The Simplest Way to Search All of Your Big Data Sources, including IMS ;)

Hélène Lyon

Distinguished Engineer & CTO, z Analytics & IMS for Europe

IBM Systems, Software Sales, Europe

IMS Technical Symposium 2015

Agenda

- **Big Data in an "Data Driven" economy**

- **Why start with z Systems**
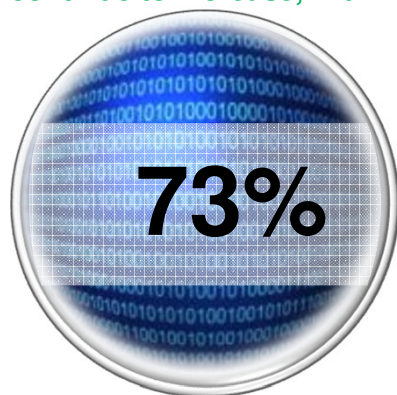
- **IMS strategies for big data**
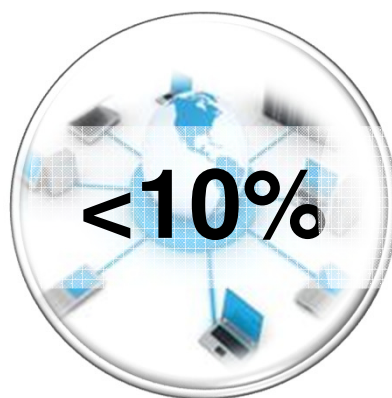
- **Summary**

Market Opportunities and Challenges

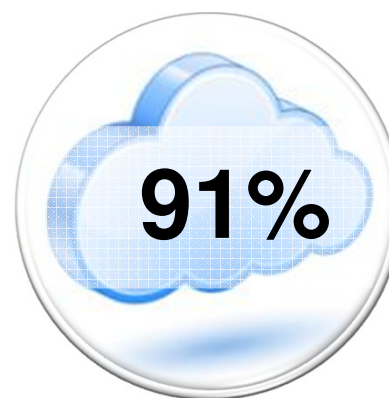## Investments in Big Data
continue to increase, with

73%

...of organizations having invested or planning to invest within 24 months.[1]
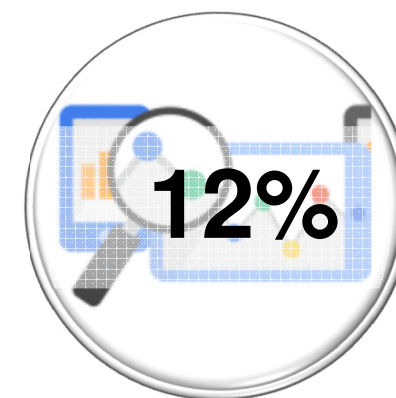
Infrastructure Ready

<10%

...organizations fully prepared for new digital trends like cloud, mobile, social and analytics (CAMS)[2]

Cloud Ready

91%

...software will be built for cloud delivery in 2014 [3]

Data Analyzed

12%

... of the data companies already have, leaving 88% of it on the cutting-room floor [4]

Sources
1. Gartner Inc, Research Note G00263798 - Survey Analysis: Big Data Investment Grows but Deployments Remain Scarce in 2014 . Nick Heudecker, Lisa Kart, Date Published 09/09/2014
2. IBM_Institute_for_Business_Value - Date published : 01/09/2014
3. IDC, "Directions 2014, Information is Everywhere", Robert Mahowald, Date published: 17/03/2014
4. Forrester Research, Inc., "The Forrester Wave(TM): Big Data Hadoop Solutions, Q1 2014" Mike Gualtieri and Noel Yuhanna,Date published: 27/02/2014

Why is it happening?

Cloud

private	public	SaaS

Data is…
- ✓ Leaving the Data Center
- ✓ Stored on shared drives
- ✓ Hosted by 3rd party
- ✓ Managed by 3rd party

Consumerization of IT

Mobile

BYOD	Apps	Social

Data is…
- ✓ Generated 24x7
- ✓ Used Everywhere
- ✓ Always Accessible
- ✓ On private devices

Everything is Everywhere

BigData

Hadoop	No-SQL	Files

Data is…
- ✓ Produced in high volumes
- ✓ Stored unstructured
- ✓ Analyzed faster/cheaper
- ✓ Monetized

Data Explosion

✓ There is more data
✓ Data is leaving the data center
✓ Data is consumed everywhere
✓ Data is worth more than ever before

4

New paradigms impacting Big Data & Analytics

Cloud

ZBs of transaction data

Social

Big Data

Mobile

Internet of Things

We've moved into a new era of computing …

Radical Flexibility

12 terabytes

of Tweets create daily

Extreme Scalability

5 million

trade events per second

Volume

Velocity

Variety

Veracity

100's

Of video feeds from surveillance cameras

Information from everywhere

Only **1 in 3**

Decision makers trust their information
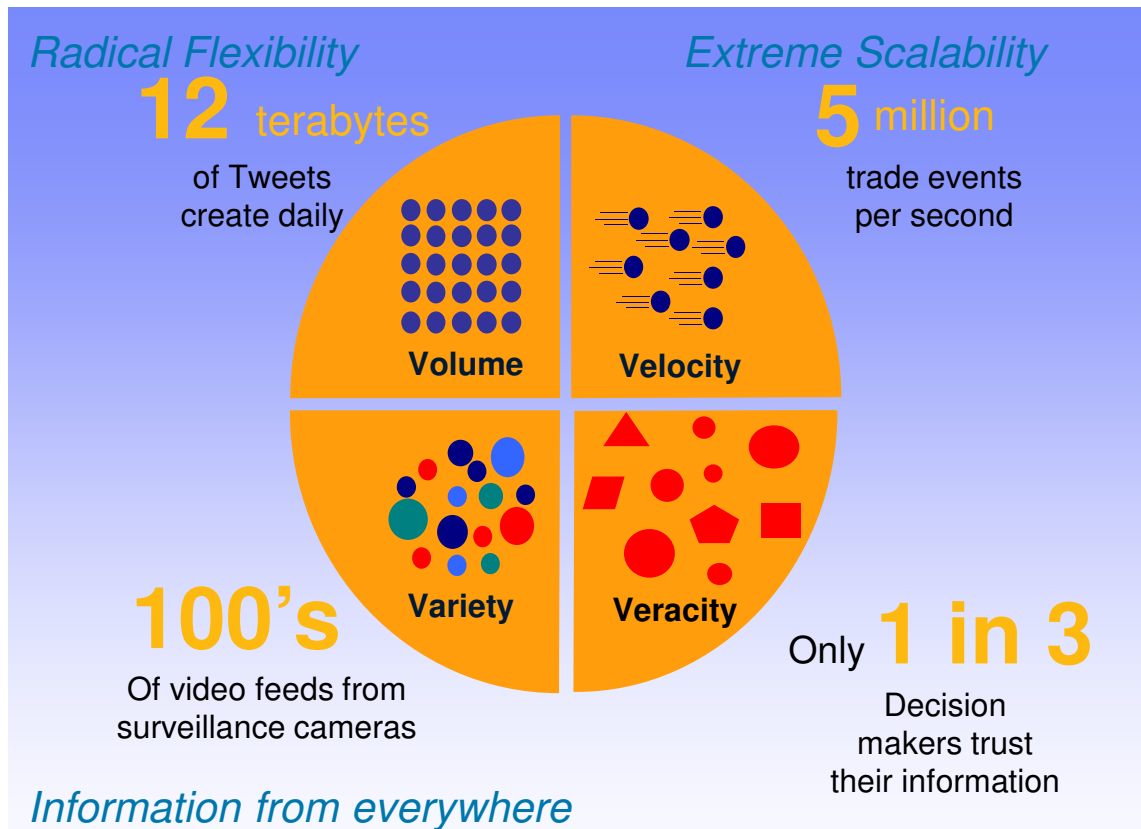
"We have for the first time an economy based on a key resource [Information] that is not only renewable, but self-generating.

