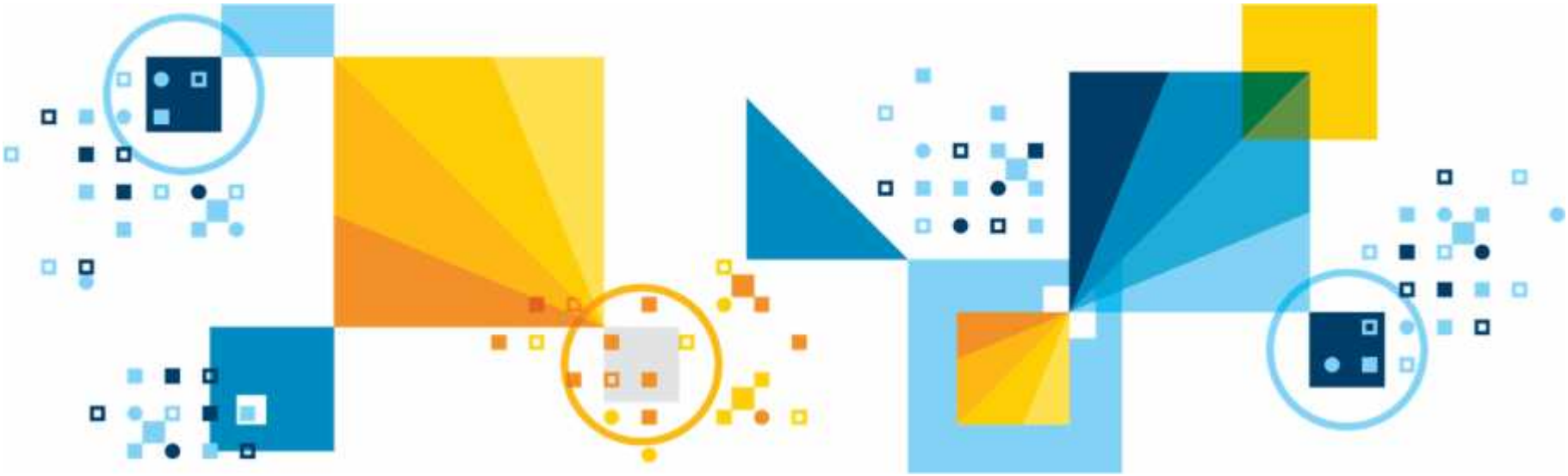
**Les King**
**Director, Hybrid Data Management Solutions**
**March, 2018**
**lking@ca.ibm.com**
**ca.linkedin.com/pub/les-king/10/a68/426**

# 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

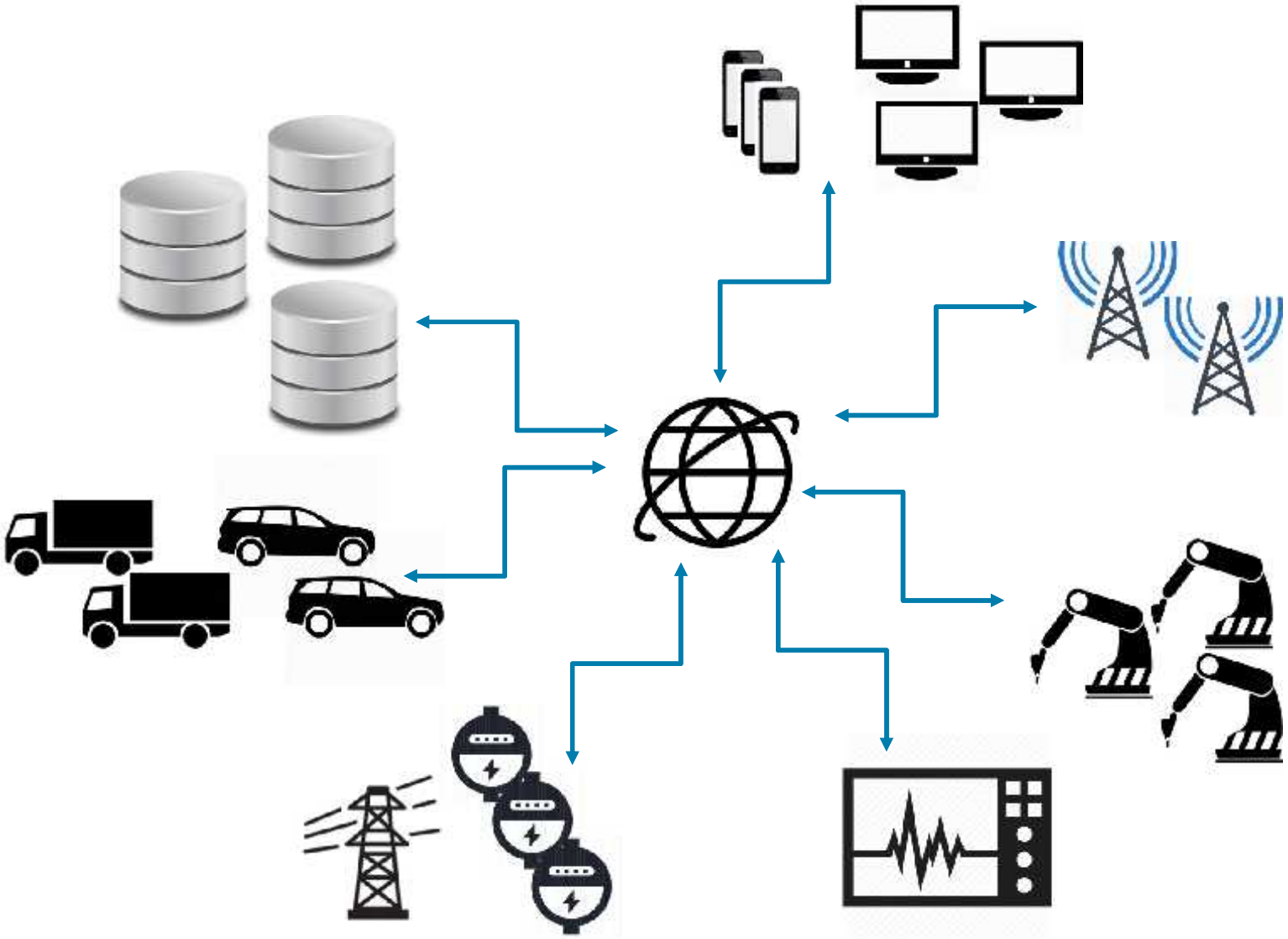
- 27 years of Information Management, Database and Analytics
- Technical sales
- Technical customer support
- Software development
- Product / Offering management
- Product Marketing
- Product Sales
- Taught mathematics at University of Toronto
- Teaching data warehousing, big data and Db2 at Seneca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

# Data is Everywhere

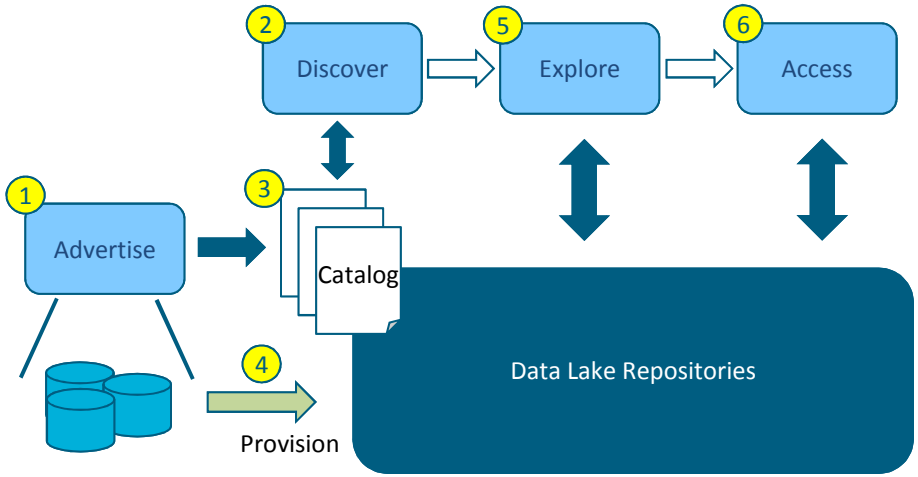


# All businesses have become data driven


# Market Observations - #1

## There is increasing pressure to perform analytics where data gets created

*"Point-of-decision HTAP promises to simplify the information infrastructure by removing unnecessary data marts and, potentially, data warehouses."*
*– Gartner*



# Market Observations - #2

## **Event-driven applications** will enable new analytic use cases

*"Event-driven real-time digital business is poised to become a priority for mainstream business " - Gartner*

*"In-process HTAP could potentially redefine the way some business processes are executed" - Gartner*

*"Perishable Insights: Insights that can provide exponentially more value than traditional analytics but the value expires and evaporates once the moment is gone" Forrester – Mike Gualtieri*






# Market Observations - #3

## Hybrid cloud capabilities of software support economies of scope – and - Private cloud needs cloud-scale convenience

*"Public cloud adoption has stalled for the time being, signaling enterprises are moving to the hybridization phase of their IT transformations." TBRI 2H 2016*

*"By end of 2016 38% of the IT Market spend will be private hosted or private on Prem Cloud with On-Demand Convenience future growth point within private cloud." - IDC*

*"Skills, timing or cost to effectively procure, assemble, run, manage disperse infrastructure resource require integrated versatile platform offerings with appliance-like simplicity" – A client*

**1 | Public**

Maximize on cloud economics and agility.

**2 | Dedicated**

Everything is dedicated and connected to you — agility of public cloud, yet feels like home.

**Seamless Experience**

Regardless of which combination you choose, you can expect a single, seamless experience.

**3 | Private**

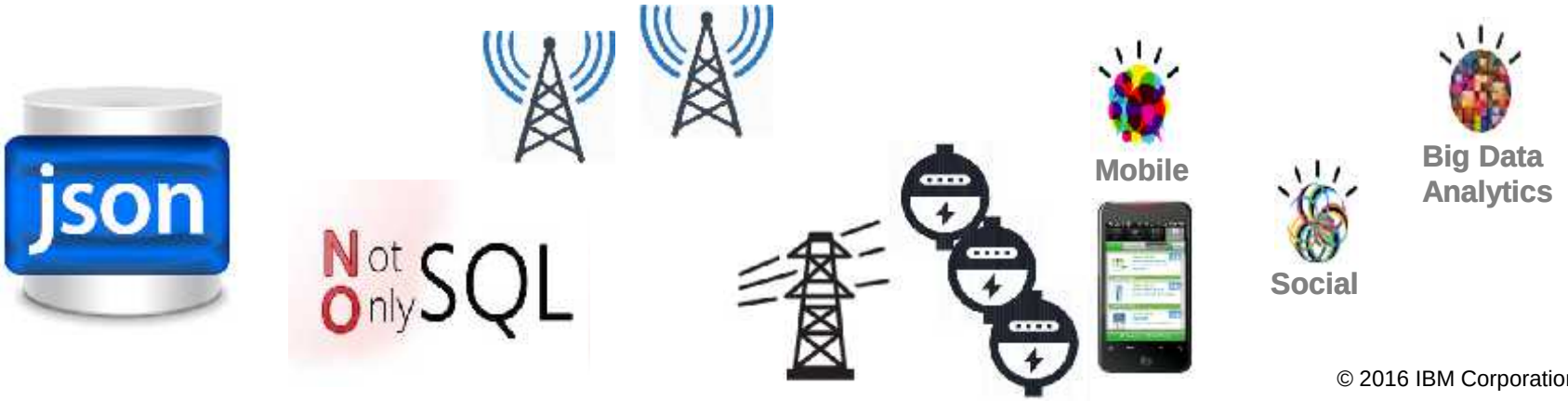
Behind the firewall for the most sensitive workloads.

