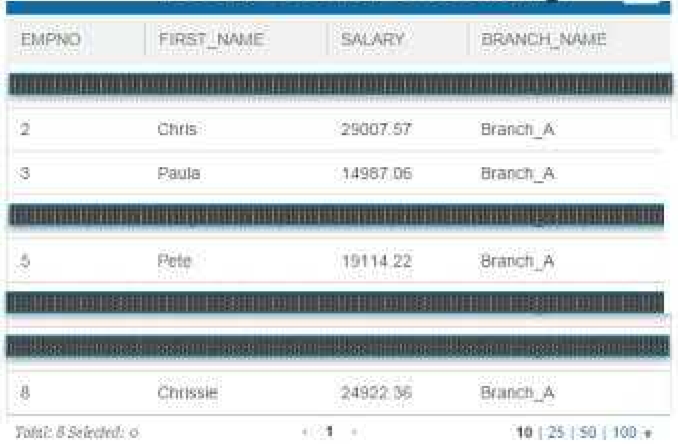
Big SQL is the most secure analytical engine that offers row and column level access control (RCAC) among other security settings

Data engineer



Role Based Access Control enables separation of Duties / Audit

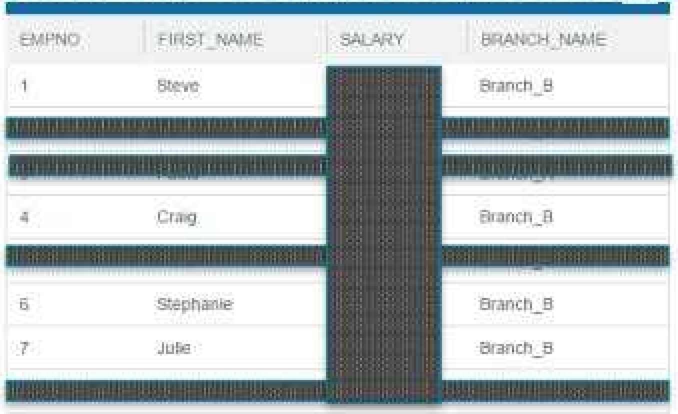
Row Level Security



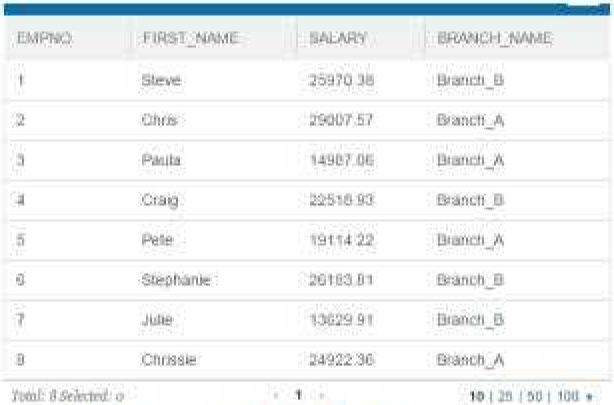
EMPNO	FIRST_NAME	SALARY	BRANCH_NAME
[redacted]	[redacted]	[redacted]	[redacted]
2	Chris	29007.57	Branch_A
3	Paula	14987.06	Branch_A
[redacted]	[redacted]	[redacted]	[redacted]
5	Pete	19114.22	Branch_A
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
8	Chrissie	24922.36	Branch_A

Total: 8 Selected: 0 ‹ **1** › **10** | 25 | 50 | 100 ▾

Row and Colum Level Security



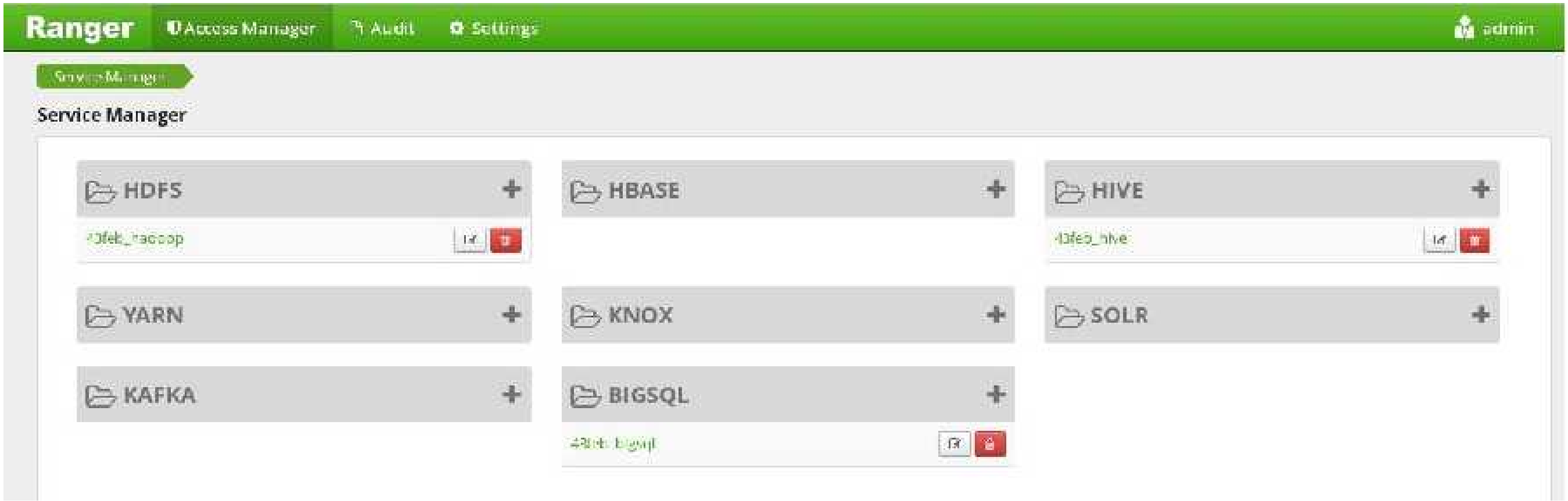
EMPNO	FIRST_NAME	SALARY	BRANCH_NAME
1	Steve	[redacted]	Branch_B
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
4	Craig	[redacted]	Branch_B
[redacted]	[redacted]	[redacted]	[redacted]
6	Stephanie	[redacted]	Branch_B
7	Julie	[redacted]	Branch_B
[redacted]	[redacted]	[redacted]	[redacted]



EMPNO	FIRST_NAME	SALARY	BRANCH_NAME
1	Steve	25970.38	Branch_B
2	Chris	29007.57	Branch_A
3	Paula	14987.06	Branch_A
4	Craig	22516.93	Branch_B
5	Pete	19114.22	Branch_A
6	Stephanie	26183.81	Branch_B
7	Julie	13629.91	Branch_B
8	Chrissie	24922.36	Branch_A

Total: 8 Selected: 0 ‹ **1** › **10** | 25 | 50 | 100 ▾

Big SQL and Apache Ranger Integration



- **Setup ACLs for access to Big SQL tables:**
  - create, alter, analyze, load, truncate, drop, insert, select, update, and delete.
- **Supports Ranger Audit**
  - Big SQL access audit records written to HDFS and/or Solr
- **If also using Ranger Plugin for Hive – operates independent of Big SQL plugin**

Big SQL Tables over S3 Object Storage

```
CREATE HADOOP TABLE staff ( … )
LOCATION
        ’s3a://s3atables/staff’;
```

- **Create Tables over Data residing in Object Store directly (no copy required into Hadoop)**

- **Once configured, Object Store tables work like any other table in Big SQL**

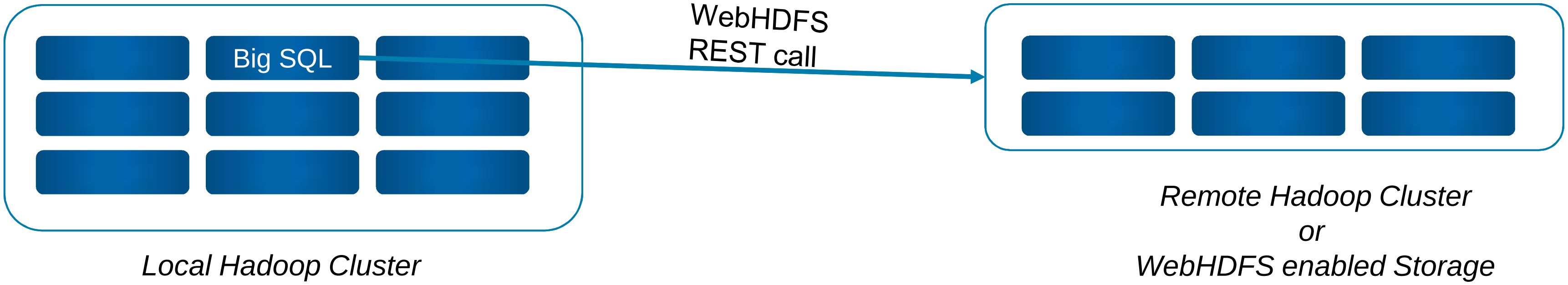
- **Benefits:**
  - No need to copy data into Hadoop first! Query data where it resides.
  - Partitioning supported!

