Data Management & Data Virtualization with Cloud Pak for Data

Legal Disclaimer

The information contained in this publication is provided for informational purposes only. While efforts were made to verify the completeness and accuracy of the information contained in this publication, it is provided AS IS without warranty of any kind, express or implied. In addition, this information is based on IBM's current product plans and strategy, which are subject to change by IBM without notice. IBM shall not be responsible for any damages arising out of the use of, or otherwise related to, this publication or any other materials. Nothing contained in this publication 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 the applicable license agreement governing the use of IBM software.

References in this presentation to IBM products, programs, or services do not imply that they will be available in all countries in which IBM operates. Product release dates and/or capabilities referenced in this presentation may change at any time at IBM's sole discretion based on market opportunities or other factors, and are not intended to be a commitment to future product or feature availability in any way. Nothing contained in these materials is intended to, nor shall have the effect of, stating or implying that any activities undertaken by you will result in any specific sales, revenue growth or other results.

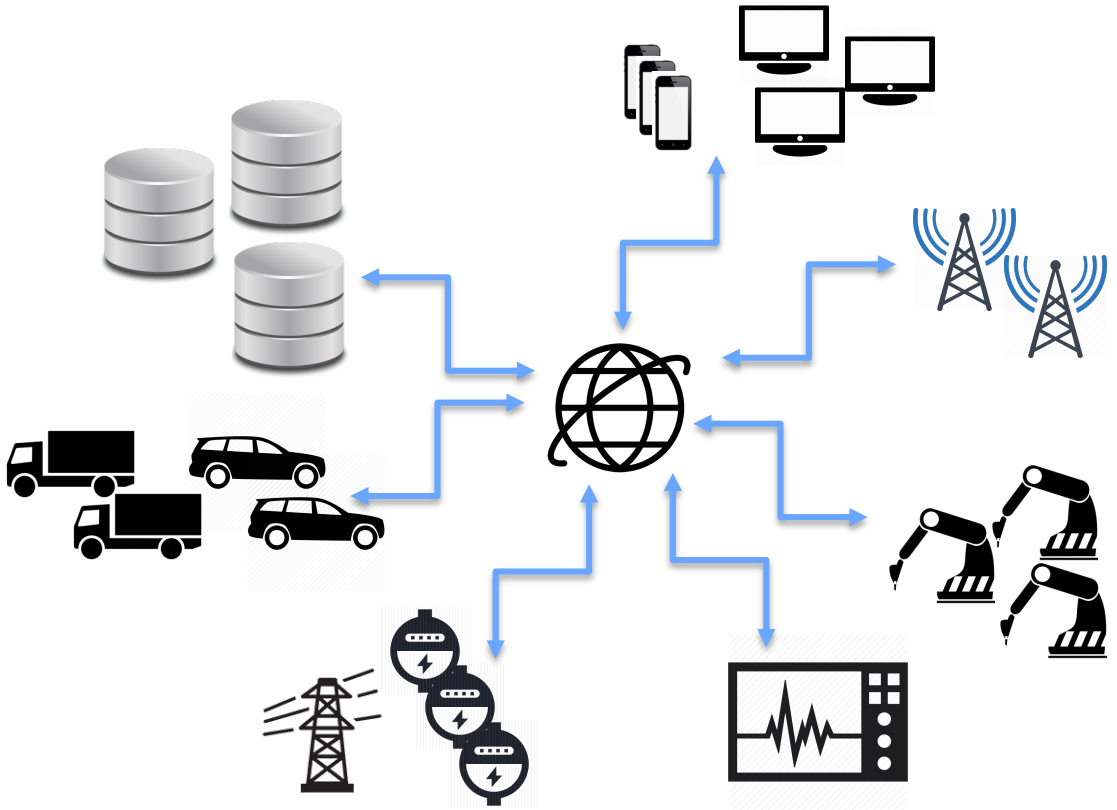
All customer examples described are presented as illustrations of how those customers have used IBM products and the results they may have achieved. Actual environmental costs and performance characteristics may vary by customer.

Data is Everywhere

Number of Sources and volume rapidly increasing

More and more heterogeneous

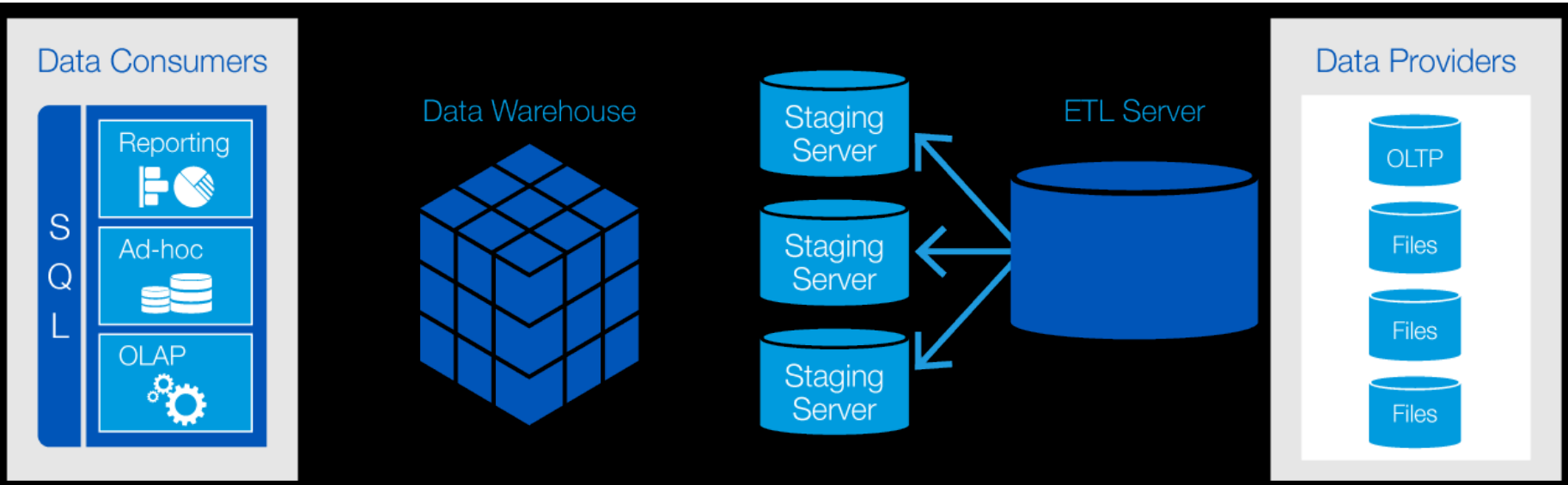
Highly distributed – internal and external



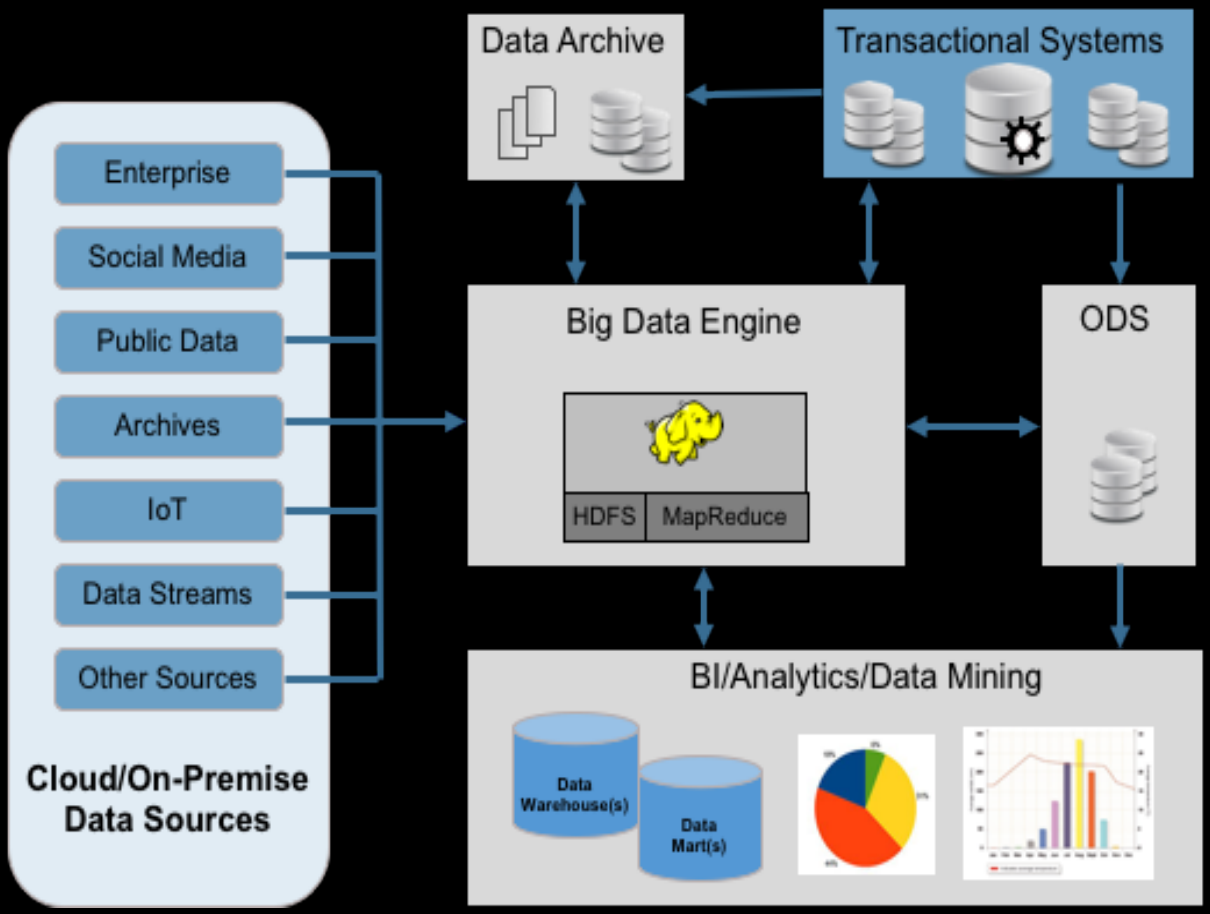
Traditional Data Integration

Not Viable to continue to Move or Copy *all* Data
(using extract, transform, load ETL)

- Risk to data security
- Data inconsistency & error prone
- Rigid and limits business agility
- High cost and latency
- Finite scalability



Current Data Architectures



- Numerous ETLs
- Unnecessary duplication and data replication as business users demand more data views
- Data governance issues accelerating across the enterprise

Physics have driven Cost & Complexity
And impede Productivity

A Single View for Self-Service

Study Reference : “From Data to Insight: Work Practices of Analysts in the Enterprise”, Kandogan, Balakrishnan, Haber, Pierce, submitted to IEEE Computer Graphics and Applications, Special Issue on Business Intelligence Analytics

The growth of Big Data compounds complexity, resulting in an increase in *friction* that already exists between IT and business consumption

Studies report **friction between elements** of the ecosystem lead to major inefficiencies.

- Real-time data access for real-time decisions
- Finding data is hard
- Need metadata for lineage, quality, currency
- Need for virtualized access to persistent data

What is Data Virtualization?

The ability to view, access, manipulate and analyze data without the need to know or understand its physical format or location, and without having to move or copy it.

What is the value proposition ?

What if you could tap into all of your critical data assets no matter where they physically are?

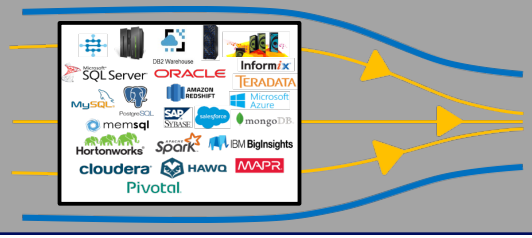
What if you could query 2 or 2,000 data systems with a single query?