# Market Observations - #4

## Diverse data sources support an ecosystem of innovation

*"Top relational database solutions are now offering a wide range of new features to combine structured and unstructured data types " …. Database decision-makers need to look at investing in these database technologies. – Forrester*

*Many have added open-source and NoSQL products to their portfolio in an attempt to capture a new generation of buyers – Established Vendors*



# 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

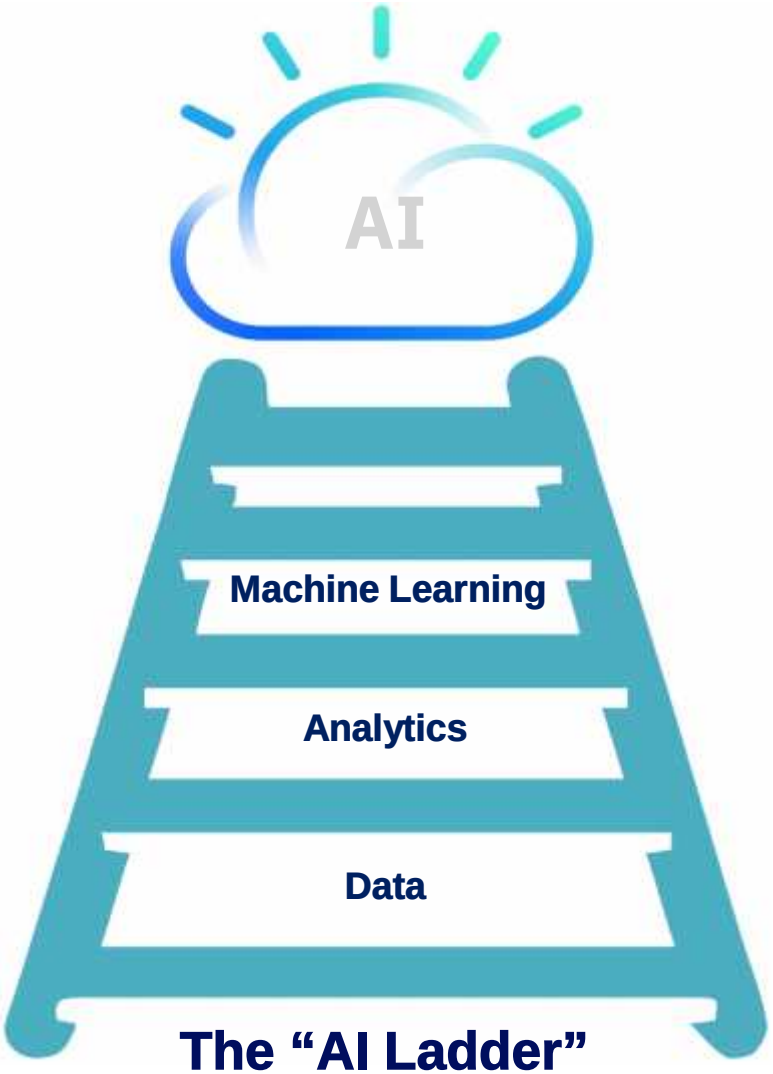
- Storing all events implies large data sets
- Storage costs are significant, and must be managed

## 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

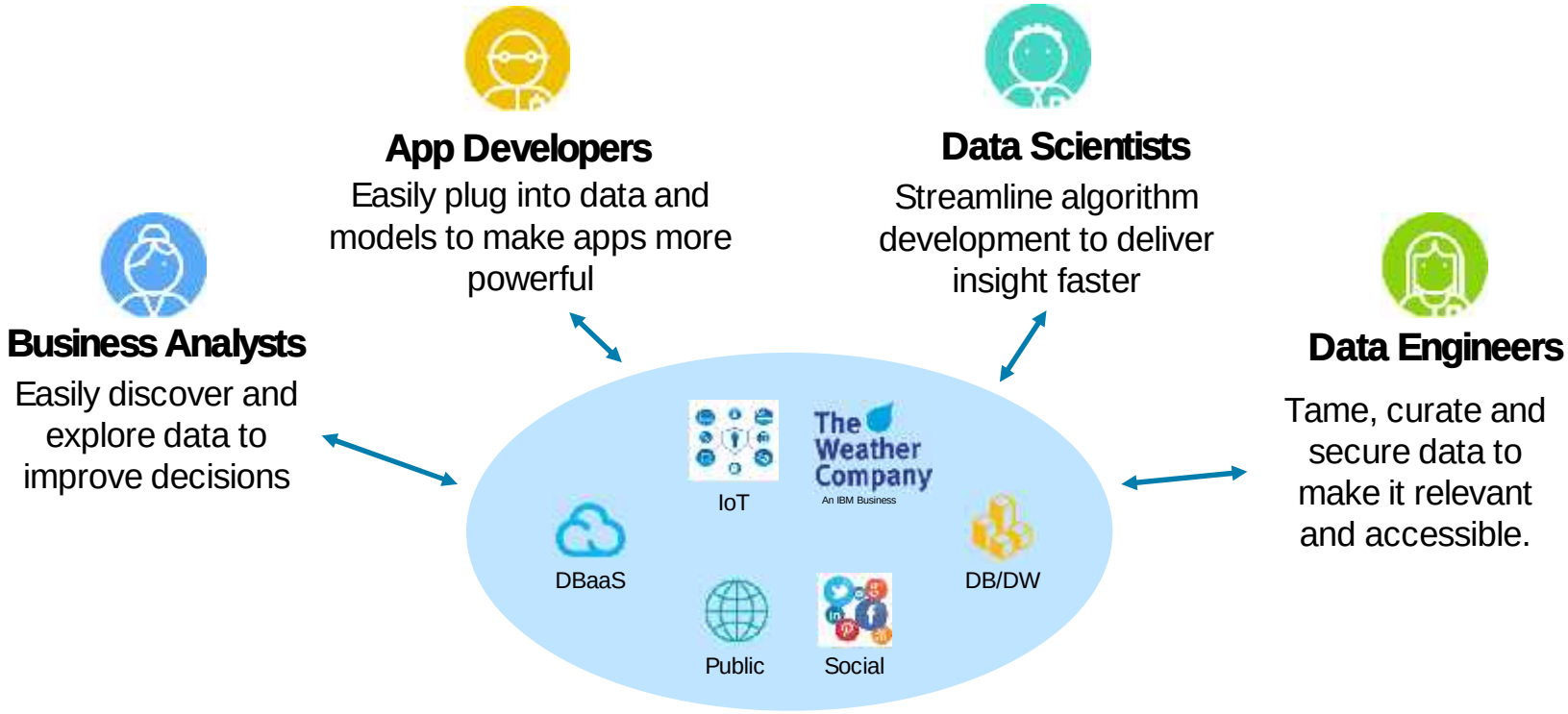
- Data must be available to the entire organization without requiring replication or duplication
- Maintain data in open format for future-proofing



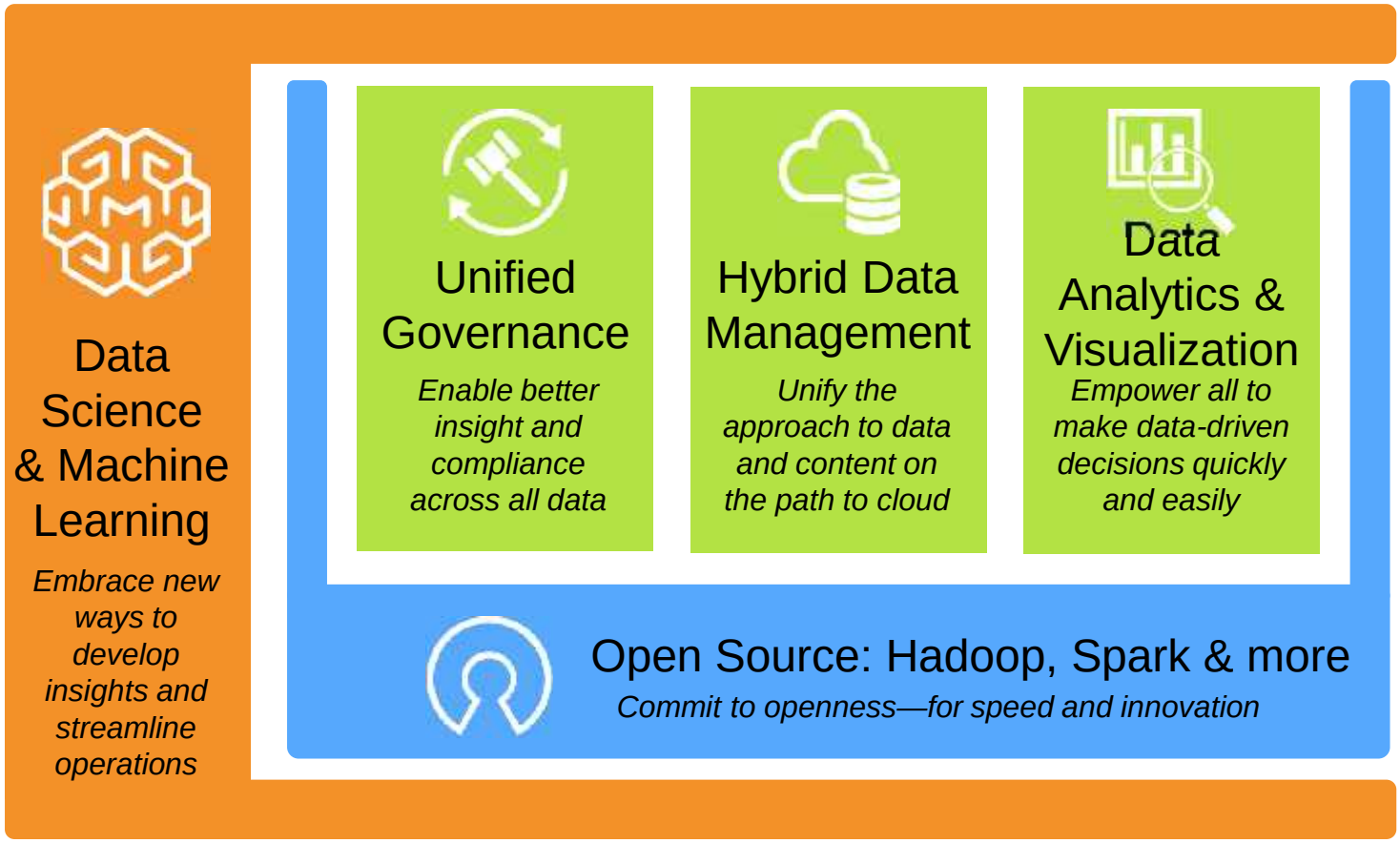

The "AI Ladder"

