Do data science faster

IBM®

Essential elements of a hybrid data management strategy

Access your data
All sources and all types

Democratize access
Provide data-driven decisions to everyone

Flexibility
Support all data types, all workloads, all consumption models

Analytics for all data
Easily query all of your data

Intelligent analytics
Make better decisions, provide smarter capabilities

Cloud journey
Support your data regardless of location

Gain value from your data, without limits

Write once, run anywhere, with a Common SQL Engine

IBM Hybrid Data Management solutions

Anchored by a **Common SQL Engine** enabling true, highly scalable hybrid data warehousing solutions with portable analytics

Managed public Cloud DBaaS	Software defined warehouse (on-premises or in cloud)	Dedicated analytics system	Custom deployable database	Open source Hadoop with Hortonworks
Db2 on Cloud	Db2 Warehouse	Integrated Analytics System	Db2	Big SQL
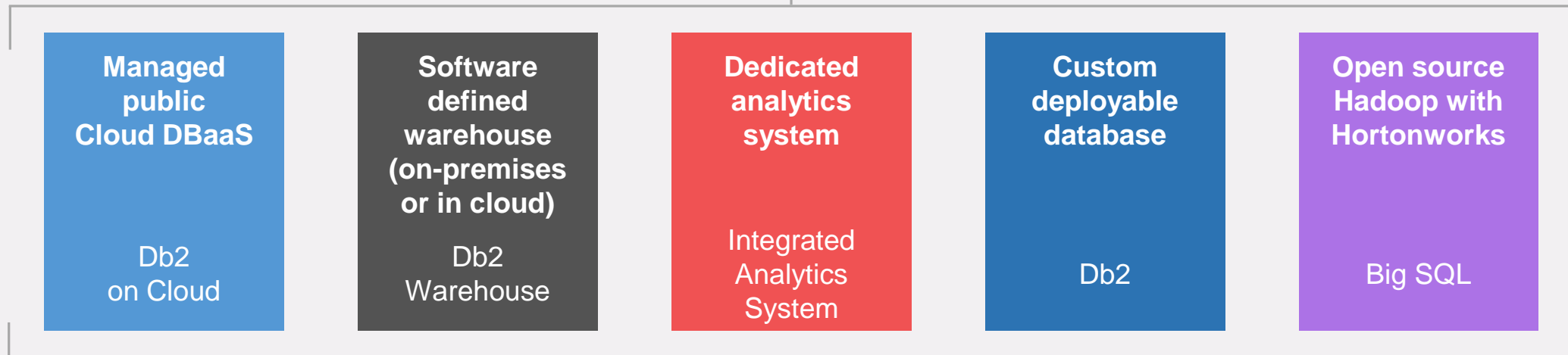
- **Application compatibility**
  Write once, run anywhere
- **Operational compatibility**
  Reuse operational and housekeeping procedures

- **Licensing**
  Flexible entitlements for business agility and cost-optimization
- **Integration**
  Data virtualization capabilities for query federation and data movement

- **Standardized analytics**
  Common programming model for in-DB analytics
- **Ecosystem**
  One ISV product certification for all platforms

Introducing the IBM Integrated Analytics System

Do data science faster

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

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

Real time analytics with **machine learning** that accelerates decision making, bringing new opportunities to the business

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

Cloud-ready to support multiple workload deployment options

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

IBM

Analytics Platform Convergence

Db2 Warehouse in the Cloud
(formerly dashDB for Analytics)

PureData Systems for Analytics
(Netezza)

Cloud first, in-memory, SIMD acceleration for analytics (BLU)

Enhanced elasticity, UDX integration and managed service

Further UDX integration and optimization

PureData System for Analytics
N3001

Integration with Hadoop for high speed query routing and bulk data movement

Continued enhancements to caching algorithms, optimizer, workload optimizations, etc.

