Data Virtualization for the Enterprise

New England Db2 Users Group Meeting
Old Sturbridge Village, 1 Old Sturbridge Village Road, Sturbridge, MA 01566, USA September 27, 2018

Milan Babiak
Client Technical Professional
Analytics on IBM Z Systems
Milan.Babiak@ca.ibm.com

Title: Data Virtualization for the Enterprise
Real-time universal access to mainframe and non-mainframe data

ABSTRACT:
Large amounts of data accumulated over decades on variety of data management system present a wealth of information for your enterprise analytics, but to satisfy current analytics needs - modern APIs, web services, mobile, real-time analytics, security - presents challenges.

The concept of Data Virtualization addresses these challenges in three key areas:
1. Data Access Modernization for Business Intelligence and Data Science
2. Real Time Analytics
3. Data Access Optimization and Data Security
They provide a universal, transparent data access for modern application developers, data scientist, financial analysts and many other enterprise data consumers while saving time and ETL costs.

IBM Data Virtualization Manager for z/OS implements this concept on a highly secure, reliable mainframe platform, with the ability to access not only traditional mainframe data - IMS, VSAM, Db2 for z/OS, SMF, RMF, ... , but also non-mainframe data sources - Oracle, Microsoft SQL Server, Hadoop, and many other data sources.

Speaker's Bio:
Milan is Client Technical Professional at IBM. He creates and presents solutions and product demonstrations for IBM clients. His specialty is Analytics on IBM Z Systems - Mainframe. He earned his Master of Computer Science degree from Slovak Technical University Bratislava, Slovakia. Milan is a regular presenter and motivational speaker at conferences, IBM educational seminars, customer workshops, and Toastmasters meetings in Canada, US and Europe. His passion is explaining complex technical topics in a simple, understandable language to wide audiences. He is also an active member and club executive at IBM Toastmasters in Ottawa, Ontario, Canada.
LinkedIn: https://www.linkedin.com/in/milanbabiak/ Twitter: https://twitter.com/BabiakMilan

AGENDA

1. Data Processing in Historical Perspective
2. Hardware and Software Modernization
3. Data Processing Challenges and Needs
4. IBM Data Virtualization Manager (DVM) Solution
5. IBM Advanced Analytics for z/OS - Integrated Solution

Q & A

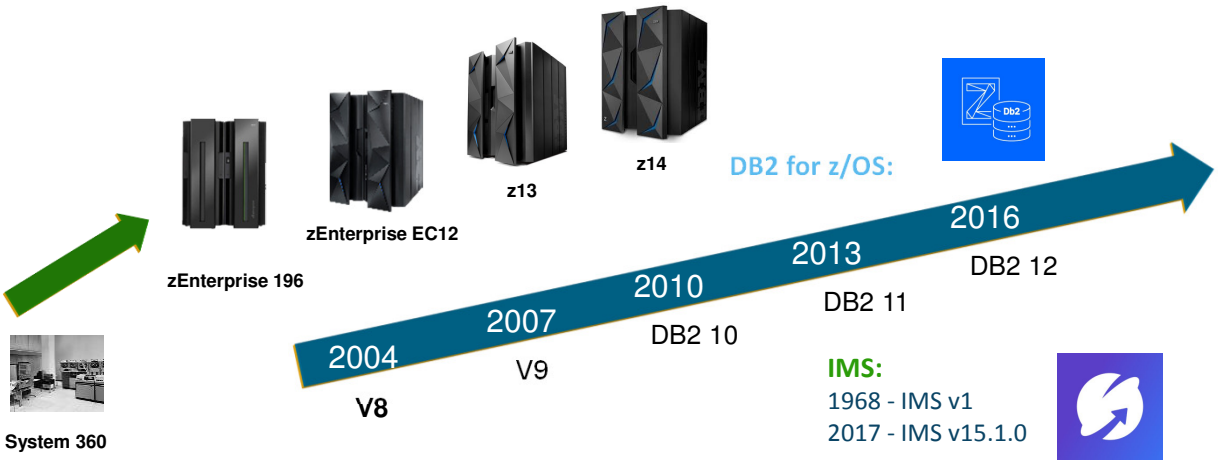
SUMMARY

Data Processing in Historical Perspective

IBM System 360	1964
IBM IMS	1968
VSAM	1970s
ADABAS	1971
SMF/RMF system data	1980s
DB2 for z/OS	1983
DB2 UDB LUW	1993