- **Tradeoff:**
  - Expect reduced performance relative to HDFS local tables

LOAD FROM
Object Store
also supported!

Big SQL Tables over WebHDFS (Technical Preview)

```
CREATE HADOOP TABLE staff ( …  )
   PARTITIONED BY (JOB VARCHAR(5))
LOCATION 'webhdfs://namenode.acme.com:50070/path/to/table/staff';
```



- **Transparently access data on any platform implementing WebHDFS**
  - Examples: Microsoft Azure Data Lake (ADL) service
- **Once setup, WebHDFS tables work like any other table in Big SQL**
- **Technical Preview Limitations:**
  - WebHDFS via Knox not supported
  - Performance not well understood. Reduce performance expected.

LOAD FROM WebHDFS also supported!

Big SQL – Integration with Yarn and Spark

Big SQL is a self-tuning memory management SQL engine that integrates with Spark 2.1

Data engineer

Share data in memory
Big SQL worker
Spark executor
HDFS

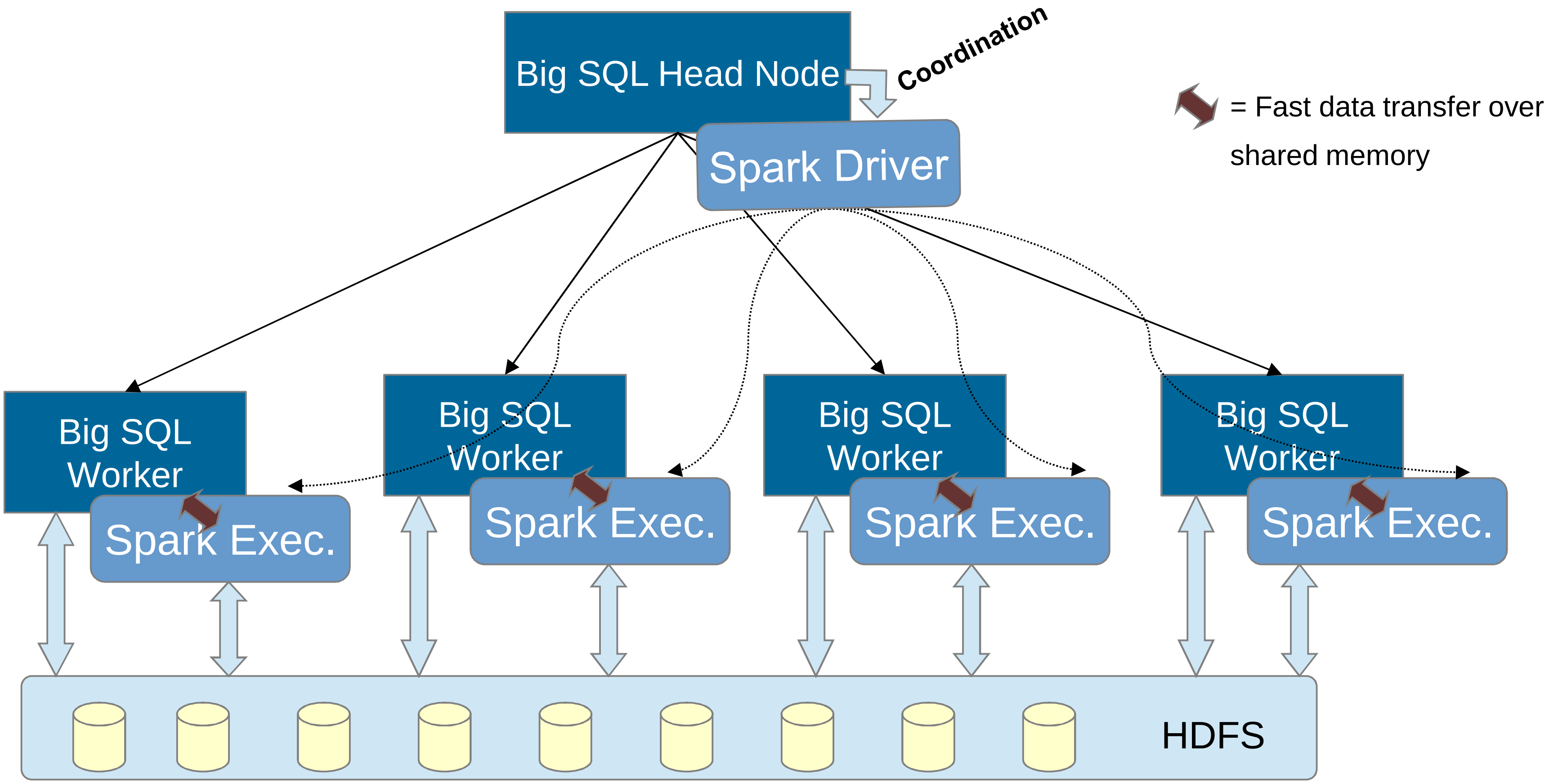
Spark 2.1 is a powerful analytic co-processor that complements the rich SQL functionality of Big SQL

Bi-directional integration allows Spark jobs can be executed from Big SQL

[illegible]

Tight integration with Spark enables Big SQL worker and Spark Executor to communicate in memory without writing to disk

Big SQL – The ONLY engine with Deep Integration with Spark



Exploit Big SQL from Spark

```
import org.apache.spark.sql.Dataset;
import org.apache.spark.sql.Row;
…
Dataset<Row> tableDf = sqlCtx.read()
  .format("jdbc")
  .option("driver", "com.ibm.db2.jcc.DB2Driver")
  .option("url", "jdbc:db2://server1.foo.bar.com:32051/BIGSQL")
  .option("user", "joe")
  .option("password", "joespwd")
  .option("dbtable", "myshcema.mytable")
  .load();

tableDf.createOrReplaceTempView("myTable");
Dataset<Row> queryDF =
    spark.sql("SELECT col2, col3 FROM myTable WHERE col1 > 100");
```

Big SQL secures data for self-service data exploration.

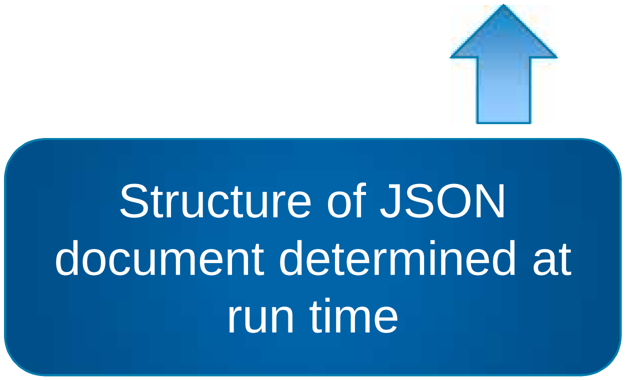
Used this way, Spark users are subject to Big SQL row/column security

- **Requirements:**
  - db2jcc.jar must be added to the classpath of the Spark application (found in /home/bigsql/java/)

Exploit Spark from Big SQL

Example: Spark Schema Discovery for JSON

```
SELECT doc.*
FROM TABLE(
        SYSHADOOP.EXECSPARK( class => ’DataSource’,
                                           load => ’hdfs://host.port.com:8020/user/bigsql/demo.json’)
) AS doc
WHERE doc.language = ’English’;
```