Running out of it is not a problem, *but drowning in it is.*"

– John Naisbitt

And Analytics is becoming the Keystone of every organization …

- **Analytics derive insight from data**
 - To help optimize business performance
  - To build new innovative services
  - To fight against fraud
  - To make all customer interaction personal!
  - …

- **Analytics become Business Critical!**
  - Analytics services are tightly integrated with business critical applications and data
    - Often hosted in z/OS transaction and batch systems
    - Often relying on copies or aggregation of transaction and application data
  - Analytics is part of the flow of the business.
  - Decision processes have to be improved with new business insight derived from real time or near real time data.
  - Failure of these applications for any length of time can result in lost business or reputation.
  - Analytics solutions need to support a large concurrent user population with high volumes of requests.

- **Analytics are only as good as the underlying data foundation**
  - Data governance & Security & Performance
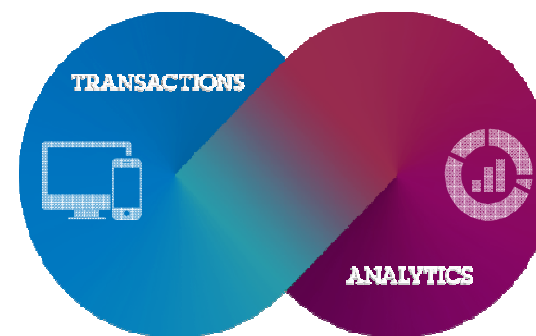
Agenda

- **Big Data in an "Data Driven" economy**

- **Why start with z Systems**

- **IMS strategies for big data**

- **Summary**

The Big Data Starting Point

Which types of big data does your organization currently analyze?

Source : *Gartner Inc, Research Note G00263798- Survey Analysis: Big Data Investment Grows but Deployments Remain Scarce in 2014 . Nick Heudecker, Lisa Kart, Date Published 09/092014*

	2013 Survey (n = 218)	2014 Survey (n = 120)
Transactions	70	79
Log data	60	58
Geospatial/location data	26	34
Social media profile data	34	29
Free-form text	28	26
Emails/documents	33	26
Social media chat/interaction data	34	25
Sensor/machine-generated data	23	22
Images	18	18
Audio	7	12
Video	9	10
Other	15	12

Percentage of Respondents

■ 2013 Survey (n = 218)   ■ 2014 Survey (n = 120)
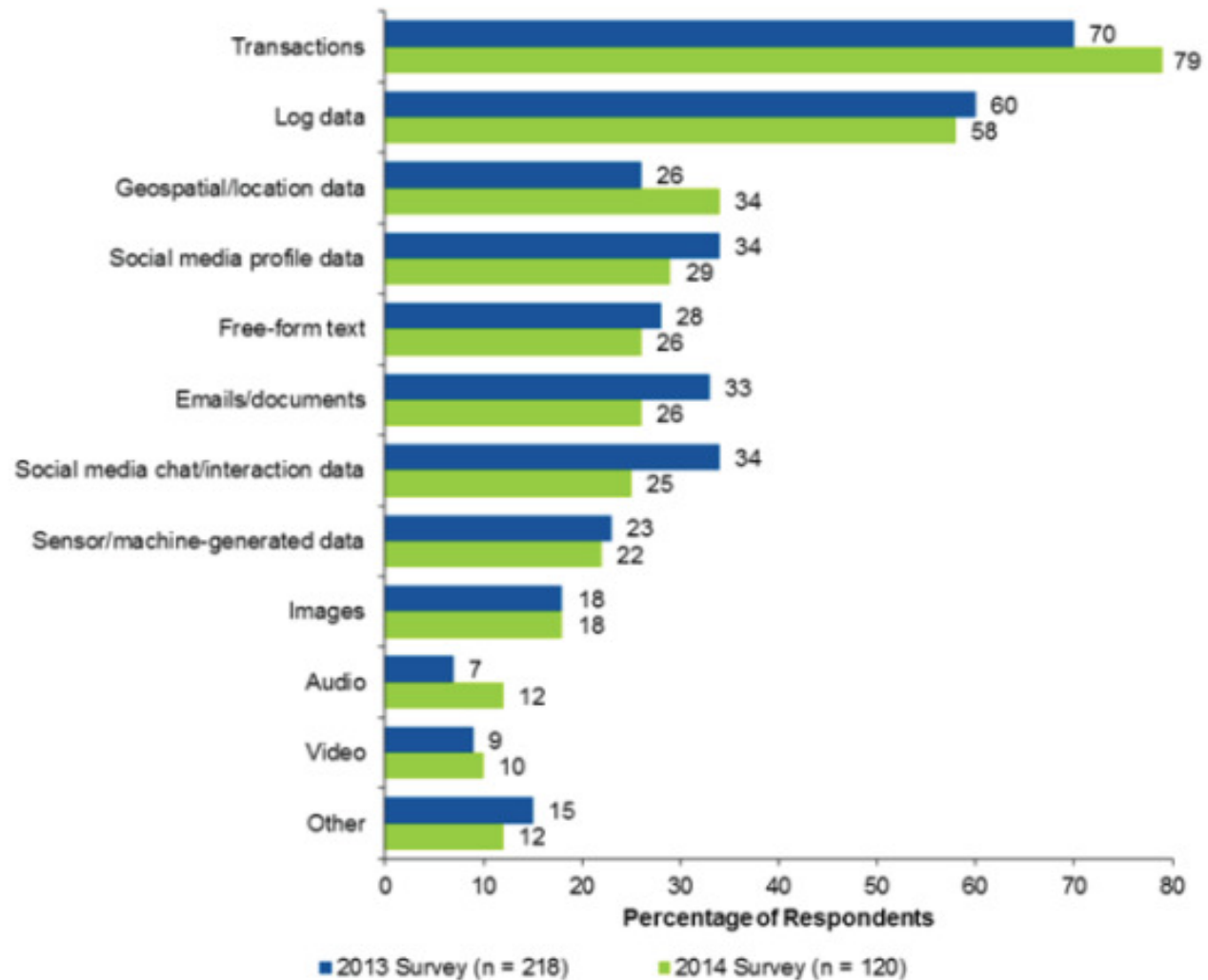
Base: Respondents invested in big data technology
Multiple responses allowed

Source: Gartner (September 2014)

The z Systems advantage

- **80% stored or originates** on IBM z Systems

- **2/3 of business transactions** for US retail banks run directly on mainframes

- Businesses that run on z Systems
  - **92 of the top 100** worldwide banks
  - **23 of the top 25** U.S. retailers
  - **10 of the top 10** global life / health insurance providers
  - **23 out of the 25** largest airlines

- **EAL5+ encryption** and cryptographic hardware to secure data in motion and at rest

- **Run over 8000 virtual Linux images** -Virtualization of services for cloud implementations

