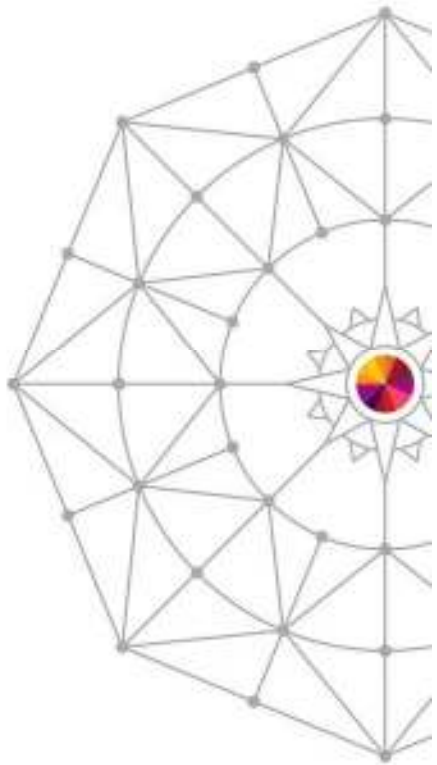
Big Data Strategies with IMS

#16103

Richard Tran

IMS Development

richtran@us.ibm.com



Insert Custom Session QR if Desired.

#SHAREorg

SHARE is an independent volunteer-run information technology association that provides **education**, professional **networking** and industry **influence**.

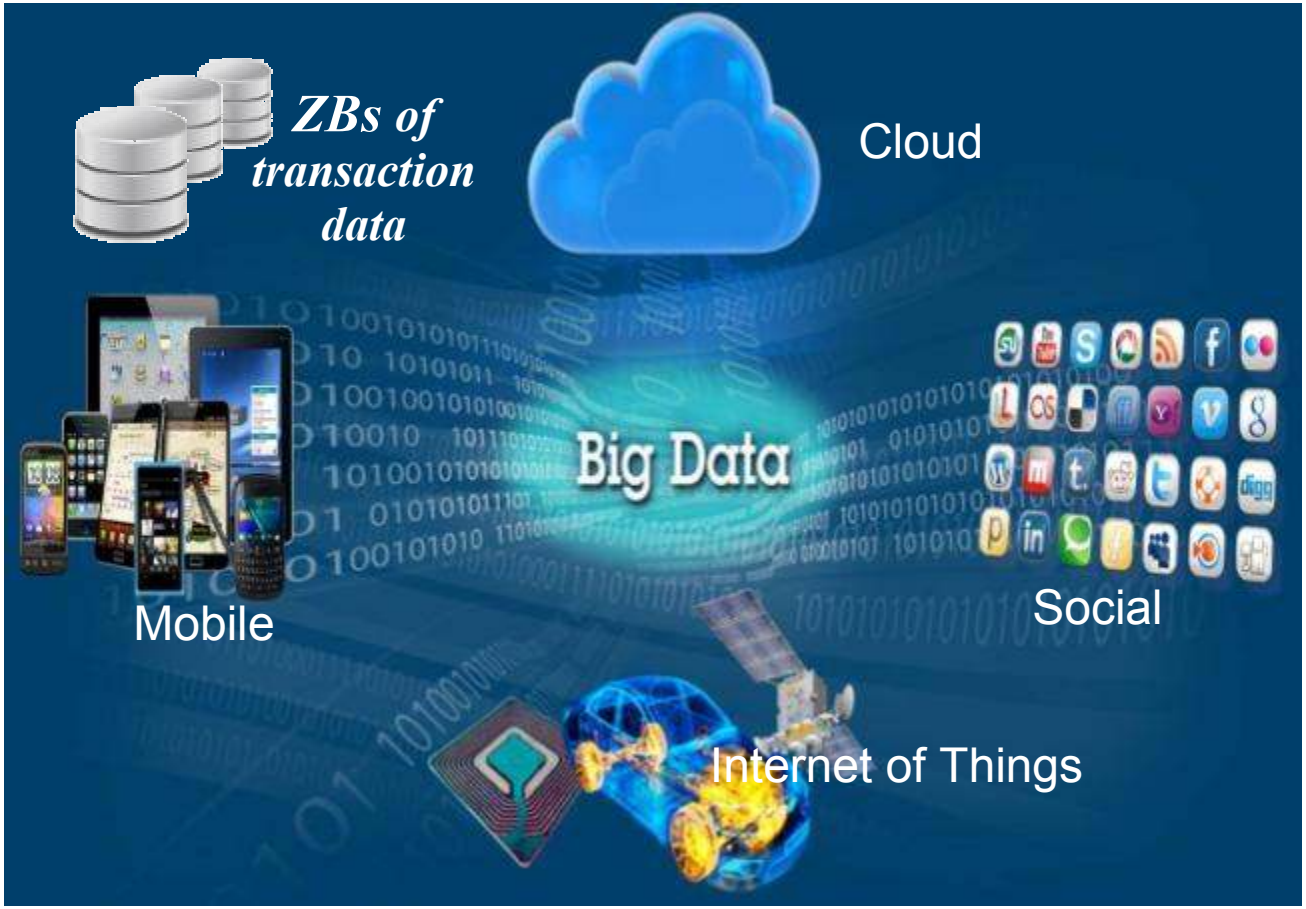
Copyright (c) 2014 by SHARE Inc. Except where otherwise noted, this work is licensed under http://creativecommons.org/licenses/by-nc-sa/3.0/

SHARE in Pittsburgh 2014

Agenda

- Big Data in an Information Driven economy
- Why start with System z
- IMS strategies for big data
- Summary / Call to action

On a Smarter Planet, Unprecedented Changes are Occurring

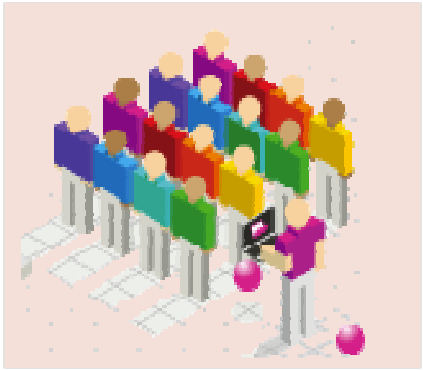


- Business models under constant pressure
- Customers are more demanding and connected
- Great relationships trump great products

Complete your session evaluations online at www.SHARE.org/Pittsburgh-Eval

And leaders are responding by…

Providing a Great Experience

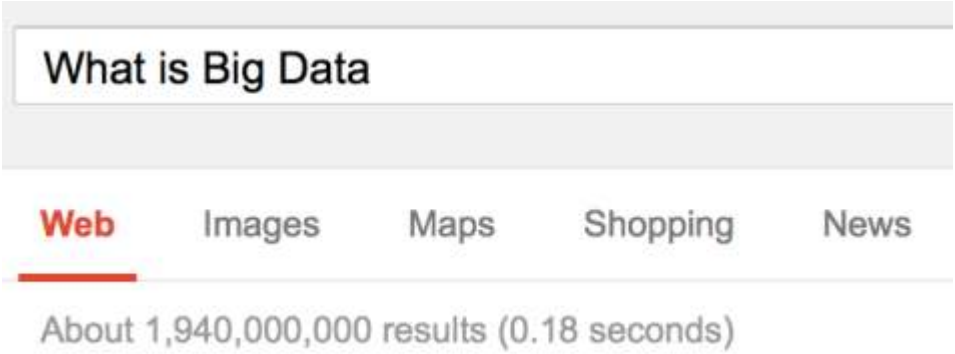


Offering Value In Every Interaction

Innovating Across the Ecosystem

But what is Big Data?

- ***Google can give you nearly 2 Billion options***
- ***Vendors have even more definitions***



Here is how Gartner defines Big Data

- *Big data is high-volume, high-velocity and high-variety information assets that demand* ***cost-effective, innovative information processing for enhanced insight and decision making****.*

Gartner research note "Survey Analysis - Big Data Adoption in 2013 Shows Substance Behind the Hype" Sept 12 2013
Analyst(s): Lisa Kart, Nick Heudecker, Frank Buytendijk

We've moved into a new era of computing - *V⁴*

Radical Flexibility

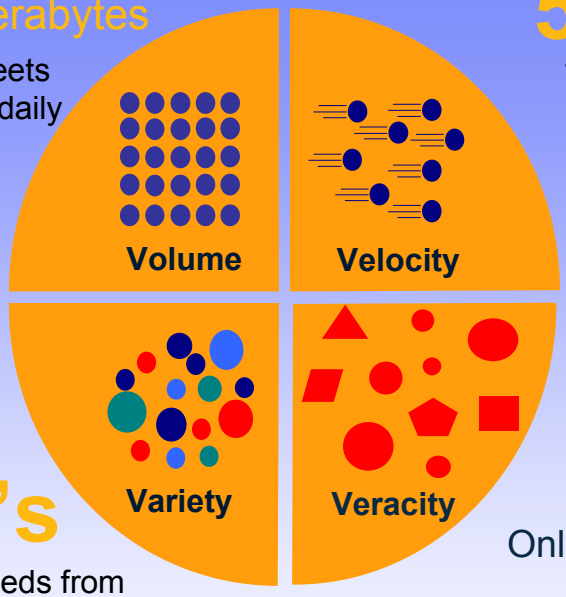
12 terabytes

of Tweets create daily

Extreme Scalability

5 million

trade events per second



100's

Of video feeds from surveillance cameras

Only **1 in 3**

Decision makers trust their information

Information from everywhere

"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

Complete your session evaluations online at www.SHARE.org/Pittsburgh-Eval

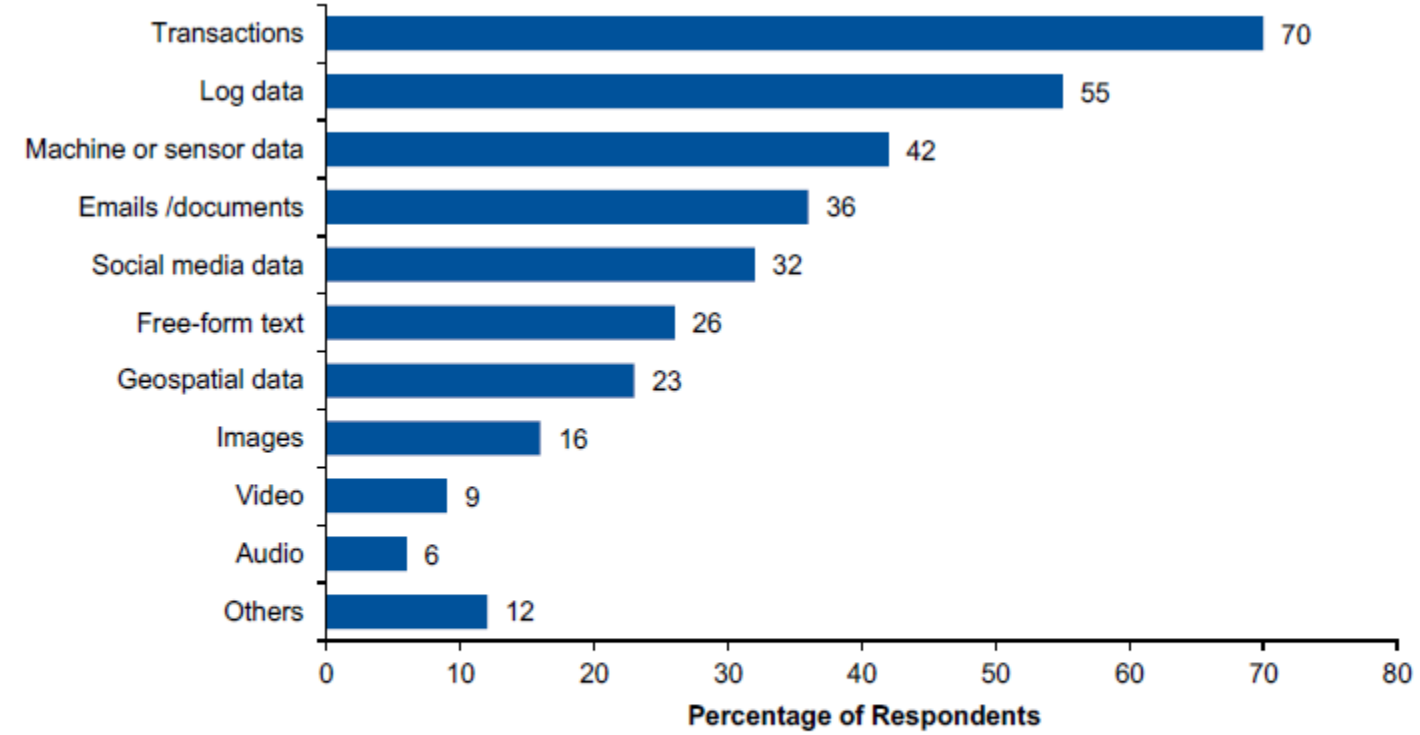
Agenda

- Big Data in an Information Driven economy
- Why start with System z
- IMS strategies for big data
- Summary / Call to action