Data virtualization key Use Cases

Driven by patterns needing low data latency, high flexibility with transient schema

- On-demand Virtual Data Marts to save cost of standing up EDW
- EDW prototyping and migration (mergers/acquisitions)
- Virtualization with Big Data (Hadoop, NoSQL, and Data Science)
- EDW augmentation (offload workload)
- Data discovery for ”what if” scenarios across hybrid platforms
- Data Caching of combined data for frequently accessed data
- Combine MDM with IoT for Systems of Insight (IT/OT)
- Data integration preparation tool to complement ETL
- Master data hub extension to enrich 360 View (e.g. multi-channel CRM)

✓ single View across your business
✓ real-time analytics without moving data
✓ fast TTV with high ROI



How does it work ?

Some Definitions

Data Federation - technology

Federation is the underlying approach for defining access/authorization to logically mapped remote data sources and query technology used for the execution of distributed query processing against multiple data sources.

Data Virtualization - platform

Any approach to data management that allows an application to retrieve and manipulate data without requiring technical details about the data, such as how it is formatted at source, or where it is physically located, and can provide a **single customer view** of the overall data. Typically necessitates a platform or information architecture. (Data Federation is embedded)

Data Integration enhanced

Data integration solutions traditionally **move a copy of the data** from disparate sources into a new consolidated source. Data Virtualization complements by providing a view of the integrated data while **leaving the source data exactly where it is**.

IBM Data Virtualization

Query across multiple data sources:

- Oracle, Db2, SQLServer, Informix, Netezza, MySQL, PostgreSQL, Big SQL, Apache Hive, HDP Hive, Cloudera Impala and more!

Scale! 1 or 1,000 at once

- More than 10x better for several important use cases

Schema discovery and folding

- Automatically find and match tables across systems so you can query them as a single virtual table.

Rich application capabilities

- Connect to Data Virtualization with your favorite SQL apps and tools
- RStudio, Jupyter Notebook, Cognos, Tableau, Microstrategy

Secure!

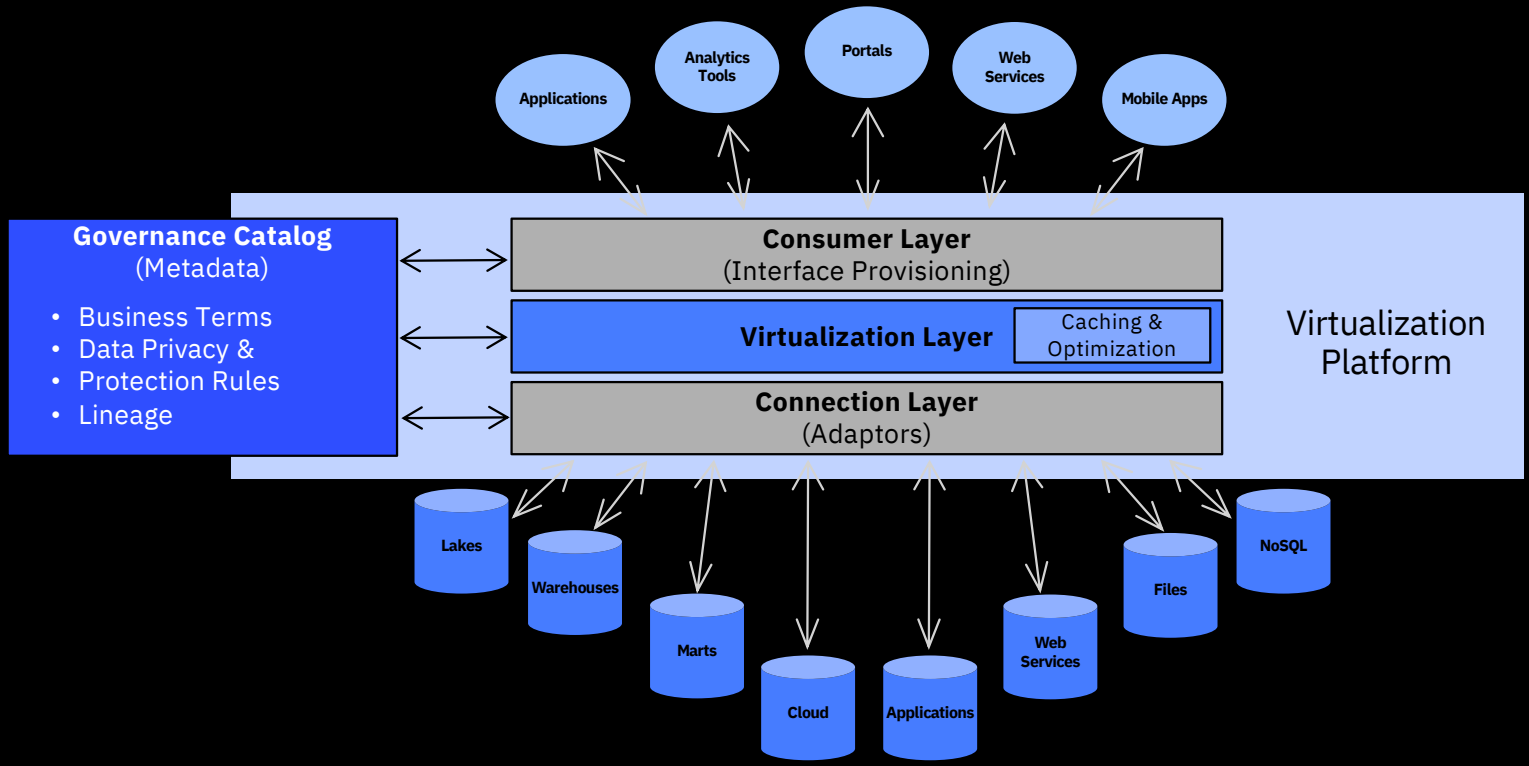
- Strict access controls
- Fully encrypted communications

Deeply integrated w Cloud Pak for Data

- Enterprise Data Catalog, governance and security. e.g. Automatic publishing of virtualized data into the data catalog.
- Immediate access via Cognos and Watson Studio

Data Virtualization in Cloud Pak for Data

The ability to view, access, manipulate and analyze data without the need to know or understand its physical format or location, and without having to move or copy it.



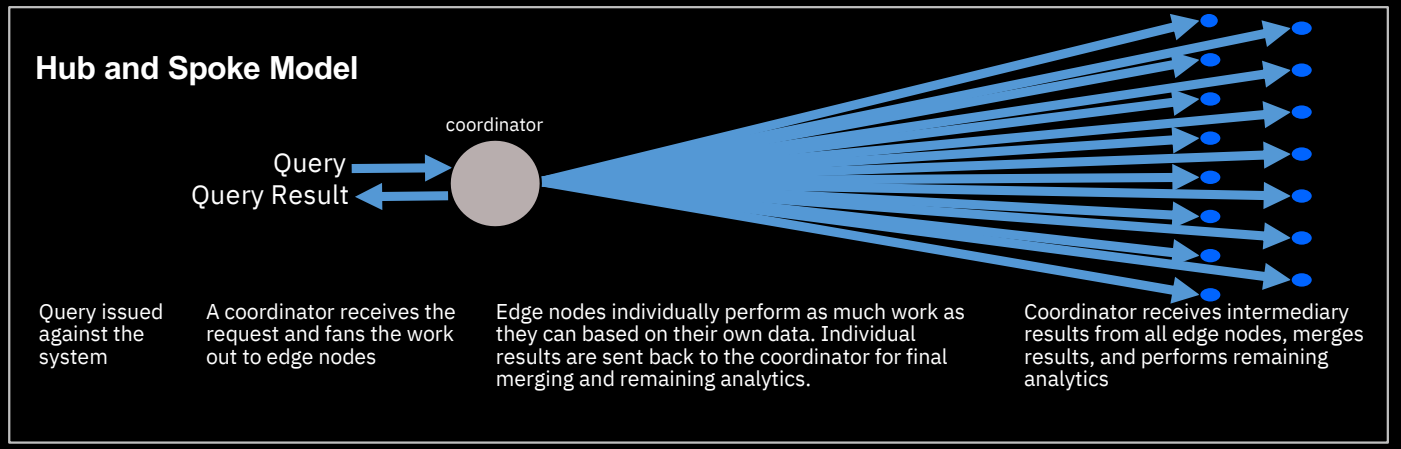
Key Architectural Differentiation

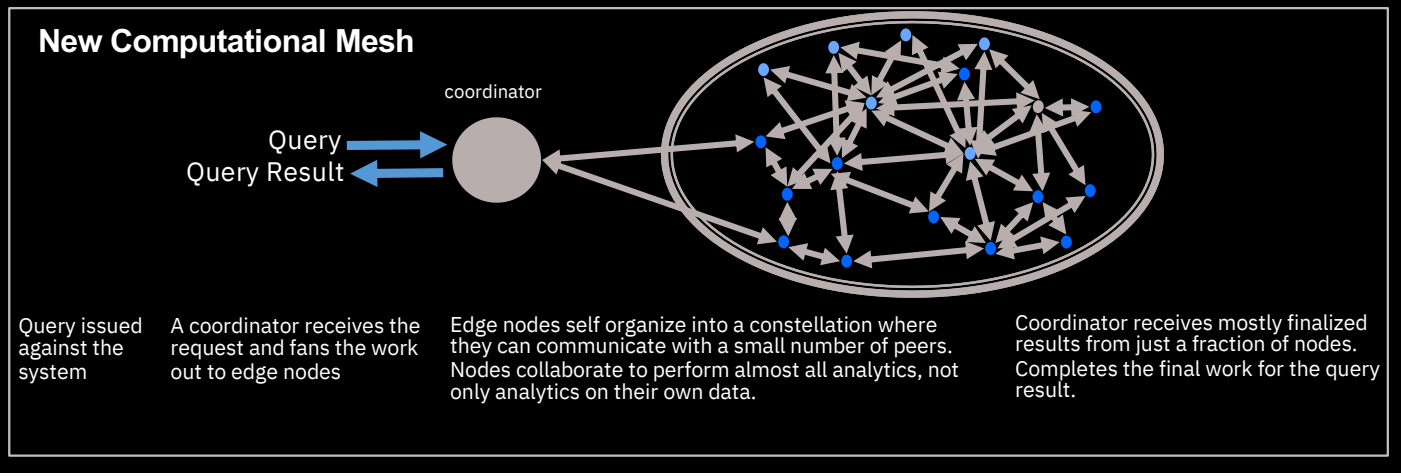
Hub and spoke execution models:

- Lacks scalability
- Performance constrained
- Basis for Federation and our competitors

IBM is first to market with a parallel processing model:

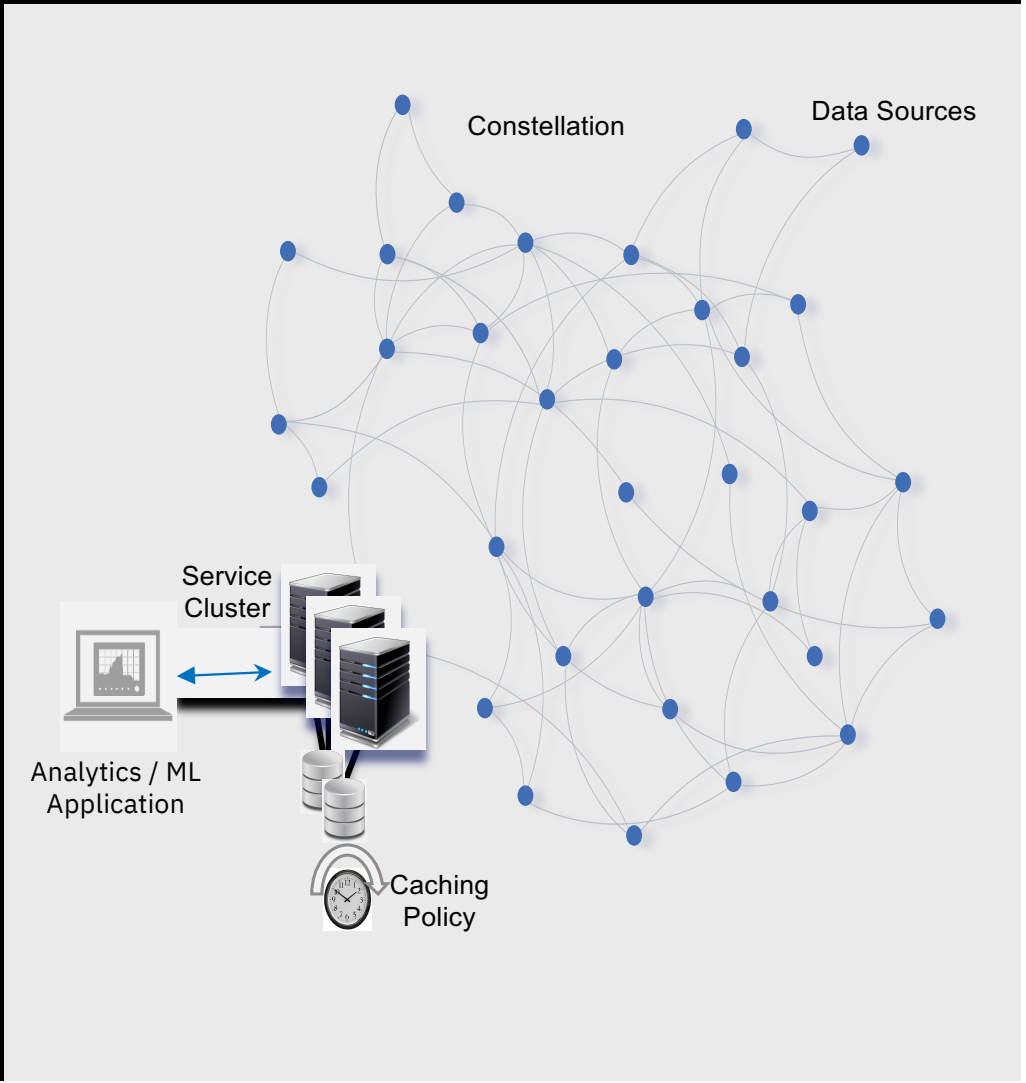
- Theoretically unlimited scalability
- Ease of addition/removal of sources
- Execution pushed down into the constellation mesh





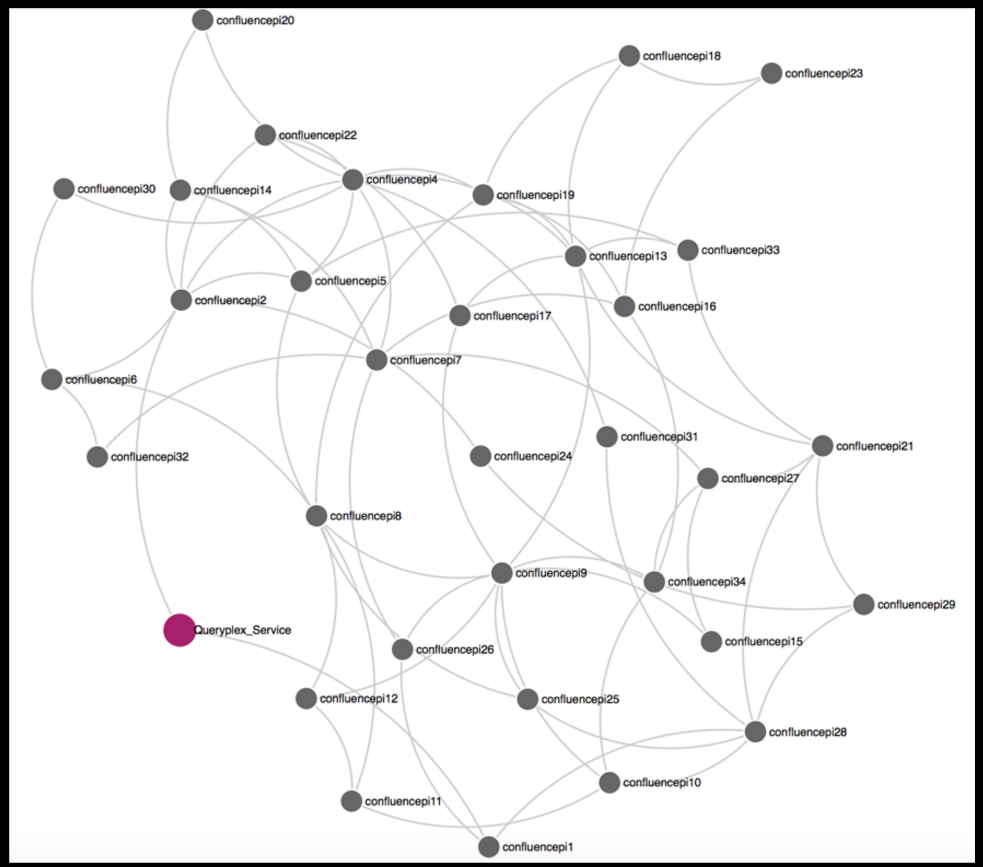
IBM’s Unique Approach to Data Virtualization

- Parallel processing mesh providing execution performance and scalability:
  - *Quickly deliver analytics results and easily evolve with new data source demands*
- Service Cluster scalability and Enterprise robustness:
  - *Reliability and ability to quickly adapt to increasing business demand*
- Governance Integration and Security:
  - *Controlled, governed and secure access to virtual data sets within the Cloud Pak for Data Platform*
- Common SQL engine, rich set of SQL dialects, application portability
  - *Retain the use of existing applications and tools within the business*
- Richness of automation and discovery of Data:
  - *Enable more self service and increase productivity (while retaining access control)*
- Underlying platform flexibility with Cloud Pak for Data:
  - *Grow or move your Analytics and Virtualization platform with any environment*



Creating the Constellation

- Small packet of Data Virtualization software deployed on each data node, approximately 50MB.
- Automatic dynamic and resilient organization of data sources.
- Query compiler builds a collaborative query plan, forcing the nodes to collaborate



Broad support for common data source types

More to be added in the future.

Note - IBM's statements regarding its plans, directions, and intent are subject to change or withdrawal without notice at IBM's sole discretion.

Cloud Pak for Data (Nov 2019)

- Db2 family for HDM
- Db2 for iSeries, zSeries
- Db2 for z/OS
- Big SQL
- IIAS, PDA (Netezza)
- Informix
- Derby
- Oracle
- SQL Server
- MySQL
- PostgreSQL
- Apache Hive, HDP Hive
- Cloudera Impala
- Teradata*
- MongoDB
- Hive
- Excel, CSV, Text*
- Sybase
- MariaDB
- Snowflake
- Z Data Sources through IBM DVM Integration
  - VSAM, IMS, CICS, Adabas
- Map-R (Hive)

In the roadmap pipeline

- BigQuery (patch in 2019, GA Q1, 20)
- SAP HANA
- SAP BW
- Amazon Redshift
- Salesforce
- SAS
- Interbase
- Apache Drill
- Amazon Dynamo, Aurora
- Generic use of any JDBC access
- CouchDB
- Stream / MQ
- Apache Spark SQL
- Apache Kafka
- Cloudant
- Pivotal Greenplum
- Cassandra

Thank you

Data Virtualization Manager for z/OS

Data Virtualization Manager for z/OS

More than 35 supported data sources

IBM Z

ADABAS

VSAM

IMS

IBM DB2

CICS

IDMS

CA IDMS

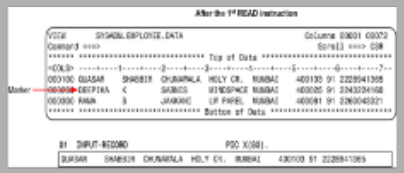
Systems Management Facility SMF

Data for IT Operational Analytics

Syslogs log streams

LOG

Sequential-file



File Systems

zFS

HFS

SFM

Magnetic Tape Virtual Tape

Non-IBM Z

IBM Db2 Family, Informix

hadoop

BigInsights

ORACLE

HIVE

mongoDB.

Microsoft SQL Server

20+ DRDA data sources

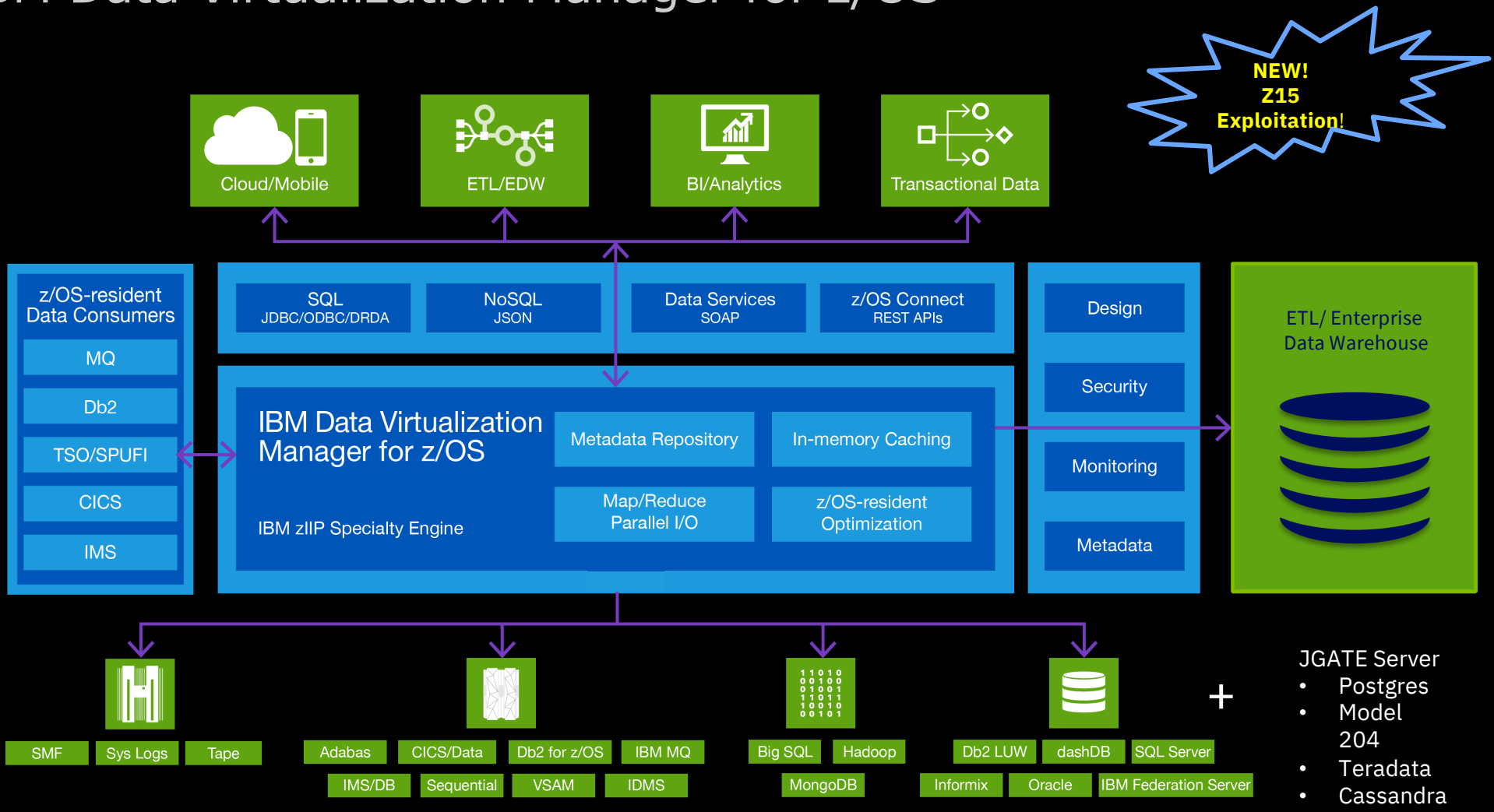
Other DBs

THE Open GROUP

... exposed to data consumers

in SQL (JDBC/ODBC/DRDA) and NoSQL (JSON) formats

IBM Data Virtualization Manager for z/OS



DVM for z/os Use Cases

- **Expand Db2 applications beyond just Db2**
  - New features making Db2 as the Datahub for non-Db2 data sources
  - ”Any Db2 application to any data”
  - Large Insurance company

