



Big Data Strategies for IMS Session #17752

Poonam Chitale pchitale@us.ibm.com August 11, 2015







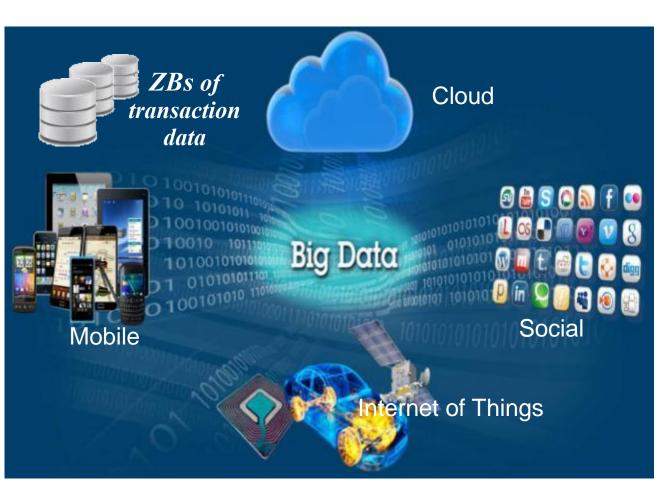


- Big Data in an Information Driven economy
- Why start with System z
- IMS strategies for big data
- Summary / Call to action



On a Smarter Planet, Unprecedented Changes are Occurring





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



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

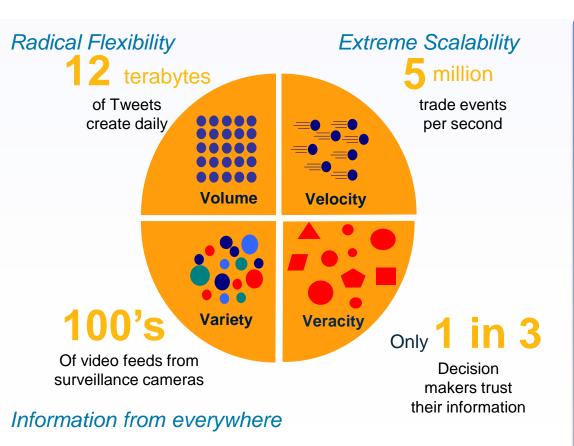
What	What is Big Data						
Web	Images	Maps	Shopping	News			

About 1,940,000,000 results (0.18 seconds)

- 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.

We've moved into a new era of computing





"We have for the first time an economy based on a key resource [Information] that is not only renewable, but selfgenerating.

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

- John Naisbitt



Agenda



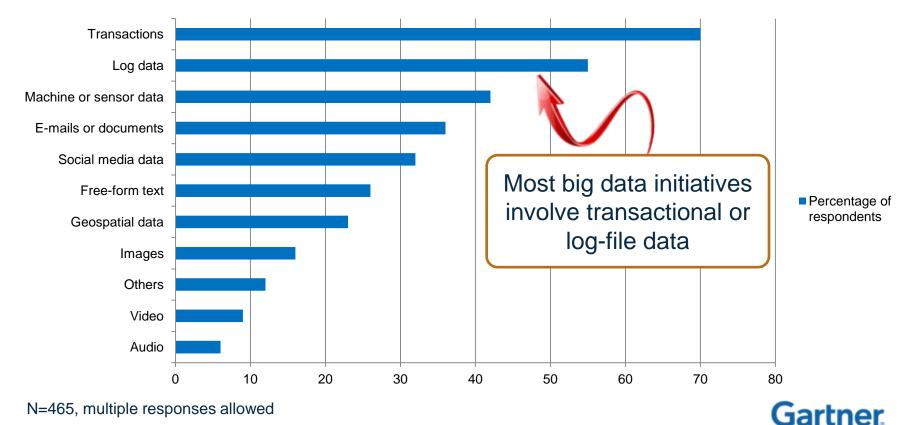
- Big Data in an Information Driven economy
- Why start with System z
- IMS strategies for big data
- Summary / Call to action



The Big Data Starting Point



Where are organizations getting the most return on Big Data projects?