The Big Data starting point

Types of Data Analysed

Transactional sources are the dominant data types analyzed in big data initiatives



N =465 (multiple responses allowed)

Source: Gartner (September 2013)

Gartner research note "Survey Analysis - Big Data Adoption in 2013 Shows Substance Behind the Hype" Sept 12 2013

Analyst(s): Lisa Kart, Nick Heudecker, Frank Buytendijk

The Big Data starting point

Types of Big Data Analysed by Industry

Transactional sources are the dominant data types analyzed in big data initiatives

	Manufacturing and Natural Resources	Media/ Communications	Services	Government	Education	Retail	Banking	Insurance	Healthcare	Transportation	Utilities
Transactions	73%	62%	67%	67%	54%	93%	83%	81%	75%	79%	80%
Log data	44%	57%	58%	59%	54%	40%	66%	61%	33%	71%	60%
Machine or sensor data	53%	38%	35%	33%	31%	27%	27%	48%	42%	50%	40%
Emails /documents	27%	43%	43%	41%	46%	27%	34%	39%	17%	29%	20%
Social media data	32%	52%	39%	26%	54%	73%	27%	13%	-	50%	-
Free-form text	17%	24%	28%	30%	31%	20%	34%	35%	67%	21%	40%
Geospatial data	27%	14%	19%	19%	38%	27%	27%	26%	8%	29%	40%
Images	19%	24%	17%	11%	38%	13%	5%	16%	25%	7%	-
Video	8%	29%	12%	7%	31%	13%	-	6%	8%	7%	-
Audio	10%	19%	8%	4%	8%	-	-	6%	-	-	-
Other	8%	14%	13%	15%	8%	7%	10%	16%	42%	14%	-
n =	*59*	*21**	*127*	*27**	*13**	*15**	*41*	*31*	*12**	*14**	*5**

Note: Highlighted cells indicate the top three data types by industry.
Multiple responses allowed

Source: Gartner (September 2013)

Gartner research note "Survey Analysis - Big Data Adoption in 2013 Shows Substance Behind the Hype" Sept 12 2013

Analyst(s): Lisa Kart, Nick Heudecker, Frank Buytendijk

The role of zEnterprise in Big Data analytics

- A large percent of the data that is accessed for analytics originates/resides on IBM zEnterprise
  - 2/3 of business transactions for U.S. retail banks
  - 80% of world's corporate data
- Businesses that run on zEnterprise
  - 66 of the top 66 worldwide banks
  - 24 of the top 25 U.S. retailers
  - 10 of the top 10 global life/health insurance providers
- 1,300+ ISVs run zEnterprise today, more than 275 of these selling over 800 applications on Linux
- The downtime of an application running on System z equates to approximately 5 minutes per year
- The System z mainframe can run over a thousand virtual Linux images on a single frame the size of a refrigerator

Majority of today's analytics based on relational / "Structured" Data

- Analytics and decision engines reside where the DWH / transaction data is
- "Noise" (veracity) surrounds the core business data
  - Social Media, emails, docs, telemetry, voice, video, content
- What data are you prepared to ***TRUST?***
- Where do you put your trusted Data?

"Circle of trust"

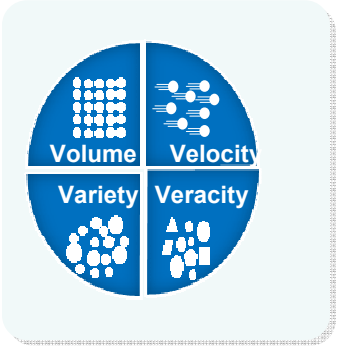
Demand for differently structured data to be seamlessly integrated, to augment analytics / decisions

- Analytics and decision engines reside where the DWH / transaction data is
- "Noise" (veractity) surrounds the core business data
  - Social Media, emails, docs, telemetry, voice, video, content
- Expanding our insights – getting closer to the "truth"
  - Lower risk and cost
  - Increased profitability

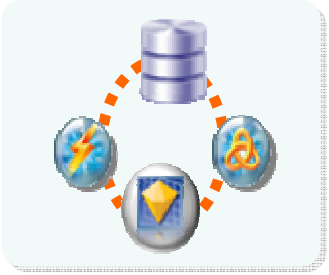
"Circle of trust" widens

Forward Thinking Organizations are Creating Value From Big Data

The power of Data coming together…



…with the power of Technology…



…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

Fraud Detection – Claiming disability allowance.

"Unable to work"

Work Status

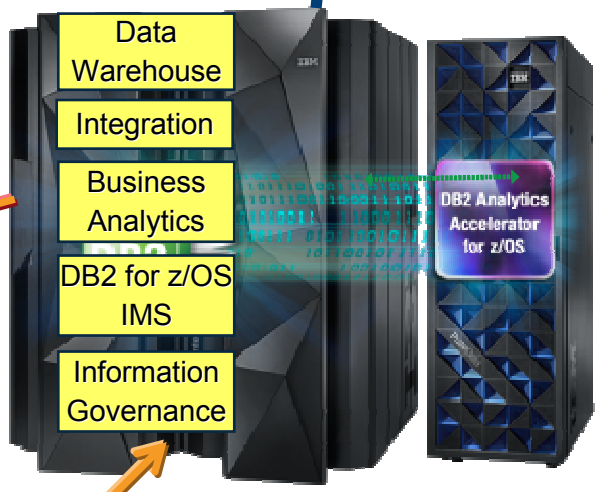
"Dude – awesome vacation"

Facebook Post

Deterrent for fraudsters - Cost Savings for the business

Make payment or investigate

zEnterprise

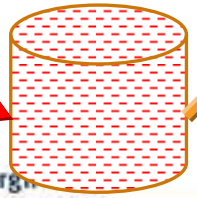


Data from Social Media sites analyzed with Text analytics

Hadoop or agency

Refined Search parameters from OLTP environment

Result Set for further processing

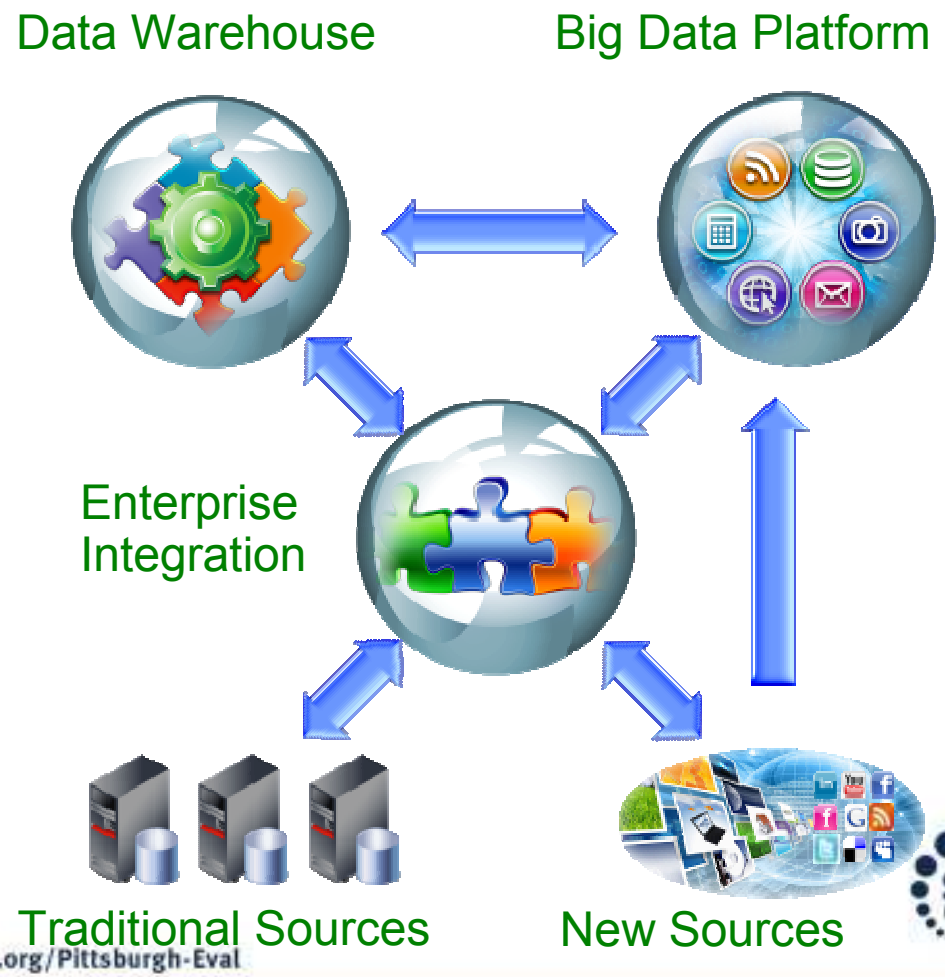


Data Warehouse + modeling applications

Result set uploaded or directly imported into OLTAP DBMS

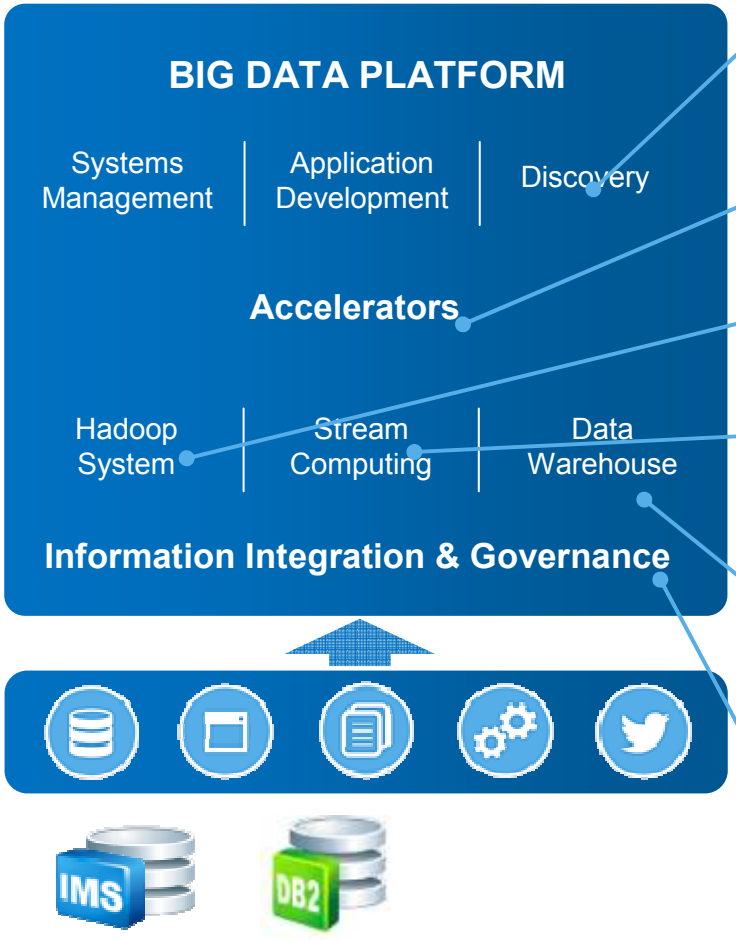
Enterprise Integration and Governance… the key to success of incorporating Big Data

- **Information Integration**
  - –Insights from big data must be incorporated into the warehouse and analytics/ decision engines
- **Information Governance**
  - –Companies need to govern what comes in, and the insights that come out