Structure of JSON document determined at run time

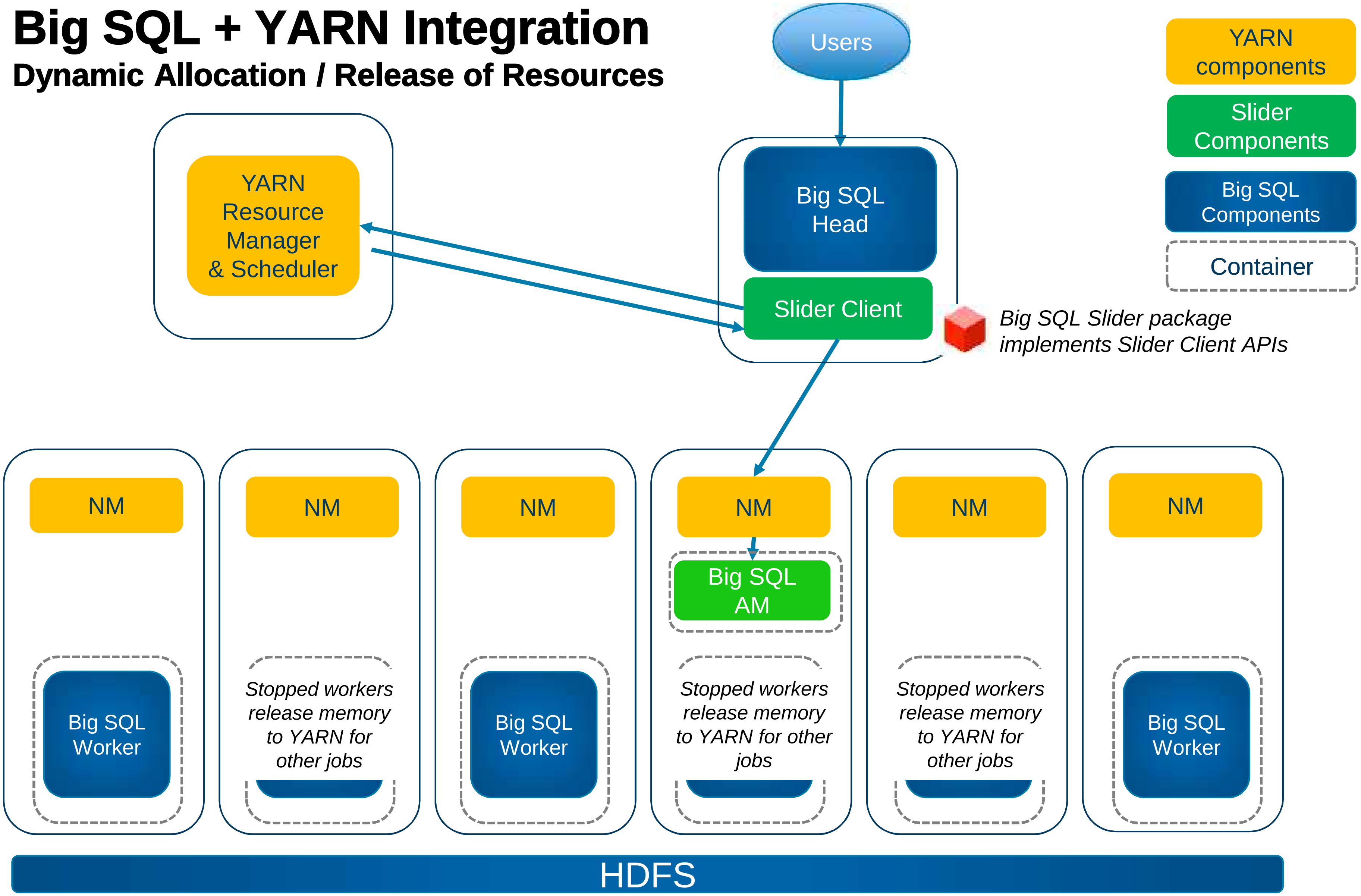
- **Bring the best of Spark into Big SQL!**
  - Machine Learning
  - Cache remote tables (Spark has rich library of connectors)
  - Graph Processing
  - General in memory processing

Apache Slider

- **Apache Slider**
  - Enables long running services (e.g. Big SQL) to integrate with YARN (similar to HBase)
  - Provides:
    - Implementation of Application Master
    - Monitoring of deployed applications
    - Component failure detection and restart capabilities
    - Flex API for adding/removing instances of components of already running
- **Apache Slider does not yet have a GUI nor Ambari integration.**
  - Big SQL operations for Slider can be executed through two methods:
    - Big SQL Service Actions in Ambari
    - Command line scripts