Convergence
SQL, Analytic, Application, ETL and Client

Db2 Warehouse on Cloud – Managed Service

Db2 Warehouse – Private Cloud

IBM Integrated Analytics System – NEW!

- Common SQL Engine
- Built-in data virtualization
- Machine Learning with Apache Spark
- Built-in IBM Data Science Experience
- Operational "mixed" workloads
- Configured for performance
- In-place expansion*
- Multi-temperature storage*
- Real-time data ingest

* Planned in 2018

IBM

For your analytics workloads

Built-in Data Science Experience

Tools to connect directly to your organization's data

In-place machine learning

Bring performance and simplified process to organization-changing analytics

Operational "real time" workloads

Ensures insights are matched with performance

Data virtualization

Bring data across multiple locations to a single query

In-place expansion and cloud-ready

Elasticity to grow and scale the system
Option to seamlessly move workloads to the cloud

High availability and disaster recovery

Maintain service level agreements and stability for analytics core to your organization's success

Less admin, more analytics

NO requirements for armies of administrators	Load data, **start querying within hours**, not days
NO "trapping" your data on the platform	**Built-in data** migration tools
NO tuning, no configuration	**Single management** console
NO storage administration	System is **performance optimized** for analytics
NO physical data modeling needed	In-place **expansion** capabilities
NO indexes and tuning necessary	**Single support** number

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)

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

Performance optimized hardware

Power architecture

- Higher performance across fewer nodes
- CPU acceleration with multi-core and Single Instruction Multiple Data (SIMD) parallelism
- Increased reliability and availability

Flash storage, standard

- Near real-time latency for higher transfer speeds 99.999% reliability and operational efficiency

Hardware acceleration with Power architecture

Power Systems

Processors	**Memory**	**Cache**
flexible, fast execution of analytics algorithms	large, fast workspace to maximize business insight	ensure continuous data load for fast responses
4X	**4X**	**6X**
threads per core vs. x86	**memory bandwidth vs. x86[1]**	**more cache vs. x86[2]**
(up to 1536 threads per system)	(up to 32TB of memory)	(>19MB cache per core)
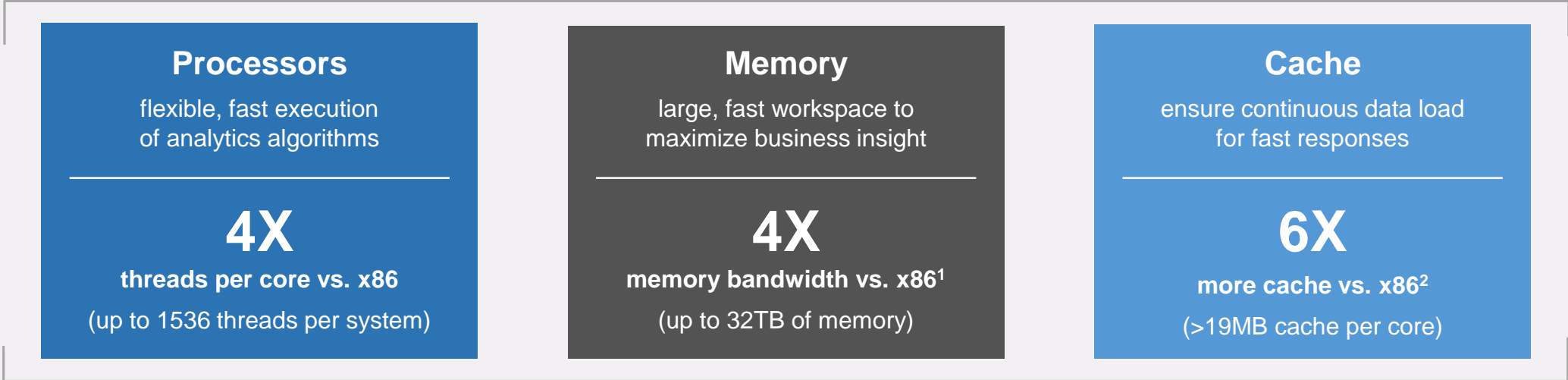
Continuous data load

Massive IO bandwidth

Parallel processing

Flash for extreme performance

Large-scale memory processing

1. Up to 4X depending on specific x86 and POWER8 servers being compared
2. Up to 6X more cache comparing Intel e7-8890 servers to 12 core POWER8 servers. See speaker notes for more details.