IBM BIG DATA PLATFORM

Logical platform with many physical deployment options



Watson Explorer
Find, navigate, visualize all data

Accelerators
Speed time to value with analytic and application accelerators

InfoSphere BigInsights
Bringing Hadoop to the enterprise

InfoSphere Streams
Analytics for data in-motion exploration

InfoSphere Data Warehouse
Delivers deep insight with advanced database analytics & operational analytics

Information Integration and Governance
Governs data quality and manages the information lifecycle

Complete your session evaluations online at www.SHARE.org/Pittsburgh-Eval

Core data management solutions for the 21st Century

DB2 11 for z/OS

IMS 13

Unmatched availability, reliability, and security for business critical information

- **Up to 40% CPU reductions and performance improvements** for (OLTP), batch, and business analytics
- **Improved data sharing performance** and efficiency
- **Integration with InfoSphere BigInsights™ / Hadoop and noSQL support**
- **Improved utility performance** and additional zIIP eligible workload

Delivering the highest levels of performance, availability, security, and scalability in the industry

- **Breaking through 100k TPS** 800% greater than IMS 12
- **CPU reductions up to 62% for Java Apps**
- **SQL access to IMS data from both .NET and COBOL applications**
- **Greater flexibility and faster deployment** for new applications with database versioning
- **Big Data Ready** exploitation of Hadoop / Big Insights, MDA, Watson Explorer…

IBM PureData System for Hadoop

Accelerate Hadoop analytics with appliance simplicity

Accelerate Big Data projects with built-in expertise
- Explore new ways to use all data
- Unlock new insights from unstructured data
- Establish a cost efficient on-line data archive

Simplify with integrated system management
- InfoSphere BigInsights software
- Compute and Storage hardware

Ensure production grade security and governance

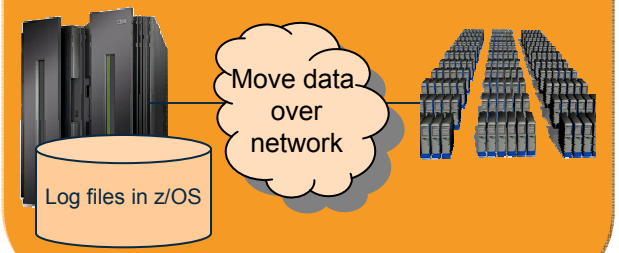
Easily integrate with other systems
in the IBM big data platform

CIO:Insight Apr 29 2013 *...Issues surrounding how long it takes to get a Hadoop application into production coupled with a lack of real-time capabilities are proving to be important barriers to deployment. As a result, the respondents are reporting that both the number of Hadoop applications and the size of the overall Hadoop environment remain relatively small.*

Approaches

1

Processing done outside z
(Extract and move data)

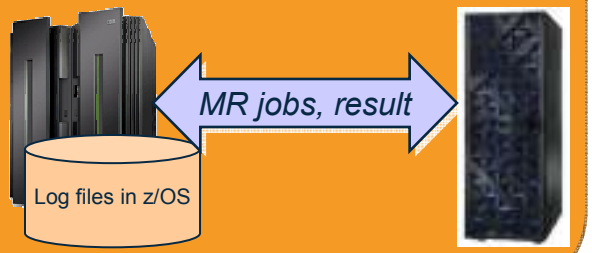


$$$$

Additional infrastructure.
Challenges with scale, governance, ingestion.

2

Processing in Appliance
(z remains the master)



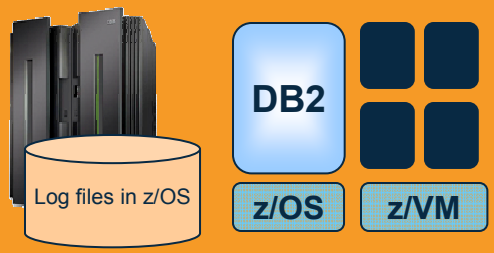
$$

Appliance approach with PureData System for Hadoop.
High speed load.
z is the control point.

3

Hadoop for Linux on System z
• Apache site
• Veristorm - zDoop

Processing done on System z
(MR cluster on zLinux)



$$$

Provision new node quickly
Near linear scale.
High speed load.
z is the control point.

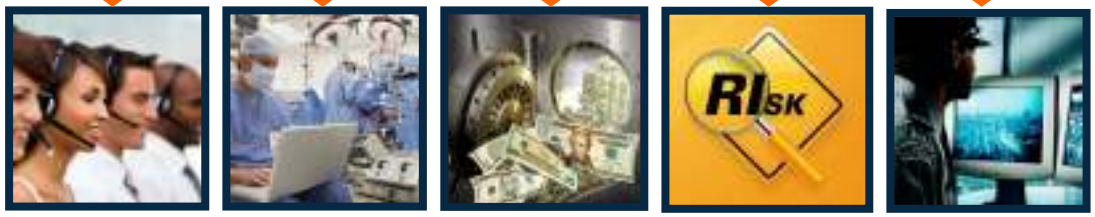
Data Explorer : visualization & discovery across all your data sources : "*Integration at the glass*"

InfoSphere Data 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 System z 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

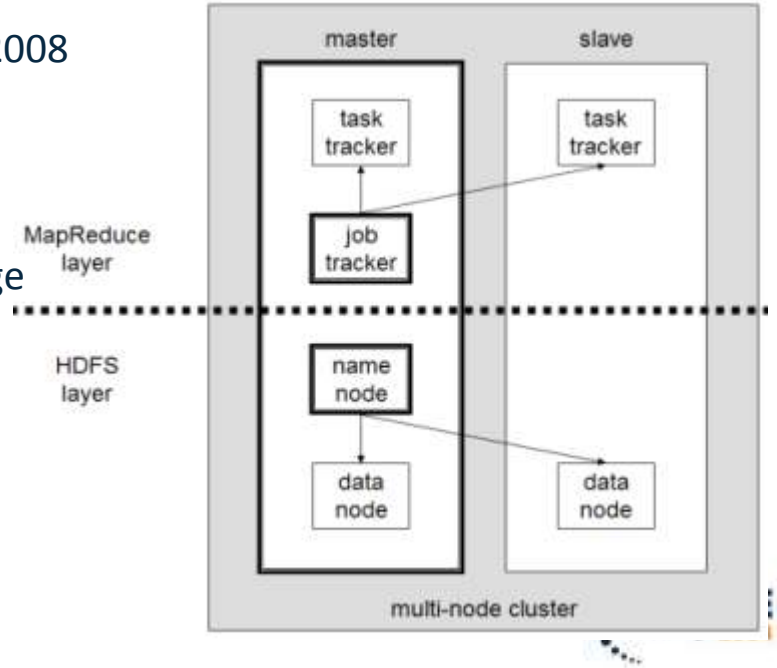
Create unified view of ALL information for real-time monitoring

Agenda

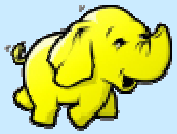
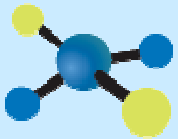
- Big Data in an Information Driven economy
- Why start with System z
- IMS strategies for big data
- Summary / Call to action

What is hadoop

- An open source software framework that supports data-intensive distributed applications
  - **High throughput, batch processing**
  - **runs on large clusters of commodity hardware**
    - Yahoo runs a 4000 nodes Hadoop cluster in 2008
- Two main components
  - **Hadoop distributed file system**
    - self-healing, high-bandwidth clustered storage
  - **MapReduce engine**

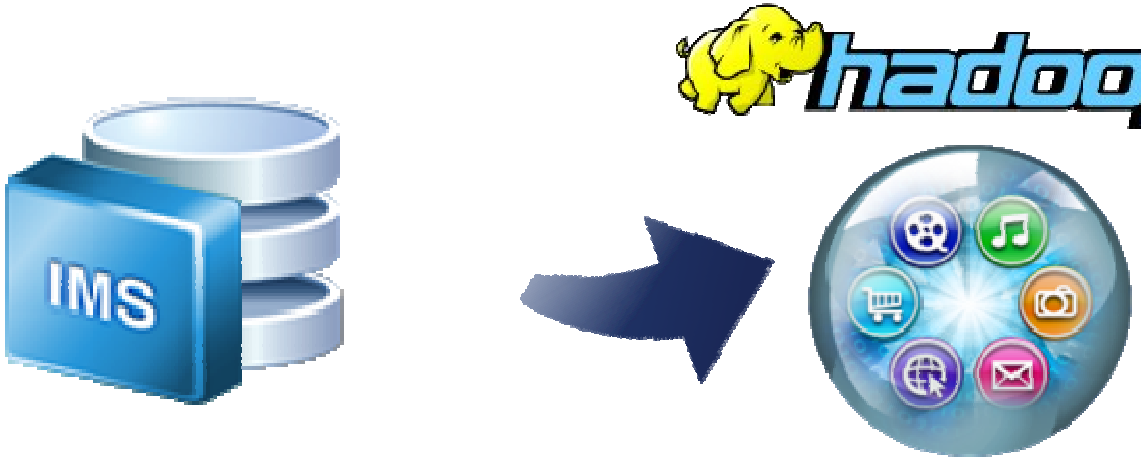


BIG DATA is not just HADOOP

Understand and navigate federated big data sources		Federated Discovery and Navigation
Manage & store huge volume of any data		Hadoop File System MapReduce
Structure and control data		Data Warehousing
Manage streaming data		Stream Computing
Analyze unstructured data		Text Analytics Engine
Integrate and govern all data sources		Integration, Data Quality, Security, Lifecycle Management, MDM

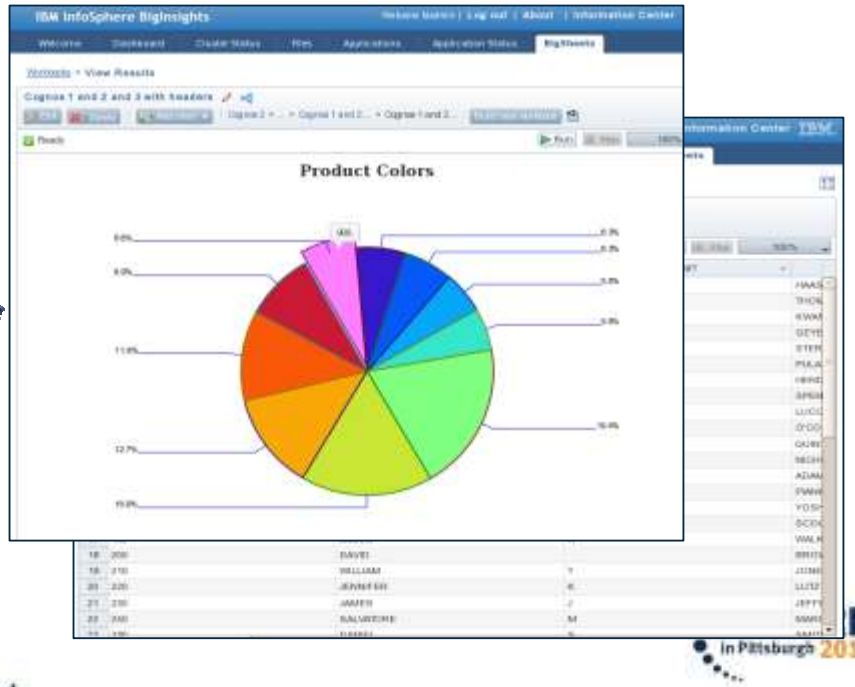
Enhancing IMS analytics on System z with Big Data

- **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 System z data sources**
- **IMS provides the connectors and the DB capability to allow BigInsights v2.1.2.0 to easily and efficiently access the IMS data source**
  - **BigInsights v2.1.2.0 is available on 3/13/2014**