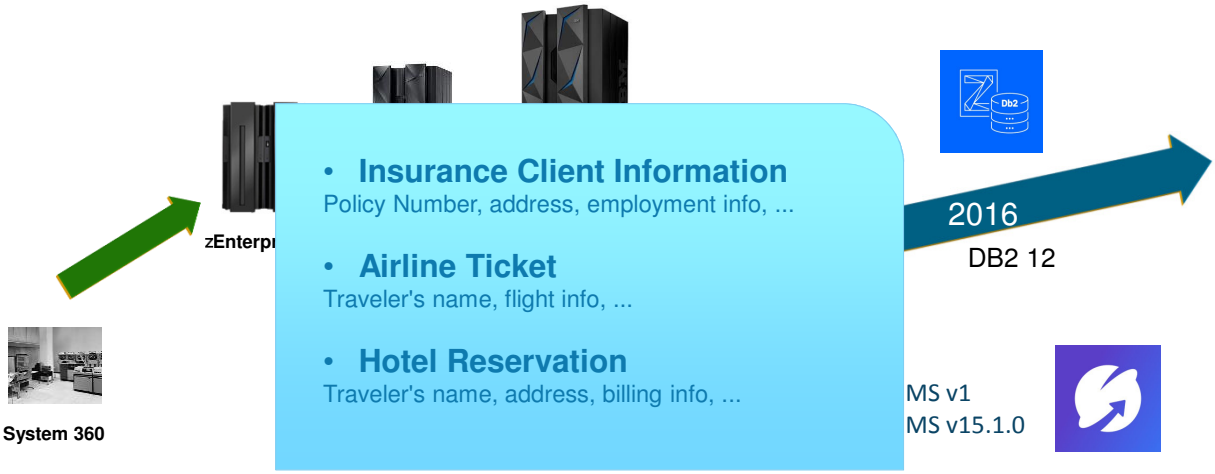
Increasing volume of valuable enterprise data