N=465, multiple responses allowed

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

in Orlando 2

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 <u>differently</u> 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"

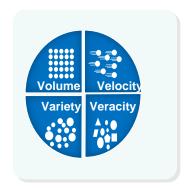
Complete your session evaluations online at www.SHARE.org/Orlando-Eval



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 360° View of the Customer



3. **Optimize operations** *with Operations Analysis*



4. Gain IT efficiency and scale with Data Warehouse Augmentation

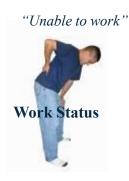


5. Prevent crime

with Security and Intelligence Extension in Orlando 2015

Fraud Detection – Claiming disability allowance







"Dude – awesome vacation"

Facebook Post



Make payment or investigate



Deterrent for fraudsters -**Cost Savings** for the business

Accelerator

zEnterprise

Data

Warehouse

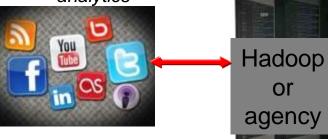
Integration

Business Analytics

DB2 for z/OS

IMS





Refined Search parameters from **OLTP** environment

Result set uploaded or dire imported into OLTAP DBMS

Result Set for further processing



Information Governance

Data Warehouse + modeling applications

Enterprise Integration and Governance... The key to success of incorporating Big Data

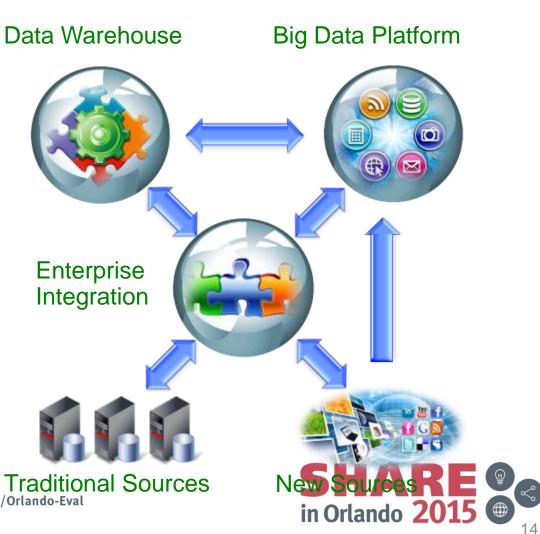


Information Integration

 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



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



Agenda

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



IMS and Analytics Use case

Solution



BI, dashboarding, reporting of IMS data



Cognos 10.2 BI



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



IBM InfoSphere BigInsights



Bring analytics to the data



IBM DB2AnalyticsAccelerator



 Visualize entire big data landscape



IBM WatsonExplorer



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

IMS and Analytics



- Predictive analytics
- Decision management
- Fraud detection



- SPSS
- Operational Decision Manager







Cognos 10.2 BI with IMS Data





Get Started Today!

developerWorks article available here

IBM. developerWorks. IBM Business Analytics Proven Practices: Best Practices for Using IMS Data in IBM Cognos BI Reports

- Certified against IMS 13 using IMS Open Database technology
 - Universal JDBC driver
- Real-time analytics



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 <u>merge</u> 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





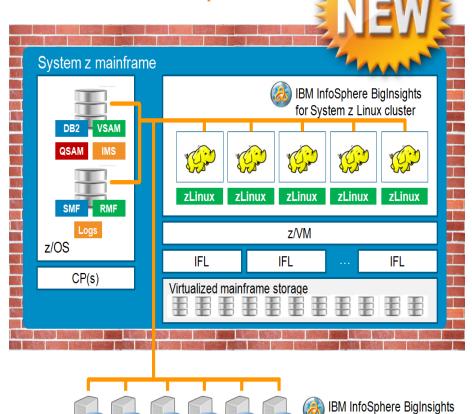
Complete your session evaluations online at www.SHARE.org/Orlando-Eval