Large Financial Institute Chooses IIAS over Teradata!

Current Environment	• Siloed data sets across organization on SQL Server/SSIS/SSRS • Semi "manual processes" for data in multi dimensional Excel – product sales, financials, customer trends & operational dashboard • Some SAS but not part of ecosystem
Problem	• Unhappy with level of granularity with SQL Server, SSIS,SSRS (limited analytics aggregations and summarization) • Spend one month processing manually to prepare data for final report in Excel • Missing much information because working only on summarized data • Data volume growing (eBanking grew 38% YOY)
Driver	• Performance and Cost
Competition	• Teradata 2750 vs Mako ½ rack • Teradata claimed that Mako is obsolete. Customer decided on Teradata "verbally" and during negotiations the original cost estimates changed. What looked promising with TD before, was now overpriced with consulting fees and a change of structure to per hour pricing. Customer was not comfortable, so they re-evaluated and chose IBM
Use case	• e-banking – single customer single transaction reporting services • Retail and institutional banking
Solution	• 2/3 rack of IIAS • GBS vendor project assessment, special fixed price contract • Deal included: Db2 Warehouse, InfoSphere DataStage and Big SQL (include Hadoop in future phases of implementation to analyze social media data) • Going forward they are retaining SAS and using IIAS as a sandbox for new models and data mining
Follow up	• ML deal for next year - Currently no ML, but it is a potential now that 5 DSX licenses are included with IIAS

A Large American Entertainment Company
Media and Entertainment industry

- **Current IBM Environment:**
  - PDA Twinfin24 to optimize audience measurement algorithms with Nielsen AMRLD data
    - For team of research analysts and to feed Db2 Warehouse on Cloud Portal
  - 'Bridge to Cloud' with Db2 Warehouse on Cloud (7 nodes) for client media planning portal fed by Twinfin
  - Cognos and Machine Learning with Monte Carlo algorithm

- **Problem:**
  - Scalability and Performance - too much CRM and Nielsen AMRLD data to push to cloud. Network bandwidth and Db2 Warehouse on Cloud 7 node cluster cannot accommodate performance requirements
  - Machine Learning Loop Problem - Current Spark environment on TF24 has too many ML algorithms that include a Monte Carlo library. Client needs more spark executors which require more memory/CPU power. This capability will help analysts and ad/cable buying customers predict and better understand ad campaign/viewership/commercial effectiveness/ROI
  - Flexibility to separate Db2 Warehouse on Cloud Client Portal from internal network analysts when needed
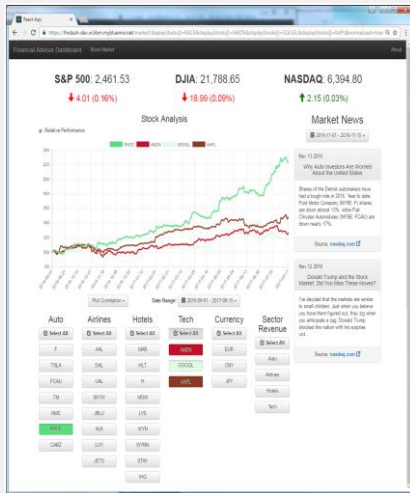
- **IIAS Solution:**
  - IIAS is the scalable and performant solution to accommodate client's increasing data sets, automated workload management, intense algorithms and ML simulations
  - Ad buying customers are better able to understand ROI and optimize media plans on client's Portal fed by IIAS
  - Client research team is able to run faster reports and more exploratory analytics (mining/predictive/etc.)
  - Client is able to optimize marketing campaigns and negotiate large deals with cable entities