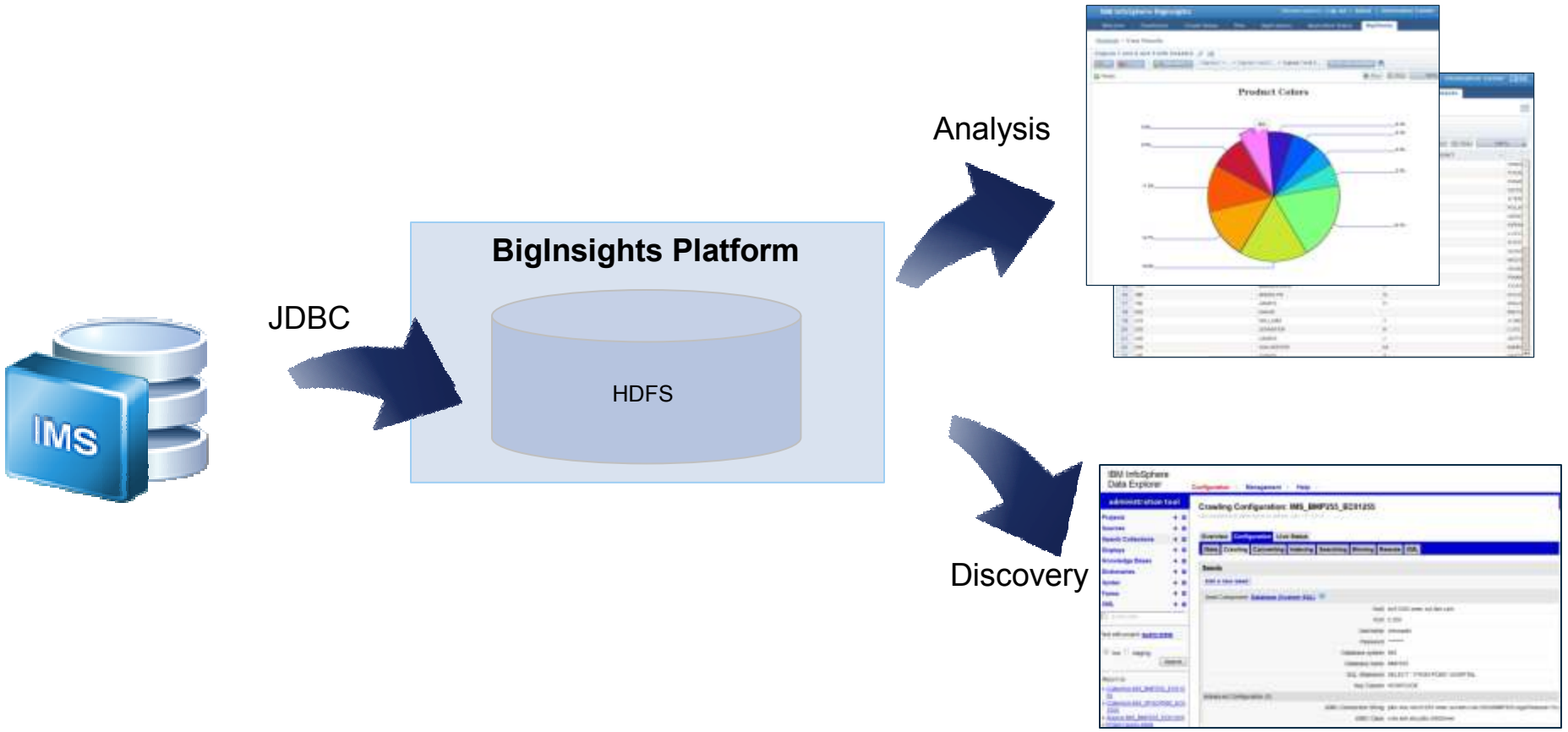


Enhancing IMS analytics on System z with Big Data

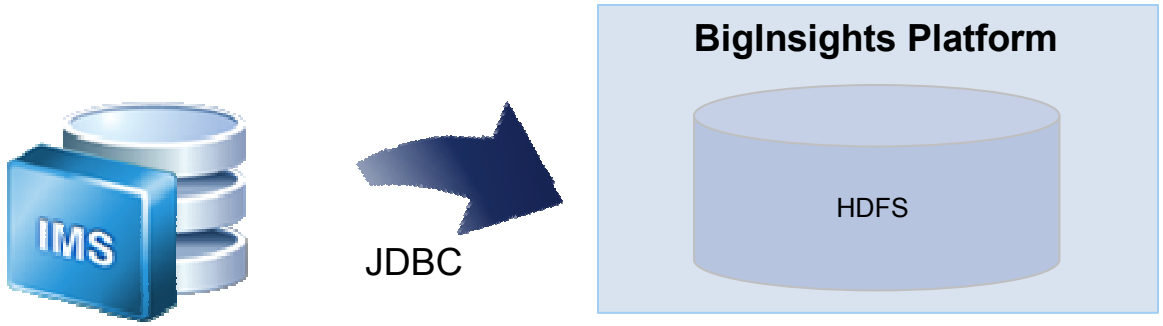
- **Observation points lead to new business opportunities**
- **Observation points gleaned from both archived data and live data**
- **Score business events, track claims evolution, and more**
- *Make the data available to people who can do something meaningful with it*



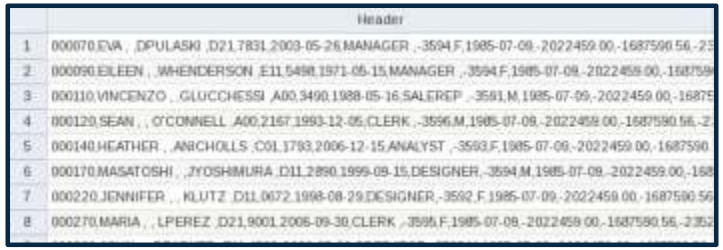
High level overview



Import options

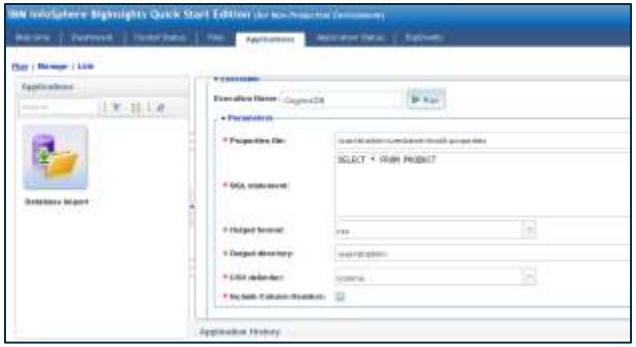


Sqoop Import



	Header
1	000070,EVA , ,DPULASKI ,D21,7831,2003-05-26,MANAGER ,-3594,F,1985-07-09,-2022459.00,-1687590.56,-2
2	000090,EILEEN , ,WHENDERSON ,E11,5498,1971-05-15,MANAGER ,-3594,F,1985-07-09,-2022459.00,-168759
3	000110,VINCENZO , ,GLUCCHESSI ,A00,3490,1988-05-16,SALESREP ,-3591,M,1985-07-09,-2022459.00,-1687
4	000120,SEAN , , O'CONNELL ,A00,2167,1993-12-05,CLERK ,-3596,M,1985-07-09,-2022459.00,-1687590.56,-2
5	000140,HEATHER , ,ANICHOLLS ,C01,1793,2006-12-15,ANALYST ,-3593,F,1985-07-09,-2022459.00,-1687590
6	000170,MASATOSHI , ,JYOSHIMURA ,D11,2890,1999-09-15,DESIGNER,-3594,M,1985-07-09,-2022459.00,-168
7	000220,JENNIFER , ,KLUTZ ,D11,0672,1998-08-29,DESIGNER,-3592,F,1985-07-09,-2022459.00,-1687590.56
8	000270,MARIA , ,LPEREZ ,D21,9001,2006-09-30,CLERK ,-3595,F,1985-07-09,-2022459.00,-1687590.56,-235

POJO Import



Complete your session evaluations online at www.SHARE.org/Pittsburgh-Eval

Sqoop Import

- Command line interface application for transferring data between RDBMS and HDFS.
- Import into Hive and Hbase
- Export from HDFS back into RDBMS
- Import:
  - Divides table into ranges using primary key max/min (can use split-by param)
  - Creates mappers for each range
  - Mappers write to multiple HDFS nodes
  - Creates text or sequence files
- Export:
  - Reads files in HDFS directory via MapReduce
  - Bulk parallel insert into database table.

Sqoop Import

- Import into HDFS using the below command:

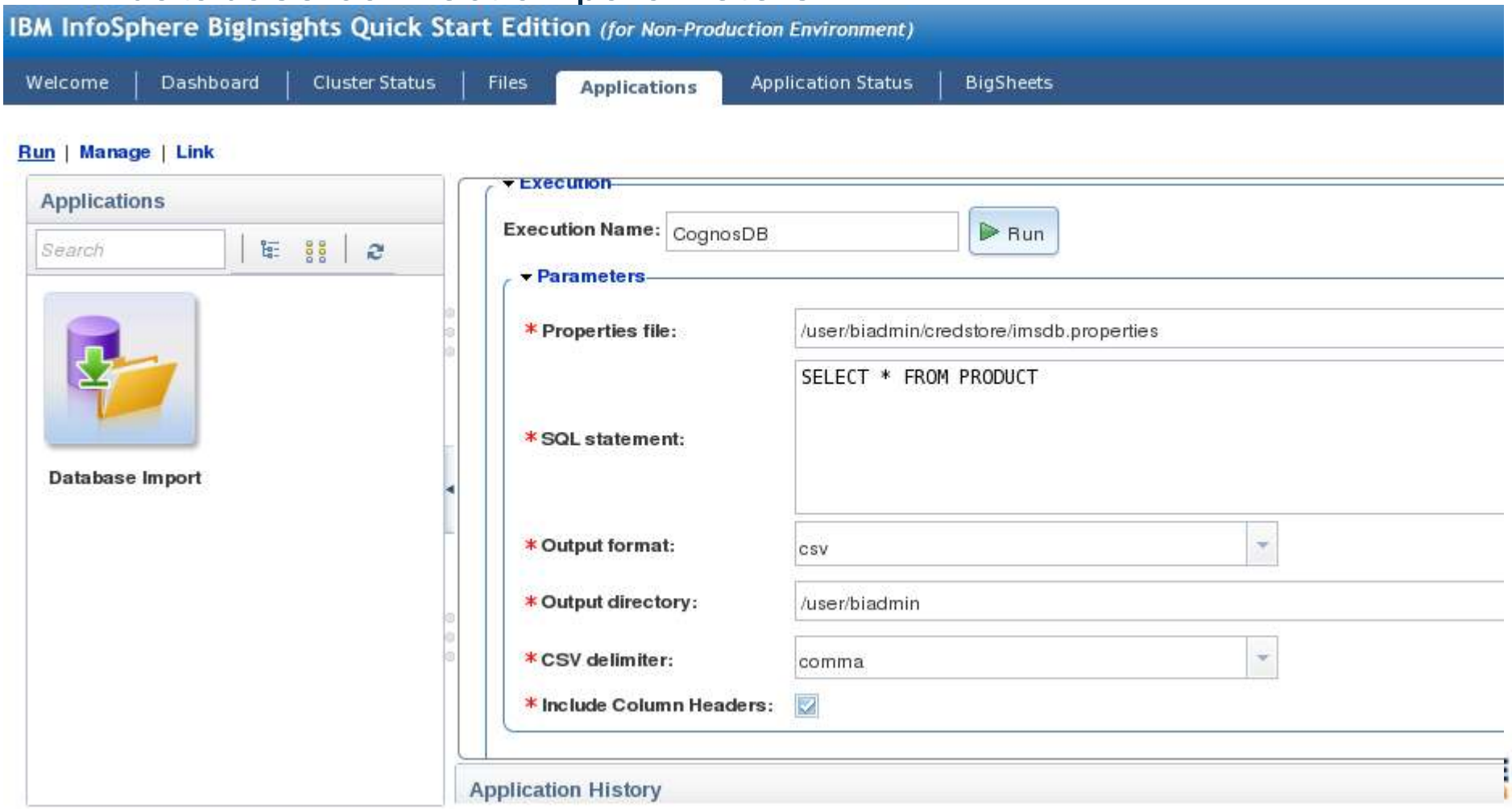
```
./sqoop import --connect
   jdbc:ims://ecwas09.vmec.svl.ibm.com:5555/BIGDATP1 --driver
   com.ibm.ims.jdbc.IMSDriver --table EMPLOYEE -m 3 --split-by
   EDLEVEL --username 'OMVSADM' -P
13/06/19 17:50:27 INFO db.DataDrivenDBInputFormat: BoundingValsQuery:
   SELECT MIN(EDLEVEL), MAX(EDLEVEL) FROM EMPLOYEE
13/06/19 17:50:46 INFO mapreduce.ImportJobBase: Transferred 5.123 KB in
   20.3762 seconds (257.4572 bytes/sec)
13/06/19 17:50:46 INFO mapreduce.ImportJobBase: Retrieved 43 records.
```