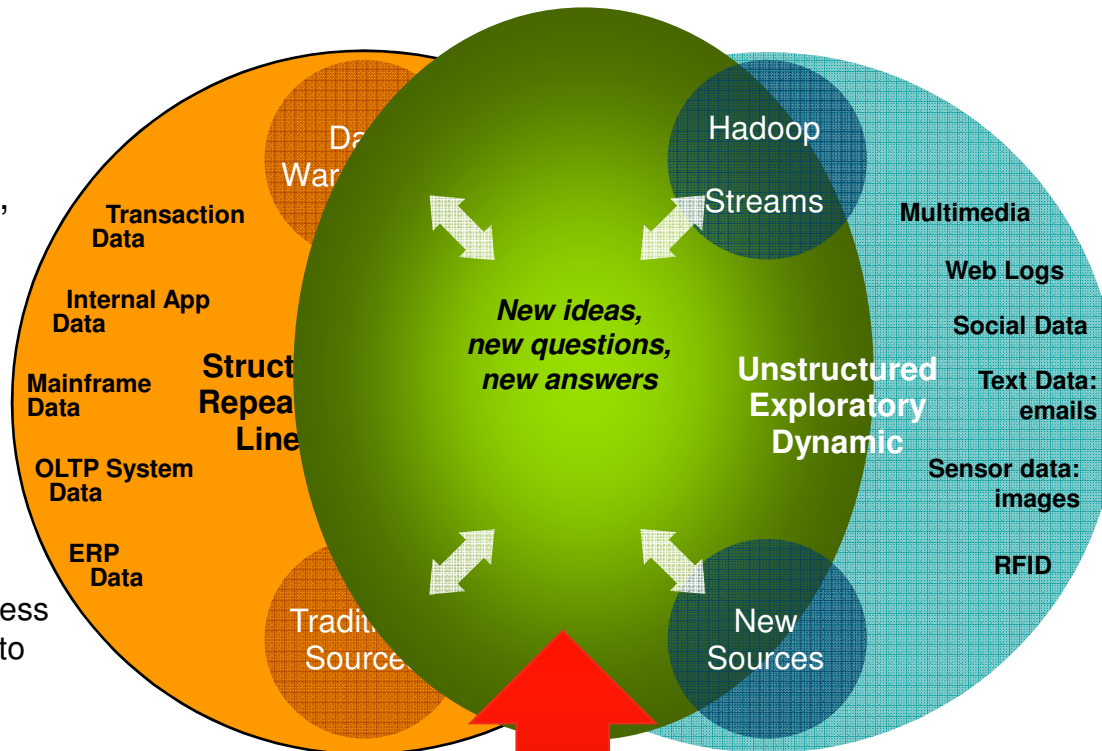- **5 minutes** per year downtime of an application running on z Systems

Imagine the possibility of leveraging all of your data assets

Traditional Approach
Structured, analytical, logical
Data: Rich, historical, private, structured
Customers, history, Transactions

The "Circle of Trust"
Data warehouse & business analytics moving closer to this data

New Approach
Creative, holistic thought, intuition

Data: Intimate, unstructured.
Social, mobile, GPS, web, photos, video, email, logs

Transaction Data

Internal App Data

Mainframe Data

OLTP System Data

ERP Data

Da... War...

Struct...
Repea...
Line...

Tradi...
Source...

New ideas, new questions, new answers

Hadoop

Streams

Multimedia

Web Logs

Social Data

Text Data: emails

Unstructured
Exploratory
Dynamic

Sensor data: images

RFID

New Sources

The real benefit is derived from integration of new data sources with traditional corporate data
- How can you query across both realms?
- How can you preserve security and lower TCO?
- How can you avoid costs and risks of offloading?

…to deliver Improved Business Outcomes

1. **Enrich your information base**
   with Big Data Exploration

2. **Improve customer interaction**
   with Enhanced 360º View of the Customer

3. **Optimize operations**
   with Operations Analysis

4. **Gain IT efficiency and scale**
   with Data Warehouse Augmentation

5. **Prevent crime**
   with Security and Intelligence Extension

12

Unfortunately for most of our clients, their data lifecycle is too fragmented to gain advantage from that data

- **Client key concerns:**
  - Cannot deliver real-time analytics
  - Inadequate performance
  - Governance model
  - Data latency
  - Data completeness
    - Not all in one source
    - Lack access to fine-grained data
    - Lack "customer intent" e.g. cancelled transactions
  - Multiple platforms, many security boundaries, many points of failure,
  - Challenging recovery scenarios

- **Multi-day workshop captured the complexity of the current architecture**

- **The picture does not show all the steps before the data gets to the off-platform warehouse**
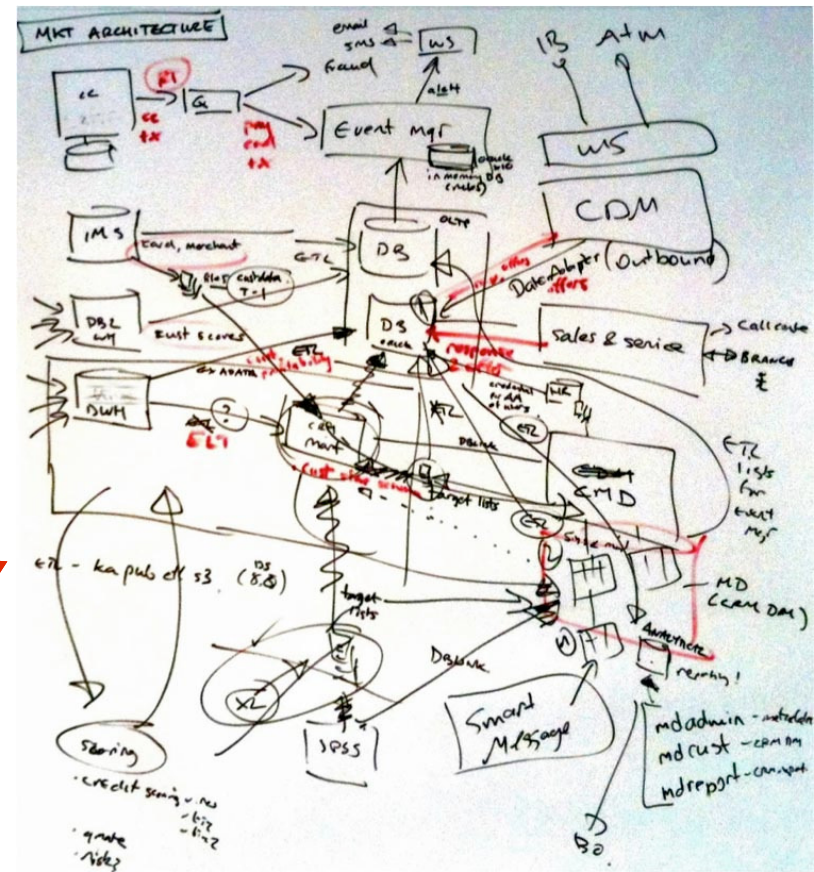
Agenda

- **Big Data in an "Data Driven" economy**

- **Why start with z Systems**

- **IMS strategies for big data**

- **Summary**

Analytics on IMS Databases & Systems

Use case	Solution

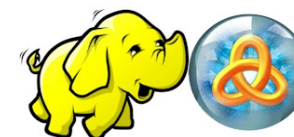- BI, dashboarding, reporting of IMS data

  > - QMF
  > - Cognos 10.2 BI

- Merge HDFS data with trusted OLTP
- IT analytics (IMS log data)