Hardware and Software Modernization

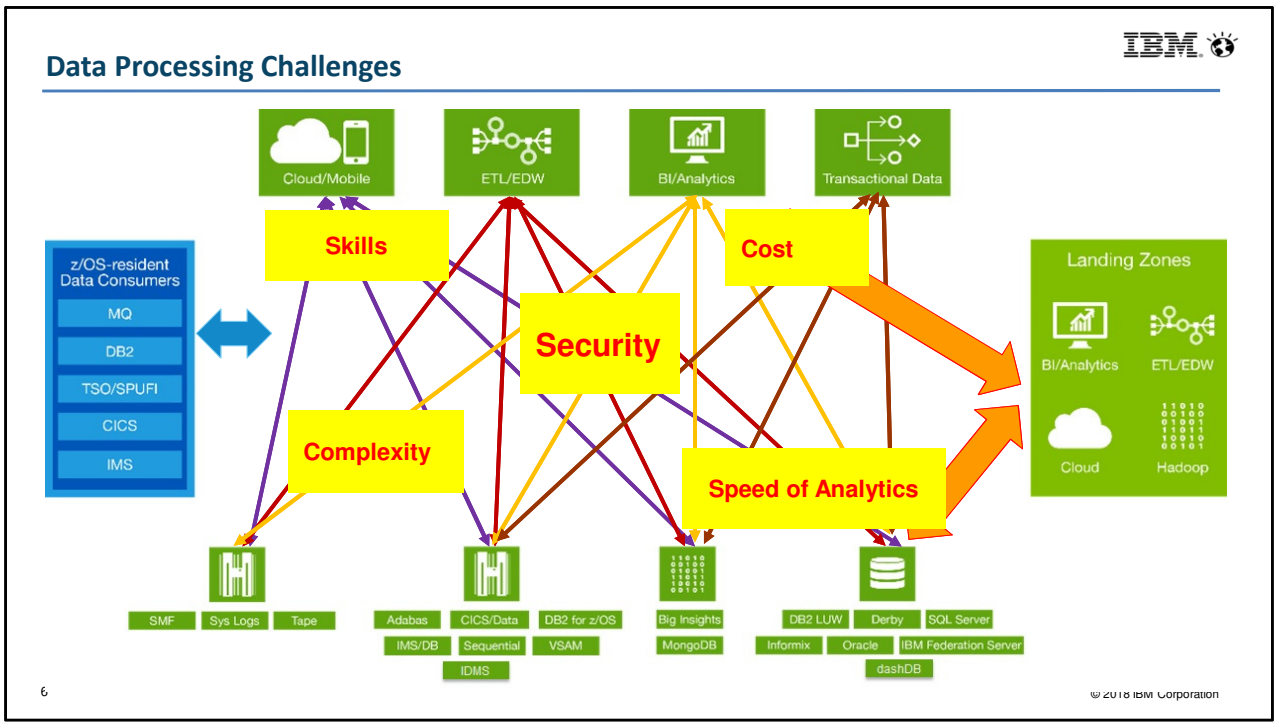


Increasing computing power and functionality

Data First, Data Gravity



The nature of data remains, format and usage grows with business needs



Challenge details:

- ***Data access complexity***

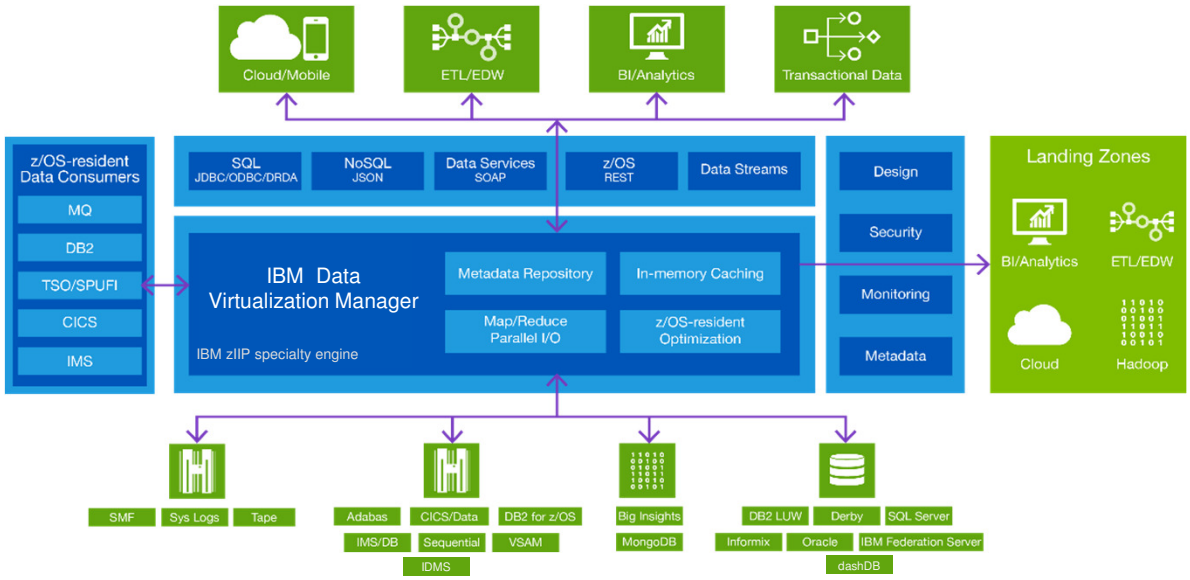
  heterogeneous data sources, numerous data connectors
- ***Speed of access***

  for real-time analytics and mobile consumers, data movement delays
- ***Skills needed to support data access***