# Data Professionals – Evolving Roles

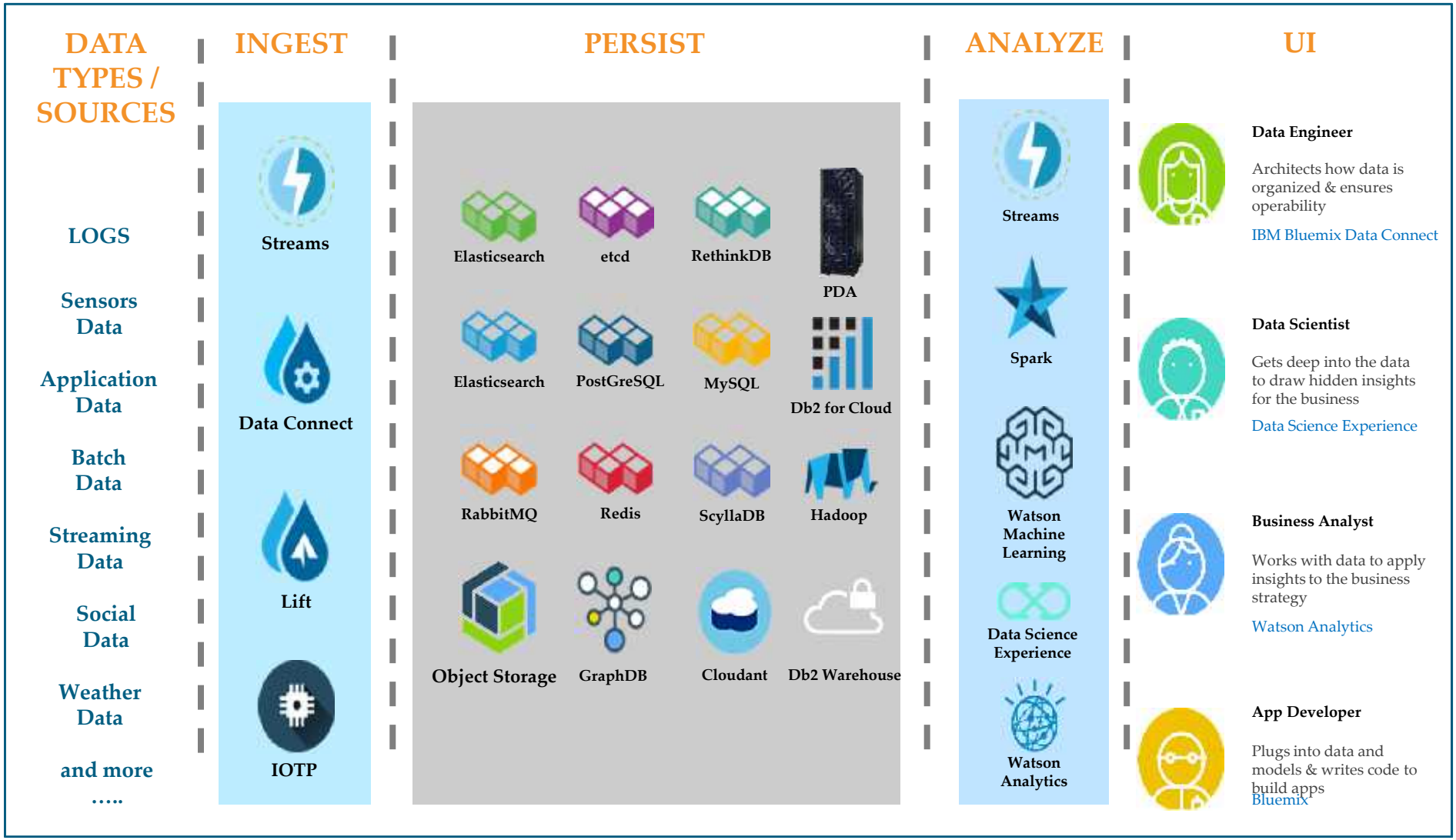
***As data maturity increases, so does the number of data professionals who are hungry to put data to work***



# Essential Elements of Data and Analytics



# Watson Data Platform



# IBM's 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 – All Workloads All Deployments

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

| Systems of Insight | Systems of Record | Systems of Engagement | Event Processing |
|---|---|---|---|
| Db2 Warehouse On Cloud, Big SQL | | Db2 On Cloud | Db2 Event Store |
| Db2 Warehouse, Big SQL | | Db2 OLTP | Db2 Event Store |
| IIAS, Big SQL, IBM Db2 | | Db2 | Db2 Event Store |

PUBLIC CLOUD

PRIVATE CLOUD

ON-PREMISES

# Common SQL Engine – Application & Ecosystem Support

IBM Db2 NETEZZA an IBM Company ORACLE SAP Spark sas python SPSS IBM COGNOS tableau MicroStrategy Best In Business Intelligence

| 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 |

# 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


# FlexPoints & HDM Offering

# Portfolio Simplification:

## *Guiding Principles*

| | |
|---|---|
| **Offering Scope** | ▪ Most on-premises analytics offerings included |
| **Offering Characteristics** | ▪ Available today for perpetual licenses |
| **“Flex Licenses”** | ▪ Consumption based model |
| **Channel** | ▪ Direct, Reseller and OEM channels to also lead with 3 Platform offerings |
| **Value Proposition** | ▪ Simplified licensing and pricing – very flexible |

# 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

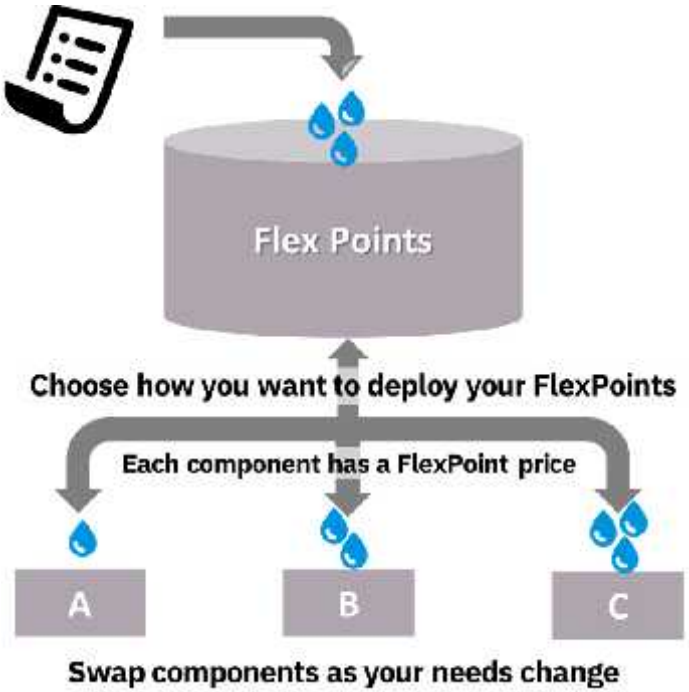
### Data Science & Business Analytics

- 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"**




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


# Db2 V11.1.2.2

# 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 |

BLU Acceleration
Speed. Simplicity. Affordability.

COGNOS

SPSS

SAP


Not Only SQL

# Db2 Version 11.1.2.2 Highlights

## Higher Availability and Core Capabilities

**Near-zero outage recovery**
- Online crash recovery
- pureScale REBUILD restore

## 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

## NoSQL Support

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


## Additional Operating System Support

Solaris Support – by exception

MacOS Support – by exception

## Db2 Tooling Capabilities

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

## Packaging Changes

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

# 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


# *Analytic Appliances*

# 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

# Less admin & more analytics

**Load and Go**

**Low TCO**

**One Touch Support**



**Simplicity**

### 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

**Data Experts, not Database Experts**

### 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

# Speed of Thought Analytics

## 2X – 5X Power

## Performance

### Powered by RedHat® Linux

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

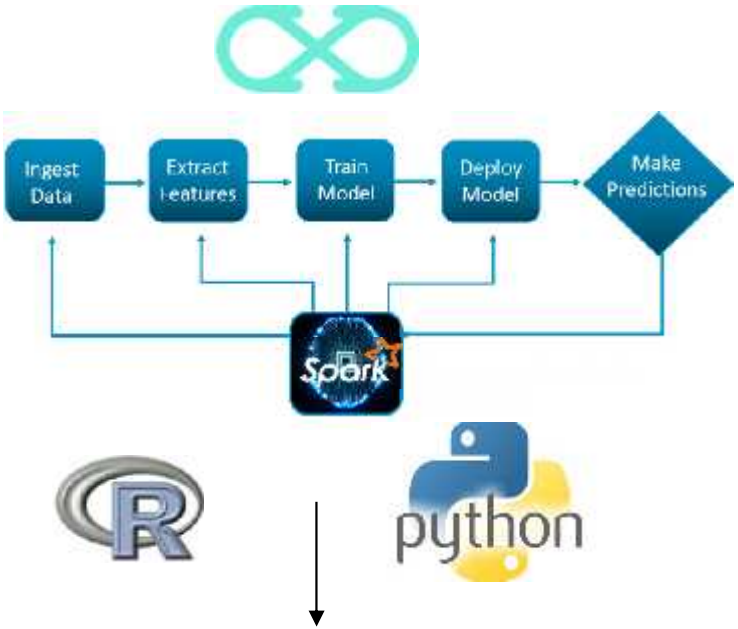
### Data Skipping

Skips unnecessary processing of irrelevant data

### Actionable Compression

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

# Rich, Self-Learning Analytics




**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

# Write Once, Run Anywhere

***IBM Data Lift***

***IBM Fluid Query***

**Hybrid**

### 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

# Unmatched multi-dimensional Flexibility

Scalable

Versatile Workloads

**Flexible**

### 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

# 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**

# 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

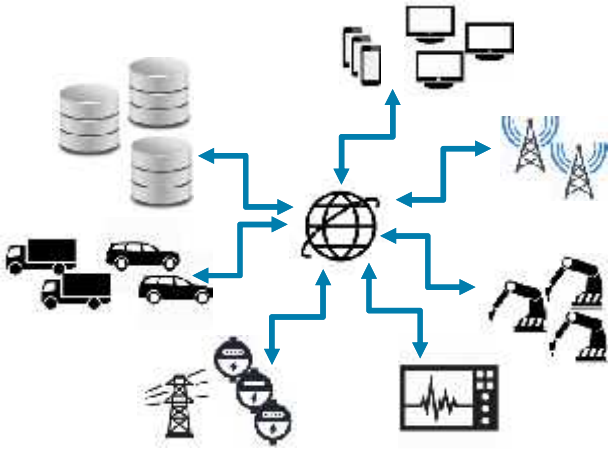

Uniform experience, simultaneous use, and easy transition between different implementations