- **Providing z data for testing scenarios**
  - DVM enabled customer’s dev tool to access IMS
  - Multi-national Telecomm company

✓ **Creating a virtual data lake**
  - System of record data stayed on z reducing ETL
  - Inventory information located across continents is virtualized with DVM
  - Multi-national Automotive company

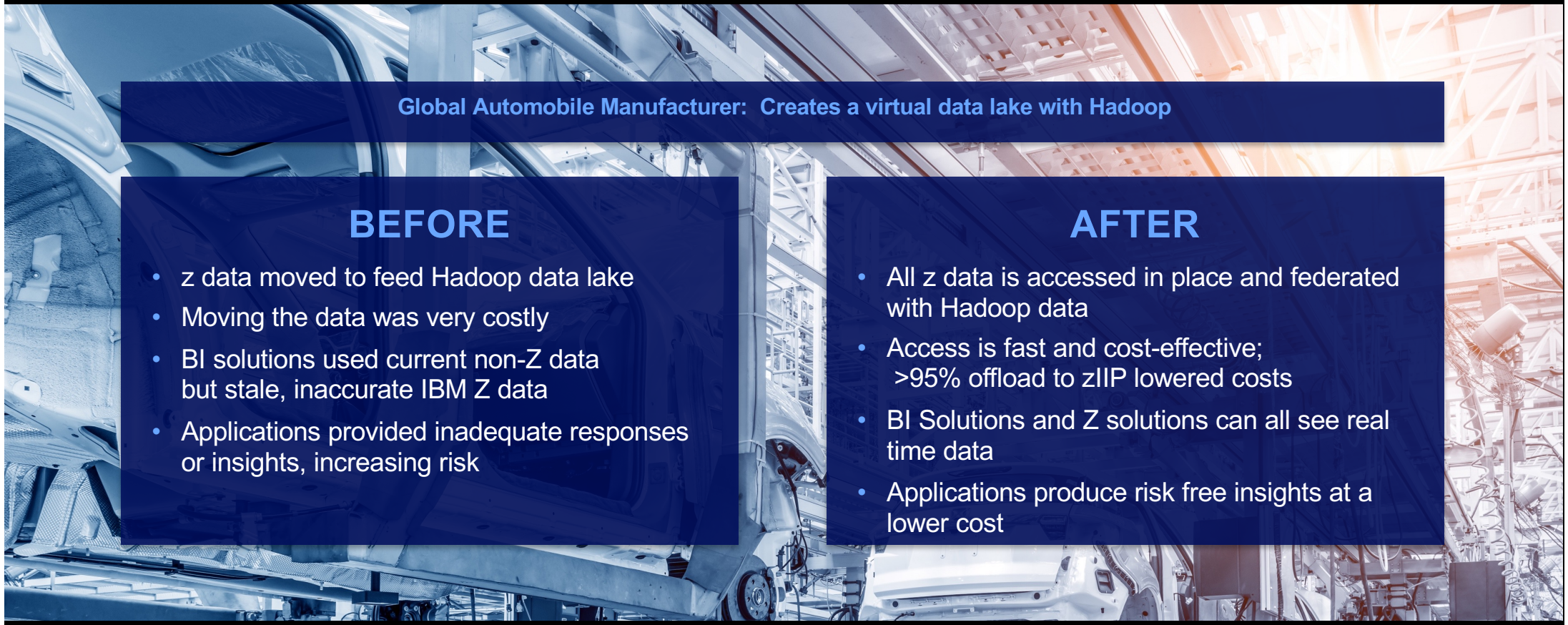
✓ **Modernization with DVM and z/os Connect**
  - DVM used to provide Rest access to z/os data (VSAM, IMS, etc) as well as Db2 z
  - DVM used to access IMS data directly as backup to unstable data warehouse on distributed platform
  - Global Airline

✓ **Shrinking app dev cycle with DVM and ETL**
  - No win situation with ETL (customer politics and deep investment)
  - Large Financial

- **Improving data access to aircraft maintenance records**
  - 4 hr ETL process to Oracle ODS reduced
  - DVM showed no significant increase in resource utilization on z
  - PoC showed accessing IMS data in place as a good option over data movement
  - Global Airline

- **IBM’s Classic Federation Upsell**
  - 5x the performance in PoC (Bank)
  - Classic Fed no longer being enhanced
  - Migration plan being developed by Rocket - - available soon

Make Data Simple & Accessible



Global Automobile Manufacturer: Creates a virtual data lake with Hadoop

BEFORE

- z data moved to feed Hadoop data lake
- Moving the data was very costly
- BI solutions used current non-Z data but stale, inaccurate IBM Z data
- Applications provided inadequate responses or insights, increasing risk

AFTER

- All z data is accessed in place and federated with Hadoop data
- Access is fast and cost-effective; >95% offload to zIIP lowered costs
- BI Solutions and Z solutions can all see real time data
- Applications produce risk free insights at a lower cost

Cloud Pak for Data DV combined with DVM for z/os

Cloud Pak for Data DV and DVM integration

- Extends the CPD DV service to the mainframe using DVM for z/OS
- Provides seamless interface and access to ALL Z data
- Auto-discovers z data, cataloging, (Gather/Manage, Understand, Govern)
- Leverages real-time persisted data for analytics and ML using z data
- Enables Developers and Business Users lacking mainframe skills
- Reduces costs and complexity of ETL when working with z data
- DVM connectivity with Integrated DRDA Facility, Db2 UDTF support (User Defined Table Function)
- ALL DVM for z/os functions (read/write, API support, data sources support, exploitation of hardware (z15!) are available – no changes to DVM

Products for Z :

- DVM for z/os (**5698-DVM**) – business as usual
- Cloud pak for Data zPPA: **D1YH6LL** (Enterprise Edition) or **D1ZXLLL** (Native Edition)

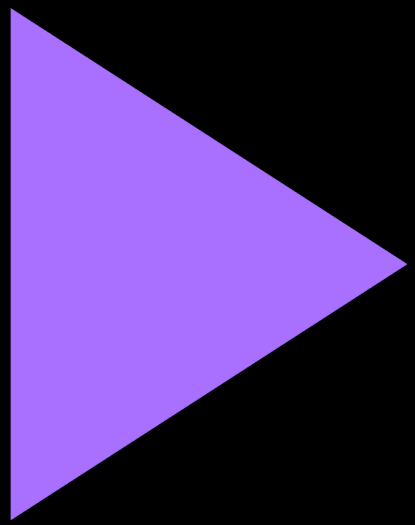
Client already has DVM?

- There is no change to their license. No additional cost..
- Action: Sell Cloudpak for Data

IBM Data Virtualization Manager– Key Enhancements

Integrated DRDA Facility	*Allows joining data from multiple data sources even when the data sources have different catalogs or are located across sysplexes. Example is joining data across IMS servers and VSAM catalogs that is spread out across sysplexes*. Technology: A DVM Server can now connect to and access virtual data from one or many DVM servers. Data can be shared between DVM servers without being on the same Sysplex.
Db2 UDTF support (User Defined Table Function)	*Allows any application to access any DVM virtualized data using DB2 via SQL. Ex: DB2 app accesses and updates VSAM records using SQL.* Technology: DVM supports the creation of a Db2 UDTF on any virtualized object defined to a DVM server within any Db2 subsystem connected to DVM. Once the UDTF has been created any connection/application to Db2 can easily access this data using Db2's SQl engine.
JAVA Gateway	*Any JAVA application can access data without needing a JDBC driver to be purchased for the data source. Customers are using this to access* ***Teradata, Postgres, Model 204 and Cassandra data****.* Technology: A DRDA application server can be placed on any system to access data via a JDBC driver and make it accessible to any application. Avoids dealing with OEM DB to drive DRDA support.

DEMO



What you'll see

Introductory video (Luis – 6 min)

- **Data Virtualization** over widely distributed data
- **SQL editor** inside Cloud Pak for Data

Caching and Governance

- Publishing the virtual objects to catalog
- Associate business terms to virtual objects
- Associate governance rule to virtual objects (allow/deny access)
- How to define cache
- Show cache usage graph (Use SQL editor + Any other tool like Tableau)

Remote connectors

- Remote connector to files from Mac
- Remote connector to Db's using discovery

Data Virtualization Roadmap - @Jan 2020

1H 2020

ICP4D Continuous Delivery

1. **Data Source Connectors to identify for Schema changes for DV Objects**
2. **Governance Enhancements for DV Objects**
   - Business Terms mapping and classification for all newly discovered data columns of a new / changes to Data assets
   - Policy Enforcement :Phase 2 – Smart Masking
   - Lineage :Phase 2 - Horizontal and Vertical
3. **Security : DV role access control enhancement- Group level Authentication /Authorization**
4. **Smart Data Caching for better Perf with Policy definition, management, incremental caching of large datasets**
5. **Additional Data Source Support**
   - Market Based Decision
1. **Monitoring enhancements**

2H 2020

ICP4D Continuous Delivery

1. **RESTful API for Virtual Object creation and updates / Other RESTful API based on market needs**
2. **Workload / Policy Management**
   - As an admin define user / resource prioritization for workload
3. **Additional Data Source Support**
   - Market Based Decision
4. **NLP Query Editor**
5. **Data Write Back capability**
   - Write to staging area
6. **Data Write Back capability**
   - Ability to write the transformed data back to the source from which the data was initially extracted
7. **High Availability of Service Node**
8. **Data Lifecycle Management**
   - Organizing data between tiers based on policy
   - Automatic migration of data between tiers

Cloud Pak for Data – Self-serve ready

Foundational “out of the box” multicloud data & AI services

Open, Extensible Platform


App Developers & Analytics Ops


Business Partners

Personalized, Collaborative Platform

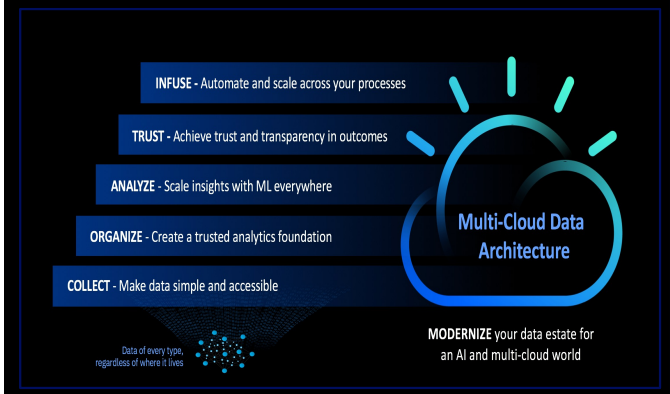

Data Engineers


Data Stewards


Data Scientists


Business Users

The Ladder to AI



MODERNIZE your data estate for AI in a multi-cloud world

APIs Integrated User Experience

Extensible : “*add-ons*”, accelerators and Solutions

***Modular**: - provision services & scale out when needed*

Collect & Connect

- Data virtualization
- Provision SQL & NOSQL Databases
- Warehouses & Marts
- Event Ingestion & Streaming Analytics
- Distributed compute – Apache Spark