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



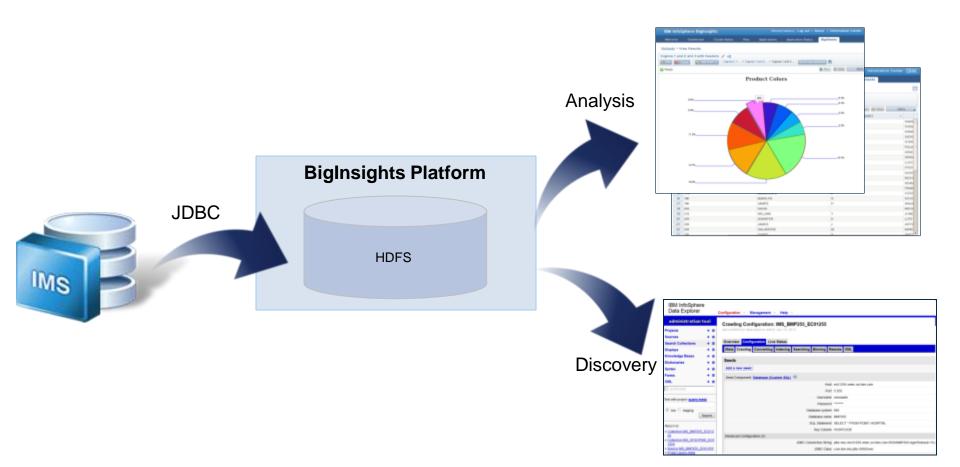
All "z" deployment for maximum security



External Cluster

IMS & IBM InfoSphere BigInsights High level overview

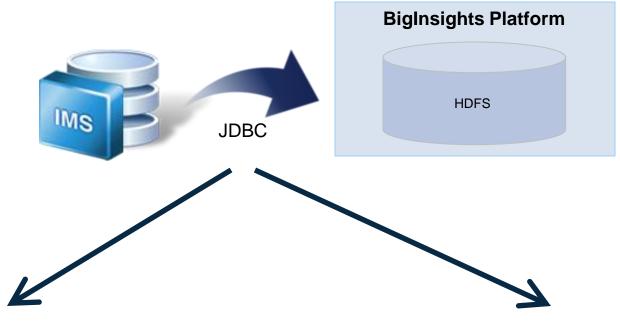




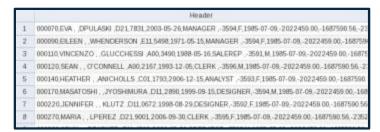


Import options

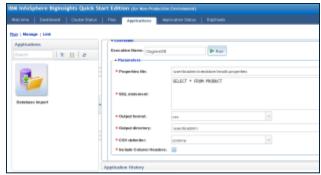




Sqoop Import



POJO Import





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.
```

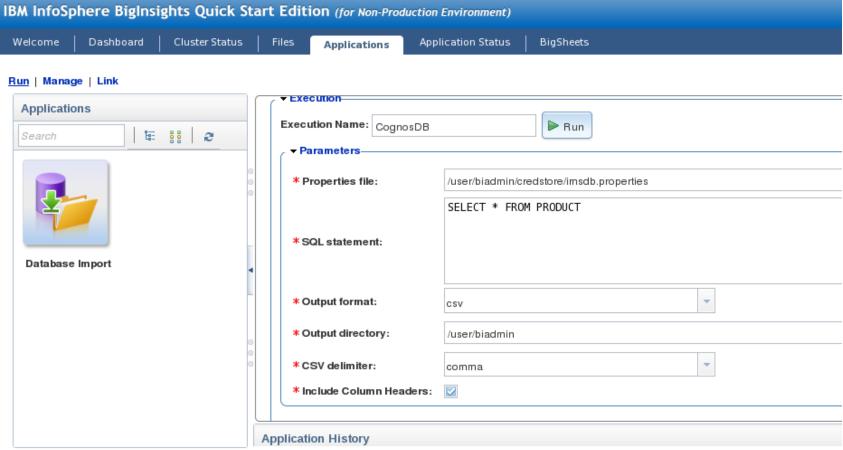
HDFS Output (below)

	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



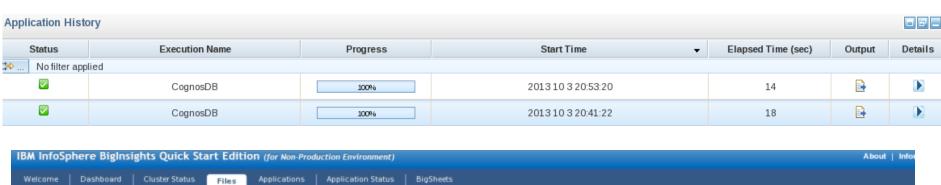
 Utilize the built in Database Import Application by providing the database connection parameters:

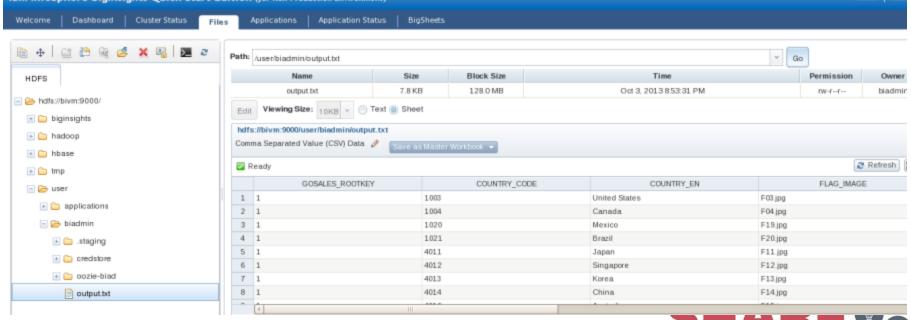


BigInsights Database Import Application



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





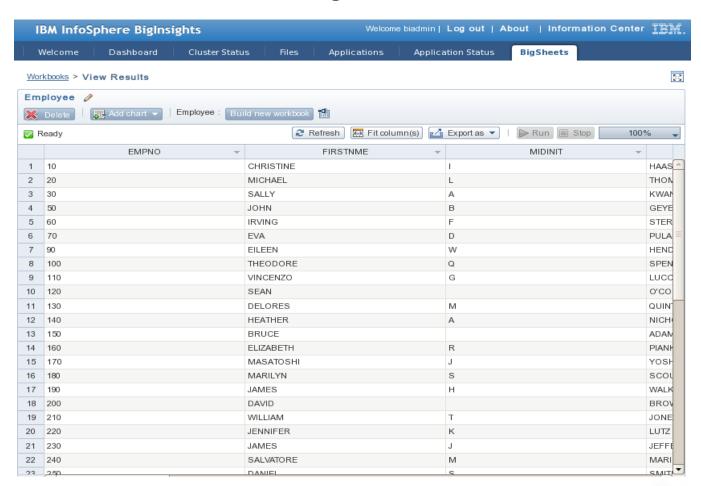
Complete your session evaluations online at www.SHARE.org/Orlando-Eval





BigInsights BigSheets

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





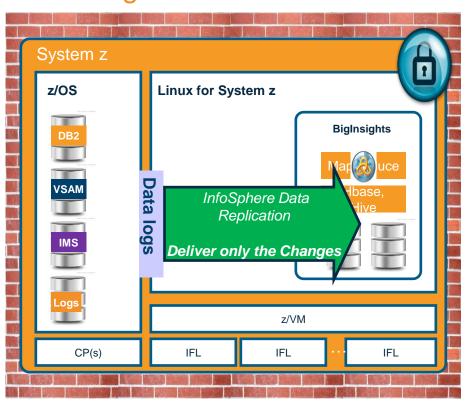
27

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!



IT Analytics for IMS