Common analytics engine across all the platforms: Db2 Warehouse


# *IBM Queryplex*

**BETA – TECHNOLOGY PREVIEW**

# 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*

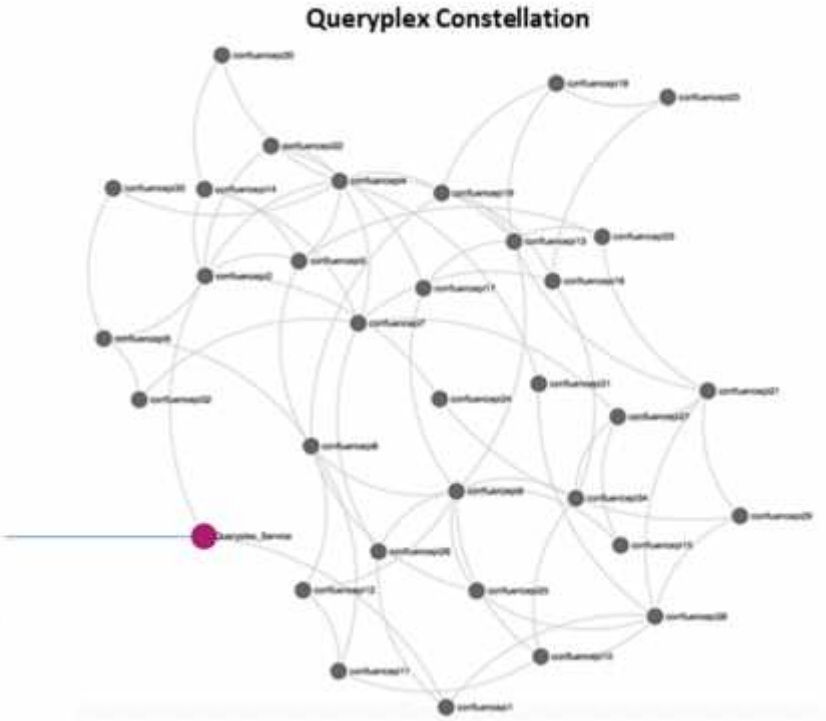
### Query anything, anywhere.

1

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

Analytics Application




### Query many sources as one with extreme simplicity.

2

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.

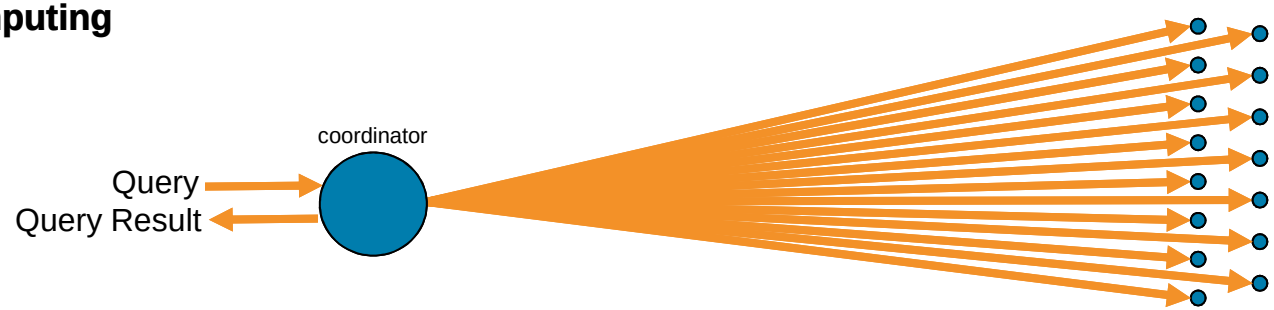
### Massive speedup.

3

**Many times acceleration** using the power of every device.

# IBM Queryplex's Computational Mesh

## Edge Computing




Query issued against the system

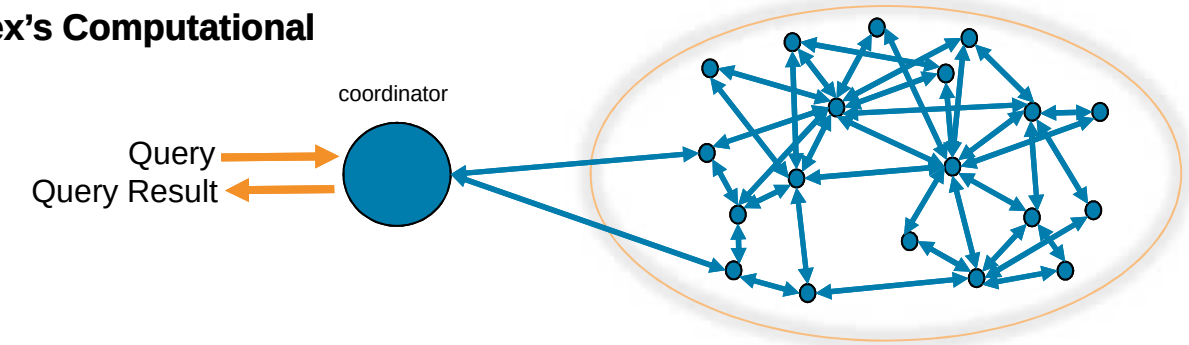
A coordinator C1, receives the request and fans the work out to edge devices

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 - Enterprise Applicability and Differentiation

- Ability to combine IoT data with trusted master data for real-time analysis
- Ability to off-load heavy operational workloads during peak time for analysis
- Ability to federate to polyglot enterprises when standard federation is limited by format and scalability
- Key performance differentiation vs. standard federation technology
- Automatic discovery for use with traditional federation
- Allows for next evolution of data modeling

# IBM Queryplex – Interested in hearing more ?

IBM Queryplex

The power of many together

http://queryplex.com


# Hortonworks Partnership

# 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***

+

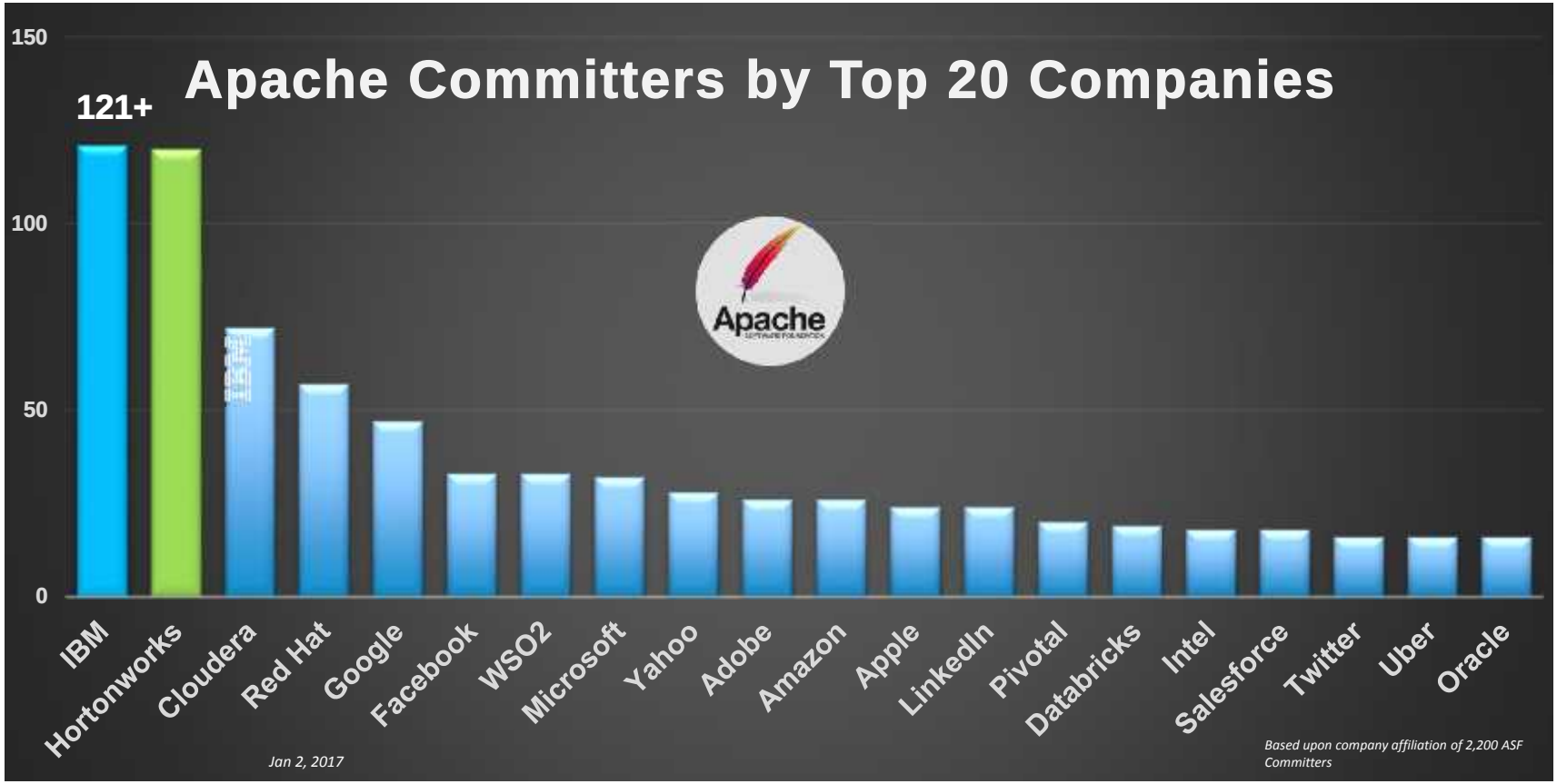

- **#1 Rank by Gartner** 2017 Data Science Magic Quadrant
- **Leader in SQL technology** for Hadoop (www.tpc.org)
- **Leader in data and analytics** solutions for hybrid cloud
- Provides Data Science **& Machine Learning**

- **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.**



# IBM Big Data High Value with Hortonworks

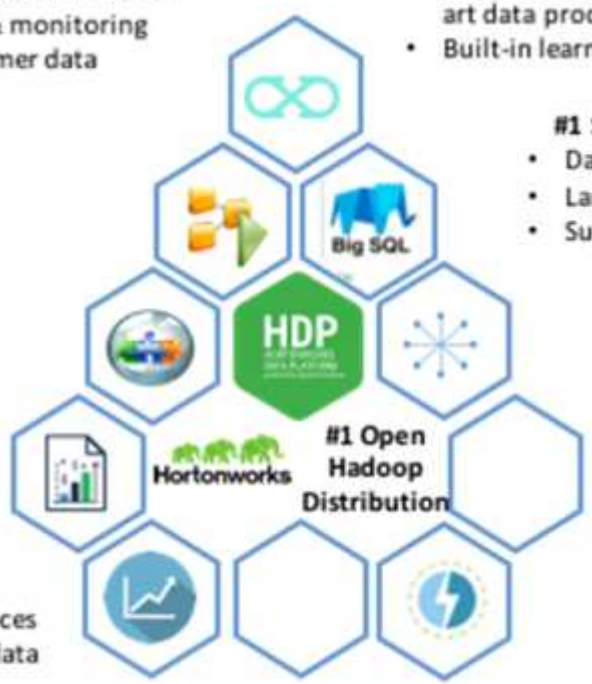
## IBM's Offerings Unlock the value of Hadoop Data