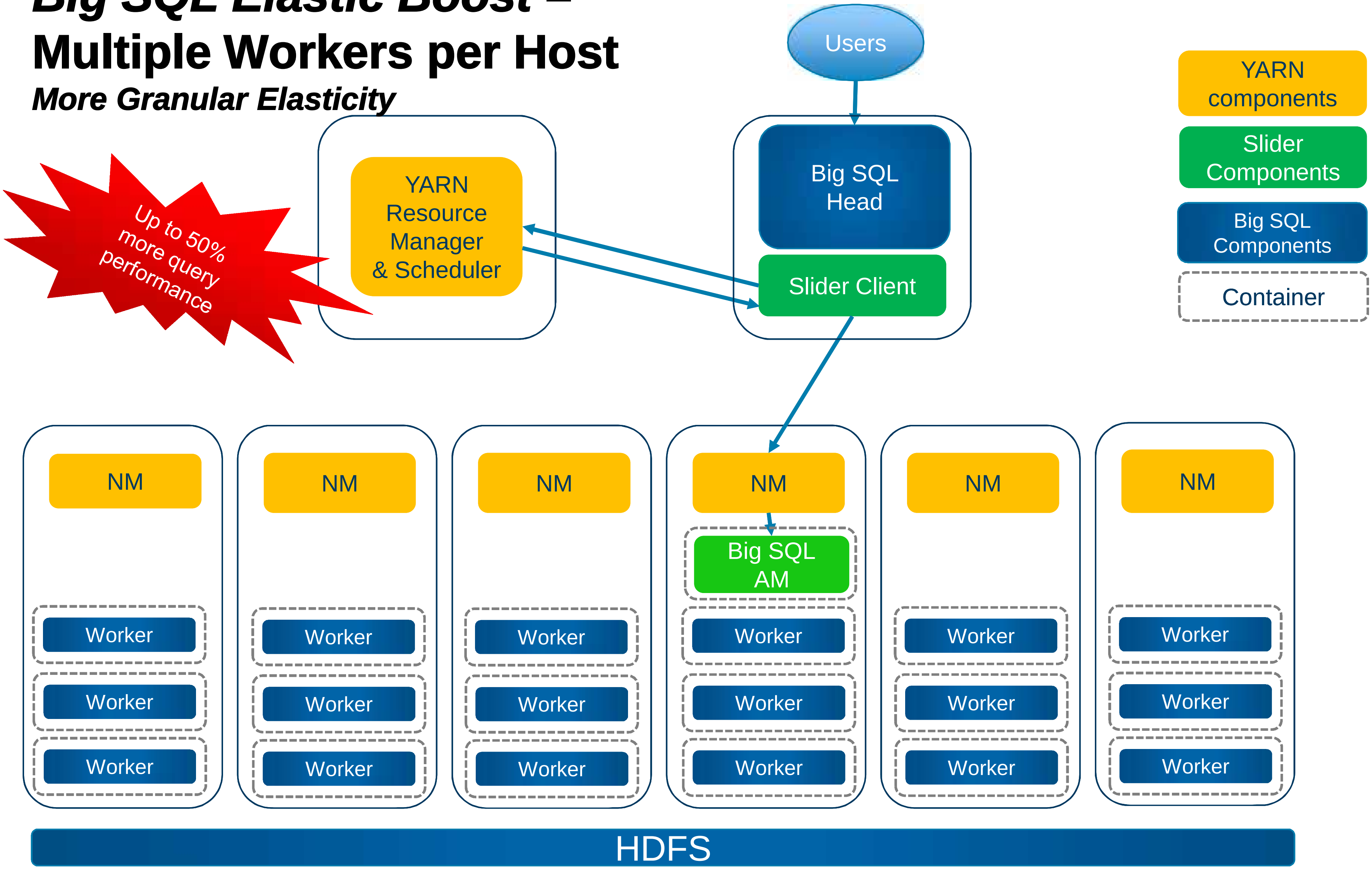
Big SQL + YARN Integration

Dynamic Allocation / Release of Resources

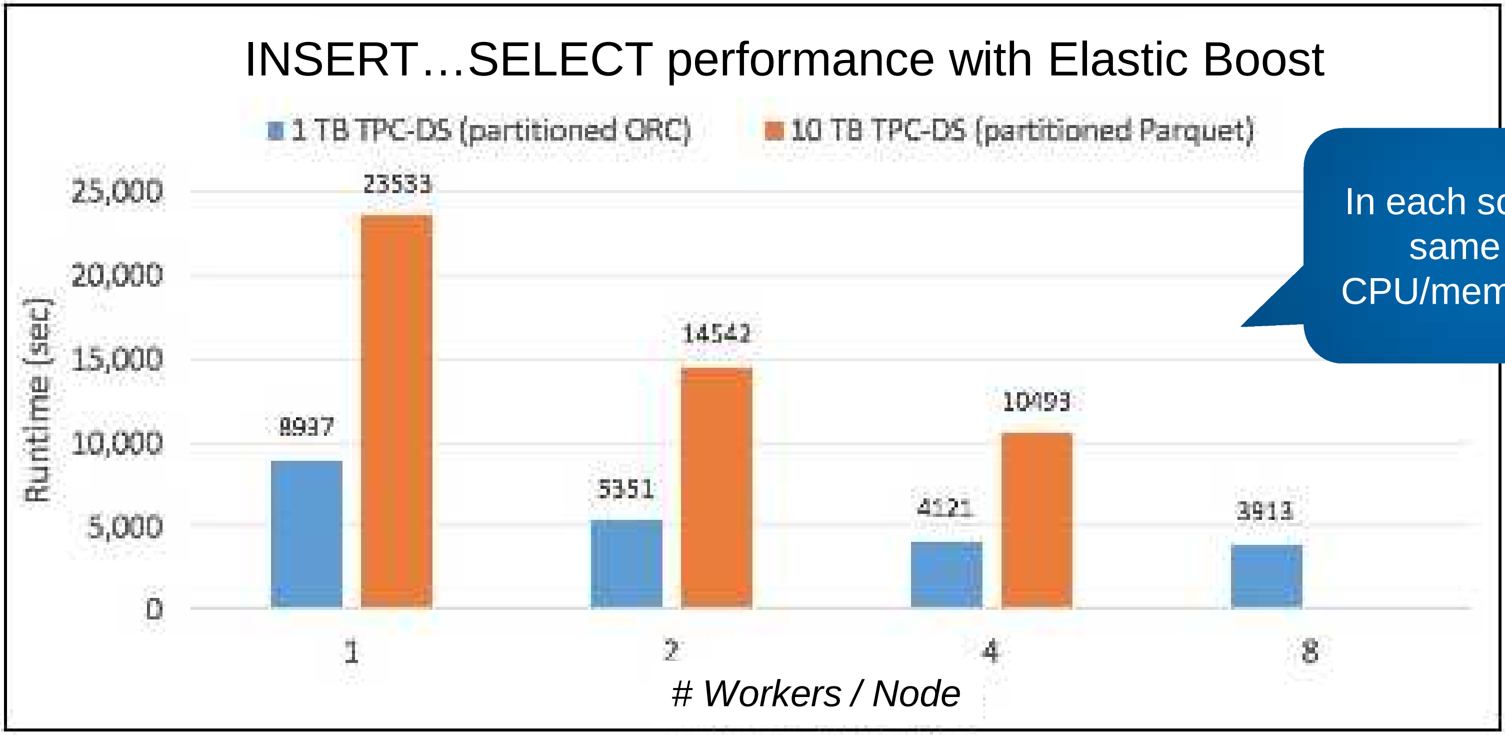

Users
YARN components
Slider Components
Big SQL Components
Container
YARN Resource Manager & Scheduler
Big SQL Head
Slider Client
Big SQL Slider package implements Slider Client APIs
NM
NM
NM
NM
NM
NM
Big SQL AM
Big SQL Worker
Stopped workers release memory to YARN for other jobs
Big SQL Worker
Stopped workers release memory to YARN for other jobs
Stopped workers release memory to YARN for other jobs
Big SQL Worker
HDFS

Big SQL Elastic Boost – Multiple Workers per Host

More Granular Elasticity



Elastic Boost Improves INSERT Performance

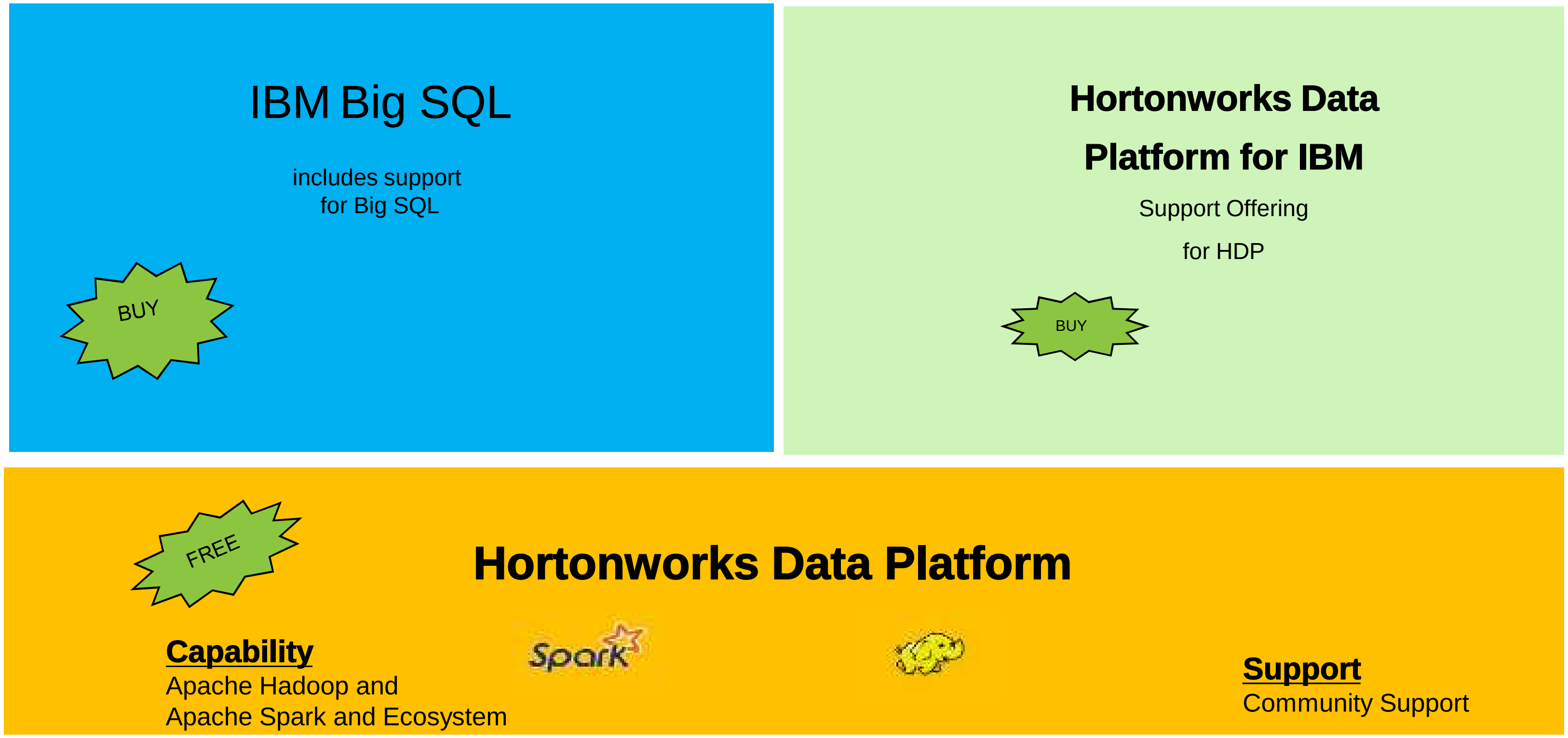


In each scenario, the same TOTAL CPU/memory is used

- **For both 1 and 10 TB TPC-DS dataset**
  - 2 Workers/Node: **1.6x speedup**
  - 4 Workers/Node: **2.2x speedup**