Organize & Integrate

- Discovery & search
- Data transformation
- Data catalogs, quality & Curation
- Business glossary
- Policies, rules & privacy

Analyze & Infuse

- Data Science & Visualization
- Dashboards & reporting
- AUTO-AI, ML deployments & operations
- AI Trust and Transparency - Explainability & Bias detection
- AI services –Chat, NLU

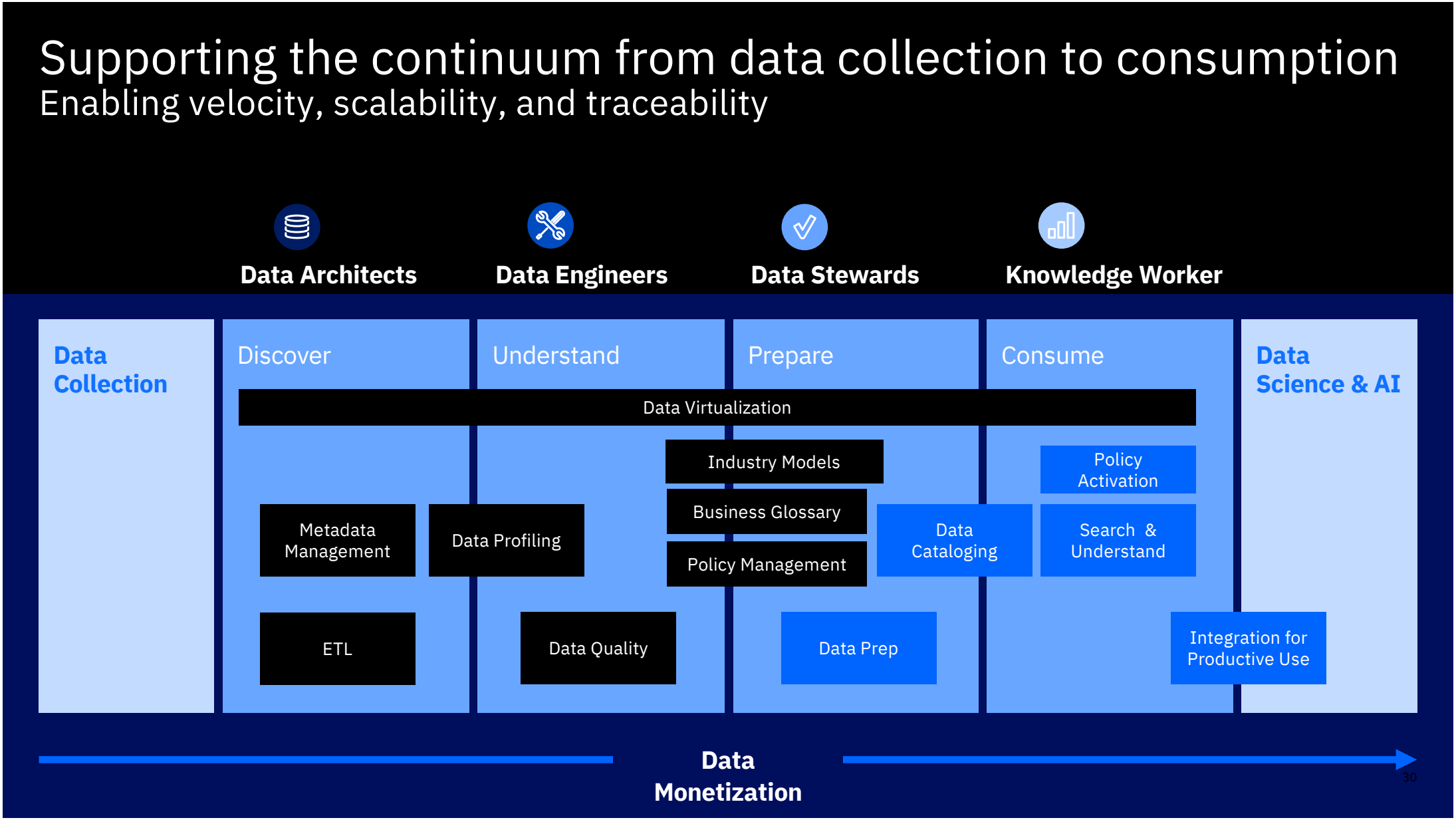
Core Services	• Logging • Monitoring	• Metering • Storage Volumes	• Auditing • Security	• Identity Access Mgmt. • Docker Registry , Helm

RED HAT OPENSHIFT

IBM Cloud aws Azure Google Cloud openstack Hyperconverged System

Supporting the continuum from data collection to consumption

Enabling velocity, scalability, and traceability



Competitors and differentiators ?

Competitive One Pager

Cloud Pak for Data Data Virtualization Service

IBM Competitive Advantages

Data virtualization in IBM® Cloud™ Pak for Data, is a unique new technology that connects tens to hundreds of data sources into a single self-balancing collection of data sources or databases, referred to as a constellation. An application submits a query that's processed near the server where the data source exists. Results of the query are consolidated within the constellation and returned to the original application. No data is copied and it exists only at the source.

Our forward facing technology with ***richness of automation*** *from data discovery to* ***self-service for data*** *preparation with* ***overlay of security and governance*** *support clients on their one single Data Fabric journey on* ***hybrid cloud platform for geographically dispersed sources.****. Accelerates their digital transformation in a more efficient, cost effective and seamless manner than our competitors.*

#1 Collaborative computing mesh providing scale and performance

- Why is it unique: *Self organizing peer to peer architecture, applications can query across multiple databases, flat files and big data repositories, individually or collectively. Make many databases — even globally distributed —appear as one to an application. Simplify data analytics with a scalable and powerful Cloud Pak for Data platform. Execution pushed down into the constellation mesh, allowing strong parallel, distributed processing.*

#2 Simplicity and Self-Serve capabilities

- Why is it unique: *Automatically find and match tables across systems so you can query them as a single virtual table. Connect to Data Virtualization with your favorite SQL apps and tools, i.e. RStudio, Jupyter Notebook, Cognos, Tableau, Microstrategy*

#3 Security and Governance

- Why is it unique: *Centralize and strong access control and governance. Fully encrypted communication. Governance through single data catalog (metadata repository) on the Cloud Pak for Data platform*

#4 Query across multiple data sources

- Why is it unique: *Supports Oracle, Db2, SQLServer, Informix, Netezza, MySQL, PostgreSQL, Big SQL, Apache Hive, HDP Hive, Cloudera Impala and more. Adding data source is easy across regions.*

Competitive Landscape (key ones)

Denodo-

- *Who are they:* Denodo is a private company, still VenCap funded, revenues ~$50M, Data Virtualization focused
- *Why a threat*: Recognized market leader, present in much of our client base, growing 60%+ annually
- *Main strength(s):* Rich GUI, robust set of data source support, graphical support for building views, some key features (e.g., caching and data masking) that are on our roadmap for 2H,19. Conceptually builds a virtual data warehouse.
- *Main weakness(es):* Some auto discovery, limited in scalability, still heavily reliant on centralized processing and hub and spoke model. Starting 2H'19 with integration of Watson Knowledge Catalog on CP4D and DV we will have very strong governance support, from business glossary, data privacy & protection rules to data lineage

TIBCO -

- *Why a threat:* Tibco has strong install base in Financial Services, but growth stalled under Cisco till recently when it came into TIBCO
- *Main strength(s):* Appear to be integrating DV into their Spotfire Analytics Platform
- *Main weakness(es*): Still to gain momentum back & lack of end to end solution

Responding to Objections

IBM Cloud Pak for Data is Industry leader in an end to end Data and AI platform allowing to simplify Data and AI, digital transformation journey. IBM has invested heavily in, custom user experience for all user persona's in an enterprise for a simplified self-serve experience. The platform is ready to help clients with their Data compliance and security journey to counter today's compliance needs. Data Virtualization data source support is growing everyday, and a comprehensive list of sources are on our roadmap. Also, IBM is ready to prioritize the sources for a client opportunity. While these IBM competitors offer only point to point solution, IBM differs in its ability to provide an end to end solution for Data and AI Journey with built-in governance.

IBM Internal ONLY

Competitive landscape matrix

Feature (reference from previous page)	**Data Virtualization Service**	**Denodo**	**TIBCO**
Collaborative computing (#1)	IBM Cloud Pak for Data	▲	▲
Simplicity (#2)	IBM Cloud Pak for Data	■	▲
Centralized Governance and One Catalog (#3)	IBM Cloud Pak for Data	▲	▲
Security (#3)	IBM Cloud Pak for Data	■	▲
Rich Application / Self-Server Capabilities (#2)	IBM Cloud Pak for Data	▲	▲
Data Source Support (#4)	IBM Cloud Pak for Data	▼ *(they have ten years lead, we are quickly adding new sources, also IBM DV can work easily with geographically dispersed sources)*	■
Optimized Performance (#1)	IBM Cloud Pak for Data	▲	▲

▲ IBM shows a favorable comparison ▼ Competitor shows a favorable comparison ■ About the same comparison

Other Vendors - Astera Software, Cambridge Semantics, Data Virtuality, Dremio, EQ Technologic, fraXses, Gluent, Marklogic, Microsoft PolyBase, SAP Smart Data Access, Stone Bond, and Teradata QueryGrid.

CP4D Experiences

Your digital selling toolkit

Glimpse into key product features

VIDEO DEMOS

Experience functionality interactively

PRODUCT TOURS

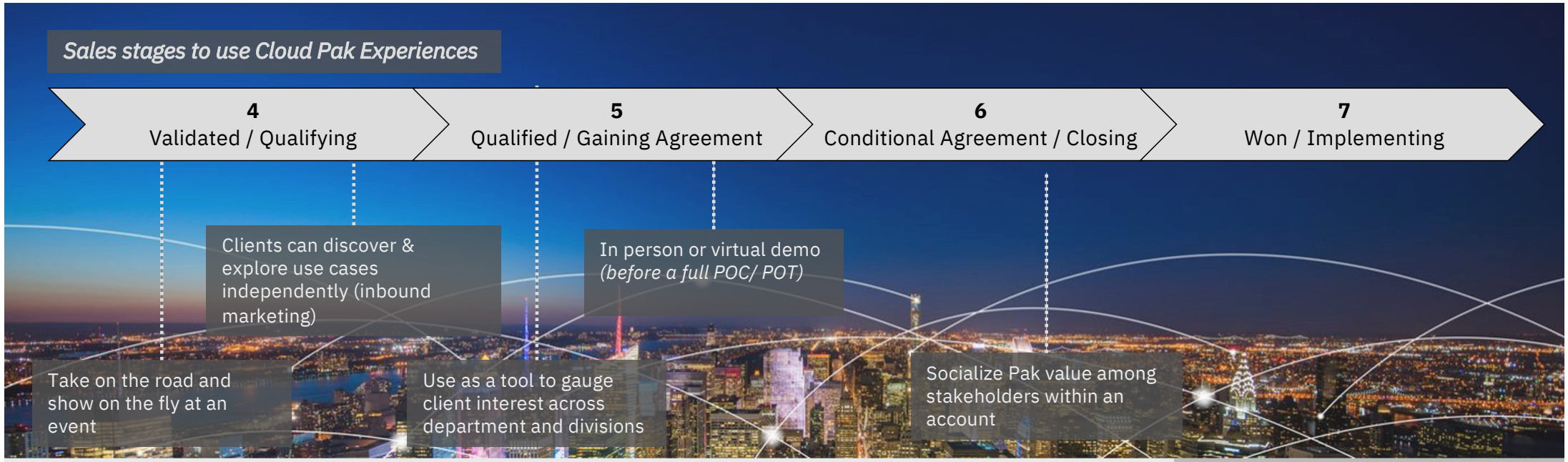
Discover 'Wow' moments in 5 minutes or less with LIVE product