- HDFS Output (below)

	Header
1	000070,EVA , ,DPULASKI ,D21,7831,2003-05-26,MANAGER ,-3594,F,1985-07-09,-2022459.00,-1687590.56,-23
2	000090,EILEEN , ,WHENDERSON ,E11,5498,1971-05-15,MANAGER ,-3594,F,1985-07-09,-2022459.00,-168759
3	000110,VINCENZO , ,GLUCCHESSI ,A00,3490,1988-05-16,SALEREP ,-3591,M,1985-07-09,-2022459.00,-16875
4	000120,SEAN , , O'CONNELL ,A00,2167,1993-12-05,CLERK ,-3596,M,1985-07-09,-2022459.00,-1687590.56,-2
5	000140,HEATHER , ,ANICHOLLS ,C01,1793,2006-12-15,ANALYST ,-3593,F,1985-07-09,-2022459.00,-1687590.
6	000170,MASATOSHI , ,JYOSHIMURA ,D11,2890,1999-09-15,DESIGNER,-3594,M,1985-07-09,-2022459.00,-168
7	000220,JENNIFER , , KLUTZ ,D11,0672,1998-08-29,DESIGNER,-3592,F,1985-07-09,-2022459.00,-1687590.56
8	000270,MARIA , , LPEREZ ,D21,9001,2006-09-30,CLERK ,-3595,F,1985-07-09,-2022459.00,-1687590.56,-2352

BigInsights Database Import Application

- Utilize the built in Database Import Application by providing the database connection parameters:

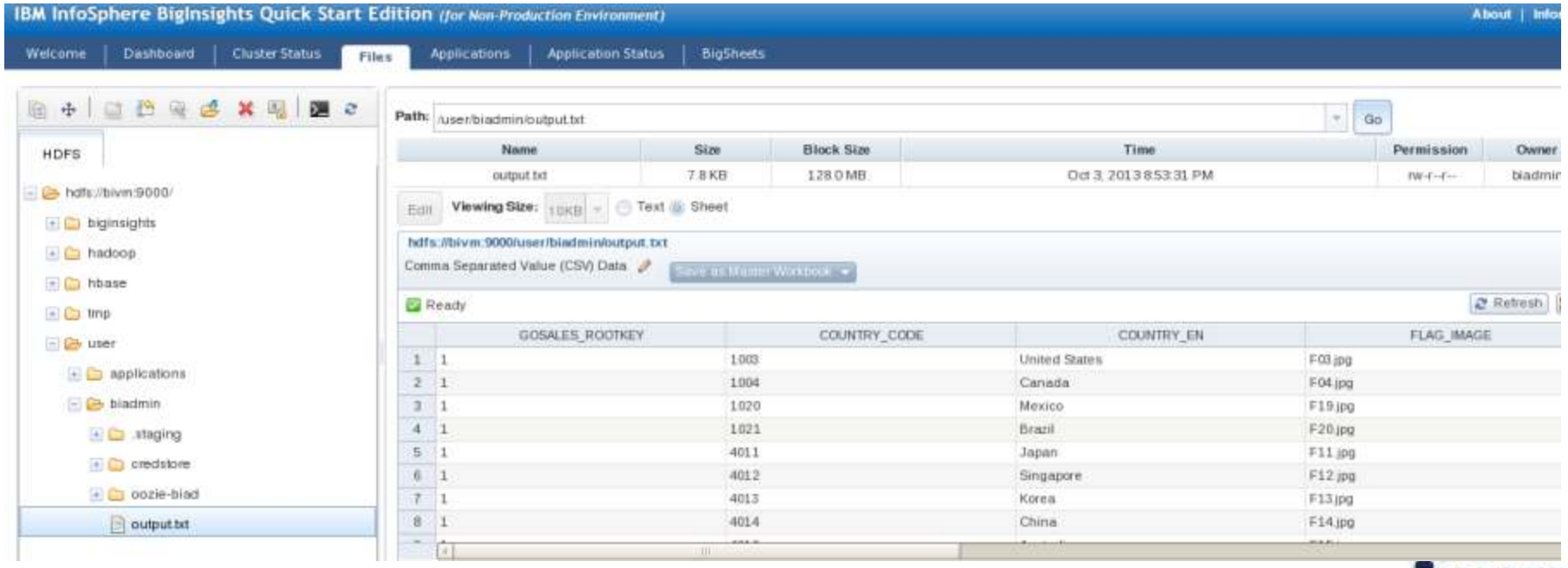


BigInsights Database Import Application

- Once the run is completed, view the data in HDFS:

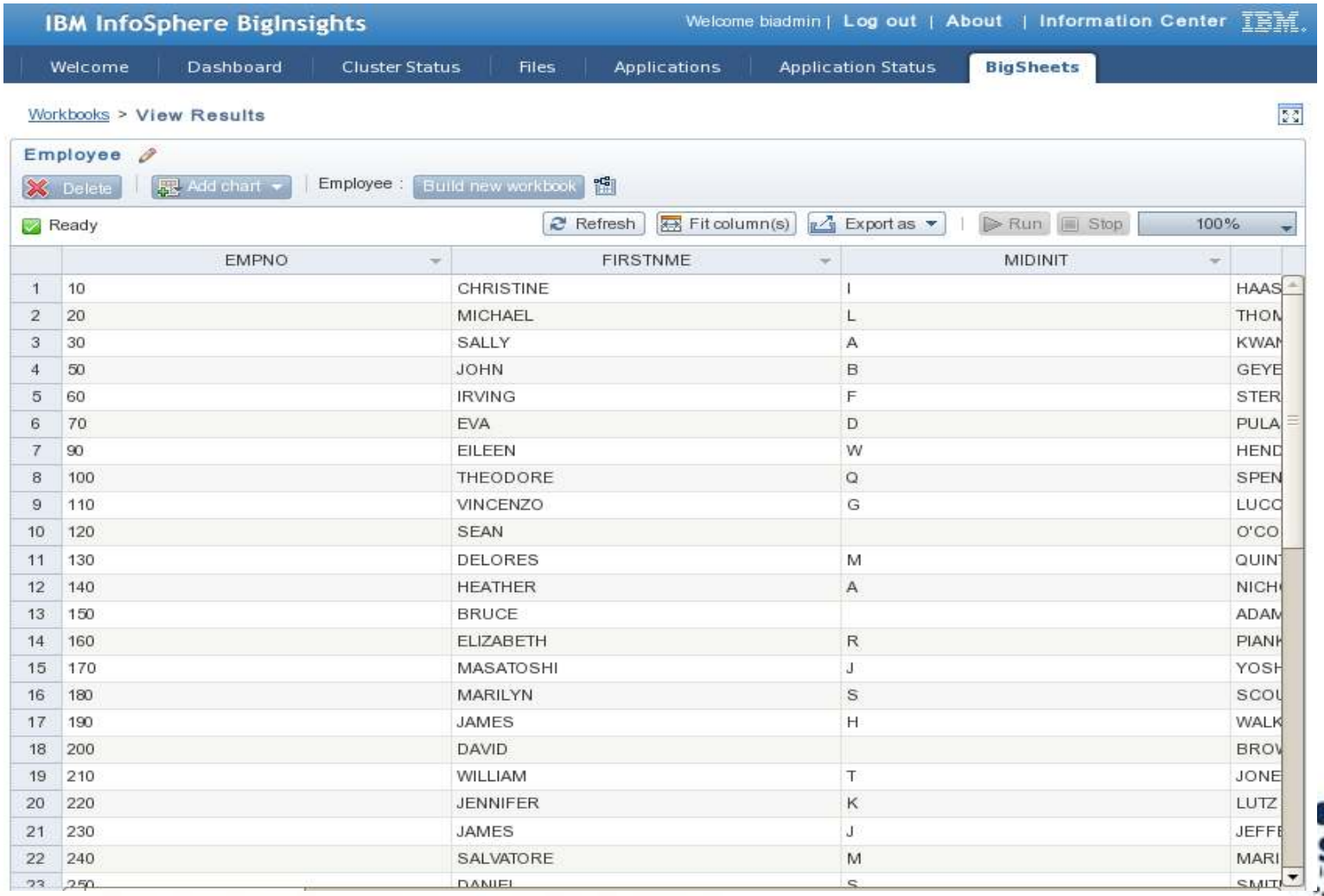
Application History

Status	Execution Name	Progress	Start Time	Elapsed Time (sec)	Output	Details
... No filter applied						
✓	CognosDB	100%	2013 10 3 20:53:20	14		
✓	CognosDB	100%	2013 10 3 20:41:22	18		



BigInsights BigSheets

- This data can be saved as BigSheets workbook for further analytics



	EMPNO	FIRSTNME	MIDINIT	
1	10	CHRISTINE	I	HAAS
2	20	MICHAEL	L	THON
3	30	SALLY	A	KWAN
4	50	JOHN	B	GEYE
5	60	IRVING	F	STER
6	70	EVA	D	PULA
7	90	EILEEN	W	HEND
8	100	THEODORE	Q	SPEN
9	110	VINCENZO	G	LUCC
10	120	SEAN		O'CO
11	130	DELORES	M	QUIN
12	140	HEATHER	A	NICH
13	150	BRUCE		ADAM
14	160	ELIZABETH	R	PIANK
15	170	MASATOSHI	J	YOSH
16	180	MARILYN	S	SCOU
17	190	JAMES	H	WALK
18	200	DAVID		BROW
19	210	WILLIAM	T	JONE
20	220	JENNIFER	K	LUTZ
21	230	JAMES	J	JEFF
22	240	SALVATORE	M	MARI
23	250	DANIEL	S	SMIT

Elevated demand for business analytics drives new requirements and focus

More aggressive requirements

Enterprise-level scale & performance

Mission critical availability

Faster access to operational data

Rapid, cost effective deployment & expansion

More integrated view of data across the environment

Driving new focuses

- Modernization
- Standardization & Consolidation
- Operational BI
- Data Governance
- Cloud Computing

Watson Explorer

- ✓ Watson Explorer is the **visualization & discovery** capability for IBM's comprehensive **big data platform**

- ✓ Watson Explorer is a **key component of all the big data use cases** with greatest impact in **Big Data Exploration** & **Enhanced 360 View of the Customer**

Complete your session evaluations online at www.SHARE.org/Pittsburgh-Eval

Watson Explorer : visualization & discovery across all your data sources : "*Integration at the glass*"

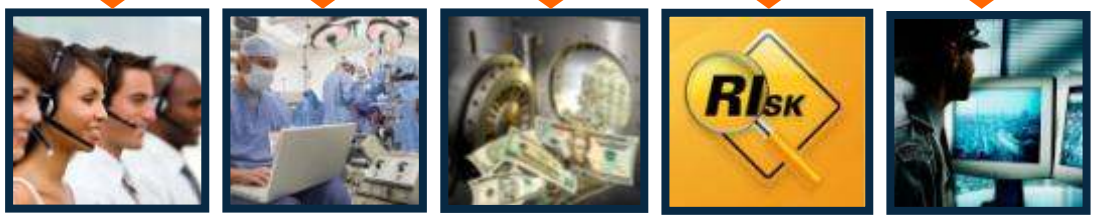
Watson Explorer

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 System z big data integration and analytics projects