  > - IBM InfoSphere BigInsights

- Bring analytics to the data

  > - IBM DB2 Analytics Accelerator

- Visualize entire big data landscape

  > - IBM Watson Explorer

Analytics in IMS Applications

Use case

- Predictive analytics

- Decision management

- Fraud detection

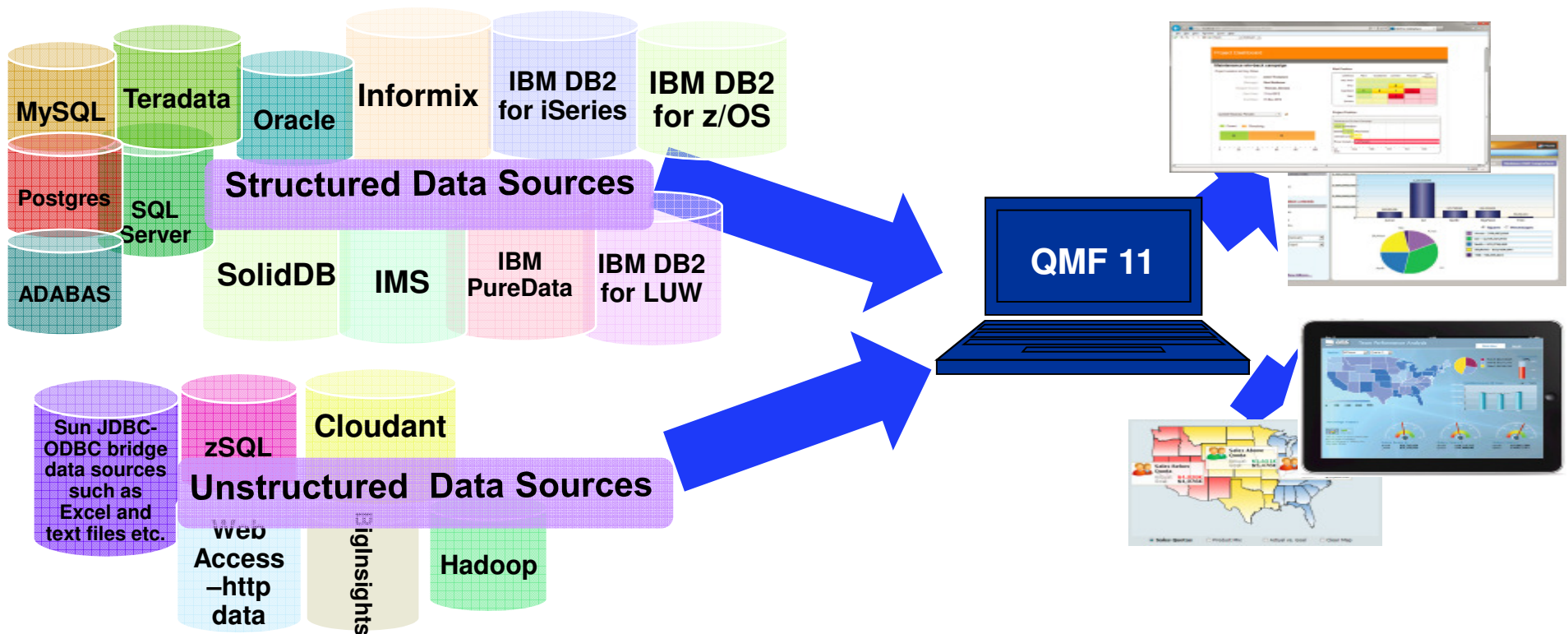## Solution

- SPSS
- Operational Decision Manager

SPSS

See Session C14 - HLyon

QMF for z/OS offers fast, simple connection to a broad spectrum of data sources

See Session A14 - SMink

MySQL

Teradata

Oracle

Informix

IBM DB2 for iSeries

IBM DB2 for z/OS

Postgres

SQL Server

ADABAS

Structured Data Sources

SolidDB

IMS

IBM PureData

IBM DB2 for LUW

Sun JDBC-ODBC bridge data sources such as Excel and text files etc.

zSQL

Cloudant

Unstructured Data Sources

Web Access –http data

BigInsights

Hadoop

QMF 11

QMF for z/OS 11 and IMS solution highlights

- **Maximize benefits from IMS environment**
  - Allow users to graphically construct ad-hoc IMS queries
  - Create reports and dashboards that draw directly from IMS data
  - Roll out web-based graphical content that blends IMS data with relational, and unstructured data sources
  - Accessible from all current interfaces,web, workstation, tablets
  - IMS support is available with QMF for z/OS. QMF Classic Edition does not support IMS as a data source

- **zIIP usage model**
  - QMF JDBC workloads are zIIP-eligible
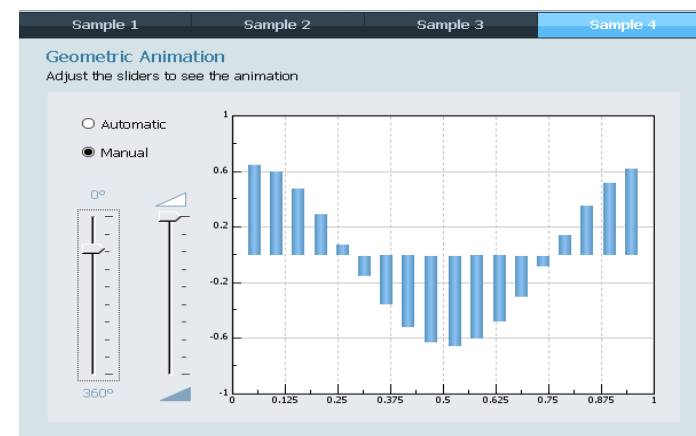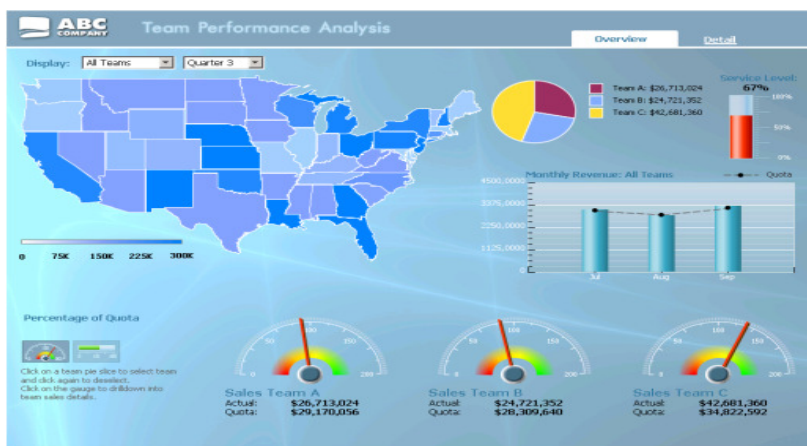
Cognos 10.2 BI with IMS Data

- **Certified against IMS 12 using IMS Open Database technology**
  - Universal JDBC driver

- **Real-time analytics**

Get Started Today!
- developerWorks article available here

Enhancing IMS analytics on Linux on z with Big Data

- **Much of the world's operational data resides on z/OS**

- **Unstructured data sources are growing fast**

- **There is a need to merge this data with trusted OLTP data from z/OS data sources**

- **IMS provides the connectors and the DB capability to allow BigInsights to easily and efficiently access the IMS data source**