Big SQL 5.0 – How it fits with Hortonworks

- **Big SQL deploys on top of Hortonworks Data Platform(HDP)**
  - *Includes:* IBM Support for Big SQL
- **Hortonworks Data Platform for IBM (Support only)**
- **Hortonworks Data Platform can be downloaded for FREE.**



IBM Big SQL

includes support
for Big SQL

BUY

Hortonworks Data Platform for IBM

Support Offering
for HDP

BUY

FREE

Hortonworks Data Platform

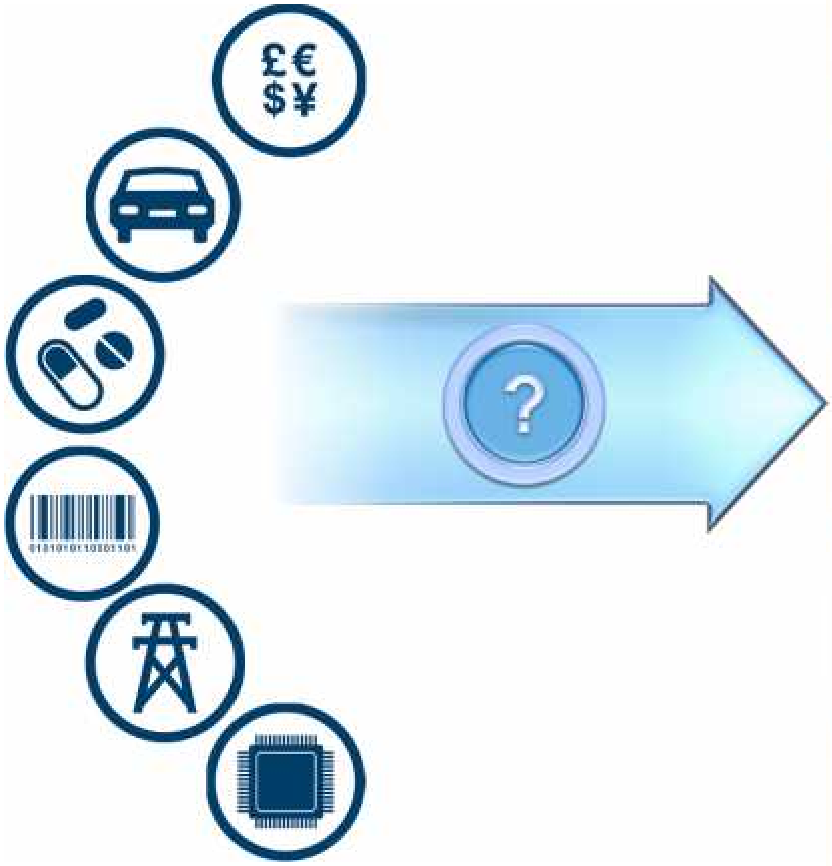
Capability
Apache Hadoop and
Apache Spark and Ecosystem

Support
Community Support

Topics for Today

- **Strategy Overview**
- **Db2 V11.1.3.3 – Introduction !!**
- **Private Cloud – Introduction !!**
- **Flex Points and HDM Offering**
- **Appliance News**
- **Hadoop and Open Source**
- **Event Processing**
- **Next Generation Data Virtualization**

Event-Driven Systems Span Many Industries



Multi-channel customer sentiment and experience a analysis

Detect life-threatening conditions at hospitals in time to intervene

Predict weather patterns to plan optimal wind turbine usage, and optimize capital expenditure on asset placement

Make risk decisions based on real-time transactional data

Identify criminals and threats from disparate video, audio, and data feeds

Industry Use Cases

Retailer Loyalty Program

Integrate streamed payment, couponing events, climate, calendar, mobile data. measure refine, deliver better couponing and loyalty system

Smart Metering/Smart Grid

Deliver a Integrated platform for optimizing energy usage, capacity and billing across a smart grid system

Banking Risk Exposure

Combine account transactions from across the bank to provide a master ledger for real-time risk exposure and fraud identification

Satellite Tracking System

Track satellites in real time and produce analytics on operations and performance

Intelligent Manufacturing

Deliver real-time monitoring framework for automated production lines, providing productivity, preventive maintenance, and reporting

Transactional to Analytics Consolidation

Capture your transactions and augment with external data into an analytics platform for deeper analytics

What is Db2 Event Store?

A unified offering for Fast Data which delivers…

IBM **Db2 Event Store**

1 Lightning Fast Ingest

- 1 Million inserts per second per node
- Ingest scales linearly with added nodes
- Data ingested quickly, then refined and enriched

2 Real-time Analytics

- Real-time analytics over ALL ingested data
- Super-fast lookups and intelligent scans
- Integrated machine learning capabilities

3 Integrated and Highly Available

- Packaged and integrated with IBM Data Science experience; available Streams sink
- Remains available on node failure
- Architected to scale to very large clusters

4 Built for Data Sharing and Efficiency

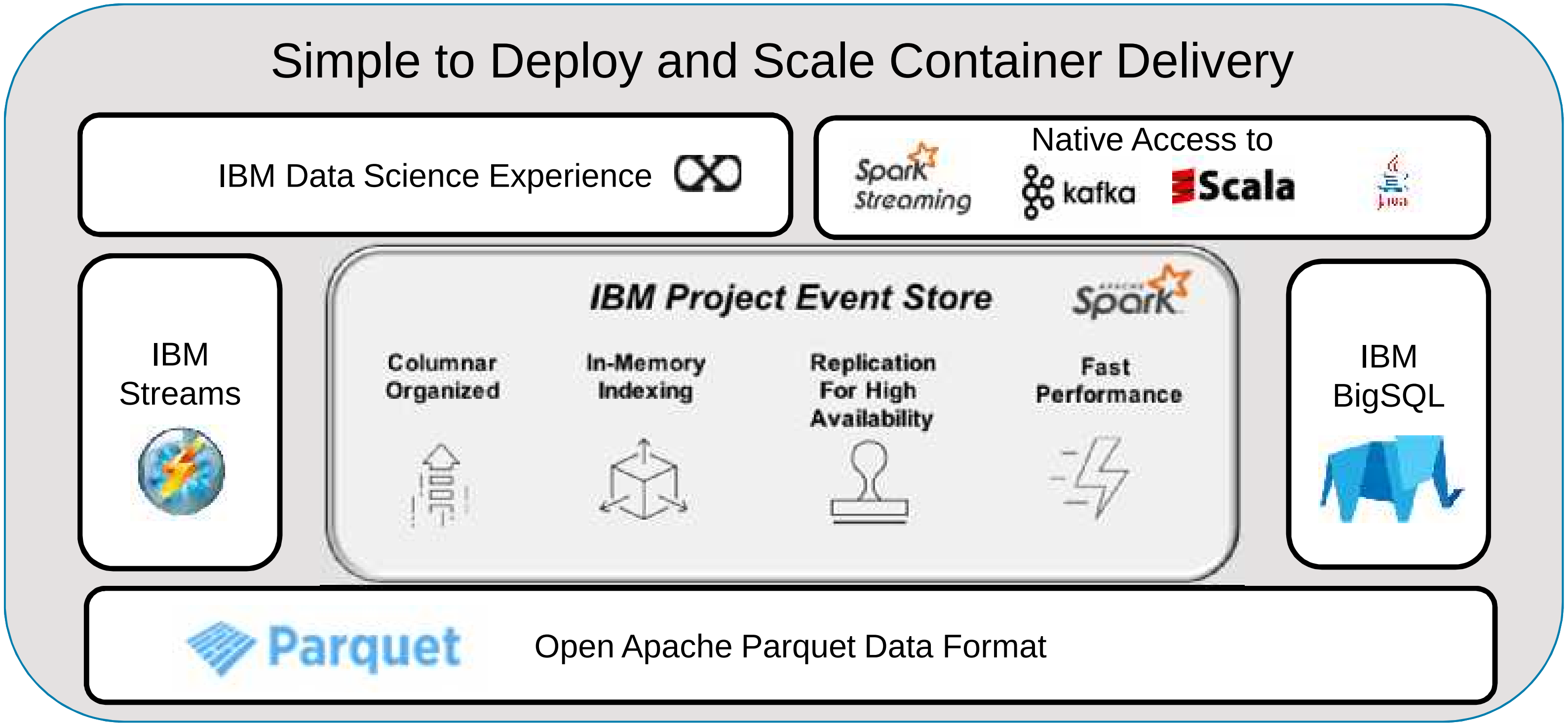
- Writes to shared storage in Parquet format
- Able to leverage low-cost object storage
- Single copy of the data