**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

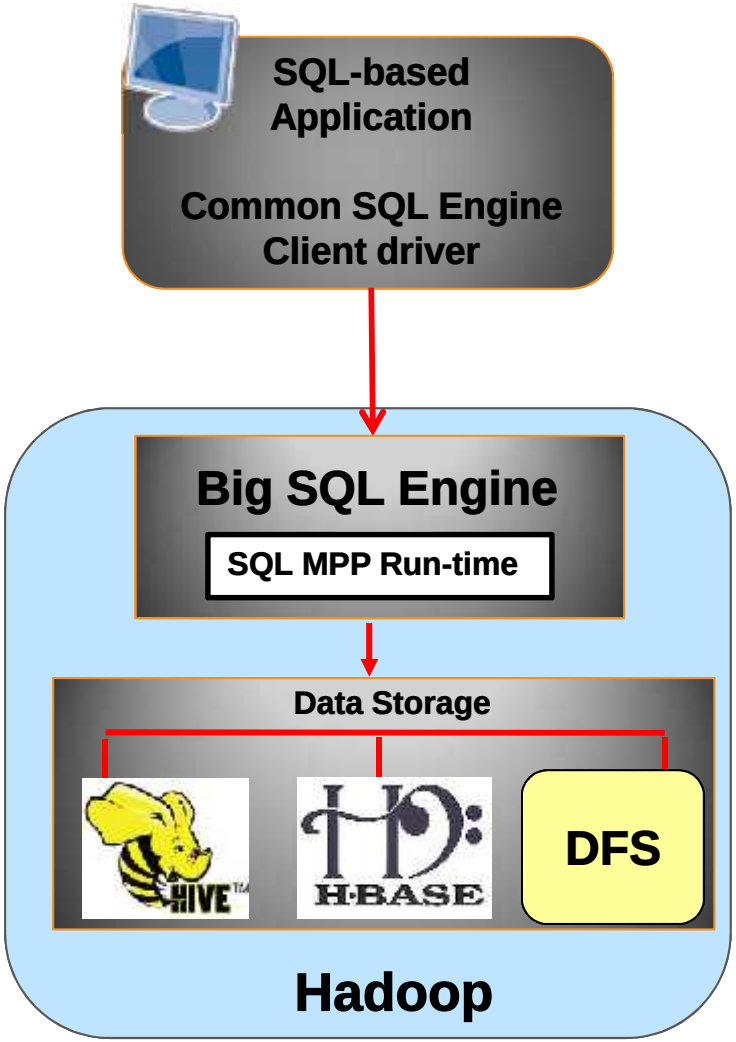
# Apache Spark

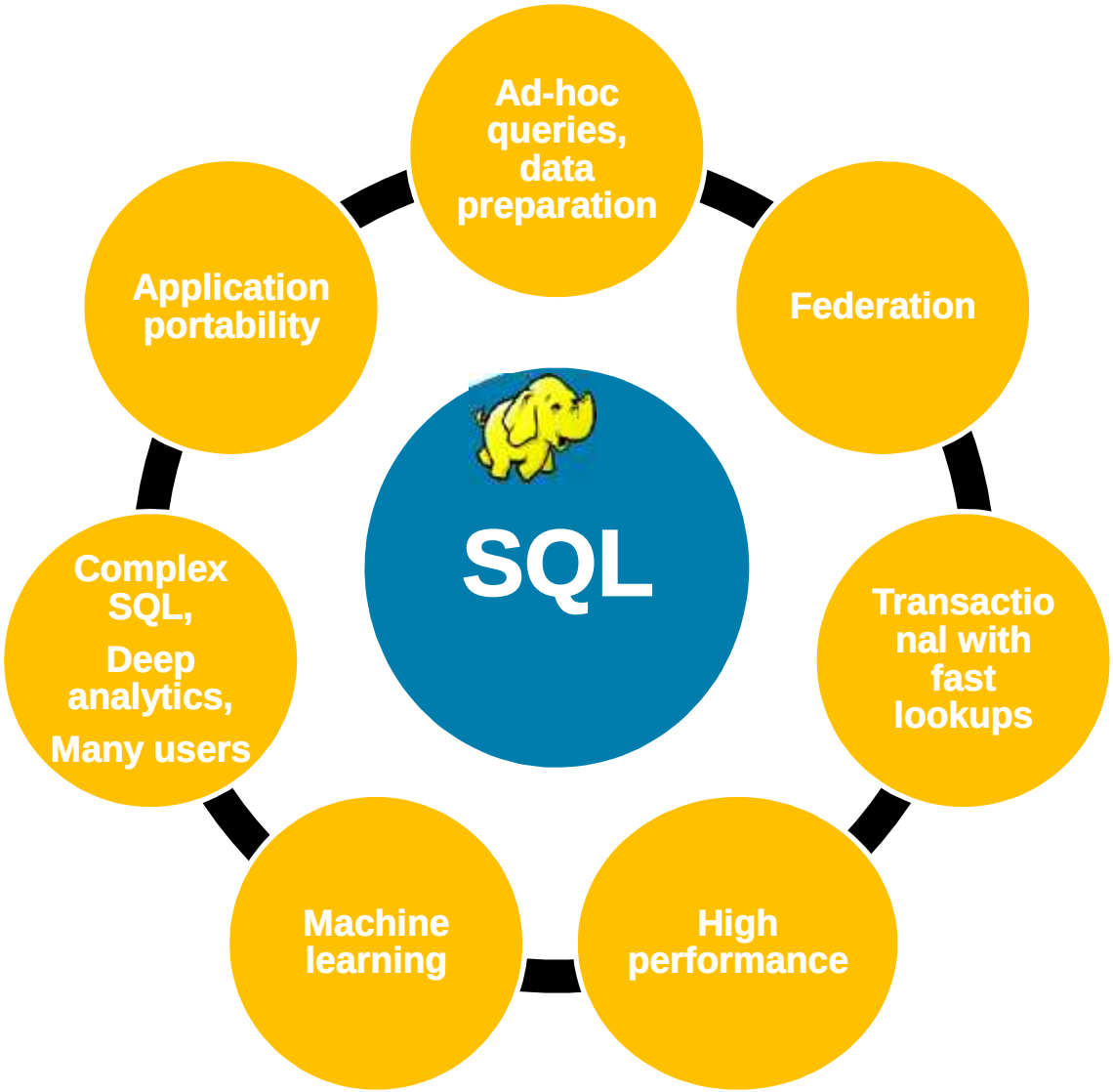
Spark's core libraries enable analytic processing of data from many sources


- Apache Spark is an **open** source, **in-memory** processing framework
- Distributed data processing & iterative analysis on **massive** data volumes
- Spark's standalone framework goes beyond Hadoop and HDFS
  - Interactive query via **SparkSQL**:
  - Accelerated predictive analytics processing via **SparkR**
  - Micro-batched event processing via **Spark Streaming**
  - Machine learning libraries via **MLib**
  - Graph processing via **GraphX**

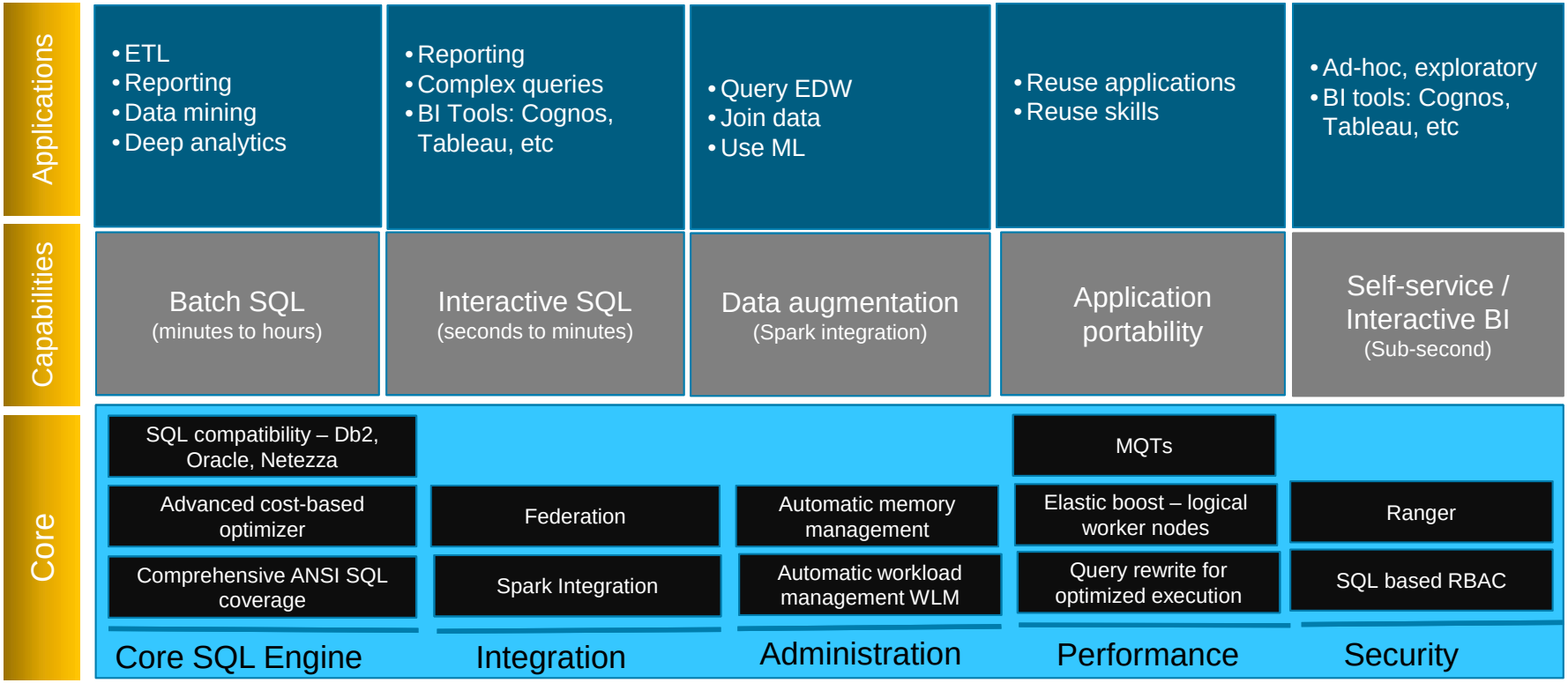

# Db2 Big SQL 5.0

# Db2 Big SQL



# Db2 Big SQL – For all Warehousing needs in Hadoop



# Combining Hadoop Technologies

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


Great for exploratory BI Data Analytics and single stream or pre-production workloads