CLOUD PAK EXPERIENCES

- Hosted environments give clients hands-on access to test drive IBM Cloud Paks
- Guided flows step users through example scenarios
- Users can explore the full live product on their own or with sales guidance
- 7 days of access, with the option to extend to 14 days

Learn a concept with in-depth of tasks and steps

HANDS-ON LABS



Client wins with Cloud Pak Experiences

NetApp®

Lilly

Associated Bank

Capgemini

Pro Tips

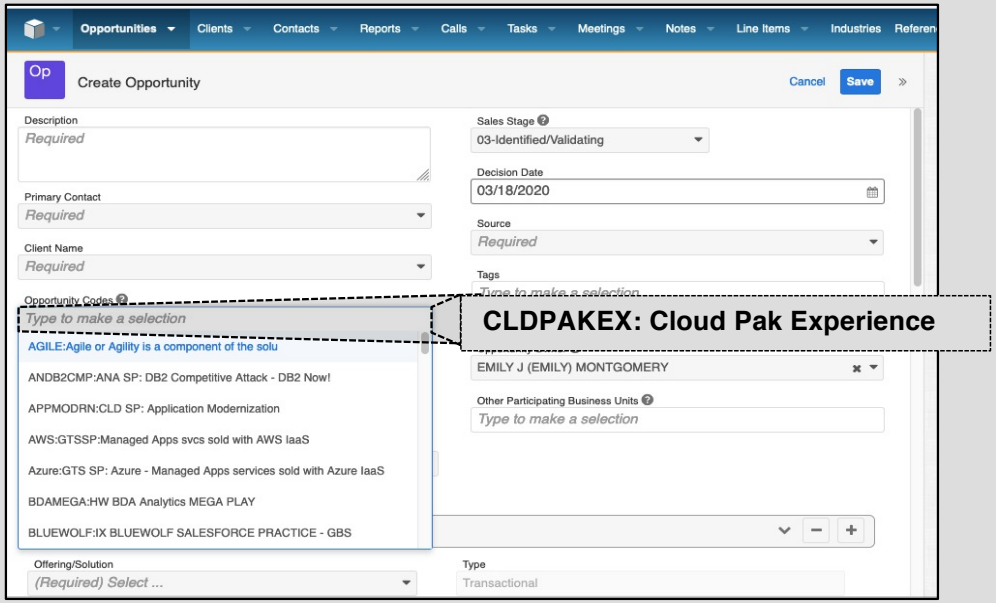
- ✓ Pre-loaded flows highlight key value propositions and common scenarios.
- ✓ Go “off-script” to craft a personalized message for your client’s needs.
- ✓ Share digitally, on the road, or in a meeting. (Firefox and Chrome are recommended.)

Cloud Pak Experiences

Add an Opportunity Code to New or Existing Opportunities

SalesConnect

1. In the “Opportunity Codes” field, type ‘Pak’ to make a selection.
2. Find “CLDPAKEX: Cloud Pak Experience”.
3. Click Save after you have entered all the other required/known values.



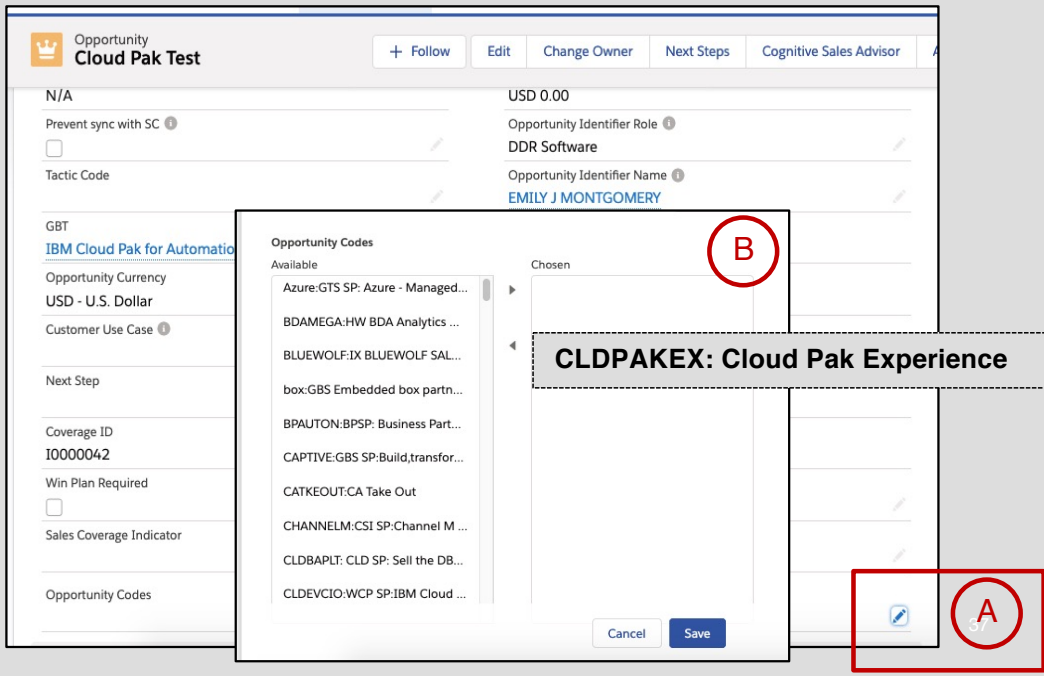
Salesforce (GAIA)

A

1. Fill the required/known fields and click Save.
2. Scroll to the Opportunity Codes field and click the Edit (pen) icon.

B

1. On the next screen, scroll through the Available pick-list and find “CLDPAKEX: Cloud Pak Experience”.
2. Add it to the Chosen list, and then click Save.



Get Started With Cloud Pak Experiences

ACCESS TODAY ibm.com/cloud/paks/experiences

RELEASE SCHEDULE

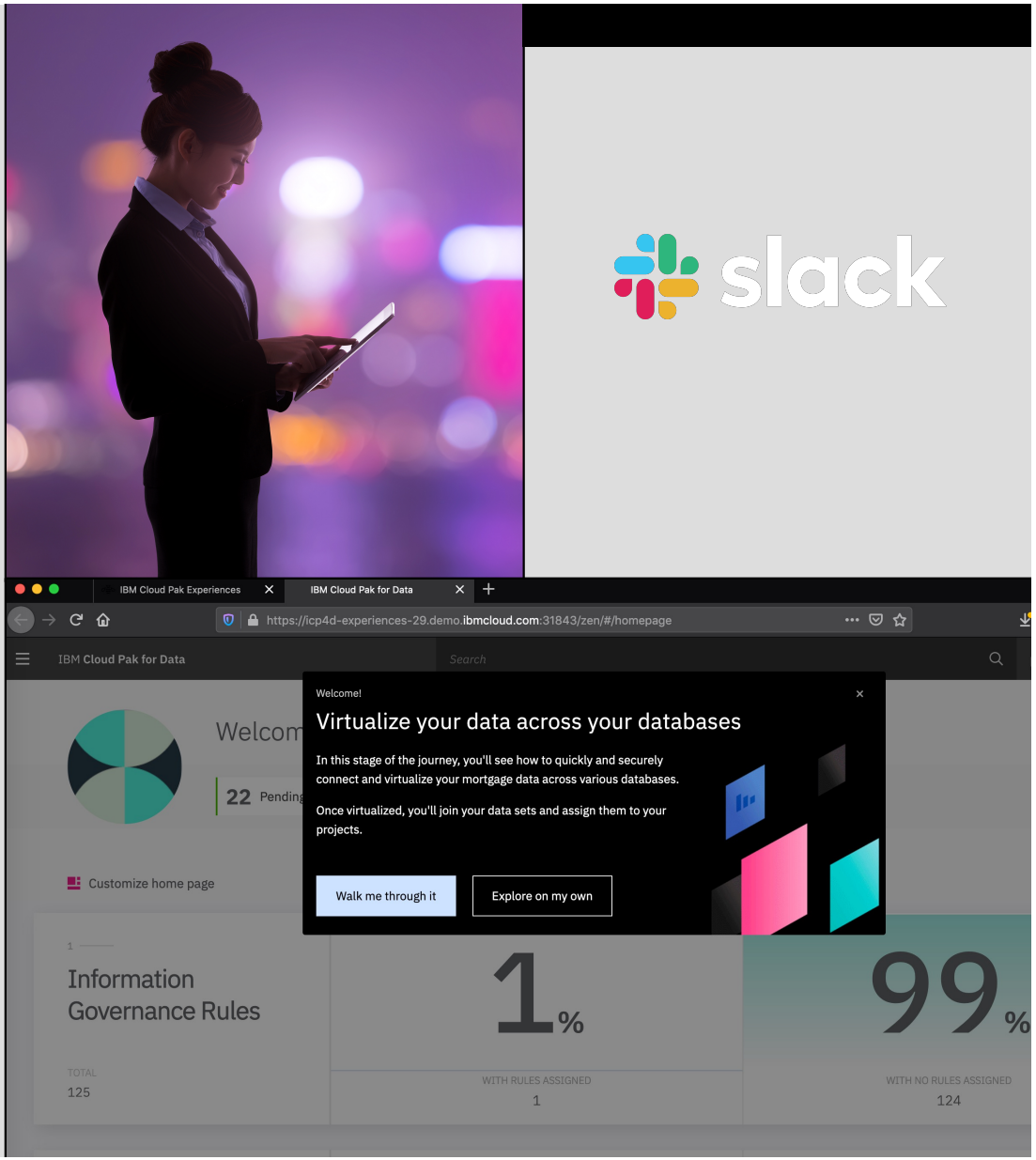
Cloud Pak for Data	
Collect & Virtualize	Available now!
Organize	Available now!
Analyze	Available now!
Update for version 2.5	*Coming 26-Jan*
Updated flow for Organize	*Coming 26-Jan*
Updated flow for Analyze	*Coming 26-Jan*
Cloud Pak for Automation	
Capture data	Available now!
Manage content	Coming 14-Feb
Cloud Pak for Applications	
Assess cloud readiness	Coming 17-Jan
Build cloud native	TBD

CONTACT US

Elizabeth Silberg
Director, IBM Cloud Paks Digital
silberg@us.ibm.com

Travis Roache
Digital Offering Manager, Cloud Pak Experiences
tgroache@us.ibm.com

Justin Herberger
Digital Offering Manager, Cloud Pak Experiences
justin.herberger@ibm.com



Additional Resources:

IBM Cloud Pak for Data

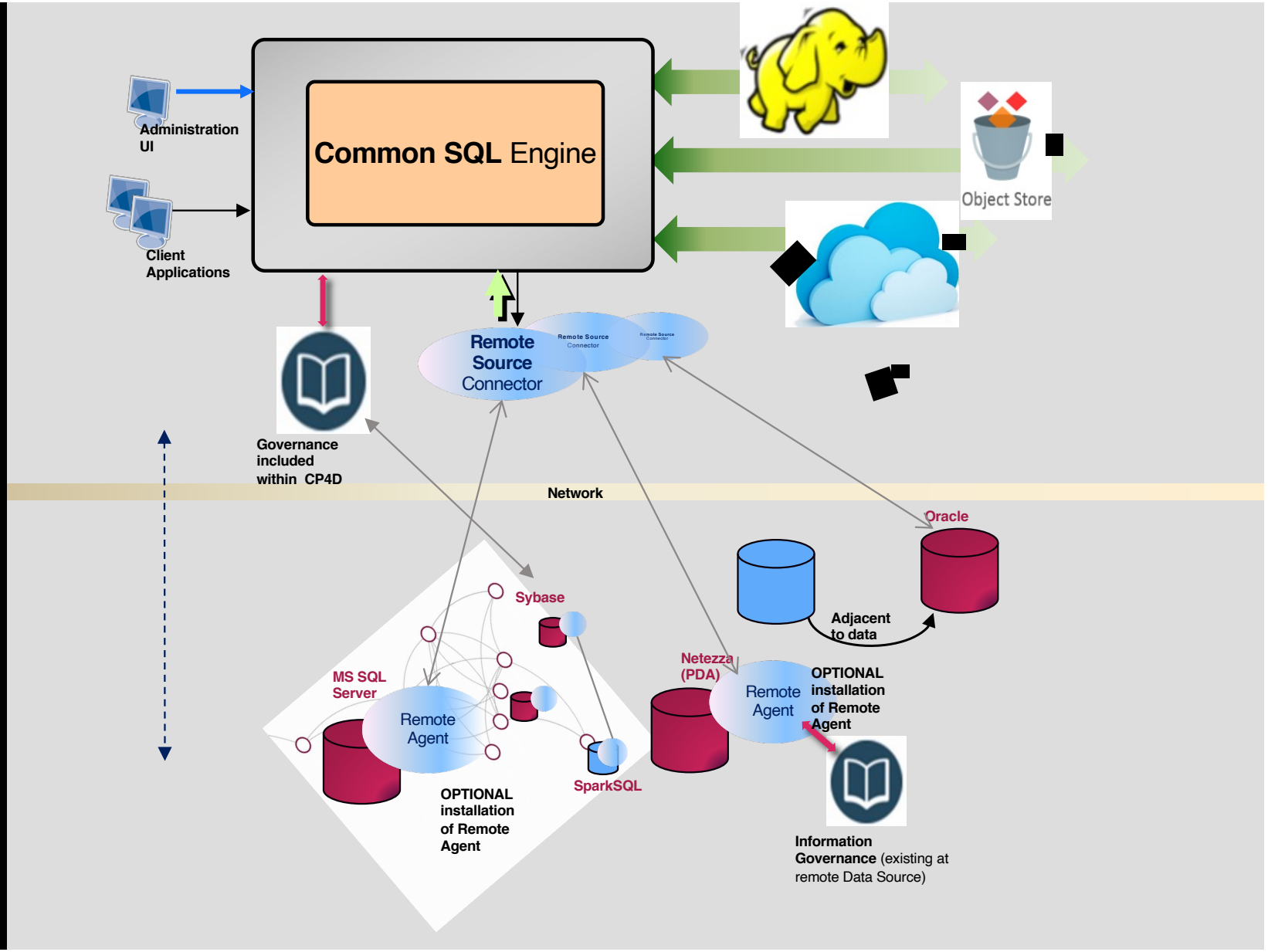
CP4D Data Virtualization
Digital Technical Engagement – Cloud Pak for Data

Data Virtualization Manager for z/OS

DVM Homepage
Digital Technical Engagement - DVM

Supplemental

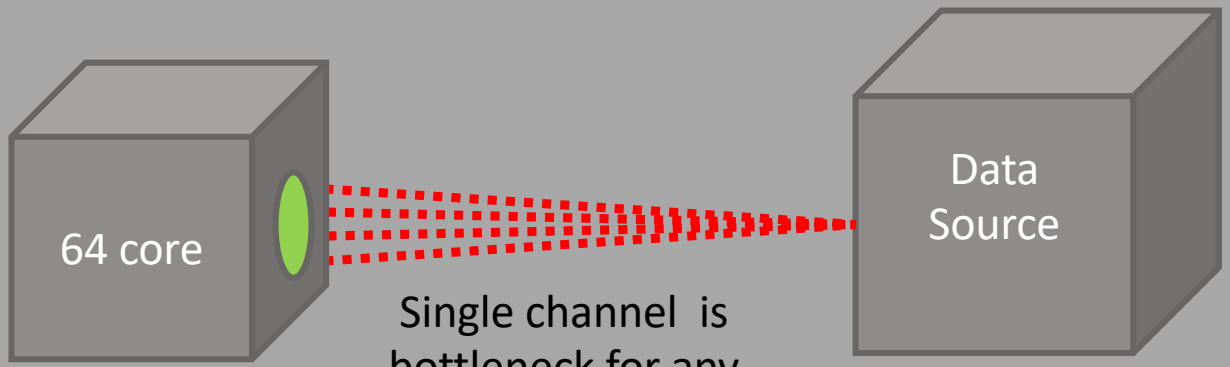
Architecture



Bringing parallelism to each data source.

Traditional processing queries the data source (on right) then processes large results on a single thread (executor) on left.

Solution: Data Virtualization queries data source (on right). Merges results in many **parallel threads** on left by leveraging parallel read streams.

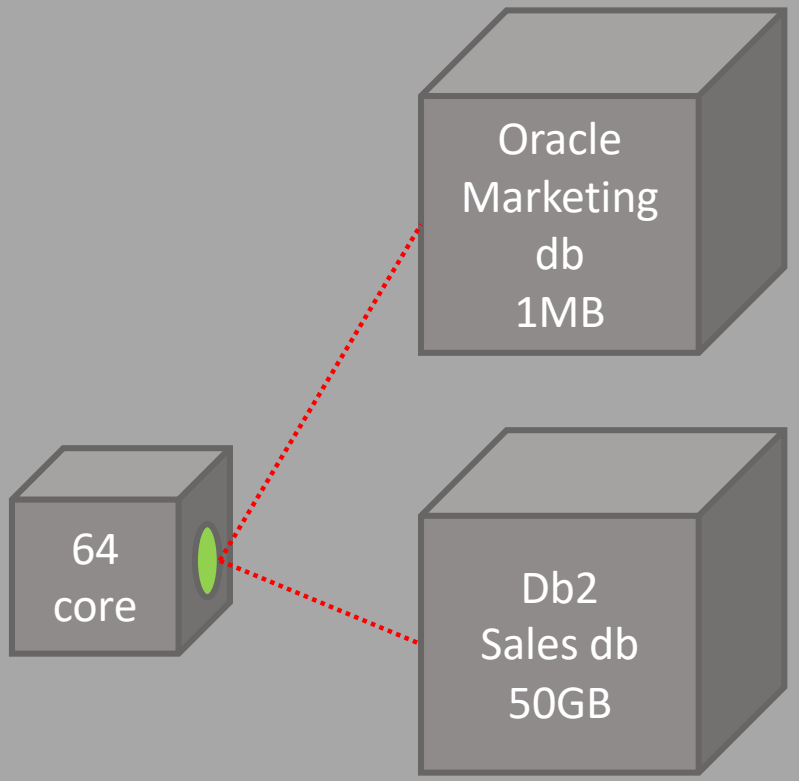


Single channel is bottleneck for any large result set. Data Virtualization resolves this through parallel read from sources.

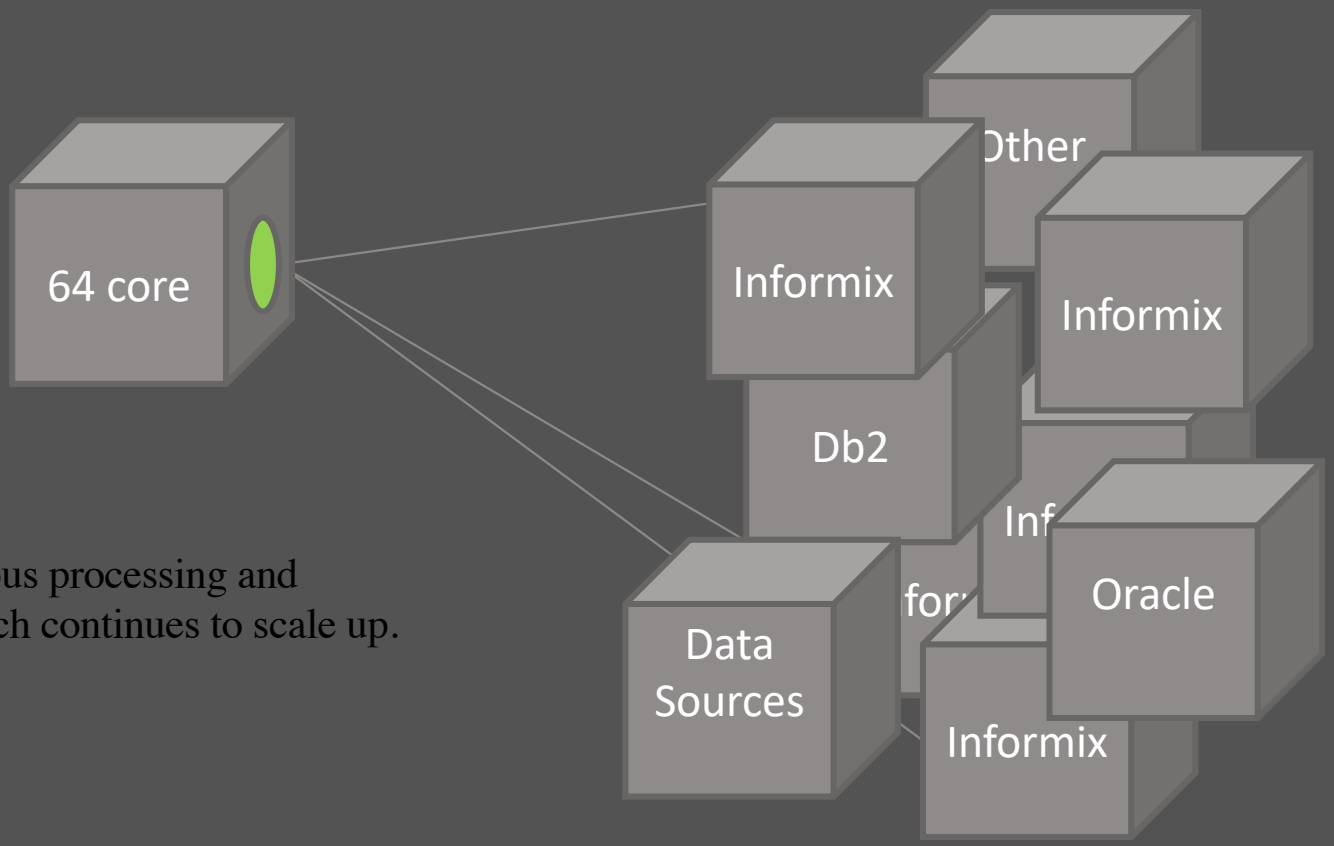
Efficient Cross Source Joins

Problem: Joining data between multiple databases is a common pattern, but brutally slow. With traditional processing, data from both tables is shipped over the network to the server, where the join is processed.

Solution: Data Virtualization uses the **early filtering** techniques to dramatically reduce transmission costs. Data from the smaller table (inner) is used to build a small filter query that is applied to the larger (outer) table before data is transmitted. In many cases this is 85% effective at filtering data that will be removed by the join before the join is processed, leading commonly to a ~10x acceleration.