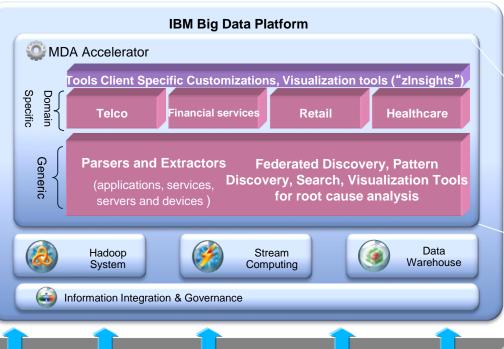


- Custom Applications

 Shrink Wrap Solutions

 Health Care Networking Insurance Telco "x2020" "Unity"
- Value: allow correlation of an entire ecosystem of application servers with IMS to provide deep insight, filtering, analytics, as well as faceted search capabilities

Log ingestion and analysis









IMS and IBM Accelerator for Machine Data Analytics

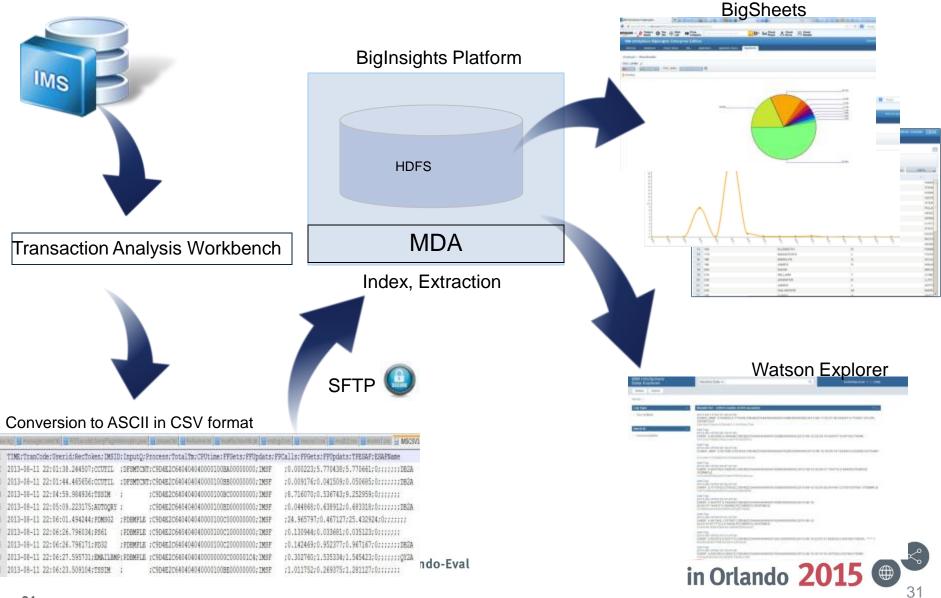


- **Consume log data produced by Transaction Analysis** Workbench
- Index and link transactions together across products (IMS, DB2, MQ, CICS, WebSphere)
- Make large amounts of IMS transactional log data available to the suite of BigInsights tools.



IMS log and MDA overview







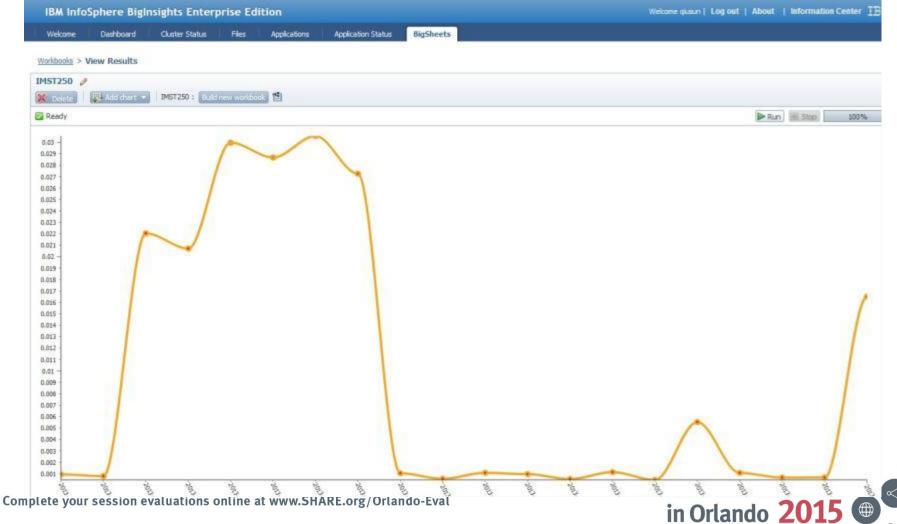


trac	log 📳 messagecontext.txt 📳 WSSecurityLibertyPlugi	nInterceptor.java 📙 pa:	sswd.bd	linuxMachineInfo.txt	imsbigd.cvs imscics1.cvs imsdb2.cvs	imsims1.cvs IMSCSV3.
1	TIME;TranCode;Userid;RecToken;IMS	ID; InputQ; Proce	ss;TotalTm;CPUti	me;FFGets;FFUpda	ts;FPCalls;FPGets;FPUpdats;TP	ESAF;ESAFName
2	2013-08-11 22:01:38.244507;CCUTIL	;DFSMTCNT;C9D	4E2C640404040000	100BA00000000;IM	SF ;0.000223;5.770438;5.77)661;0;;;;;;DB2A