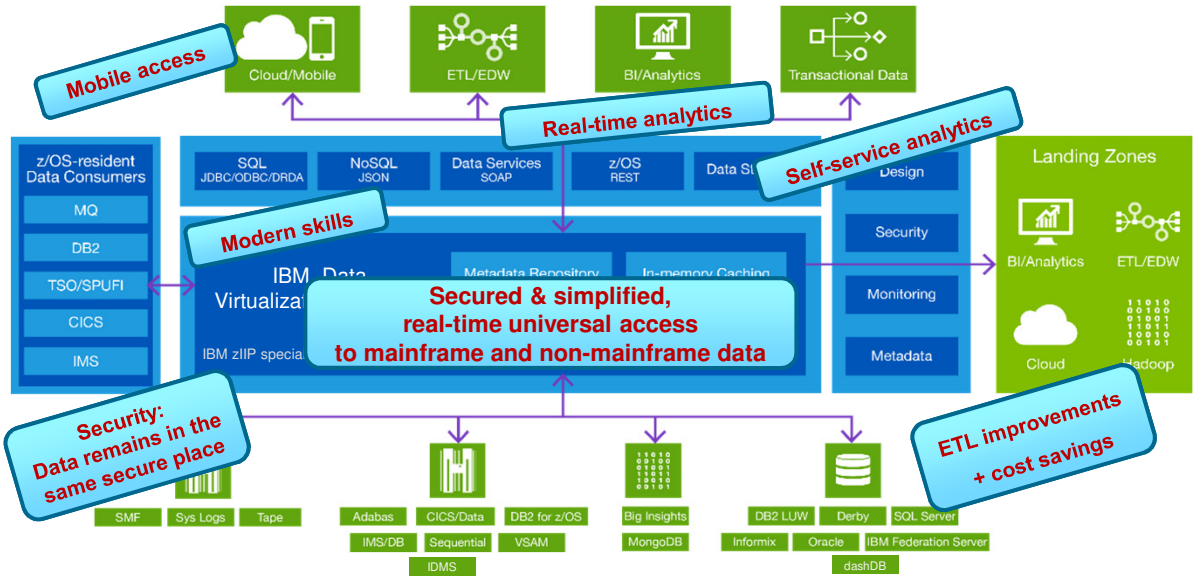
  for application developers, database administrators, ...

  Cost - Gartner study estimates that in Mainframe environments, 30% of MIPS are consumed by data movement.

IBM Data Virtualization Manager Solution



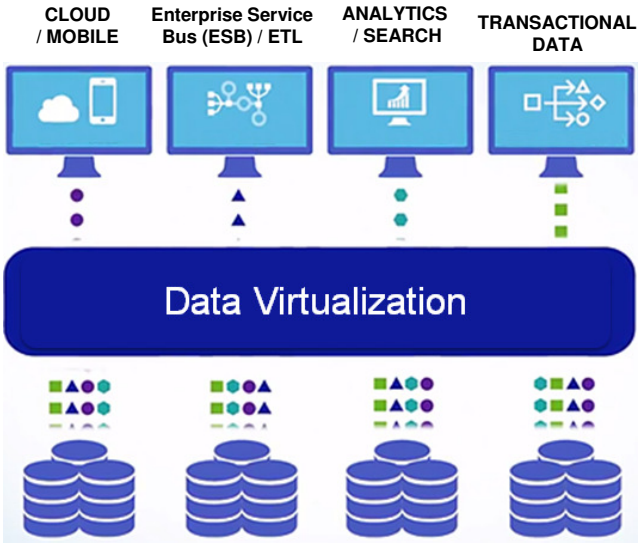
Needs and Challenges addressed by Data Virtualization Manager



Data Virtualization Concept

IBM

Enabling data structures that were designed independently, to be leveraged together, in **real time**, and **without data movement**



9

Data Virtualization: Enabling data structures that were designed independently to be leveraged together, in real time, and without data movement

Data Virtualization: a virtualized data services layer that integrates data from heterogeneous data sources and content in real-time, near-real time, or batch as needed to support a wide range of applications and processes.
Forrester Research – March 2015 - Noel Yuhanna

https://arch.simplicable.com/arch/new/when-to-use-ESB-vs-ETL
Generally, Enterprise Service Bus (ESB) is used for real-time messaging and ETL is used for high volume batch Extract, Transform, and Load.

ETL vs ESB
http://www.intricity.com/videos/etl-vs-esb-2/
These strange analogies have a very similar parallel in the data world. When were trying to move large quantities of data, often the tool of choice is an ETL

tool, which stands for extract, transform, and load. However, when we are communicating between individual application processes we often use an Enterprise Service Bus or ESB.