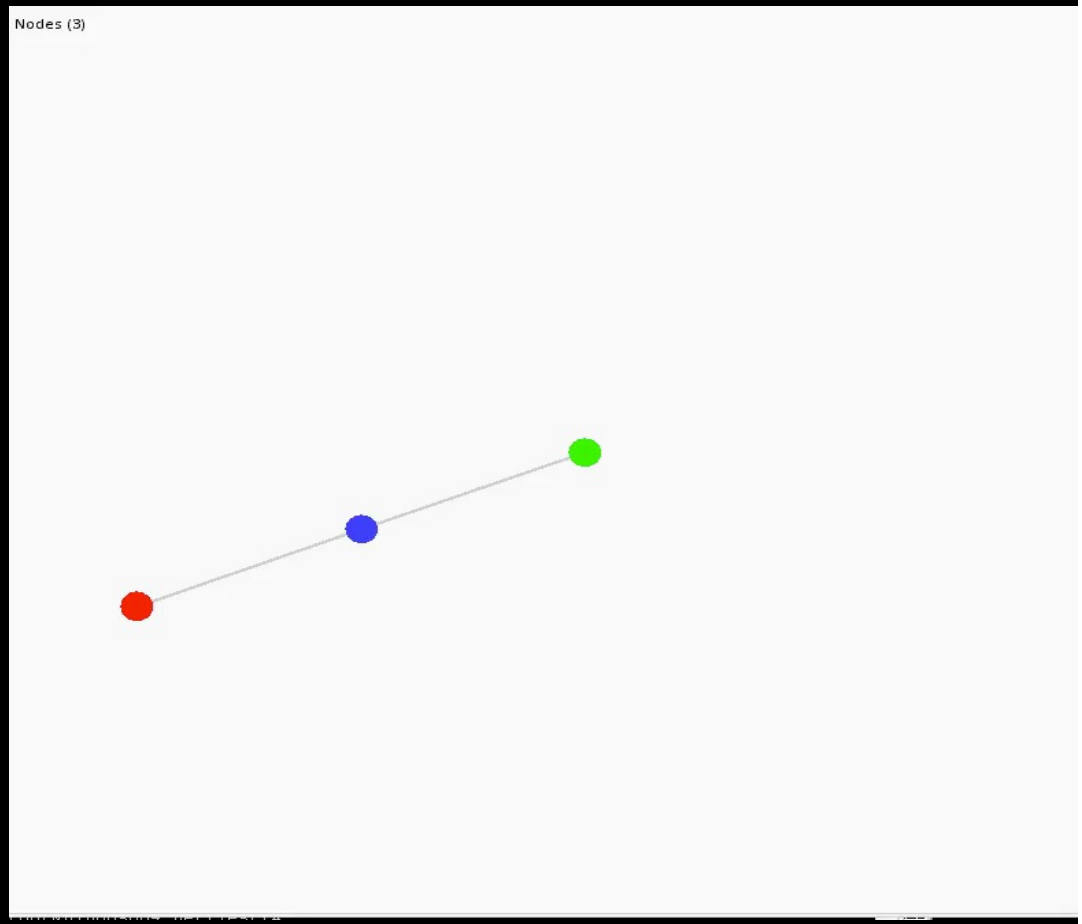
20 databases? Traditional processing gets worse as the number of databases increases



Simultaneous processing and parallel fetch continues to scale up.

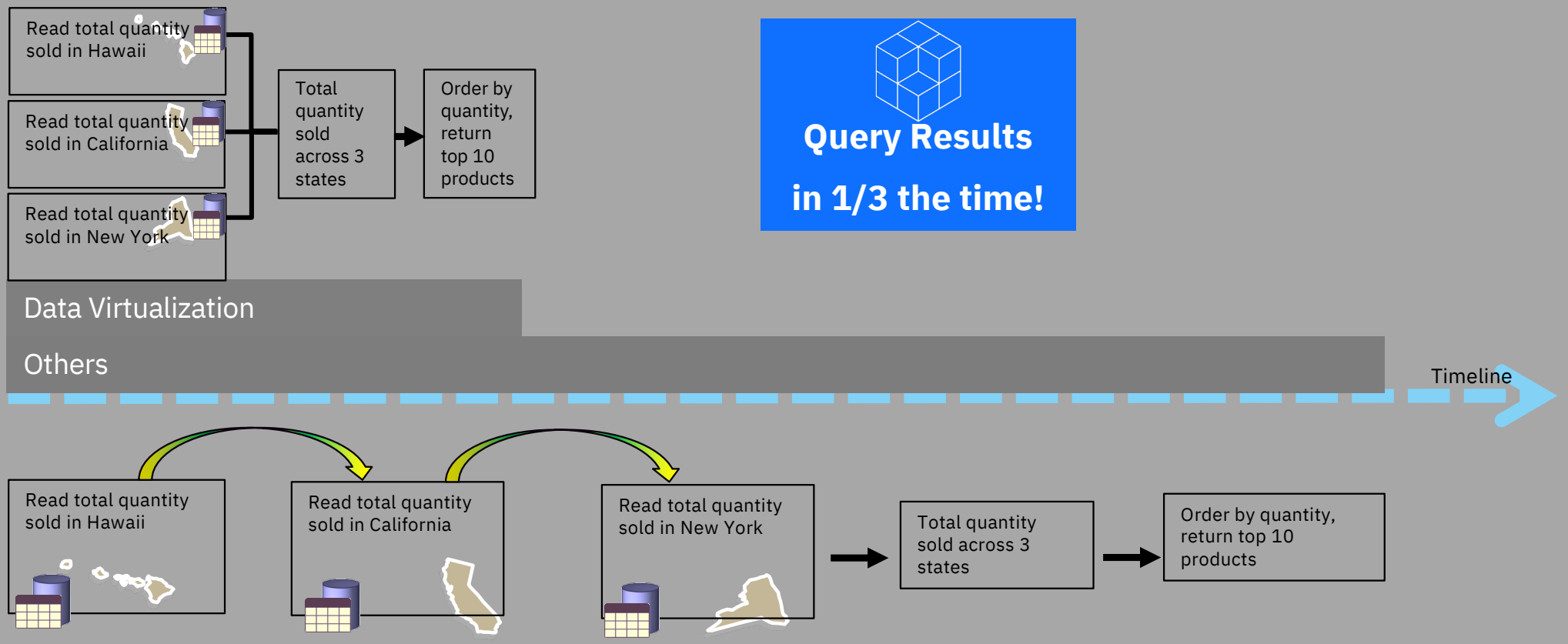
Real System test
Growing a Constellation

- Video of constellation growing to 349 Nodes.
  - Network stays compact.
  - 2 and 10 links between nodes
  - No manual configuration.
- Latency aware connection between nodes
  - Which nodes connect to which others?
  - Fastest reply strategy
- Diameter of the constellation (i.e. the number of hops between the two furthest nodes) grows logarithmically. Small diameter is ideal for communications.



Smart Query Processing

IBM Data Virtualization reduces query time by using Parallel Processing, Pushdown Optimization and Connection Pooling



DV Engine Generated SQL for Distribute Aggregation

Original

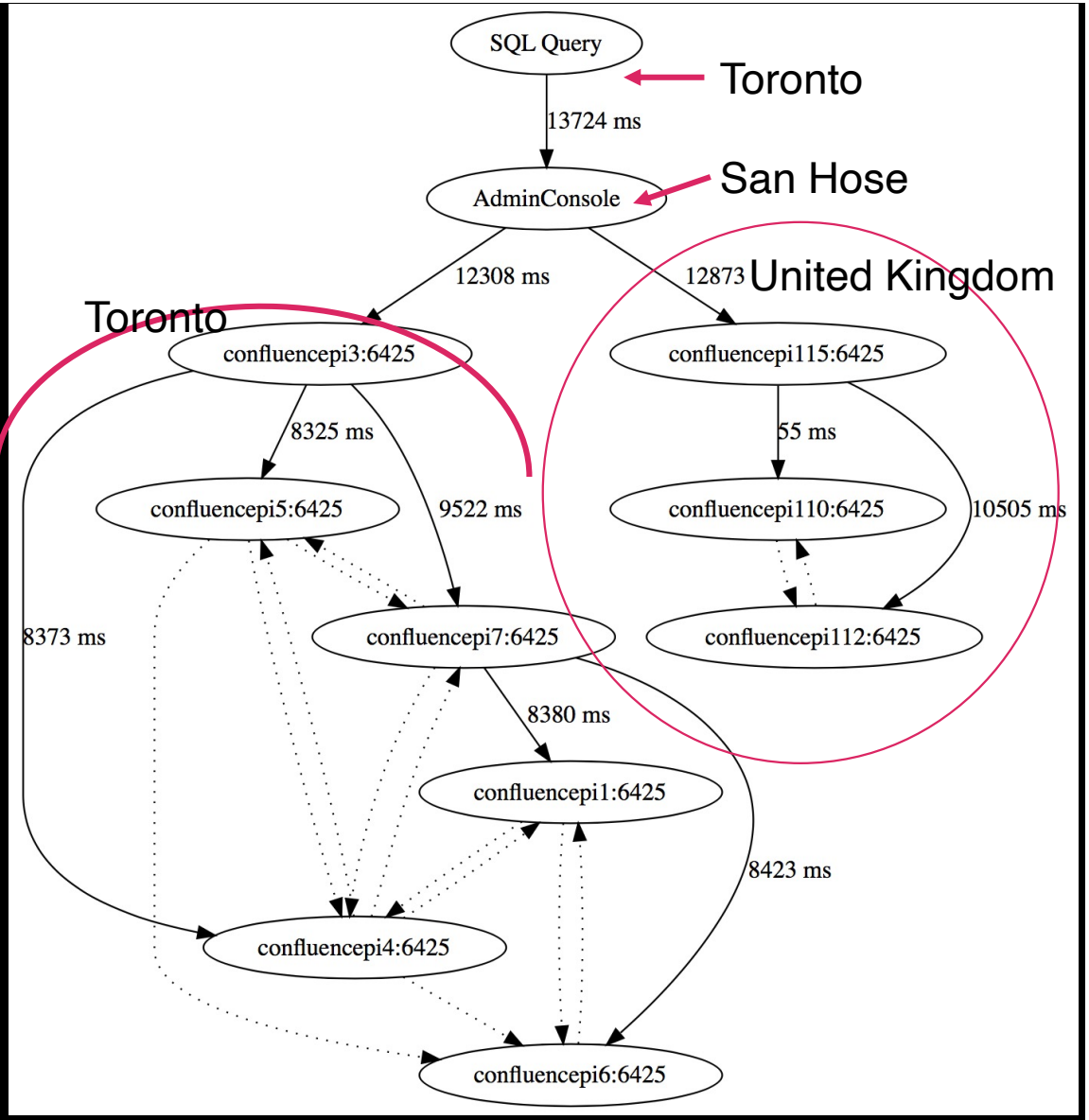
```
SELECT COUNT(PROMOVALUE2) FROM PROMOTION
```

Remote SQL

```
SELECT SUM( A0.C0)
  FROM (
    SELECT A1.C0 C0
      FROM new com.ibm.db2j.GaianQuery(
            'SELECT COUNT( A2."PROMOVALUE2") C0
               FROM new com.ibm.db2j.GaianTable(
                        ''PROMOTION'',
                        ''SOURCELIST=(MYSQL10000:"POPS_node1", MYSQL10001:"POPS_node2",
                                      MYSQL10002:"POPS_node3", MYSQL10003:"POPS_node4",
                                      MYSQL10004:"POPS_node5")'',
                        ''"PROMOKEY" INTEGER, "PROMOTYPE" INTEGER, "PROMODESC" CHAR(30),
                          "PROMOVALUE" DECIMAL(5, 2), "PROMOVALUE2" DECIMAL(5, 2),"PROMO_COST" DECIMAL(9, 2)''
                ) A2',
               'GAIAN_EXTENDER=[pAgg]','',
               'C0 DECIMAL(5, 2)') A1
       ) A0
```

Query Processing in the Constellation

- Fixed execution within the constellation is impossible because of the highly dynamic nature of the network.
- Each node instead simultaneously sends the relevant portions of the query to both the connected data source to it's peers in the network.
- Combines and process the results as they are received.
- Duplicate results are avoided by a given node only returning results to the first peer that requested them.
- Implicitly results in balanced processing of the query through the constellation.



Language Translation in Data Virtualization

Broad set of data sources supported by Data Virtualization each with unique syntax variations.

Constellation is not limited only a single data source type. A logical schema is created across all connected sources.

Multiple levels of translation as we move from the applications through the constellation down to the data source.