3	2013-08-11 22:01:44.465656;CCUTIL	;DFSMTCNT;C9D	4E2C640404040000	100BB00000000;IM	SF ;0.009176;0.041509;0.05)685;0;;;;;;DB2A
4	2013-08-11 22:04:59.984936;TSSIM	; ; C9D	4E2C640404040000	100BC00000000;IM	ISF ;8.716070;0.536743;9.25	2959;0;;;;;;
5	2013-08-11 22:05:09.223175;AUTOQR	f; ; ; C9D	4E2C640404040000	100BD00000000; IM	SF ;0.044868;0.638912;0.68	3318;0;;;;;;DB2A
6	2013-08-11 22:06:01.494244; PDMSG2	;PDBMPLE ;C9D	4E2C640404040000	100C000000000; IM	ISF ;24.965797;0.467127;25.	432924;0;;;;;;
7	2013-08-11 22:06:26.796034;PS61	;PDBMPLE ;C9D	4E2C640404040000	100C100000000;IM	SF ;0.130944;0.033681;0.03	5123;0;;;;;;
8	2013-08-11 22:06:26.796171;PD32	;PDBMPLE ;C9D	4E2C640404040000	100C200000000; IM	ISF ;0.142469;0.952377;0.96	7167;0;;;;;;;DB2A
9	2013-08-11 22:06:27.595731;EMAILB	MP;PDBMPLE ;C9D	4E2C640404040000	0000C000001C4;IM	SF ;0.302760;1.535334;1.54	5423;0;;;;;;;QY2A
10	2013-08-11 22:06:23.509104;TSSIM	; ; C9D	4E2C640404040000	100BE00000000;IM	ISF ;1.011752;0.269375;1.28	1127;0;;;;;;



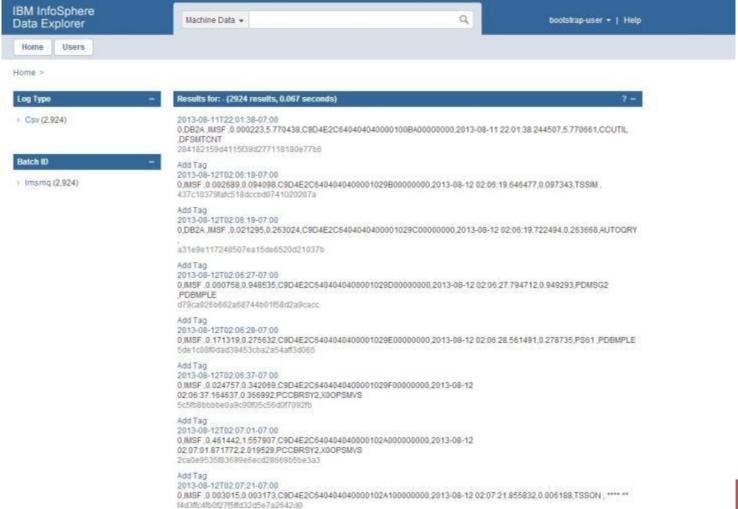


Machine Data Analytics Accelerator– Data Analytics using BigSheets





Machine Data Analytics Accelerator - Watson Explorer Search



Watson Explorer V10 delivers cognitive exploration