Parquet

Db2 Event Store

Integrated System for Managing Events

IBM Event Store



Simple to Deploy and Scale Container Delivery

IBM Data Science Experience

Native Access to

IBM Streams

IBM Project Event Store

Columnar Organized

In-Memory Indexing

Replication For High Availability

Fast Performance

IBM BigSQL

Open Apache Parquet Data Format

Db2 Event Store – Competitive Positioning

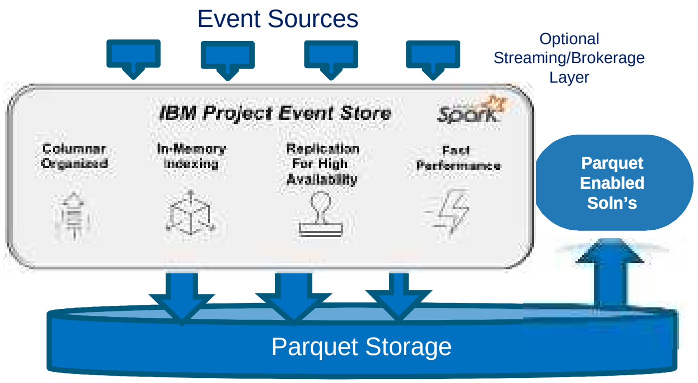
Db2 Event Store provides everything in the box

Reduced architectural components
Docker Container Delivery
Open Data Access

Complex Manual Architecture

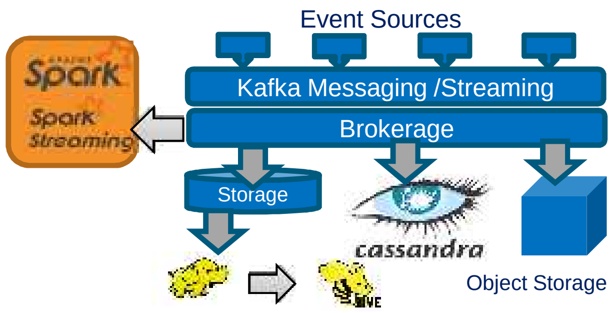
Put together your own open source components
Not everything works together
Hard to maintain

Simplified Approach With IBM Db2 Event Store



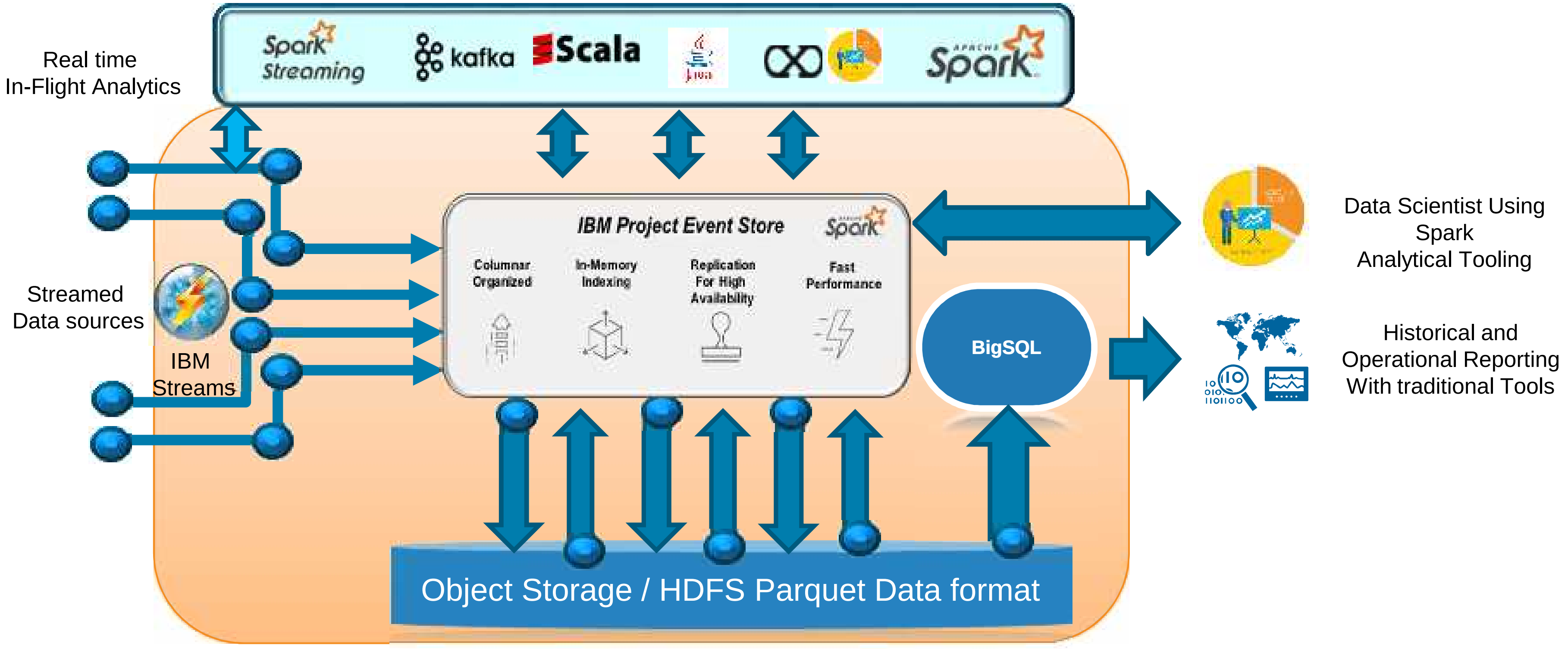
vs.