IBM

IBM Integrated Analytics System - Use Case Example

High Performance IBM Integrated Analytics System

External Data Sources

Macro Economic Data Feeds

Db2 Warehouse on Cloud

(Source: FRED)

(Structured Data Store)

Move data and federate queries with Common SQL Engine

DSX LOCAL

News Data Feeds

IBM BigSQL on HortonWorks Data Platform (Hadoop)

(Source: NASDAQ)

(Unstructured Data Store)

Stock Portfolio Analytics : Applications leverage In-database Machine Learning (ML) models and R analytics

Stock and Customer Portfolio Data

(On-Premises)

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 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

IBM Integrated Analytics System
High Powered Analytics for Lightning-fast Data Science

☐ **Data processing and analysis 2x- 5x faster than row store appliance**

☐ **Performance results:**

Category	Workloads	Query Speedup (Median)	Query Speedup (Max)
Benchmark	TPCDS-like – 1 TB	4.7x	113x
	TPCDS-like – 10 TB	2.3x	305x
	TPCDS-like – 30 TB	1.9x	210x
Customers	Large Online Advertising Co.	1.1x	3x
	Large Retail Marketing Co.	4.6x	53x
	Multinational Freight Transportation	2.8x	37x

Speed of Thought Analytics
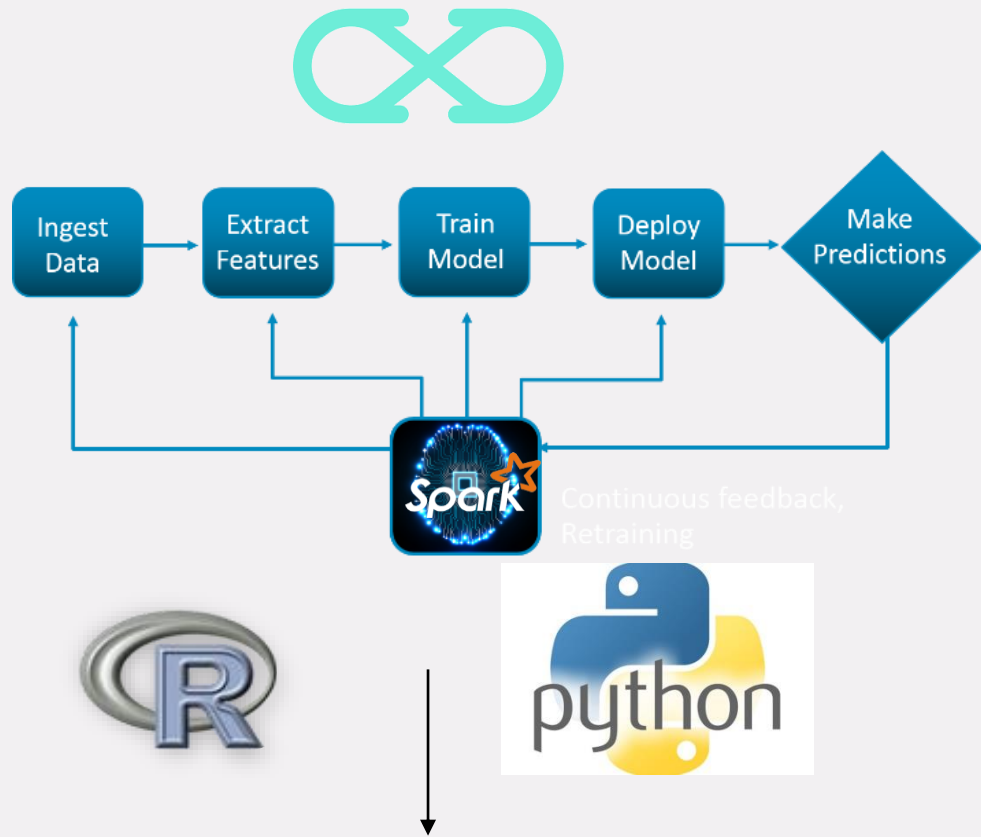
Integrated Cognitive Assist for Machine Learning
DSX for Interactive & Collaborative Data Science
Scalable ML Model Training, Deployment and Scoring with Spark embed Predictive / Prescriptive In place Analytics