Ideal for complex BI Data Analytics and enterprise-level production workloads

Ideal tool for Data Scientists and discovery

# Big SQL 5.0 - Summary

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

## Enterprise Features

- Simple Fluid Query (Federation) Setup
- Oracle PL/SQL support including stored procedures
- Performance boost for complex queries using Elastic Boost

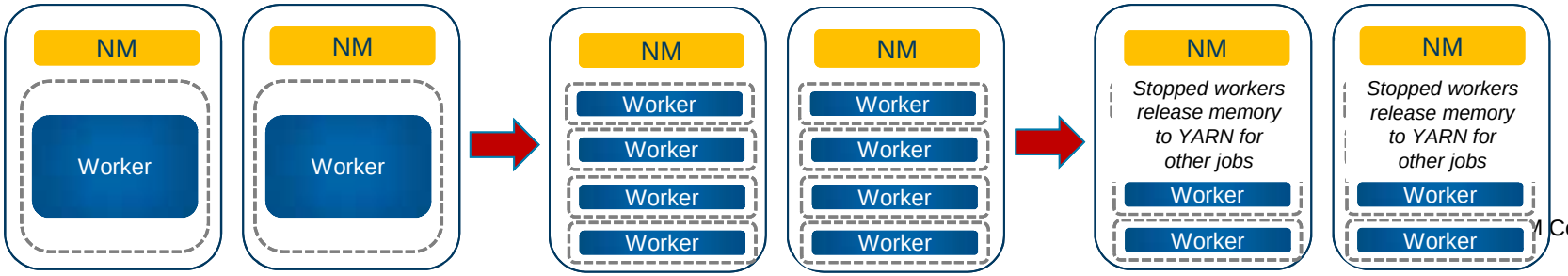
Up to 50% more query performance

## Hadoop Ecosystem

- Support for tables over Object Store (S3)
- Support for tables over WebHDFS (technical preview)
- Apache Ranger Plugin for Big SQL
- Big SQL and Spark Integration now GA
- YARN Integration with Elastic Boost

Big SQL secures data for self-service data exploration. Used this way, Spark users are subject to Big SQL row/column security

Big SQL can call Spark to determine structure of JSON document at run time

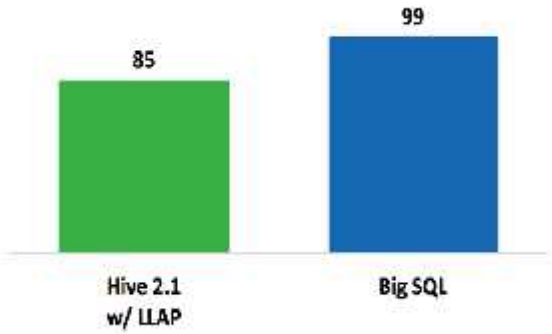


# Query Performance at a Glance – vs Hive LLAP with Tez

## HADOOP-DS @ 10TB

### 85 COMMON QUERIES

**WORKING COMPLIANT QUERIES: 6-streams**



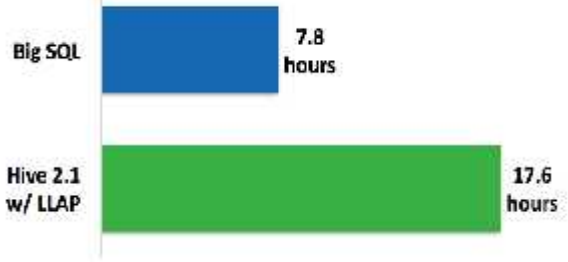

## RESOURCE UTILIZATION:

### 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)

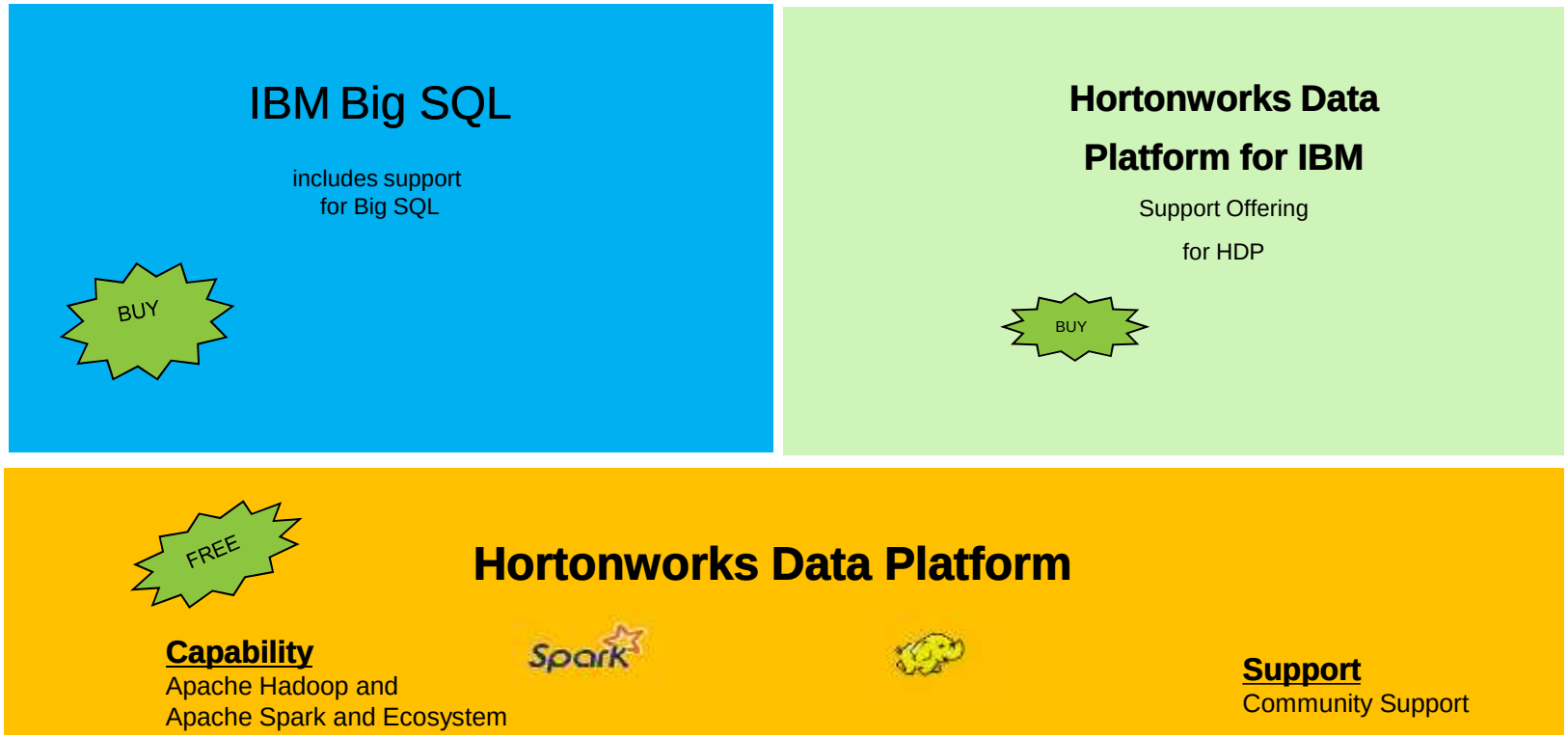
**Db2 Big SQL** FASTER FOR **80%** OF QUERIES RUN

**STACK**

HDP 2.6.1
Db2 Big SQL 5.0.1
HIVE 2.1 LLAP ON TEZ

# Big SQL 5.0 – How it fits with Hortonworks

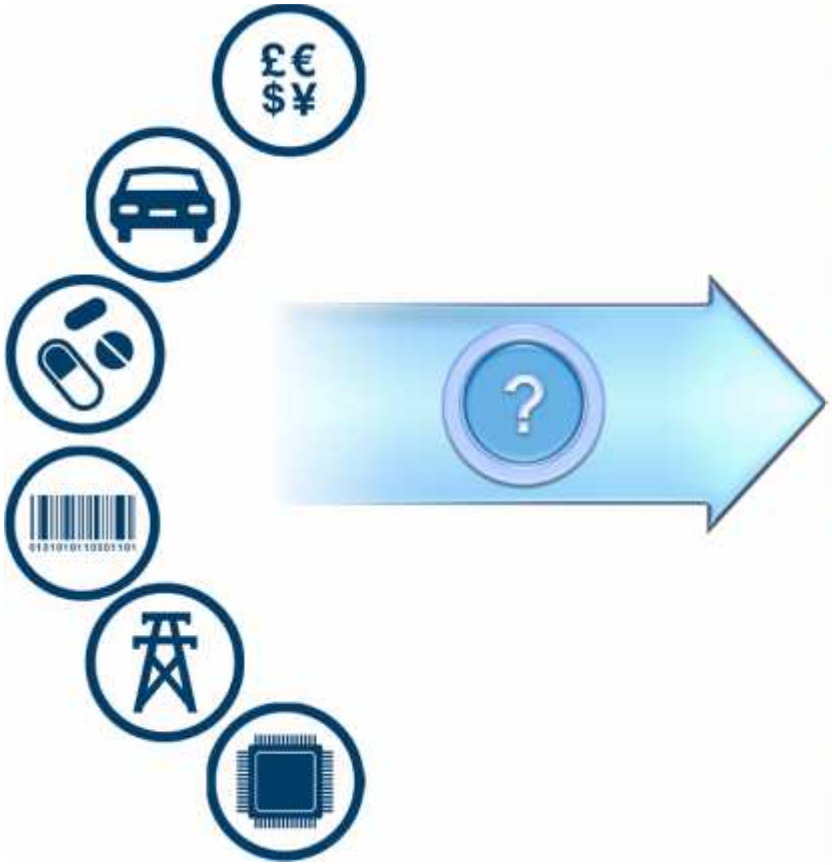
- **Big SQL deploys on top of Hortonworks Data Platform(HDP)**
  - *Includes:* IBM Support for Big SQL
- **Hortonworks Data Platform for IBM (Support only)**
- **Hortonworks Data Platform can be downloaded for FREE.**