Digital Company Example



Db2 Event Store

IBM Event Store



Real time
In-Flight Analytics

Streamed
Data sources

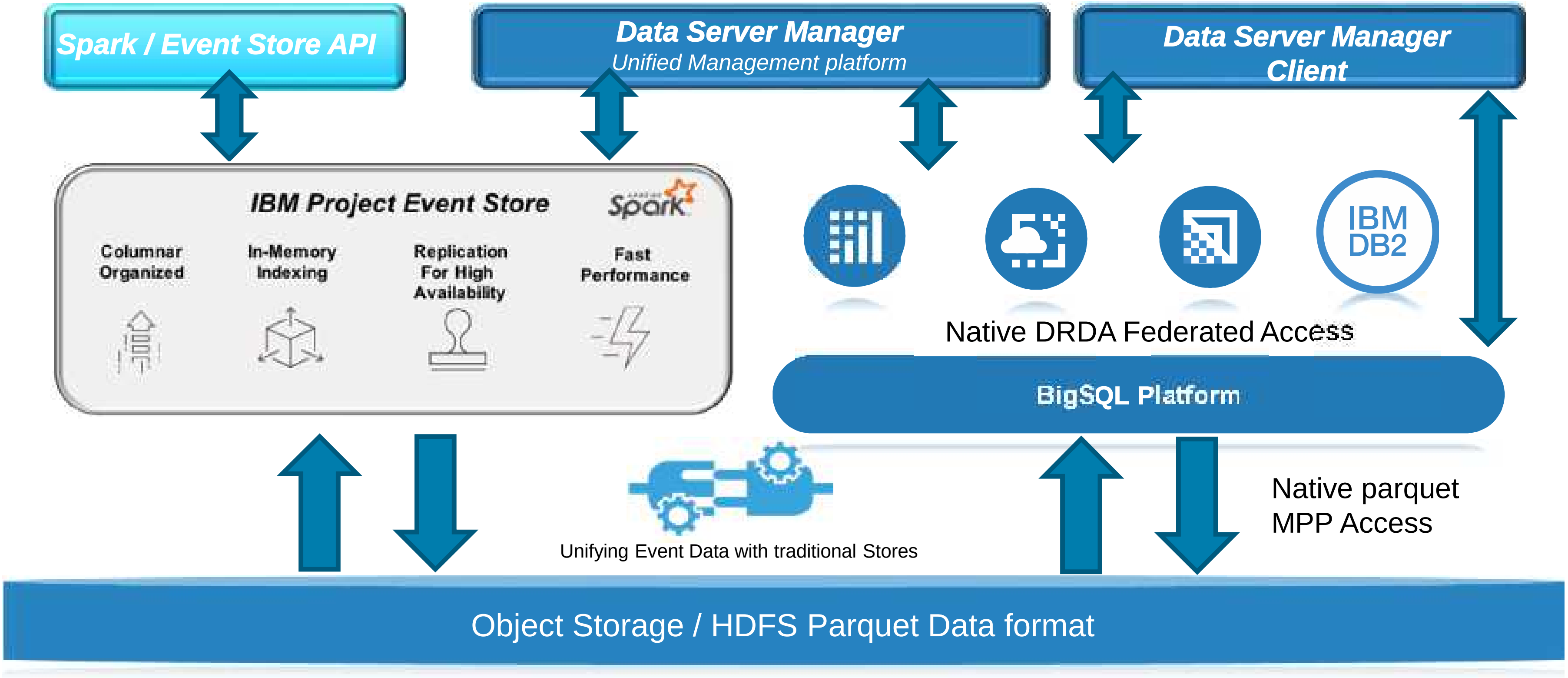
IBM
Streams

Data Scientist Using
Spark
Analytical Tooling

Historical and
Operational Reporting
With traditional Tools

84

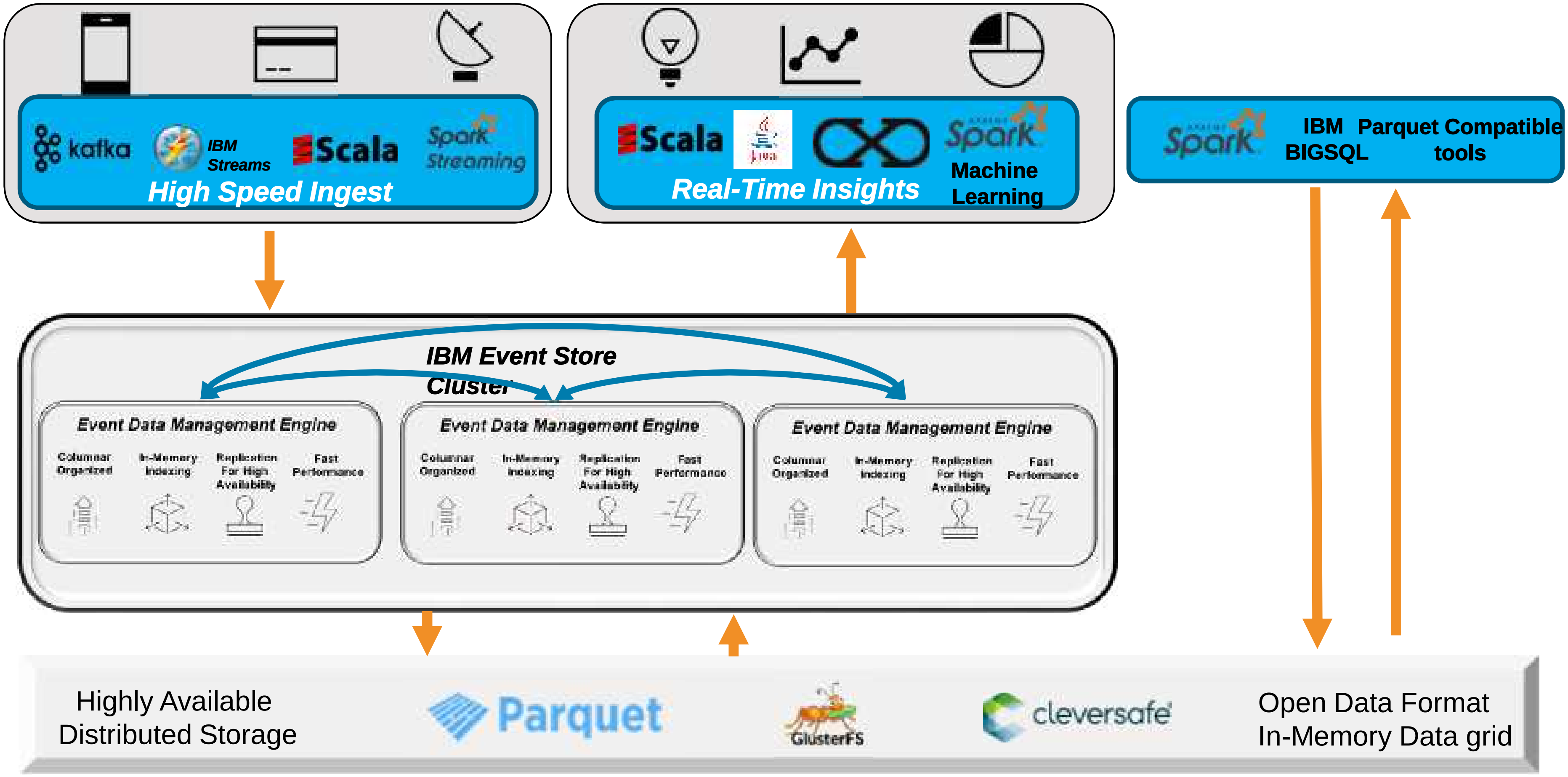
Db2 Event Store: Unified Data Access and Management



Db2 Event Store: Architecture

Understanding the Engine and Components

IBM Event Store



kafka

IBM Streams

Scala

Spark Streaming

High Speed Ingest

Scala

Java

Spark

Machine Learning

Real-Time Insights

Spark

IBM BIGSQL

Parquet Compatible tools

IBM Event Store Cluster

Event Data Management Engine

Columnar Organized

In-Memory Indexing

Replication For High Availability

Fast Performance

Event Data Management Engine

Columnar Organized

In-Memory Indexing

Replication For High Availability

Fast Performance

Event Data Management Engine

Columnar Organized

In-Memory Indexing

Replication For High Availability

Fast Performance

Highly Available Distributed Storage

Parquet

GlusterFS

cleversafe

Open Data Format

In-Memory Data grid

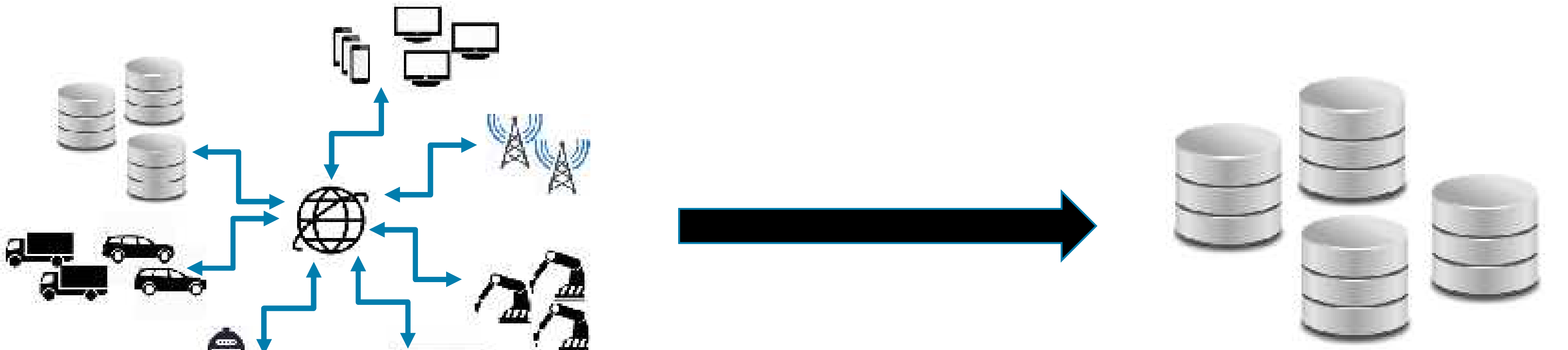
Db2 Event Store

Demo

Topics for Today

- Strategy Overview
- Db2 V11.1.3.3 – Introduction !!
- Private Cloud – Introduction !!
- Flex Points and HDM Offering
- Appliance News
- Hadoop and Open Source
- Event Processing
- Next Generation Data Virtualization