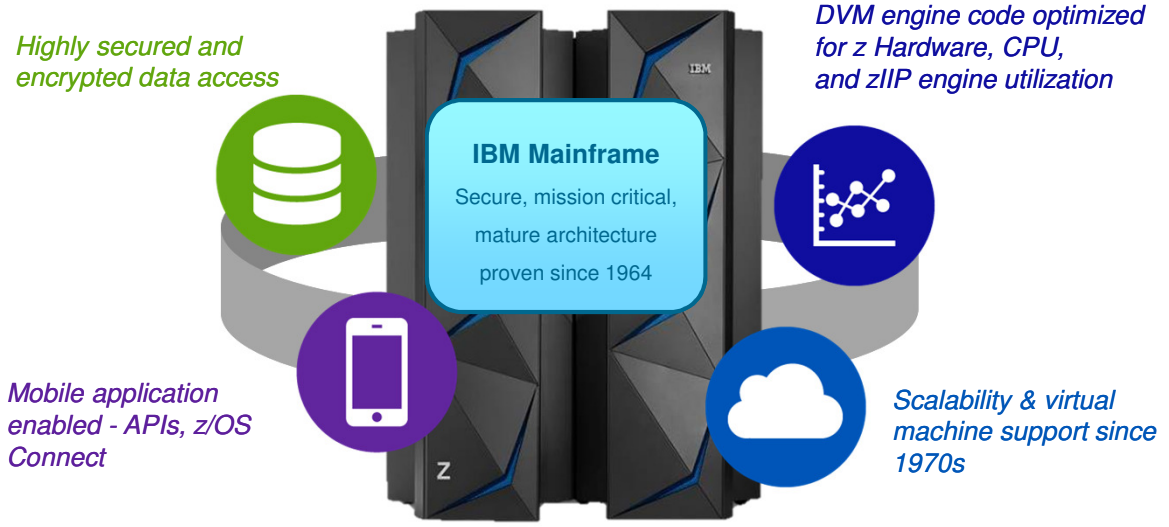
http://www-03.ibm.com/software/products/en/integration-bus-advanced
Enterprise Service Bus (ESB) provides connectivity and universal data transformation in heterogeneous IT environments.

Data Virtualization Solutions - Vendors

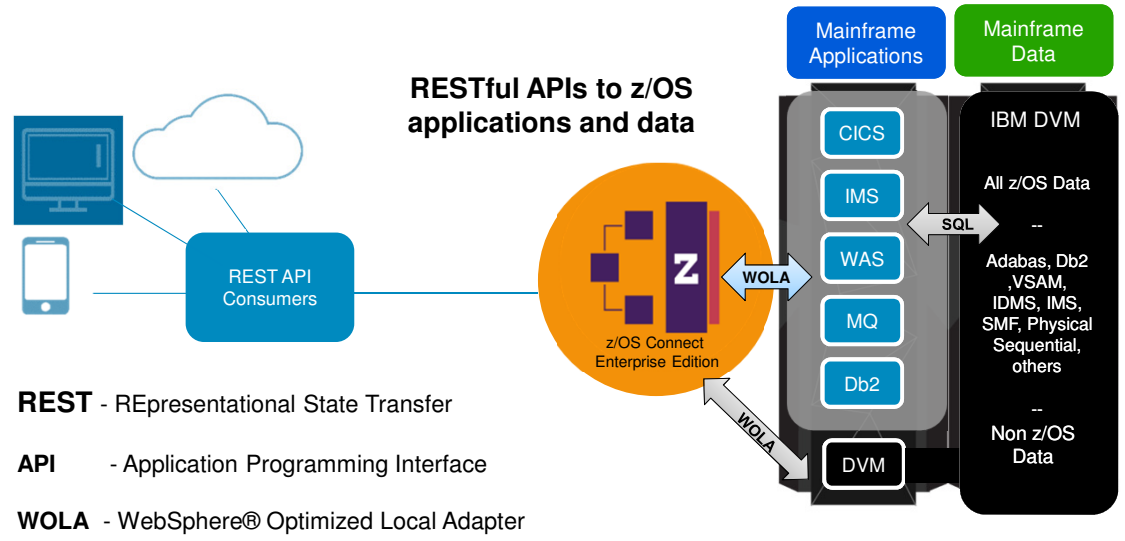
- Denodo
- Informatica
- SAP
- IBM
- Amazon Web Services (AWS)
- Cisco
- Red Hat - JBoss
- Oracle
- VMWare
- TIBCO