# *Db2 Event Store*

**ON-PREMISES TODAY**
**CLOUD COMING**

# 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

## Satellite Tracking System

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

## Smart Metering/Smart Grid

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

## Intelligent Manufacturing

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

## Banking Risk Exposure

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

## Transactional to Analytics Consolidation

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

# What is IBM 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

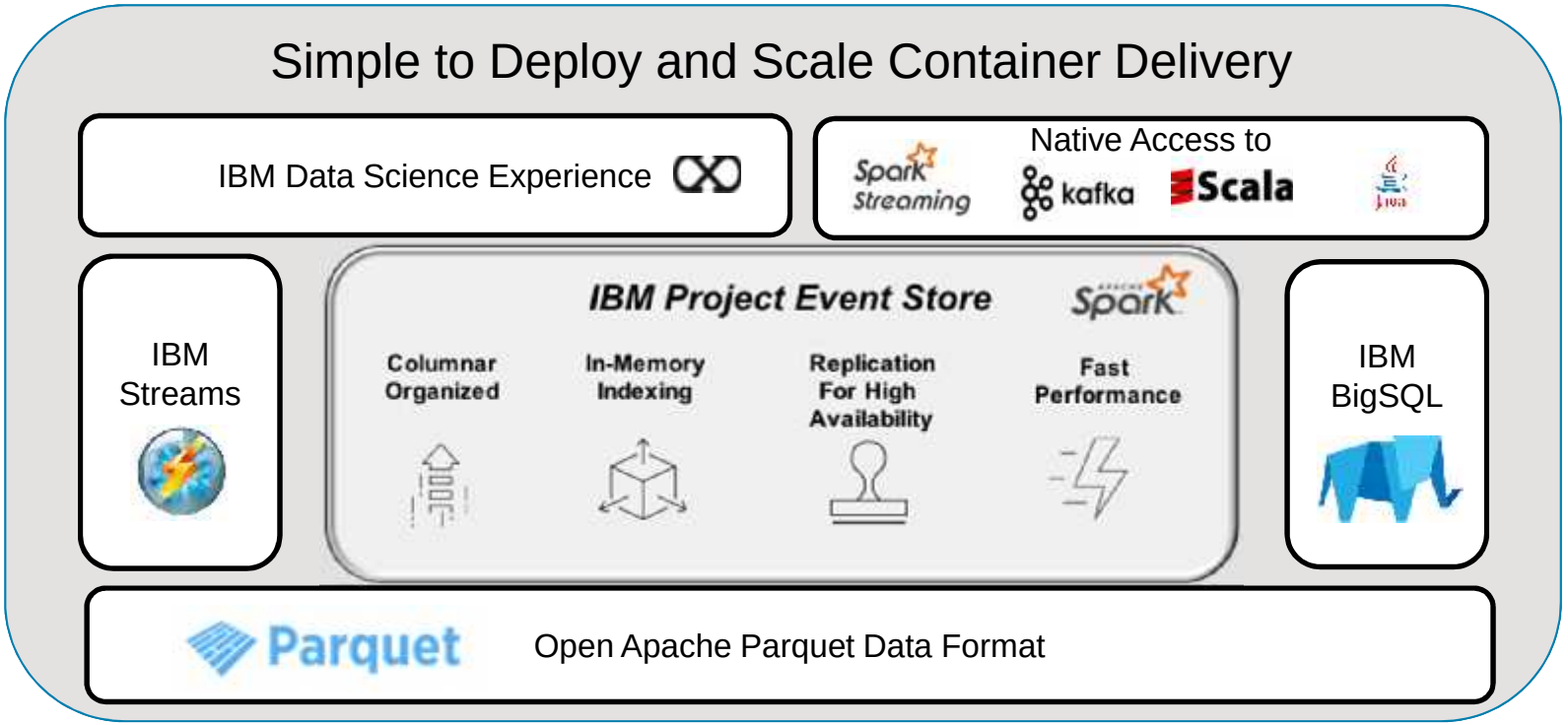
- Writes to shared storage in Parquet format
- Able to leverage low-cost object storage
- Single copy of the data

Parquet

# Db2 Event Store

## *Integrated System for Managing Events*

IBM Event Store



# IBM Db2 Event Store – Competitive Positioning

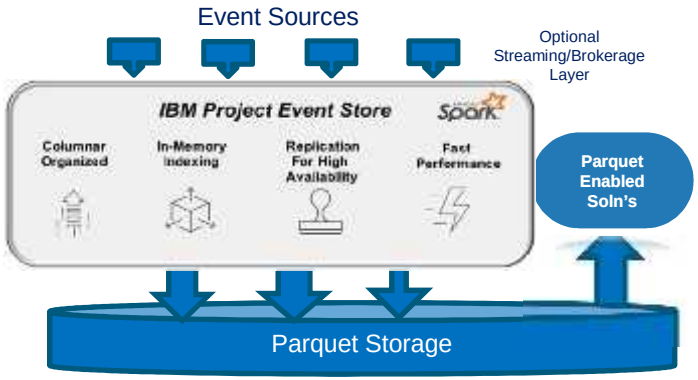
## Db2 Event Store provides everything in the box

**Reduced architectural components**
**Docker Container Delivery**
**Open Data Access**

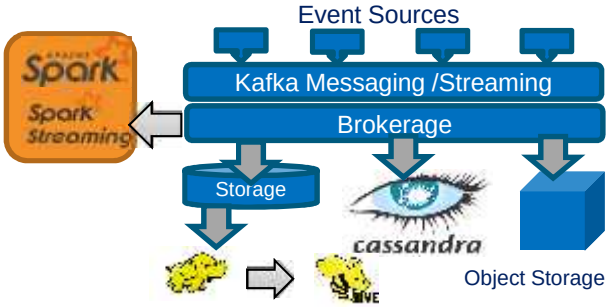
## Complex Manual Architecture

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

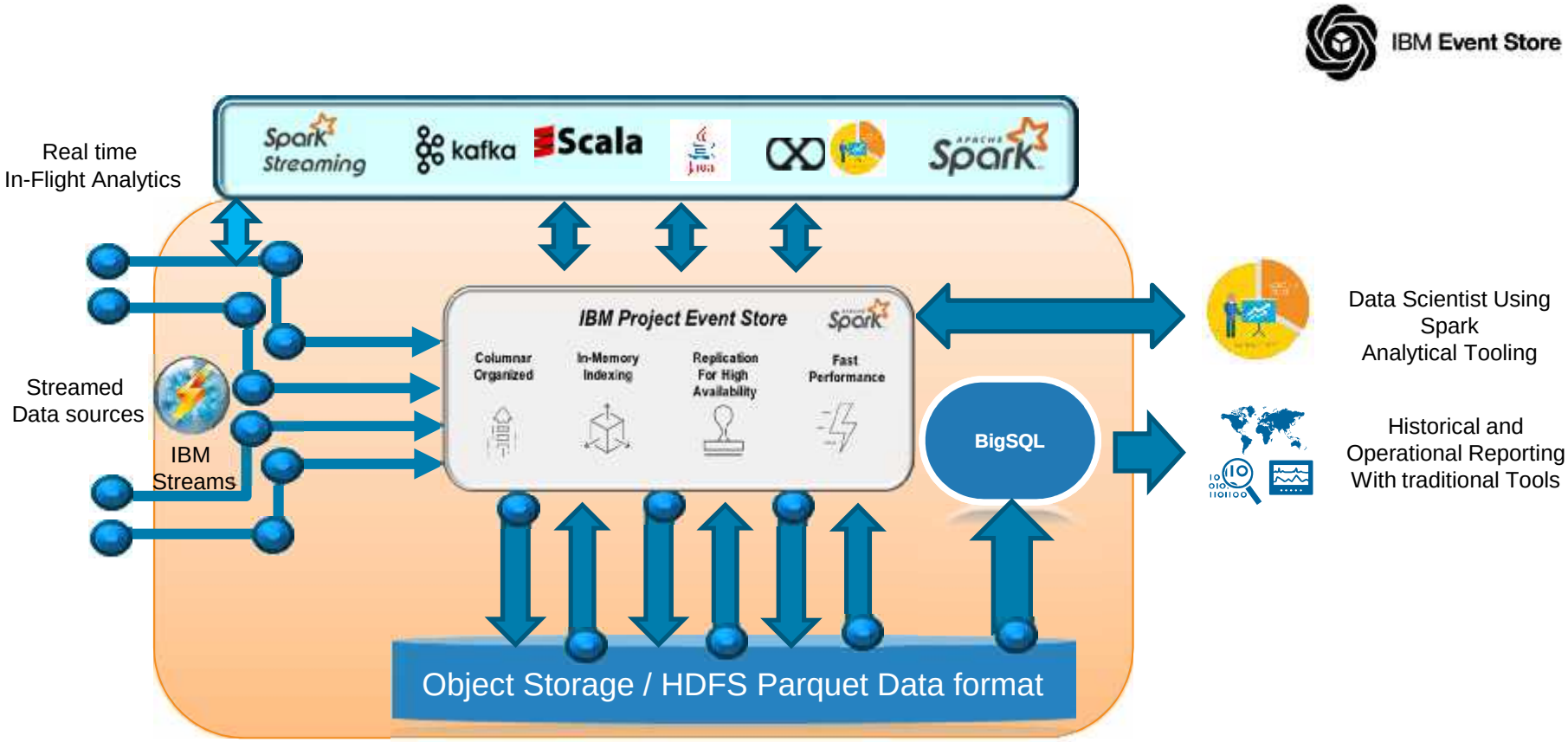
**Simplified Approach With IBM Db2 Event Store**