JDBC

Analysis

Discovery

HDFS

20
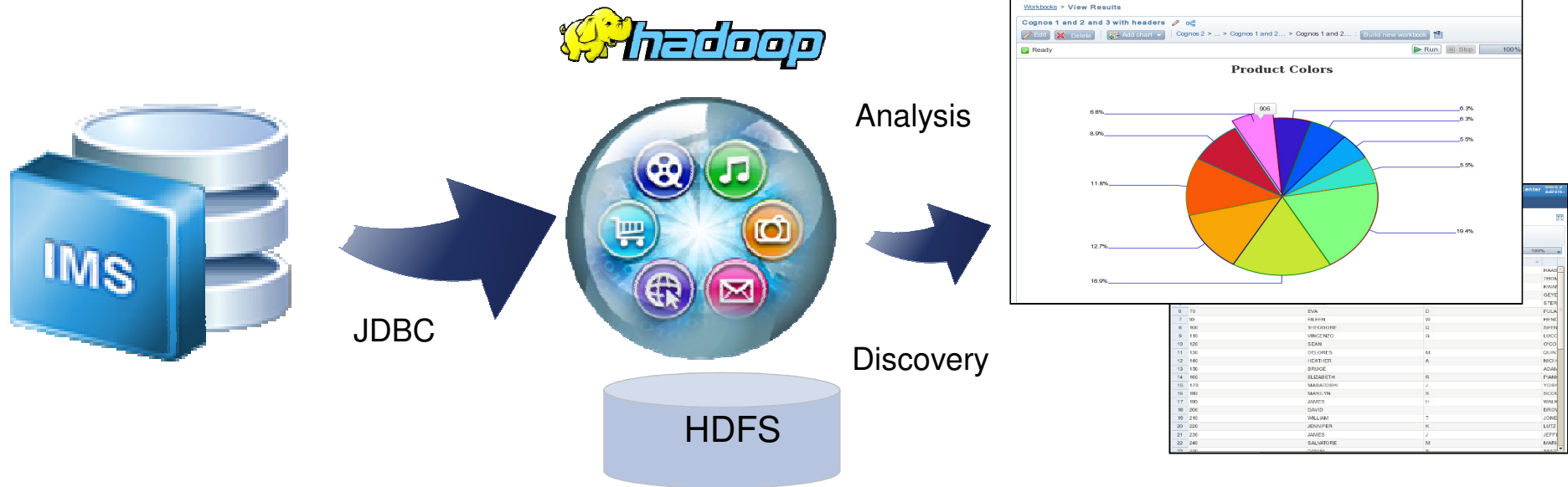
IBM InfoSphere BigInsights on System z
IBM InfoSphere System z Connector for Hadoop

- **Leverage the power of Hadoop on the mainframe**

- **Drag-and-drop extracts from mainframe sources**

- **Protect sensitive data**

- **Faster application delivery**

- **Seamless interoperability**
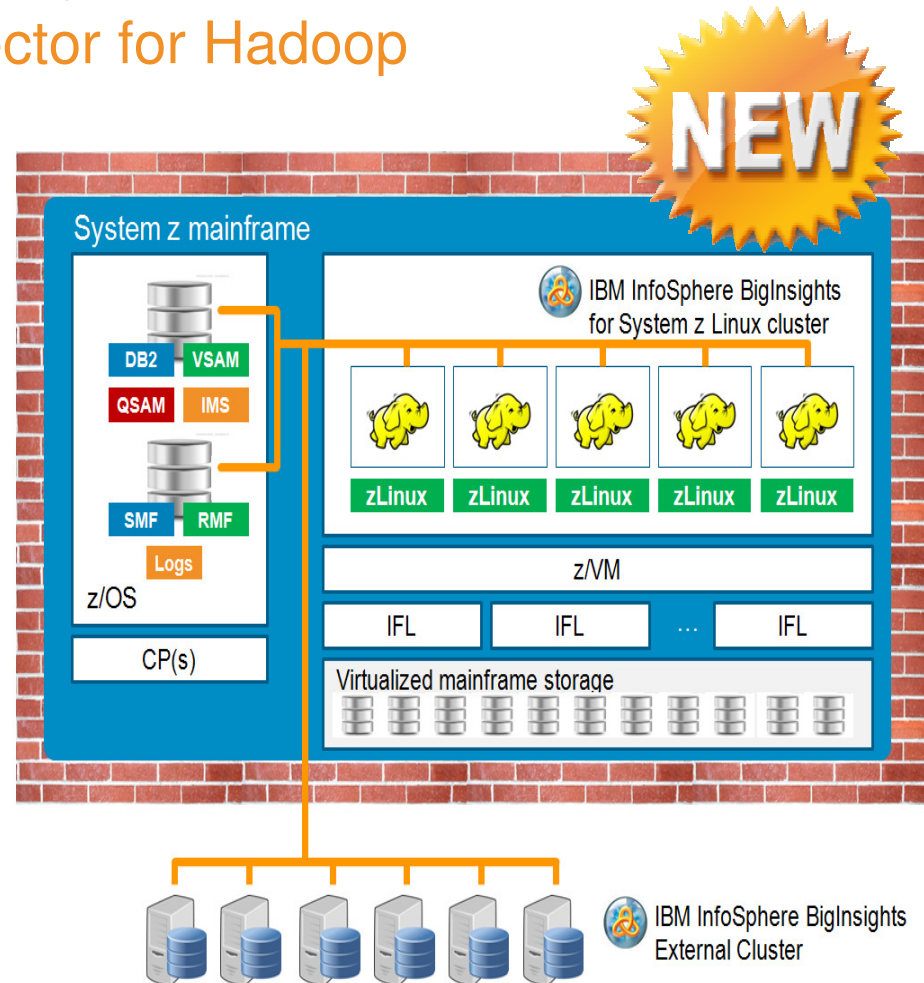
All "z" deployment for maximum security

IBM InfoSphere Data Replication Family
Expand z/OS data integration with changed data *ONLY* feeds

- **Non-intrusive Changed Data Capture and Delivery**
  - Log-based capture minimizes impact on source application environment
  - Dramatically reduces volume of data … only the changes not the entire source
  - Continuous or Periodic delivery with configurable switch of HDFS target
  - One model for z/OS and distributed data (DB2 on all platforms, IMS, Oracle, …)
  - No need to take sources "off line"
  - Fully recoverable
  - Native HDFS Apply

- **NOT for system/application logs and other sequential sources …**
  - These are always full file transfers!

z System

z/OS

DB2

VSAM

IMS

Logs

Data logs

InfoSphere Data Replication

Deliver only the Changes

Linux for System z

BigInsights

MapReduce

Hbase, Hive

z/VM

CP(s)   IFL   IFL   …   IFL
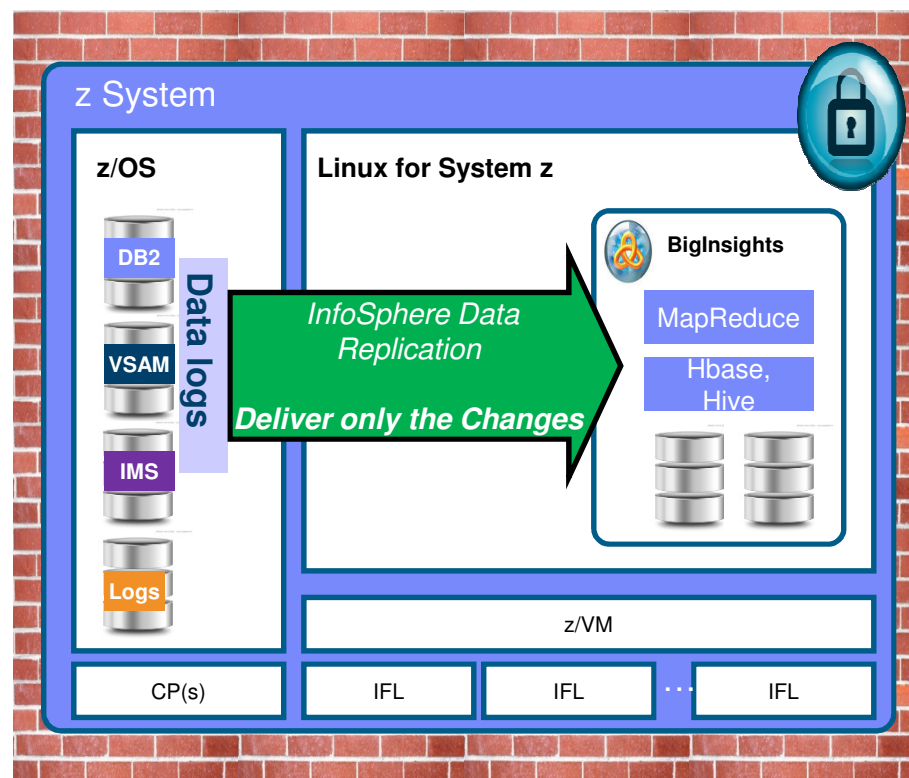
IT Analytics for IMS