Source: https://www.gartner.com/reviews/market/data-virtualization

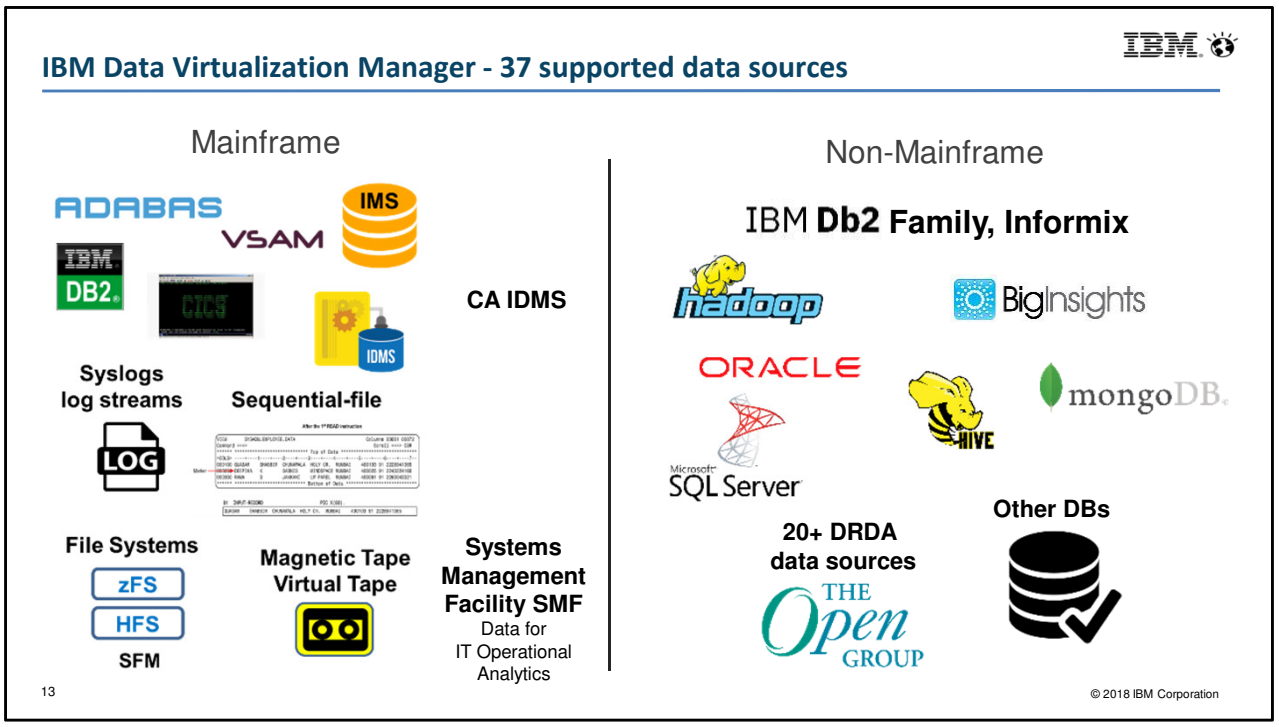
Mainframe Value Proposition for Data Virtualization



IBM z/OS Connect EE and DVM - APIs to mainframe data



RESTful web services provide interoperability between computer systems on the Internet.



Partial list of supported data sources

Data support Data source

Mainframe relational/non-relational databases and file structures

- IBM® DB2
- IBM® Information Management System

(IMS/DB)

- Native VSAM files
- Sequential files
- Software AG Adabas

Mainframe applications and screens - IBM® CICS®

- IDMS
- IBM® Information Management System
- Software AG Natural
- VSAM via IBM® CICS® Transaction Server
- IBM Rational Asset Analyzer (RAA)

Distributed data stores running on Linux, UNIX, and Windows platforms

- IBM® BigInsights Hadoop
- IBM® DB2
- Apache Derby
- IBM® Informix

- Oracle
- Microsoft SQL Server
- IBM Application Discovery and Delivery Intelligence (ADDI)

Production Customer: Self-service Analytics for Investment Advice

Simplification
Real-time direct access to data instead of FTP'ing or replicating data to a myriad of locations

Scalable data access
Ability to scale to more than 5 billion ADABAS SQL calls per month

Enhanced Developer Productivity
Focus on adding new web functionality without having to change the data source.