Improve customer service & reduce call times



Create unified view of ALL information for real-time monitoring

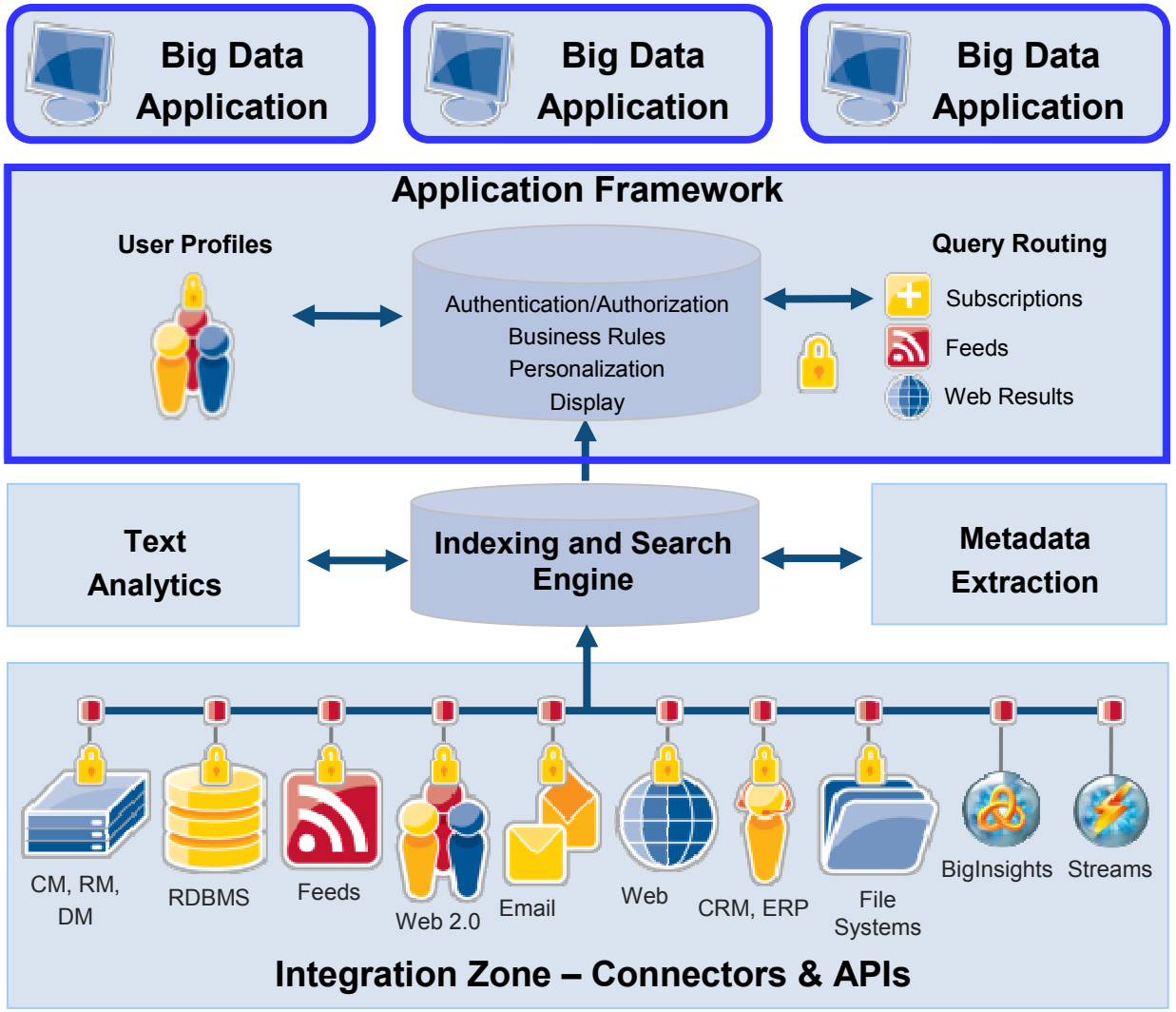
Increase productivity & leverage past work increasing speed to market

Analyze customer information & data to unlock true customer value

Identify areas of information risk & ensure data compliance

Complete your session evaluations online at www.SHARE.org/Pittsburgh-Eval

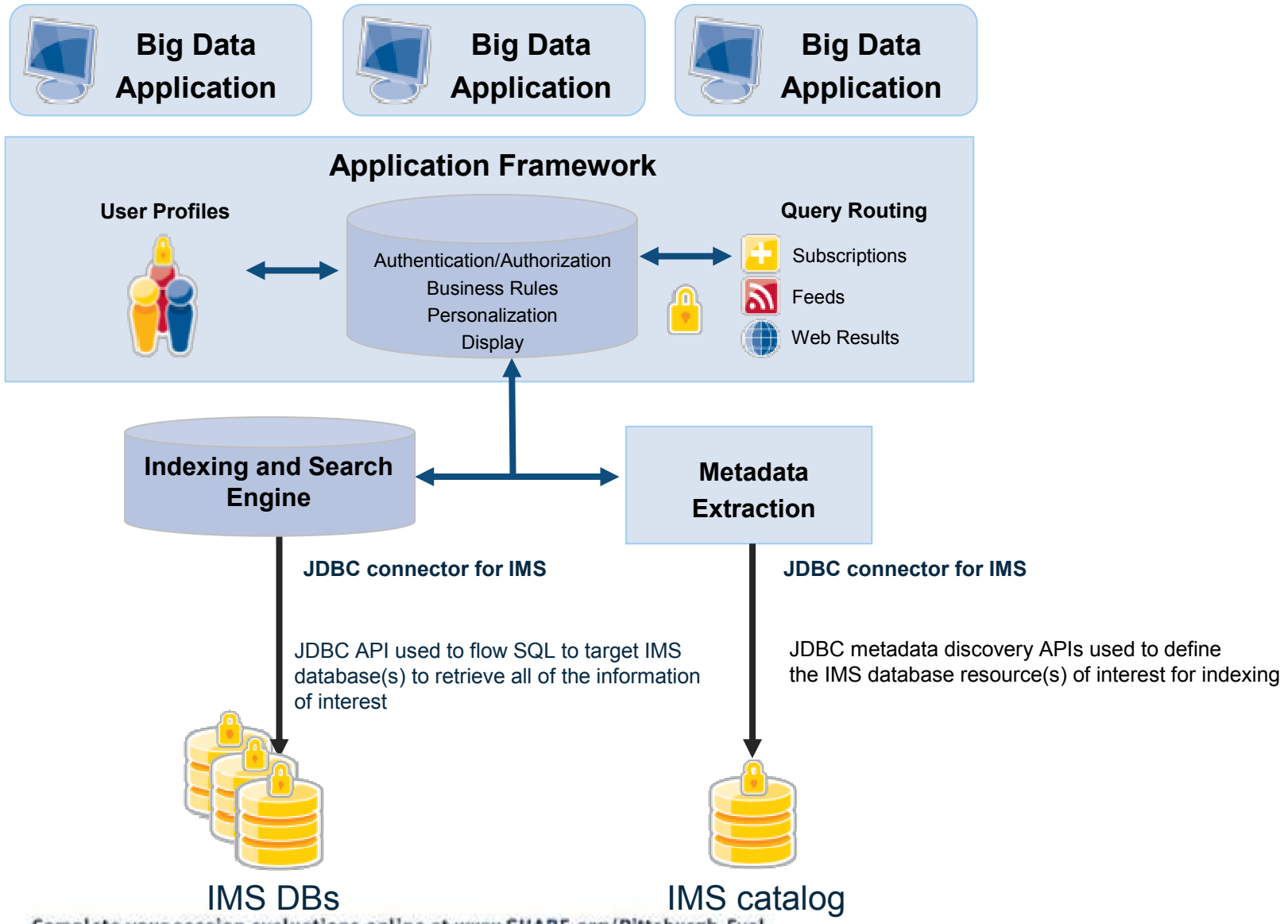
Watson Explorer product architecture and differentiators



Differentiators

- Federated discovery and navigation
- Scalable architecture
- Accurate results
- Secure connectivity
- Powerful development tools
- Unique application framework
- Fast time to value

Seamless IMS integration



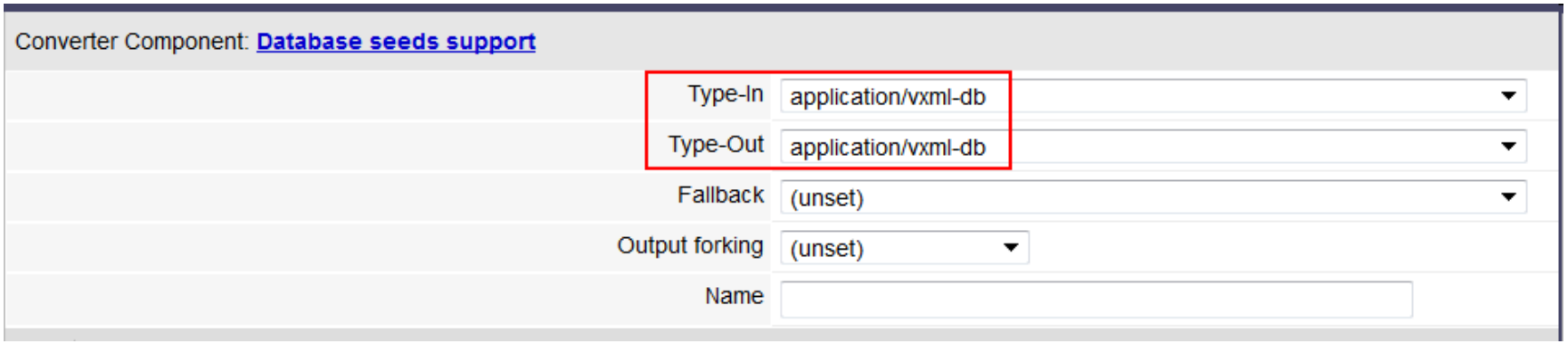
IMS + Watson Explorer -Configuring the IMS source

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

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

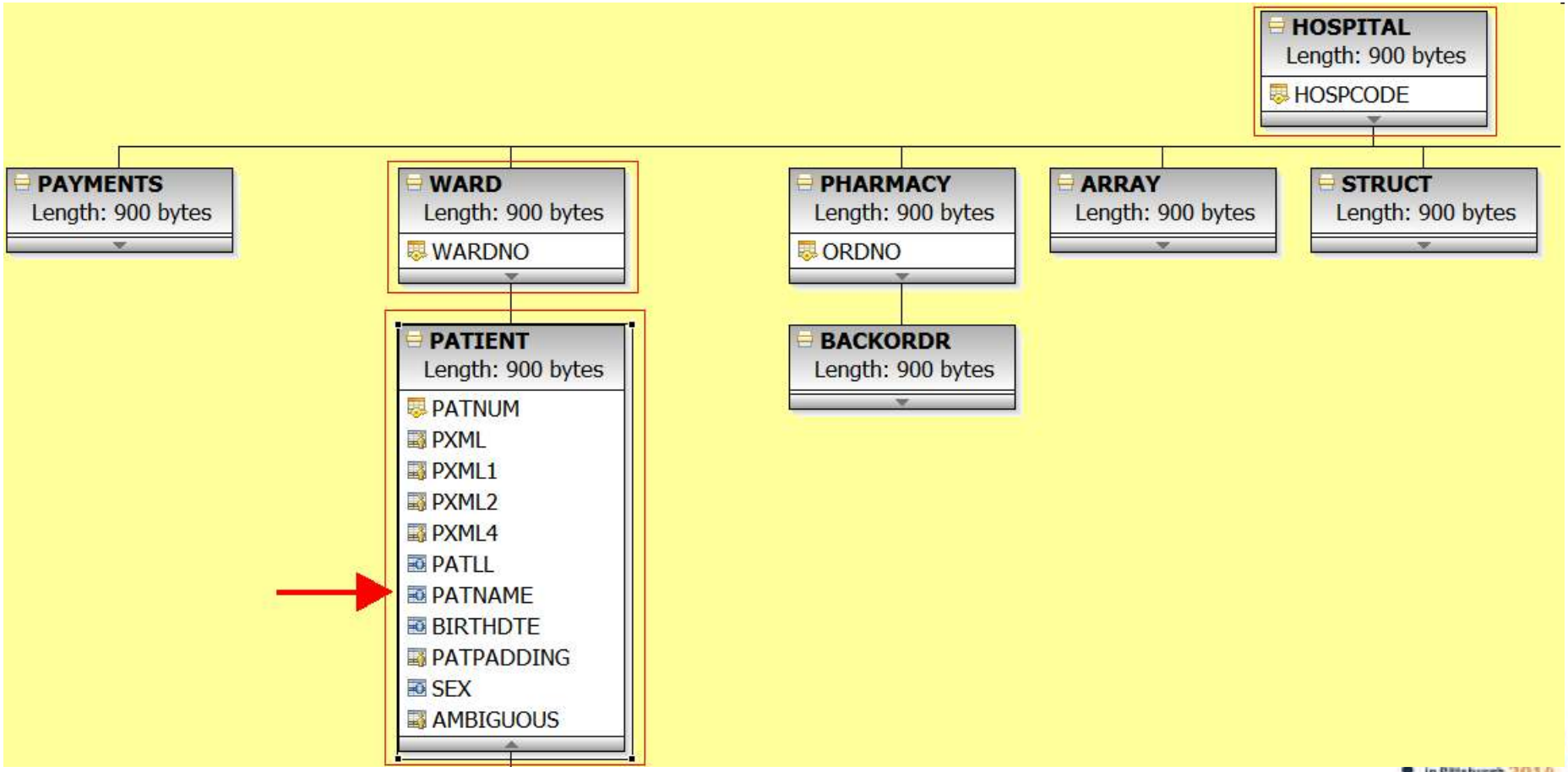
IMS + Watson Explorer
-Setting up the data transformation