- Log ingestion and analysis
- Value: allow correlation of an entire ecosystem of application servers with IMS to provide deep insight, filtering, analytics, as well as faceted search capabilities

Custom Applications				Shrink Wrap Solutions	
Health Care	Networking	Insurance	Telco	"x2020"	"Unity"

IBM Big Data Platform

⚙ MDA Accelerator

Tools Client Specific Customizations, Visualization tools ("zInsights")

Domain Specific

Telco	Financial services	Retail	Healthcare

Generic

Parsers and Extractors
(applications, services, servers and devices)

Federated Discovery, Pattern Discovery, Search, Visualization Tools for root cause analysis

Hadoop System	Stream Computing	Data Warehouse

Information Integration & Governance

IMS

IMS TM

IBM Big Data Platform

Visualization & Discovery	Application Development	Systems Management

⚙ Accelerators

Hadoop System	Stream Computing	Data Warehouse

Information Integration & Governance

IMS and IBM Accelerator for Machine Data Analytics

- **Consume log data produced by Transaction Analysis Workbench**

- **Index and link transactions together across products (IMS, DB2, MQ, CICS, WAS)**

- **Make large amounts of IMS transactional log data available to the suite of BigInsights tools.**

Transaction Analysis Workbench

BigInsights Platform

HDFS

MDA

Index, Extraction

SFTP

Search in Watson Explorer

BigSheets

Log Conversion to ASCII in CSV format
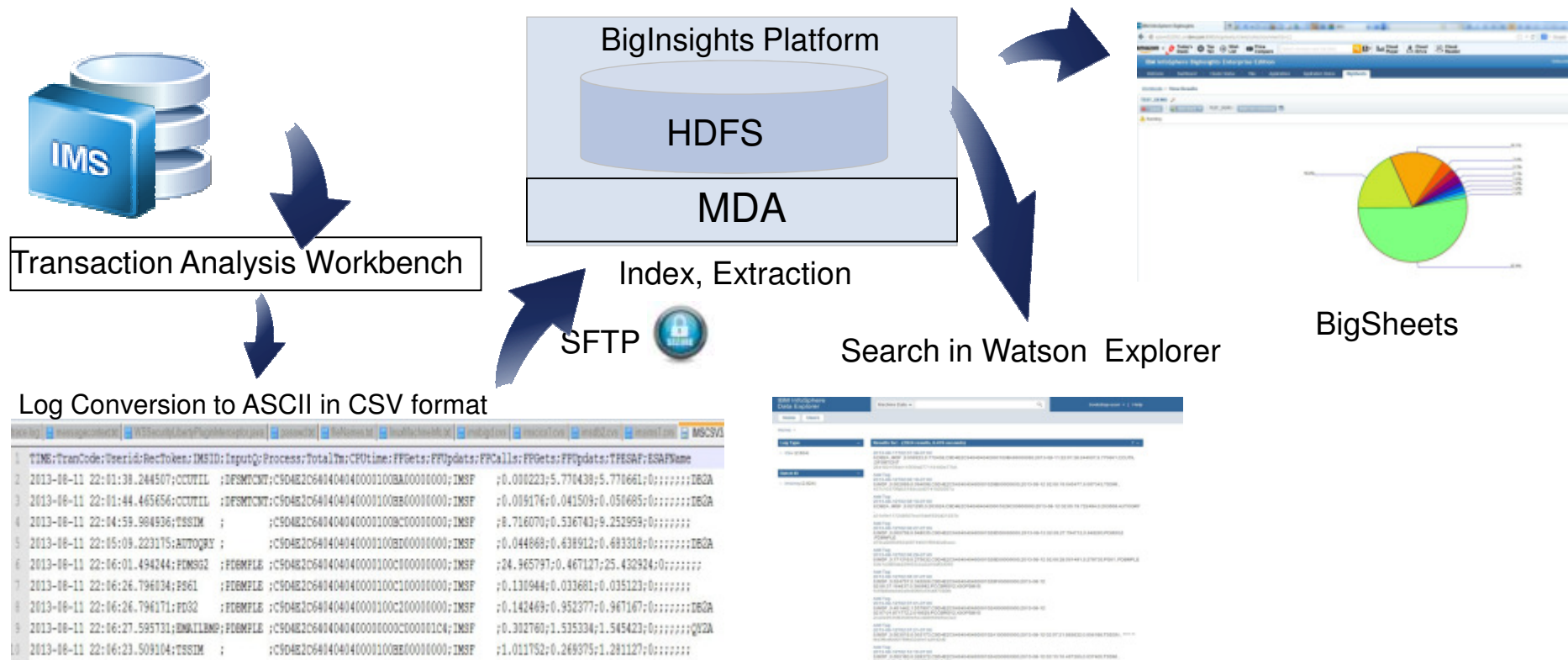
Watson Explorer - visualization and discovery across all data sources: "*Integration at the glass*"

Watson Explorer
Providing unified, real-time access and fusion of big data unlocks greater insight and ROI

Securely connect to and leverage data stored in DB2 for z/OS & IMS

Help prioritize your z System big data integration and analytics projects

Improve customer service & reduce call times

Increase productivity & leverage past work increasing speed to market

Analyze customer information & data to unlock true customer value

Identify areas of information risk & ensure data compliance

Create unified view of ALL information for real-time monitoring

IMS + Watson Explorer

- **Configuring the IMS source**
  - After deploying the IMS JDBC driver, create a new Database seed

- **Setting up the data transformation**
  - After creating a new seed, a converter needs to be configured using standard XPATH

Seed Component: **Database (Custom SQL)**

Host	ec01255.vmec.svl.ibm.com
Port	5,555
Username	omvsadm
Password	********
Database system	IMS
Database name	BMP355
SQL Statement	SELECT * FROM PCB01.HOSPITAL, WARD, PATIENT
Key Column	PATNAME

Advanced Configuration (5)

JDBC Connection String	jdbc:ims://ec01255.vmec.svl.ibm.com:5555/BMP355:dpsbOnCommit=true;
JDBC Class	com.ibm.ims.jdbc.IMSDriver

Converter Component: **Database seeds support**

Type-In	application/vxml-db
Type-Out	application/vxml-db
Fallback	(unset)
Output forking	(unset)
Name	