Solution Components

Software:
- **IBM® Data Virtualization Manager for z/OS**

14

Source:
Turning Data Into a Competitive Advantage With Data Virtualization on IBM Z
https://www.youtube.com/watch?v=XuYUnAMmyPU

The company's electronic trading division develops sophisticated proprietary front-end applications as offerings to clients using both .NET and J2EE development environments. They needed to enable flexible and frictionless access to and from mainframe business logic and data, with reduced costs and leverage mainframe assets to the fullest.

IBM DVM provides highly scalable, universal, standards-based SQL access to all three layers of the Adabas database - both the Natural presentation layer and business logic that utilizes Adabas for its database, and the Adabas database itself – all with significant reduction in mainframe TCO.

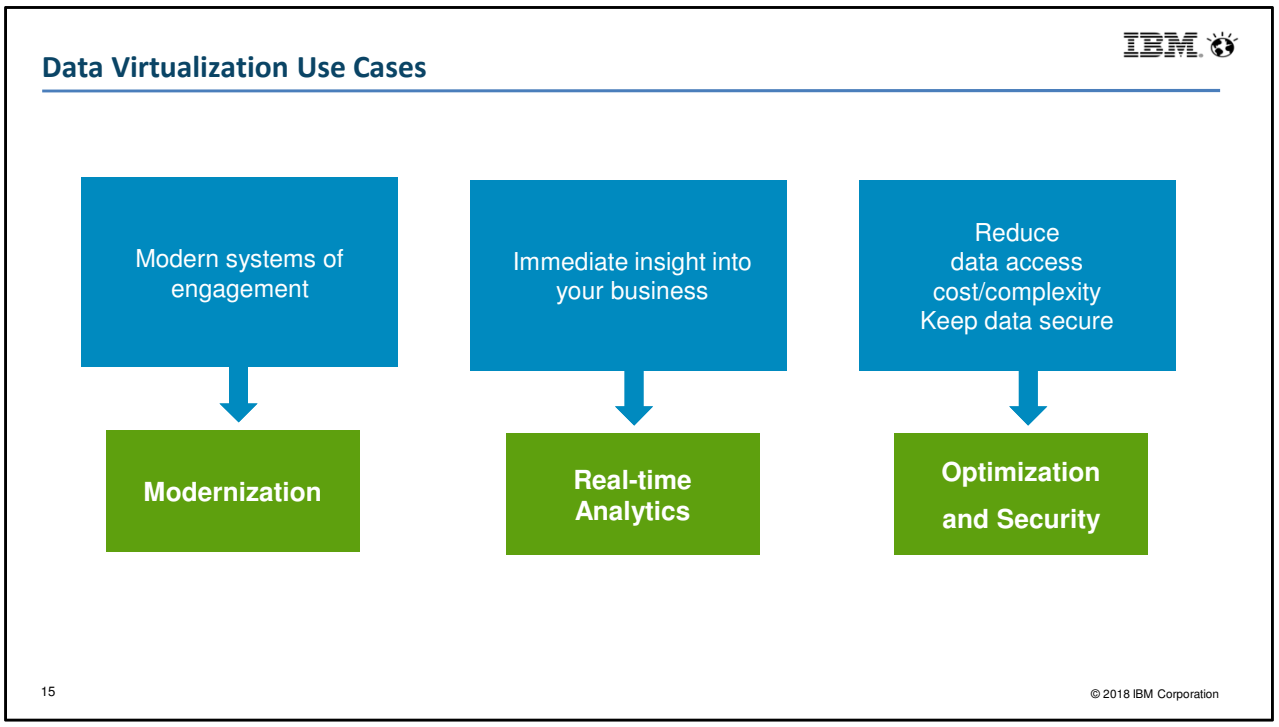
Productivity Gains – developers can focus on adding functionality and new front-end systems without changing the data source.

Simplification - real-time direct access to data instead of FTP'ing or replicating data to a myriad of locations. Scalability - ability to run 5 billion ADABAS SQL calls per month

The company's electronic trading division was developing sophisticated web portal

applications requiring fast, seamless transactional access to mainframe data within Natural applications and their Adabas database.

- IBM Data Virtualization Manager for z/OS provided highly scalable, facile method for transforming challenging Adabas data structures and older Natural applications into relational data that developers could readily use with existing skills and development, all with all with significant reduction in mainframe processing over native Adabas tools.