Embedded

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

Full integration with tools for BI & visualization

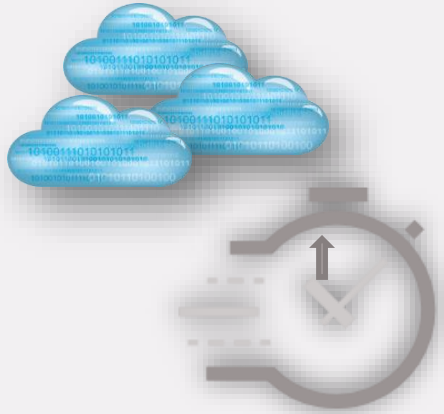IBM Cognos, Tableau, Microstrategy, Business Objects, SAS, MS Excel, SSRS, Kognitio, Qlikview

Full integration with tools for model building and scoring
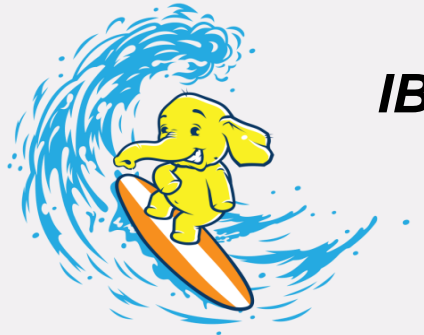
IBM SPSS, SAS, Open Source R, Fuzzy Logix

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
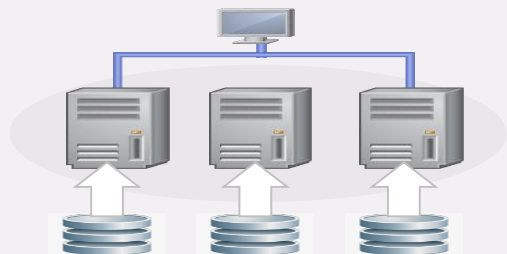
Scalable

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

Data virtualization enables IT provisioning for the business

Write your applications to pull data from IBM and our solutions
can provide transparent access to other data sources

Your applications

**IBM
Hybrid Data
Management**

**Other
data sources**

TERADATA   ORACLE   SAP

Microsoft   mongoDB

Microsoft Azure

AMAZON REDSHIFT   HAWQ   HIVE

IBM

IBM can provide what no other vendor can

More intelligent analytics

With embedded machine learning and HTAP

Deploy your data where you need it

With options for on-premises and cloud

IBM Hybrid Data Management

Go at the speed of your business

With faster analytics through HTAP and BLU Acceleration

Write once, run anywhere, from any source

With a Common SQL Engine and data virtualization

"The combination of high performance and advanced analytics – from the Data Science Experience to the open Spark platform – gives our business analysts the ability to conduct intense data investigations with ease and speed. The Integrated Analytics System is positioned as an integral component of an enterprise data architecture solution, connecting Netezza/PureData, cloud-based dashDB (Db2 Warehouse on Cloud) clusters and other data sources."

*- **Vitaly Tsivin**, Executive Vice President, at AMC Networks,*

Thank you

Q&A

IBM.

Data virtualization enables IT provisioning for the business

Write your applications to pull data from IBM and our solutions
can provide transparent access to other data sources

Your applications

→

IBM Hybrid Data Management

→

Other data sources

TERADATA   ORACLE   SAP

Microsoft   mongoDB

Microsoft Azure

AMAZON REDSHIFT   HAWQ   HIVE

IBM