Original IMS hierarchy for hospital database

- **Hierarchy: HOSPITAL->WARD->PATNAME**

- **Goal: Get a patient centric view**
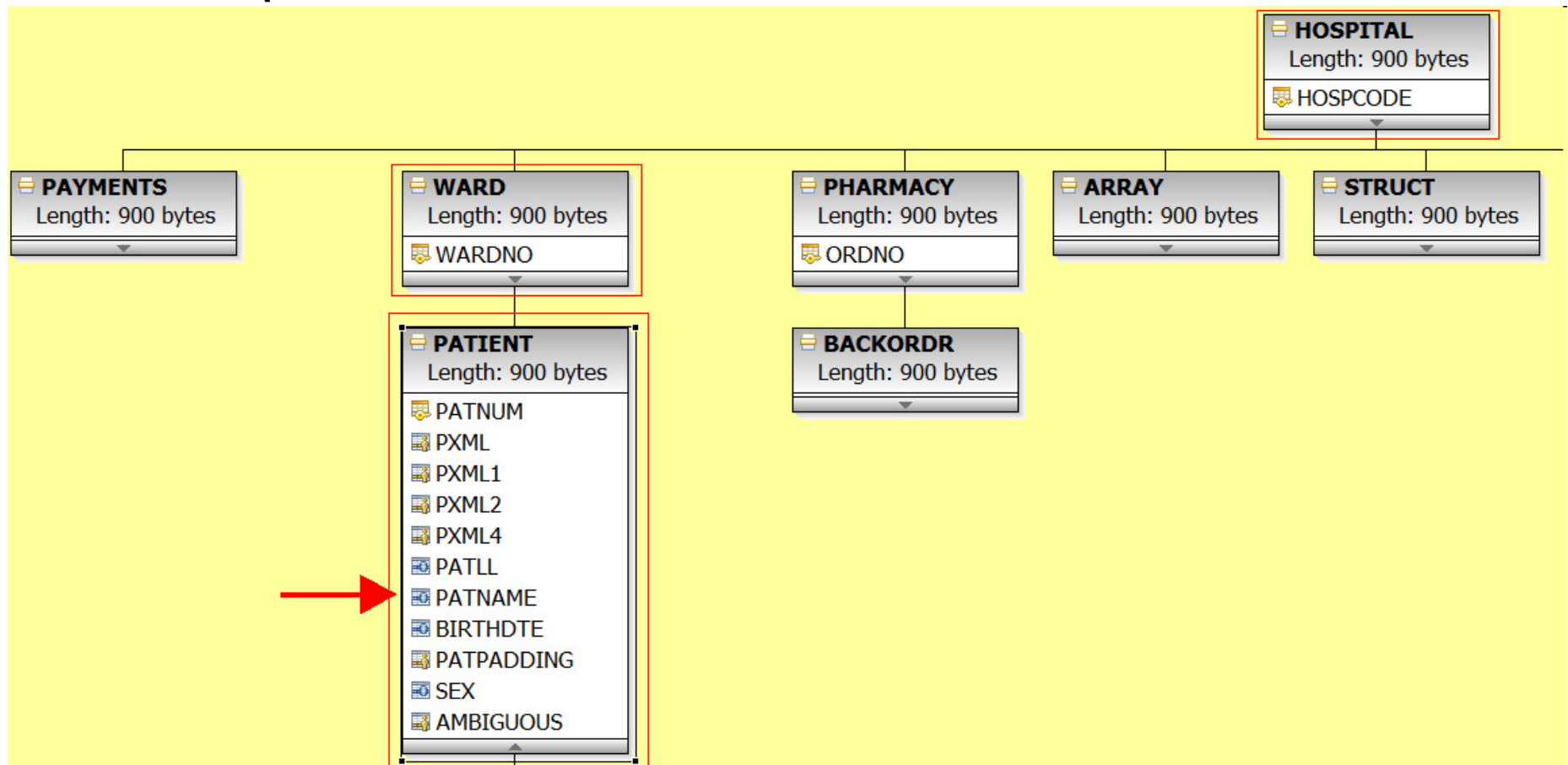
Why use Watson Explorer

- **Previously to change the schema so that the PATIENT information is at the top, a logical database needs to be created**

- **This requires a DBA to be involved and a time window when the new database resources can be brought online**

- **Watson Explorer allows indexes to be created dynamically and for better searching that is not restricted to IMS Segment Search Arguments (SSAs)**

IBM DB2 Analytics Accelerator

The hybrid computing platform on z Systems

> ➢ *Supports transaction processing and analytics workloads concurrently, efficiently and cost-effectively*

> ➢ *Delivers industry leading performance for mixed workloads*
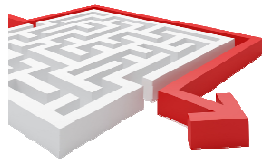
DB2 Analytics Accelerator and DB2 for z/OS

A self-managing, hybrid workload-optimized database management system that runs each query workload in the most efficient way, so that each query is executed in its optimal environment for greatest performance and cost efficiency

IDAA use cases with IMS data

Make better decisions faster

Large volume reporting of combined IMS and DB2 assets

Better understand your customers

Leverage full breadth of transactional data for analytics

Trust your data

Ensure consistency of data relationships between IMS and DB2

IBM Software Group
Technical White Paper

August 2014

IBM

Accelerate business insights with IMS transactional data

Introduction

IBM® System z® servers are home to 80 percent of the world's data,¹ a large part of which is Information Management System (IMS) transactional data. IMS applications and data continue to form the core of the world's critical online transaction processing (OLTP) workloads. IMS data on System z servers is widely viewed as the single version of the truth, from which data stores and warehouses are populated.

The System z platform offers high availability, scalability, security and performance for OLTP workloads. The data that resides on the System z platform and is targeted for analytics on the System z platform has those same qualities of service. Bringing analytics directly to the data, rather than moving data to the analytics, makes sense. It preserves these qualities of service to ensure timely, accurate, and secure insight, and reduces costs of moving data off the platform.

The insights that analytics provide must be delivered efficiently across an organization, with high quality and proper governance. Moving data off of a System z platform and across platforms is not only costly; it literally rips data out of its productive use and impedes the delivery of business insights. Where transactional data originates plays an increasingly important role in analytics. Some percentage of that 80 percent of data is not being analyzed at all due to the performance and cost challenges of moving data.

IBM

Get Started Today!

- Technical Whitepaper and "how-to" guide available here

30

© 2015 IBM Corporation

IDAA enablement steps

- **Unload IMS data and produce a DB2 UNLOAD equivalent dataset**
  - ETL platform can do this or home grown

- **Create DB2 table**

- **Add table to IDAA**
  - SYSPROC.ACCEL_ADD_TABLES

- **Initialize the table in IDAA**
  - SYSPROC.ACCEL_LOAD_TABLES

- **Enable the table for acceleration**
  - SYSPROC.ACCEL_SET_TABLES_ACCELERATION