Modernization – agile, real-time data for:

APIs to accelerate delivery of new web portals, mobile apps and cloud initiatives

Enhanced business efficiency – faster, simpler internal/external integration

Analytics – real-time customer and operational data for:

Real-time business insight into customer needs, buying preferences

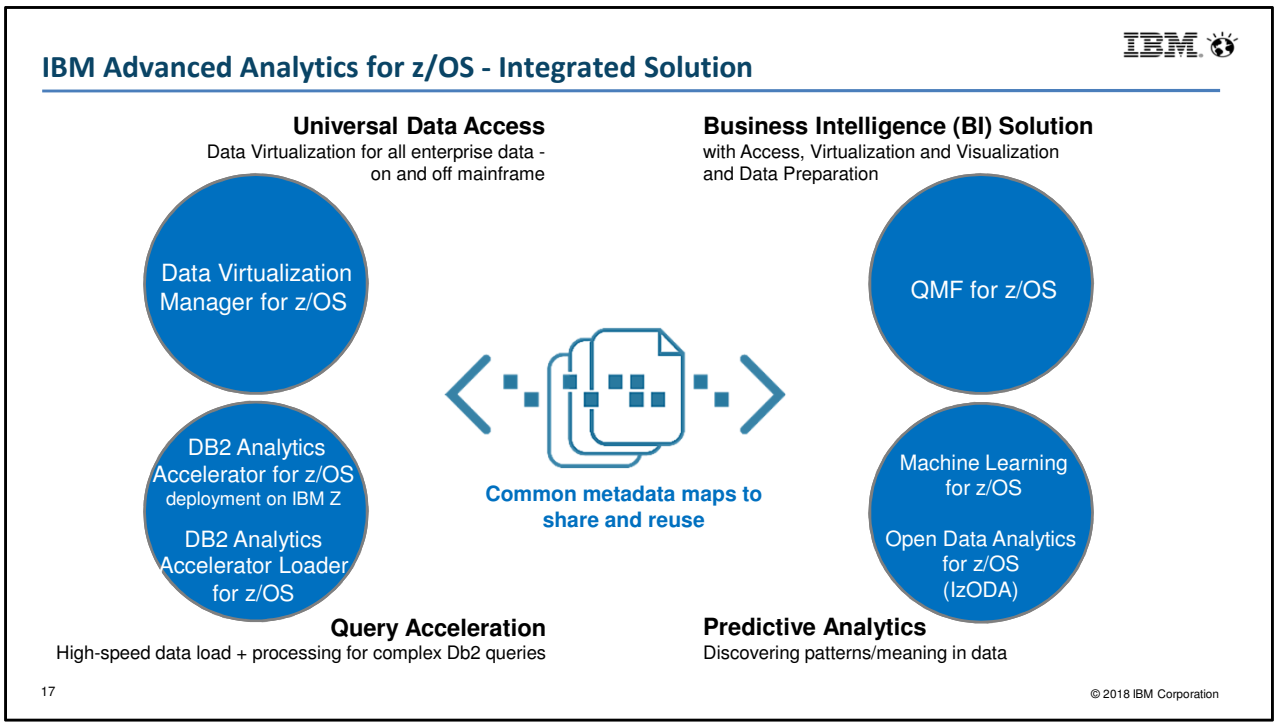
Real-time operational insight to reduce risk, fraud and improve data security

Optimization – enables data access for:

Reduce the time, cost and complexity of existing ETL processes

SUMMARY

IBM Data Virtualization Manager for z/OS



I started today saying the four cornerstones for Real Time Analytics have been set.

All four share Data Virtualization technology which means
They have common metadata maps that can be shared and re-used.

Each has its purpose and strengths (read from the slide DVM, QMF, IDAA, MLz)

In particular, DVM…

1. virtualization and APIs provide direct r/w access to differing data sources that we never designed to be together.
2. lets you use data in any format, to enrich (join) that data, requiring no movement or latency, all at high capacity
3. is integrated with z/OS Connect EE and API tooling to provide a single, simple, standard access to all z data.

SUMMARY

- ✓ 1. Data Processing in Historical Perspective
- ✓ 2. Hardware and Software Modernization
- ✓ 3. Data Processing Challenges and Needs
- ✓ 4. IBM Data Virtualization Manager (DVM) Solution
- ✓ 5. IBM Advanced Analytics for z/OS - Integrated Solution

Questions