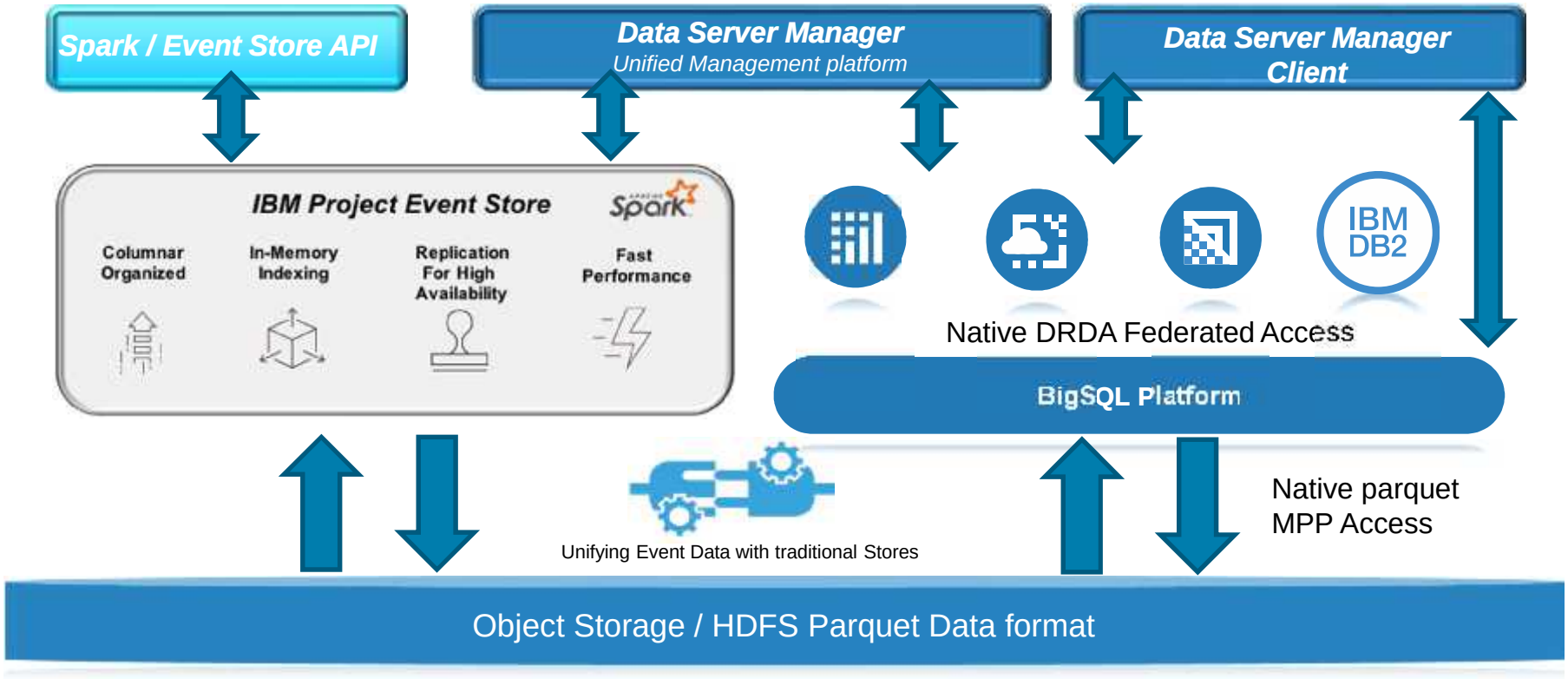

**Digital Company Example**



# Db2 Event Store



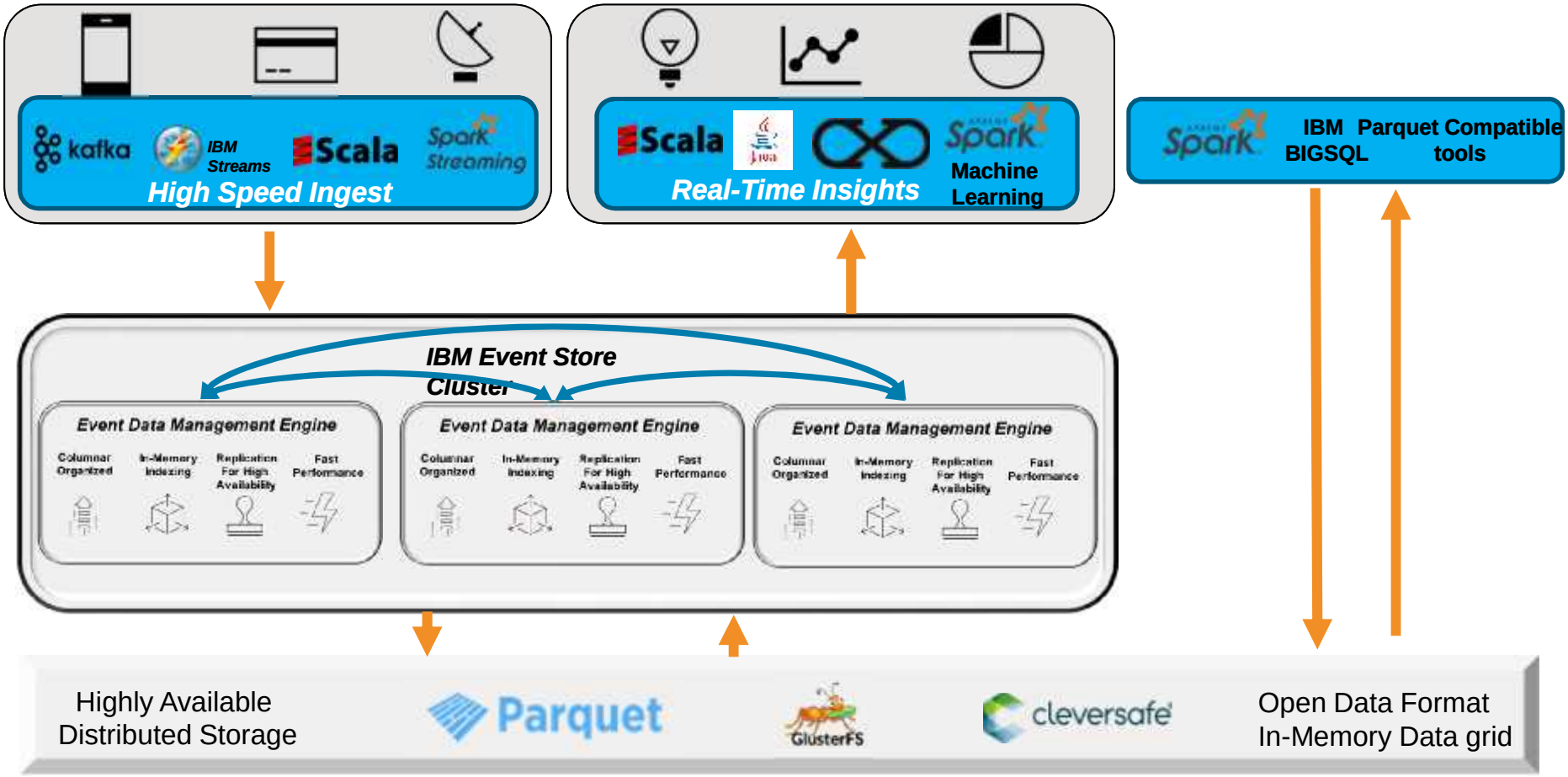
# Db2 Event Store: Unified Data Access and Management



# Db2 Event Store: Architecture

## *Understanding the Engine and Components*

IBM Event Store




# IBM Cloud Private

# Db2 and the Cloud

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

# Introducing IBM Cloud Private


Kubernetes-based container platform

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

Integrated DevOps toolchain


Catalog of integration services

API availability and management to integrate applications and data across environments


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)


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

Integration with existing systems and operations management solutions

# IBM Cloud Private – High Value

**Rapid development and deployment:**

Minutes or hours vs. days or weeks, huge catalogue of OS and IBM Services, advanced Analytics and Machine learning options …

Non-disruptive upgrade of platform integrated with enterprise network, storage, security, performance and production needs

Enablement of new and existing developers & integration with existing Dev/Ops and Security tools

Built-in support for continuous delivery, leveraging both OS and IBM tools to help speed development.

**Investment leverage:**

Infrastructure choice with complete portability and open community-based platforms for choice and flexibility, on- and off-premises

Leverage existing applications and skills while reducing TCO

No Vendor lock in !

**Application modernization:**

Modernization and optimization across multi-cloud environments – Develop Once, deploy anywhere

Reduced risk by running applications on enterprise-grade software & data platforms optimized for cloud

**Security and Compliance:**

Security and control of an untethered environment with an integrated set of management tools

**Differentiated enterprise integration:**

Set of new services available on-premises, complemented with public cloud services (Watson)

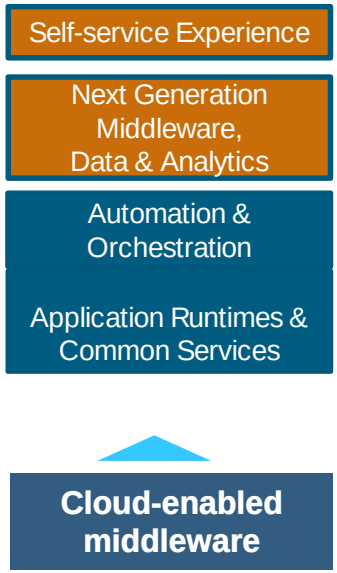
69

Integration of applications with services for operational simplicity and reduced cost

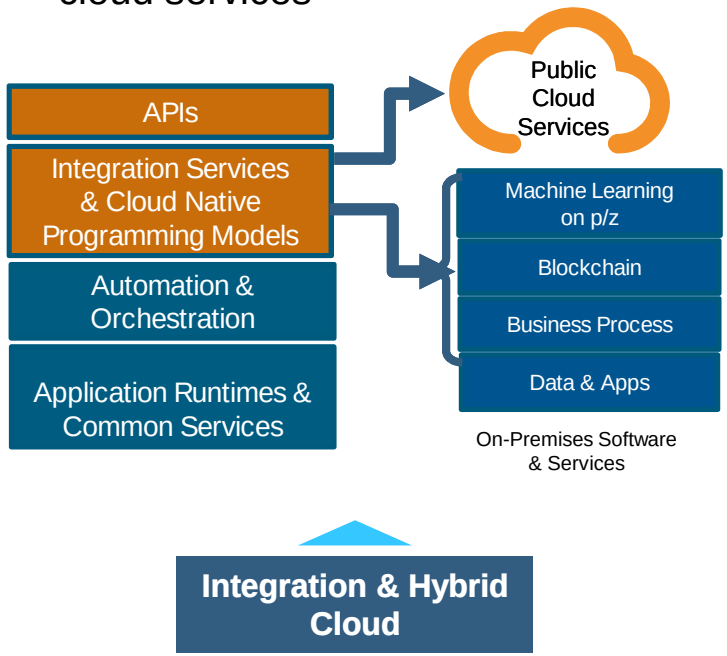
Integrated cloud management solutions to automatically provision and govern multi-cloud environments with speed and control , Coupled with IBM's expertise & Services

# Use Cases driving Private Cloud Adoption

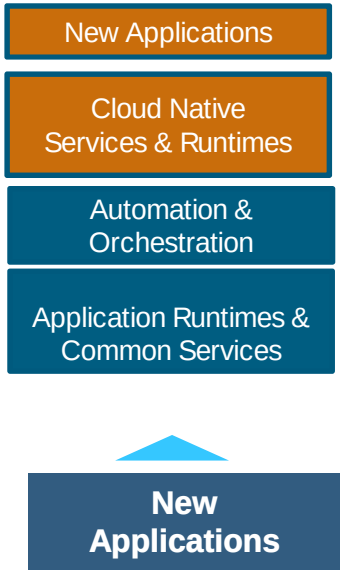
1. Optimize existing apps with cloud




2. Open your datacenter to work with cloud services




3. Create new cloud native applications



# 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 1Q

### 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 2Q

### 1. Hybrid Data Management
- Db2 OLTP MPP

### 2. Unified Governance
- WEX *

### 3. Data Science & BA
- SPSS Modeler *
- SPSS Statistics *
- Cognos *

### Common / Foundational

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

# In Summary – 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

| | |
|---|---|
| Learn More | • IBM Cloud private home: https://ibm.biz/Bdj4Bz<br>• White paper: https://ibm.biz/Bdj4UJ |
| See it In Action | • Offering demo: https://youtu.be/yzXA3qhfaq0<br>• Try It: https://ibm.biz/Bdj4UC<br>• Free Community Edition: https://hub.docker.com/r/ibmcom/cfc-installer/ |

**Les King**
**Director, Hybrid Data Management Solutions**
**March, 2018**
**lking@ca.ibm.com**
**ca.linkedin.com/pub/les-king/10/a68/426**

# Hybrid Data Management Strategy and New News !