- After creating a new seed, a converter needs to be configured using standard XPATH



Original IMS hierarchy for hospital database

- Hierarchy: HOSPITAL->WARD->PATNAME
- Goal: Get a patient centric view

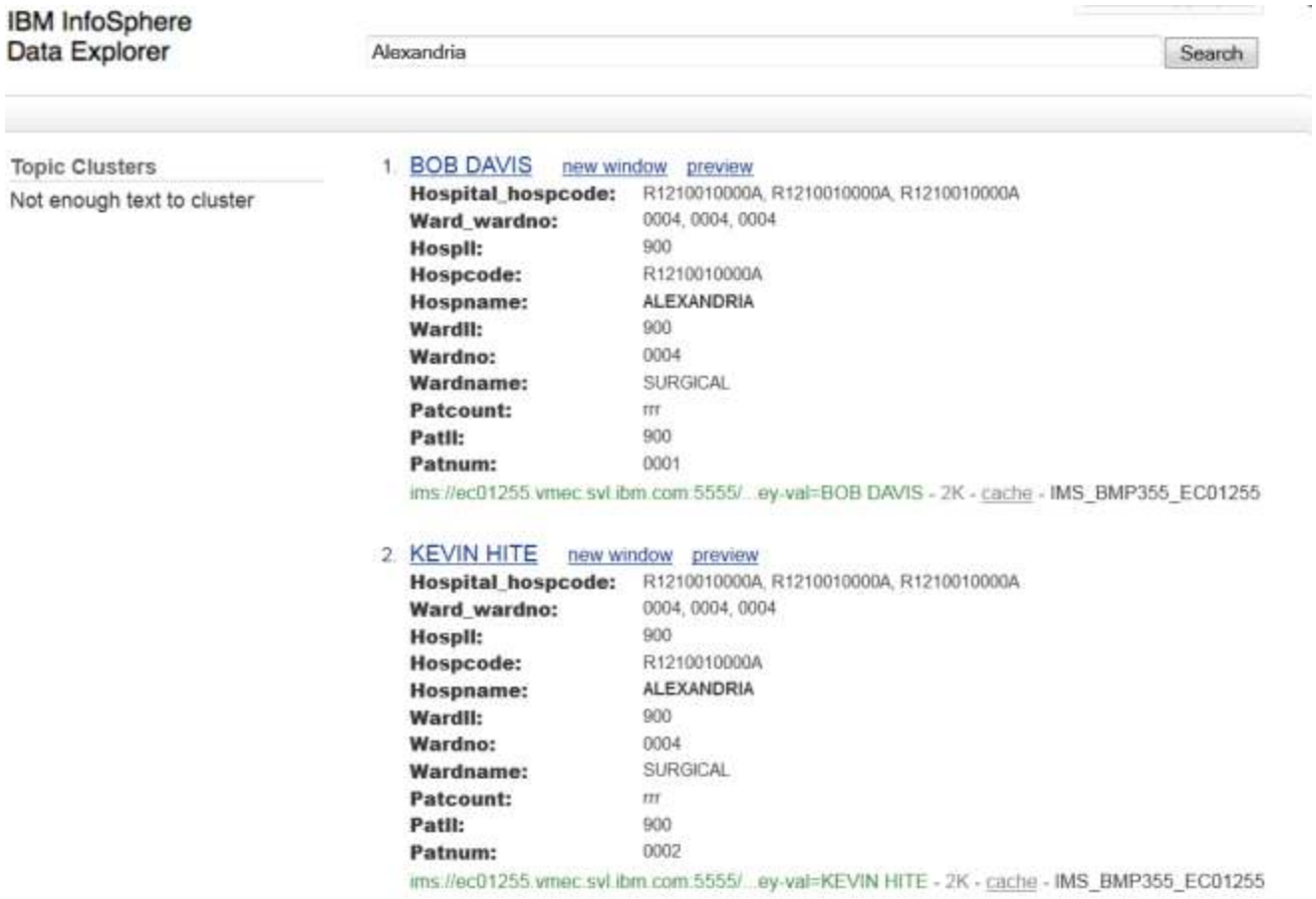


Why use Watson Explorer

- 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)

Searching the IMS database with Watson Explorer

- Query: Who are the patients in the Alexandria hospital



Searching the IMS database with Watson Explorer

- Query: Who are the patients currently in dermatology

IBM InfoSphere
Data Explorer

Dermatology

Topic Clusters

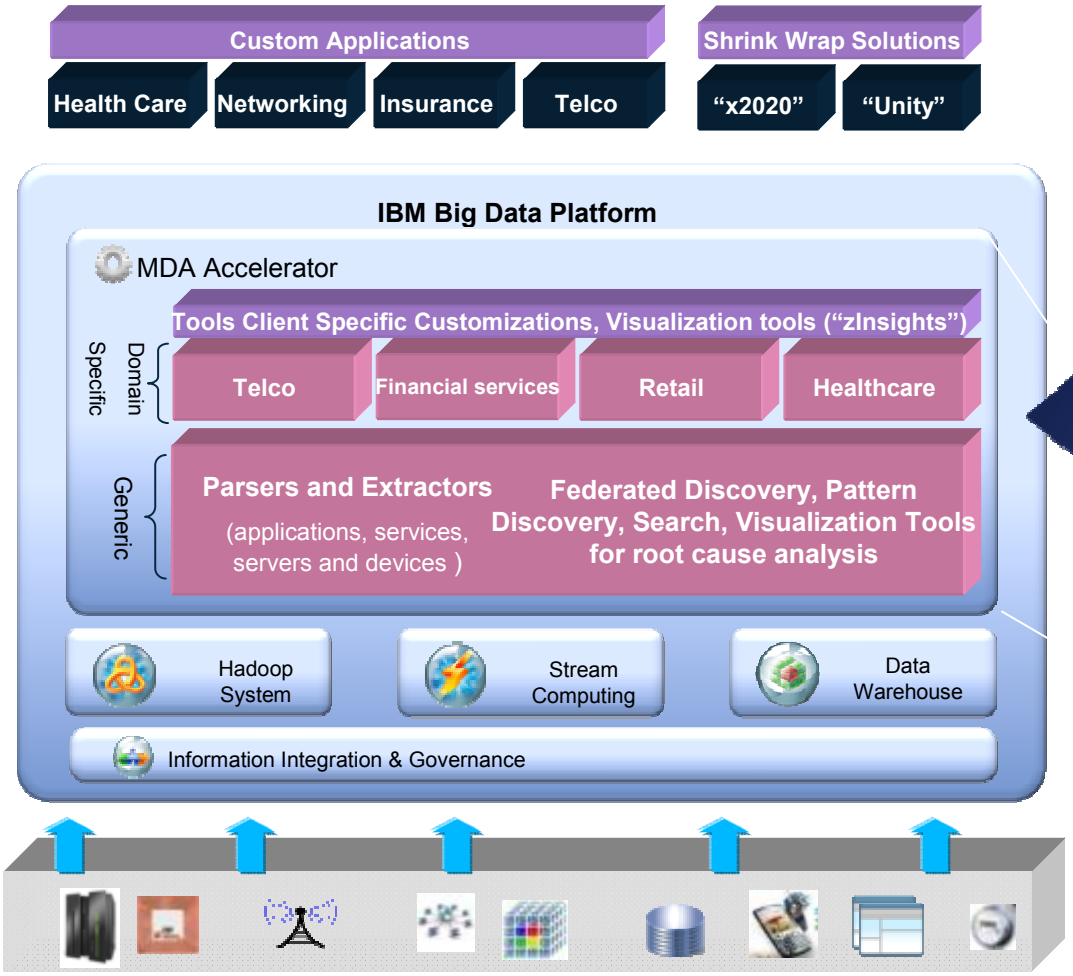
Not enough text to cluster

1. WILLIAM LI new window preview

Hospital_hospcode:	R1210020000A, R1210020000A, R1210020000A
Ward_wardno:	0002, 0002, 0002
HospII:	900
Hospcode:	R1210020000A
Hospname:	SANTA TERESA
WardII:	900
Wardno:	0002
Wardname:	**DERMATOLOGY**
PatII:	900
Patnum:	0001

ims://ec01255.vmec.svl.ibm.com:5555/...ey-val=WILLIAM LI - 2K - cache -

Machine Data Analytics Accelerator

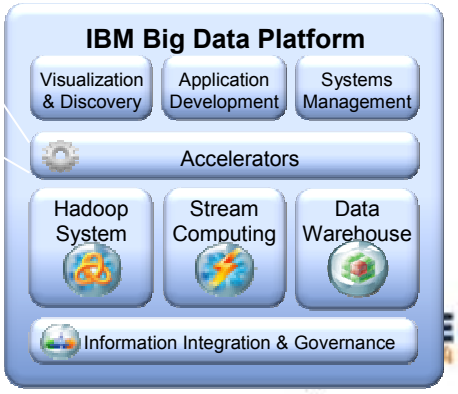


IT use cases:

- Server, performance, troubleshooting

Business use cases:

- Click stream and transaction analysis
- Optimize production, advance planning

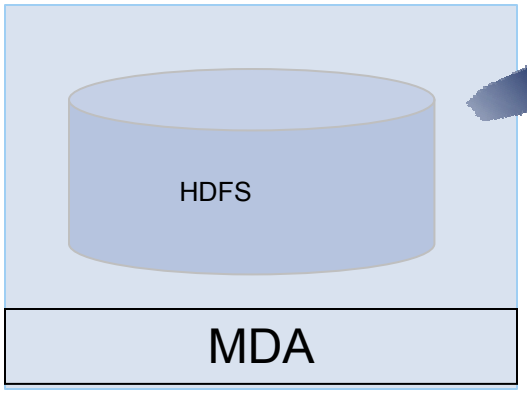


IMS and IBM Accelerator for Machine Data Analytics

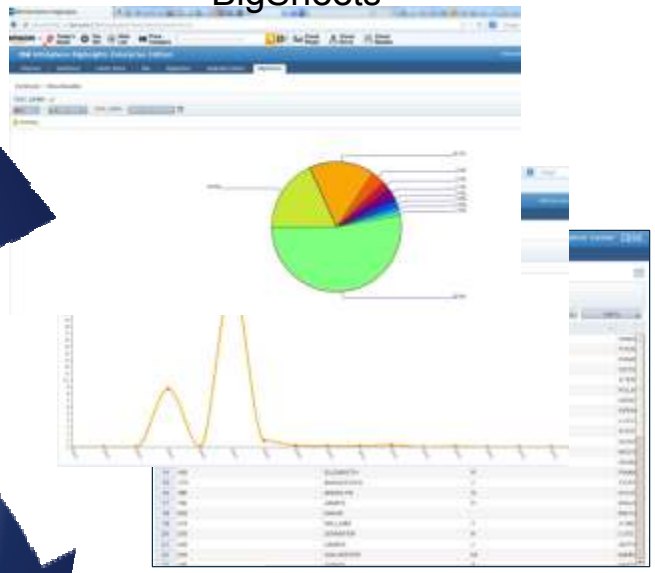
- Consume log data produced by Transaction Analysis Workbench
- Index and link transactions together across products (IMS, DB2, MQ, CICS, WebSphere)
- Make large amounts of IMS transactional log data available to the suite of BigInsights tools.