Analytics Today…

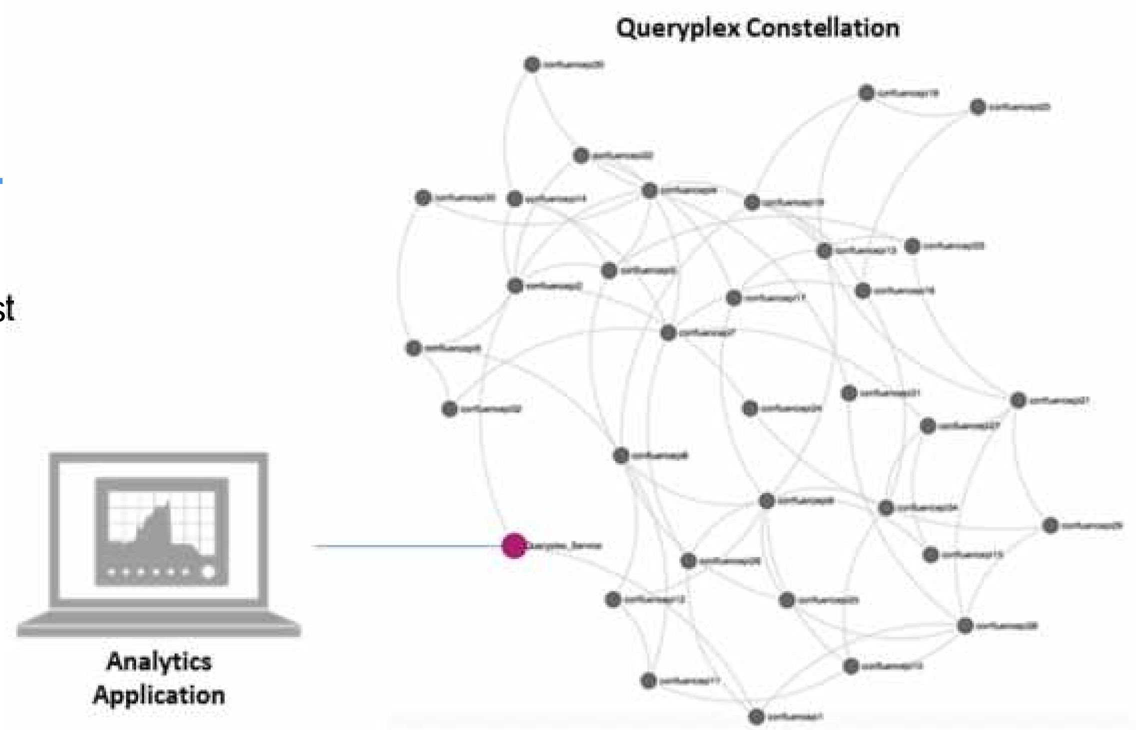


Data Lake or Data Warehouse

- Costly and Complex
- High Latency to copy and synchronize
- Available compute resources under-utilized
- Error prone and difficult to retain data integrity

IBM Queryplex

An emerging technology now in beta trial



1

Query anything, anywhere.

Query **many diverse data sources** across cloud, on-premise and mobile with advanced analytics using the most popular languages and tool

SQL, Spark, R, Notebooks, Python, Data Science Experience (DSX), Cognos Analytics, common Analytics tools

2

Query many sources as one with extreme simplicity.

Connect **few to many devices and data stores** into a single self balancing constellation. Avoid the complexity of centralized copies. Data only persists at the source.

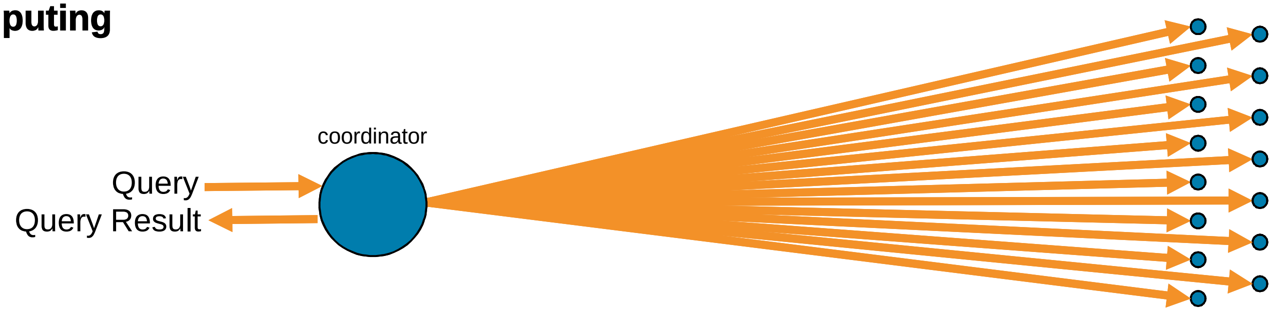
3

Massive speedup.

Many times acceleration using the power of every device.

IBM Queryplex’s Computational Mesh

Edge Computing



Query issued against the system

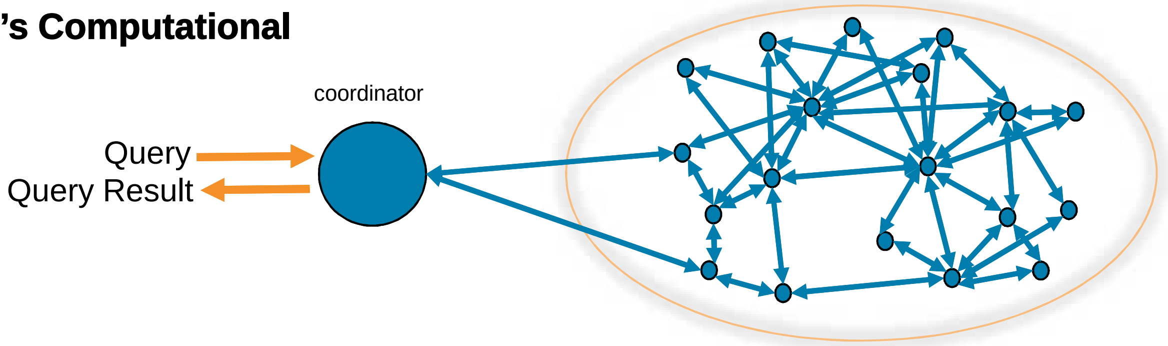
A coordinator C1, receives the request and fans the work out to edge devices

Edge devices individually perform as much work as they can based on their own data. Individual results are sent back to the coordinator for final merging and remaining analytics.

Coordinator receives intermediary results from all edge devices, merges results, and performs remaining analytics

Time

Queryplex’s Computational Mesh



Query issued against the system

A coordinator C1, receives the request and fans the work out to edge devices

Edge devices self organize into a constellation where they can communicate with a small number of peers. Nodes collaborate to perform almost all analytics, not only analytics on their own data.

Coordinator receives mostly finalized results from just a fraction of devices. Completes the final bits of the query result.

Time

IBM Queryplex - Supported Languages & Data Sources

Query Languages	
SQL (ANSI)	✓
SQL (Oracle)	✓
SQL (DB2)	✓
SQL (PostgreSQL, Netezza)	✓
Scala	✓
PL/SQL	***Future***
SQL PL	***Future***
PySpark	✓
Python	✓
R & SparkR	✓

Mix Any Combination of Data Sources			
Oracle	✓	Excel	✓
DB2	✓	CSV (delimited text)	✓
Netezza	✓	MongoDB	✓
PostgreSQL	✓	Accumulo	***Future***
Informix	✓	Redis	***Future***
MySQL	✓	Cloudant	***Future***
SQLServer	✓		
DerbyDB	✓		

IBM Queryplex - Potential Use Cases

Industry	Use Case
Telco	5G Wireless and Enterprise IoT (Devices anywhere)
Telco	Cell tower and site monitoring for Operations and Maintenance
Telco	Cell site subscriber metadata analytics for Law Enforcement
Telco	Set Top Box home applications, monitoring, Content access statistics
Energy & Utilities	Distribution network monitoring and maintenance
Energy & Utilities	Smart metering
Manufacturing & Cross/Enterprise	Time sensitive data queries
Insurance	Auto usage device monitoring
Cross/Enterprise	Data Virtualization
Cross/Enterprise	Data provisioning to untrusted external entities
Gaming	Real-time gaming queries
Media & Entertainment	Subscriber viewing and content correlation
Military	IoT Sensors

IBM Queryplex – Interested in hearing more ?

IBM Queryplex

The power of many together

http://queryplex.com

Les King

Director, Hybrid Data Management Solutions

May, 2018

lking@ca.ibm.com

ca.linkedin.com/pub/les-king/10/a68/426

Hybrid Data Management Strategy and New News !