- **Load IMS data into IDAA**
  - IDAA loader tool
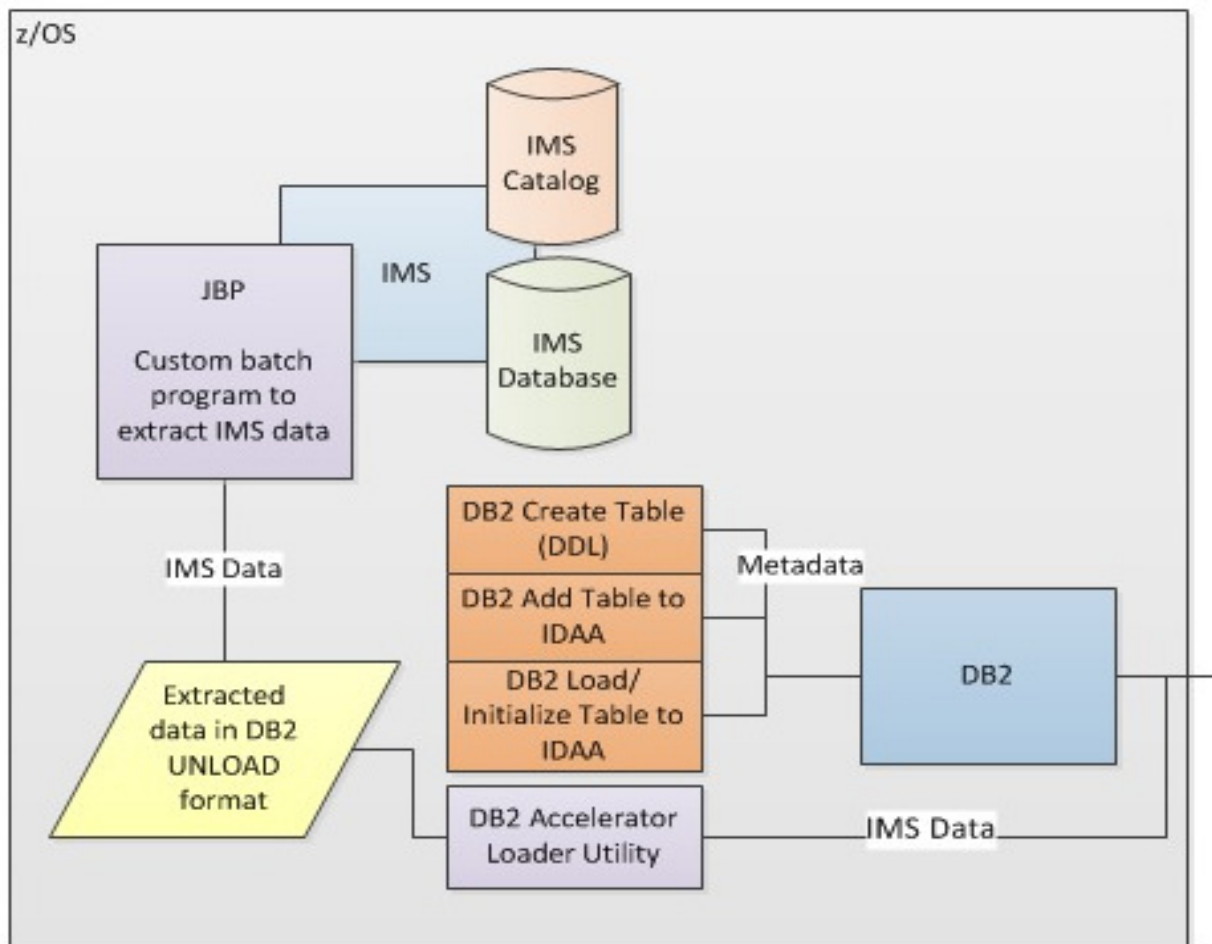
IDAA configuration with DataStage
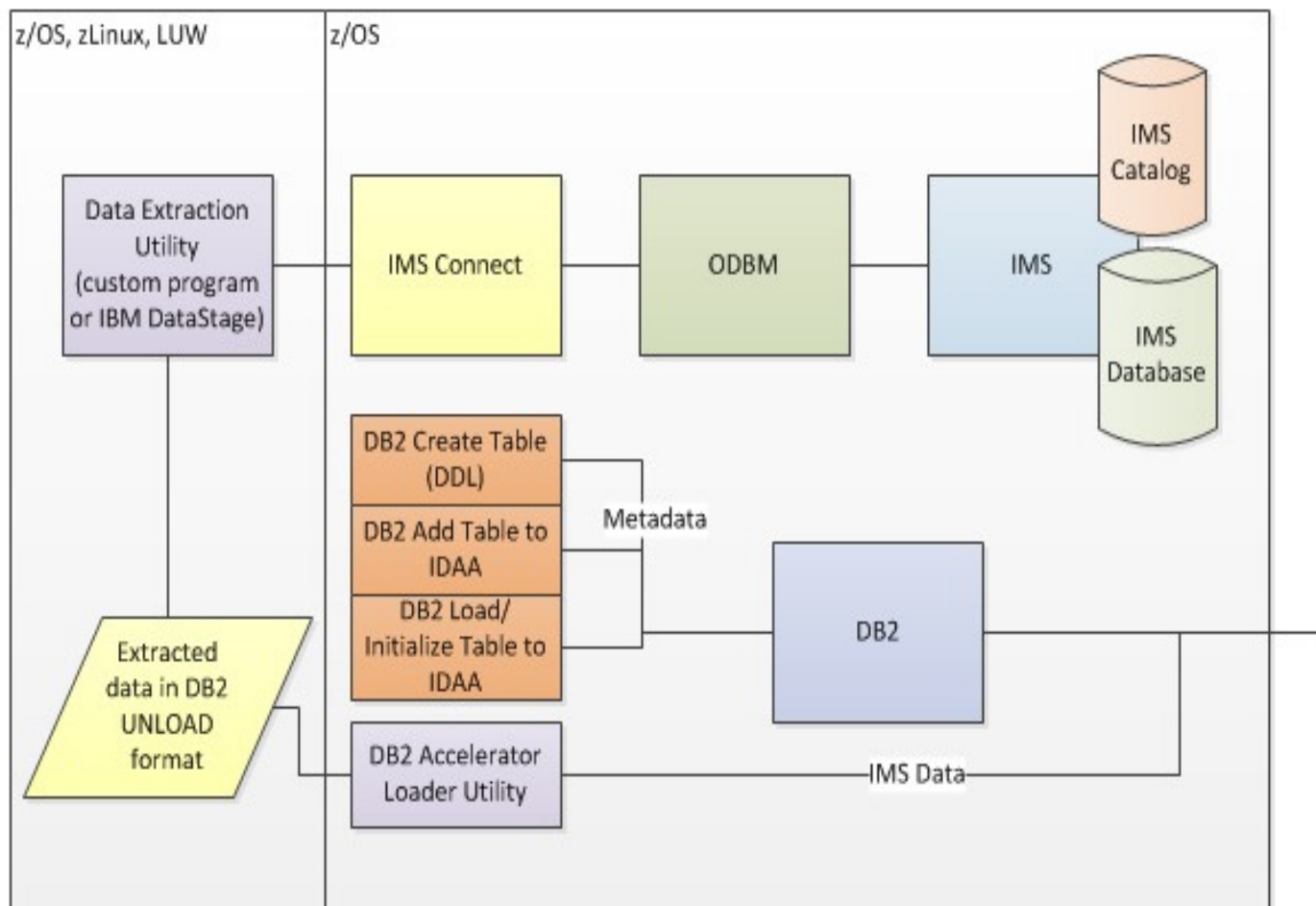
IDAA configuration with custom Java loader

Agenda

- **Big Data in an "Data Driven" economy**

- **Why start with z Systems**

- **IMS strategies for big data**

- **Summary**

Conclusion

- **Several ways to search All of Your Big Data Sources, including IMS!!**

- **For additional information including whitepapers and demos, please visit:**
    - IBM big data for z web site
    - IBM DeveloperWorks
    - For Education
        - Free online education at bigdatauniversity.com
        - 145,000+ registered students
    - Further developments
        - IBM Events and Conferences
        - World of DB2 for z/OS

- **Wanting to experiment on a big data integration project ?**
    - Partner with IBM Silicon Valley Laboratory
        - Richard Tran - richtran@us.ibm.com

- **Develop your own big data strategy**
    - Contact Helene Lyon – helene.lyon@fr.ibm.com