IMS log and MDA overview

BigInsights Platform



BigSheets



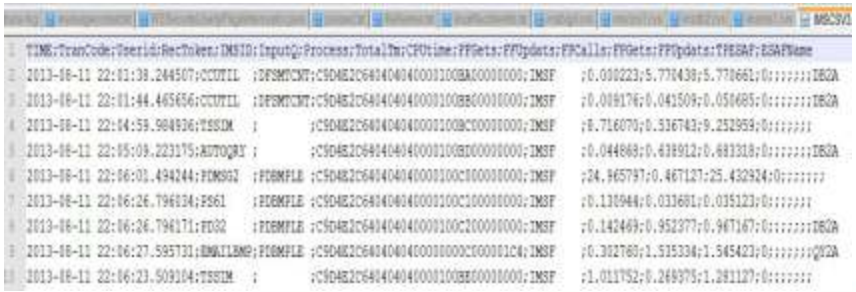
Transaction Analysis Workbench

Index, Extraction

SFTP

Data Explorer

Conversion to ASCII in CSV format



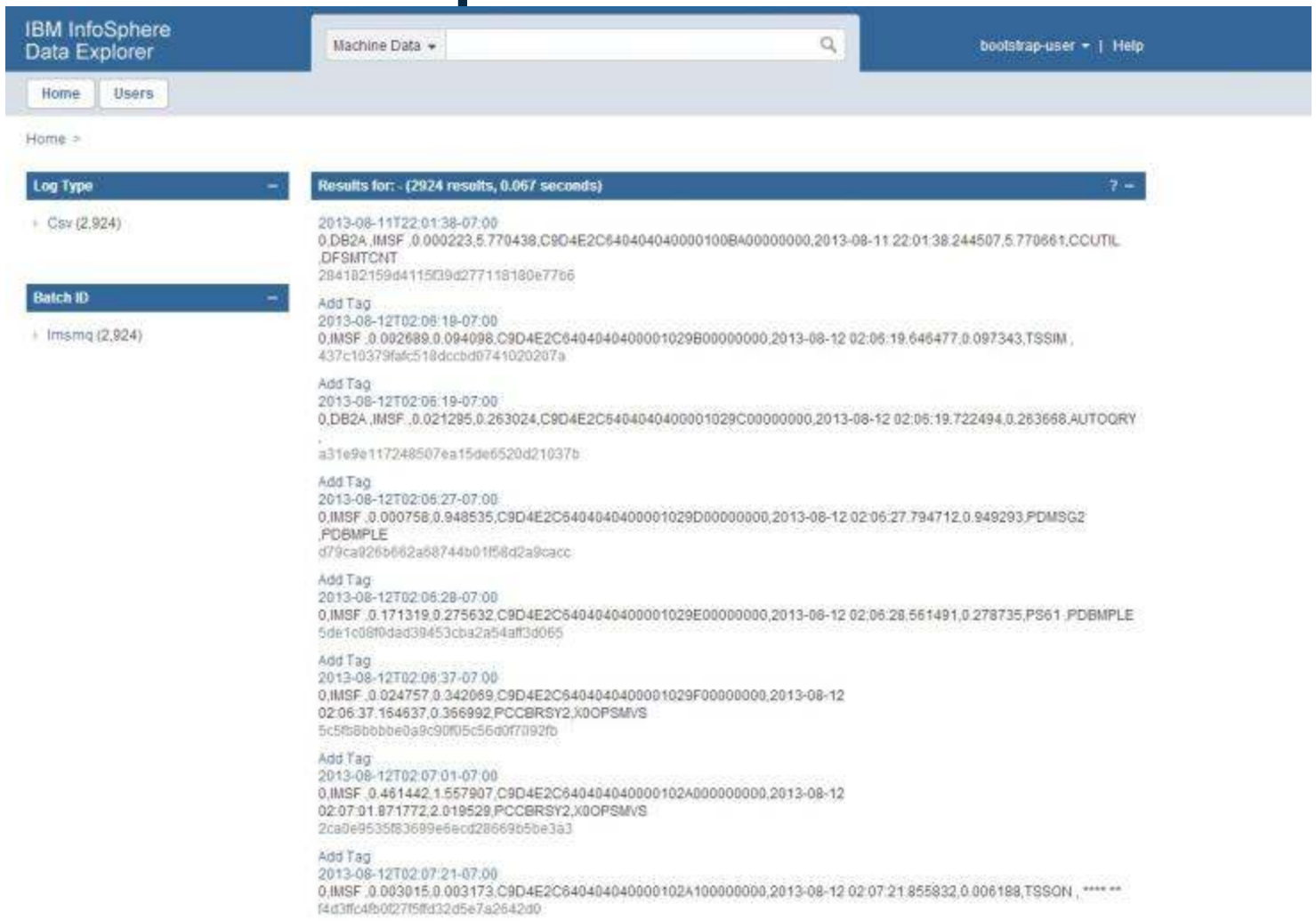
Transaction Analysis Workbench
- Log conversion

trace.log | messagecontext.txt | WSSecurityLibertyPluginInterceptor.java | passwd.txt | fileNames.txt | linuxMachineInfo.txt | imsbigd.cvs | imscics1.cvs | imsdb2.cvs | imsims1.cvs | IMSCSV3.

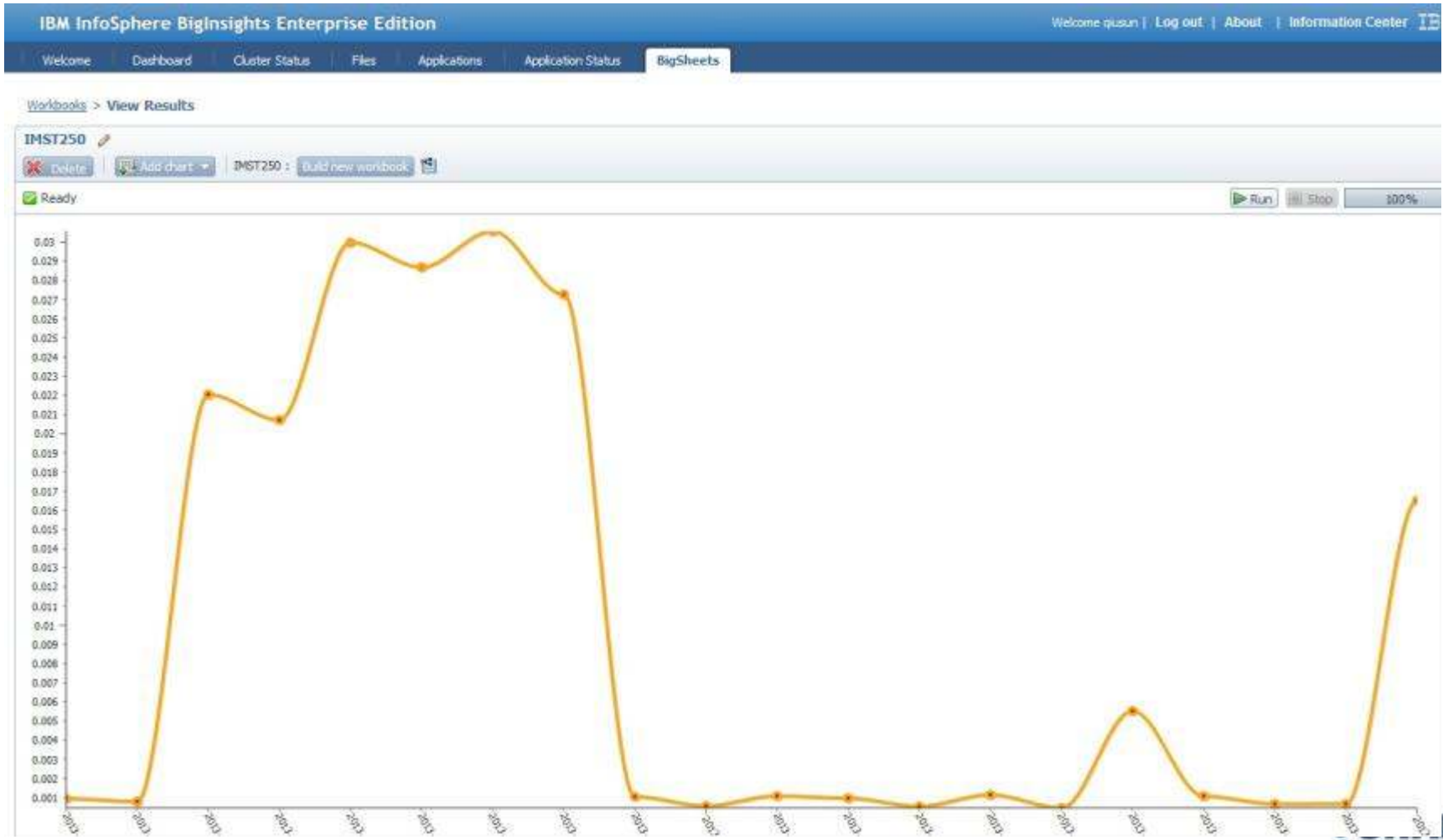
```
1  TIME;TranCode;Userid;RecToken;IMSID;InputQ;Process;TotalTm;CPUtime;FFGets;FFUpdats;FPCalls;FPGets;FPUpdats;TPESAF;ESAFName
2  2013-08-11 22:01:38.244507;CCUTIL  ;DFSMTCNT;C9D4E2C640404040000100BA00000000;IMSF    ;0.000223;5.770438;5.770661;0;;;;;;;DB2A
3  2013-08-11 22:01:44.465656;CCUTIL  ;DFSMTCNT;C9D4E2C640404040000100BB00000000;IMSF    ;0.009176;0.041509;0.050685;0;;;;;;;DB2A
4  2013-08-11 22:04:59.984936;TSSIM   ;        ;C9D4E2C640404040000100BC00000000;IMSF    ;8.716070;0.536743;9.252959;0;;;;;;;
5  2013-08-11 22:05:09.223175;AUTOQRY ;        ;C9D4E2C640404040000100BD00000000;IMSF    ;0.044868;0.638912;0.683318;0;;;;;;;DB2A
6  2013-08-11 22:06:01.494244;PDMSG2  ;PDBMPLE ;C9D4E2C640404040000100C000000000;IMSF    ;24.965797;0.467127;25.432924;0;;;;;;;
7  2013-08-11 22:06:26.796034;PS61    ;PDBMPLE ;C9D4E2C640404040000100C100000000;IMSF    ;0.130944;0.033681;0.035123;0;;;;;;;
8  2013-08-11 22:06:26.796171;PD32    ;PDBMPLE ;C9D4E2C640404040000100C200000000;IMSF    ;0.142469;0.952377;0.967167;0;;;;;;;DB2A
9  2013-08-11 22:06:27.595731;EMAILBMP;PDBMPLE ;C9D4E2C640404040000000000C000001C4;IMSF    ;0.302760;1.535334;1.545423;0;;;;;;;QY2A
10 2013-08-11 22:06:23.509104;TSSIM   ;        ;C9D4E2C640404040000100BE00000000;IMSF    ;1.011752;0.269375;1.281127;0;;;;;;;
```


Complete your session evaluations online at www.SHARE.org/Pittsburgh-Eval

Machine Data Analytics Accelerator - Watson Explorer Search



Machine Data Analytics Accelerator – Data Analytics using BigSheets



Agenda

- Big Data in an Information Driven economy
- Why start with System z
- IMS strategies for big data
- Summary / Call to action

Conclusion / Call to action:

- **For additional information including whitepapers and demos, please visit:**
  - IBM big data for z web site
  - Smarter Computing
  - Information Management System z
- **Education**
  - Free online education at bigdatauniversity.com
  - 145,000+ registered students
- **Further developments:**
  - Future webcast and announcements
  - World of DB2 for z/OS

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

- **Develop your own big data strategy** –Contact your local IBM sales representative to get started.

Take Action Now!

Thank You!

Your feedback is important to us