, visualize, and explore information from internal and external content through 360-degree information applications

Analyze, visualize, and discover insight in unstructured data through NLP and text analytics

Watson Explorer V10 Now available!



Interpret information to enhance, scale, and accelerate human expertise through cognitive capabilities

Cognitive Exploration

- Search and exploration across many different sources
- Content analytics
- Cognitive insights
- Delivered in a 360-degree information application





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

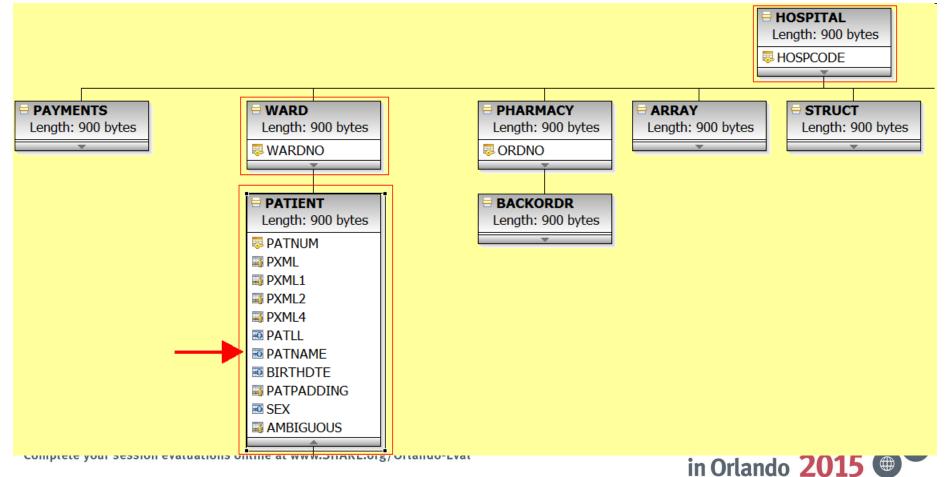
Converter Component: Database seeds support	
Туре-	In application/vxml-db ▼
Type-C	ut application/vxml-db ▼
Fallba	ck (unset)
Output forki	g (unset) 🕶
Nar	ie





Original IMS hierarchy for hospital database

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



Why use Watson Explorer



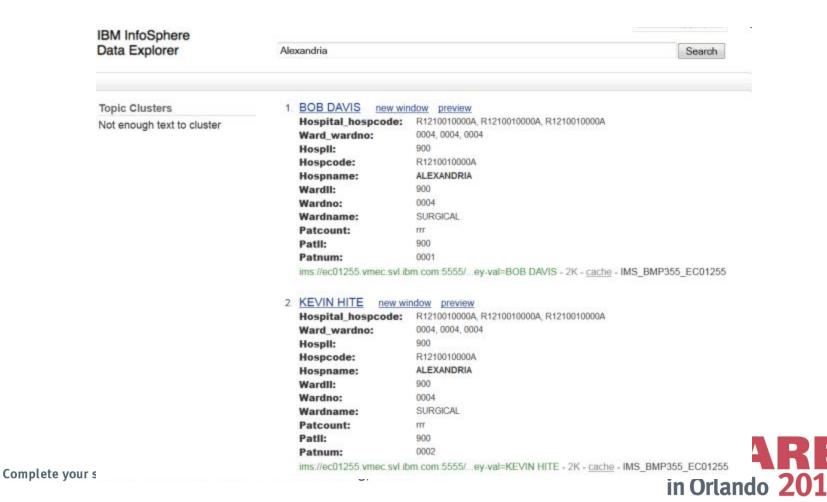
- Previously to change the schema so that the PATIENT information is at the top, a logical database needs to be created
- This requires a DBA to be involved and a time window when the new database resources can be brought online
- Watson Explorer allows indexes to be created dynamically and for better 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	Э
Data	Explorer	

Dermatology

Topic Clusters

Not enough text to cluster

1. WILLIAM LI <u>new window</u> preview

Hospital_hospcode: R1210020000A, R1210020000A, R1210020000A

Ward_wardno: 0002, 0002, 0002

HospII: 900

Hospcode: R1210020000A Hospname: SANTA TERESA

WardII: 900 **Wardno:** 0002

Wardname: DERMATOLOGY

Patil: 900 **Patnum:** 0001

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



IBM zEnterprise and DB2 Analytics Accelerator





The hybrid computing platform on zEnterprise

- Supports transaction processing and analytics workloads concurrently, efficiently and costeffectively
- 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

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

Accelerate business insights with IMS transactional data Introduction IMMS bytes of server or here to be proved of the world data? a large part of which is information in Mill parent of the world data? a large part of which is information in Mill parent of the world data? a large part of which is information and accounts on feed by remoderate data. INS applications and data continue to feed by the world or the segment (MCT) workshoth. INS data on System a server is which view of an the night works of the world removed to the world or the segment of the search of the segment of th

Get Started Today!

 Technical Whitepaper and "howto" guide available <u>here</u>



IDAA enablement steps

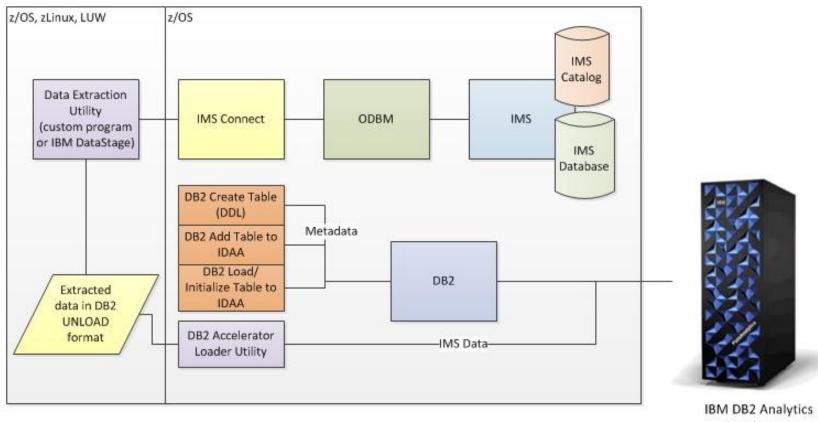


- Unload IMS data and produce a DB2 UNLOAD equivalent dataset
 - ETL platform can do this or home grown
- Create DB2 table
- Add table to IDAA
 - SYSPROC.ACCEL ADD TABLES
- Initialize the table in IDAA
 - SYSPROC.ACCEL_LOAD_TABLES
- Enable the table for acceleration
 - SYSPROC.ACCEL_SET_TABLES_ACCELERATION
- Load IMS data into IDAA
 - IDAA loader tool



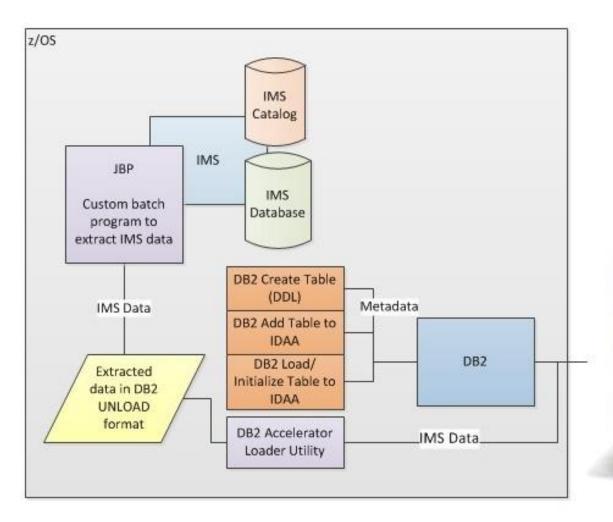
IDAA configuration with DataStage





IDAA configuration with custom Java loader







IBM DB2 Analytics Accelerator

in Orlando 2015

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. (pchitale@us.ibm.com,richtran@us.ibm.com)
- Develop your own big data strategy –Contact your local IBM sales representative to get started.

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

Take Action Now!







Thank You!

Questions? Comments?

