



7 Kasım 2012 - Çırağan Palace Kempinski

# IBM Connected 2012 Istanbul

Learn. Collaborate. Innovate.

## Cost Effective Business Analytics on System z

Avijit Chatterjee

IBM SWG Competitive Project Office



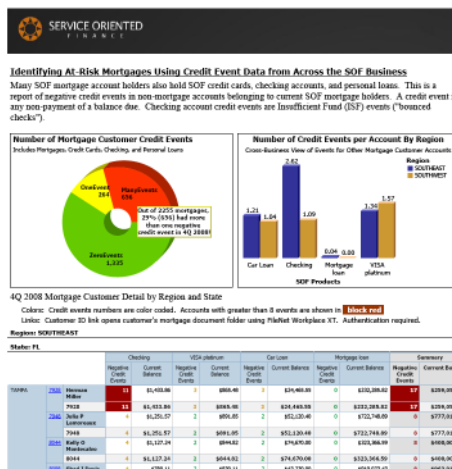
# Businesses Analytics Answers Key Questions That Drive A Competitive Edge

How are we doing?



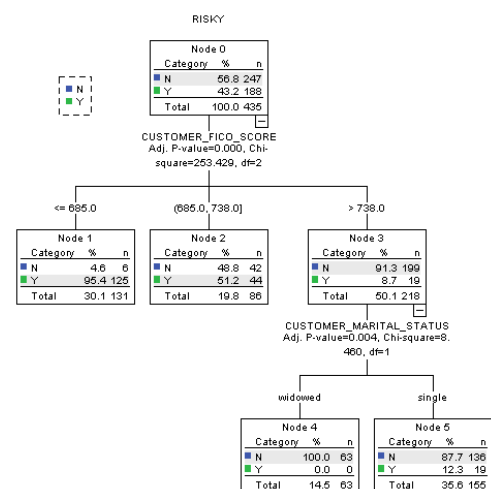
Dashboard  
Operational or Strategic

Why are we on/off track?



Query and Reporting  
Exploratory Analysis

What should we do next?



Analytics Statistics  
and Predictive Analytics



# Businesses Benefit By Using An Analytic Approach Over Intuition



**40% decline**  
in homicide rates



**600% increase**  
in cross-sell  
campaign



**\$13.8 Million**  
in cost savings

The more analytics a  
business uses, the better it  
performs



**PRIMERICA**  
**1000's of Reps**  
run their daily business using  
IBM Business Analytics



**80% decrease**  
in reporting time on top of  
Oracle e-business suite

OmnicomGroup



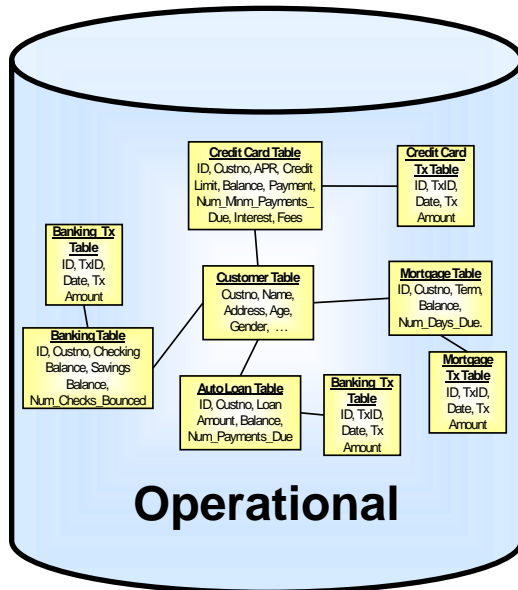
**\$200 Million**  
increase in cash flow

# There Are Two Approaches To Analytics

Operational Business Intelligence

Simple Queries

“How are we doing?”

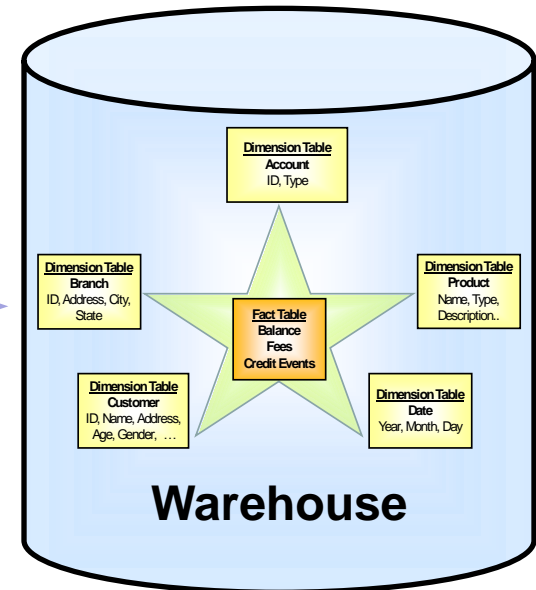


Schema organized to efficiently process simple queries  
Normalized data, no repetitions

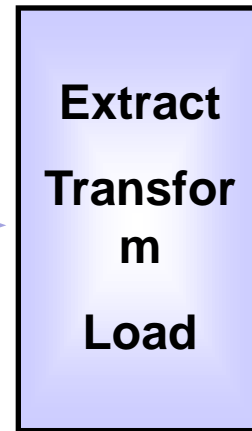
Deep Analytics

Longer Queries

“Why?”



Schema organized to efficiently answer longer queries  
De-normalized data, records repetition

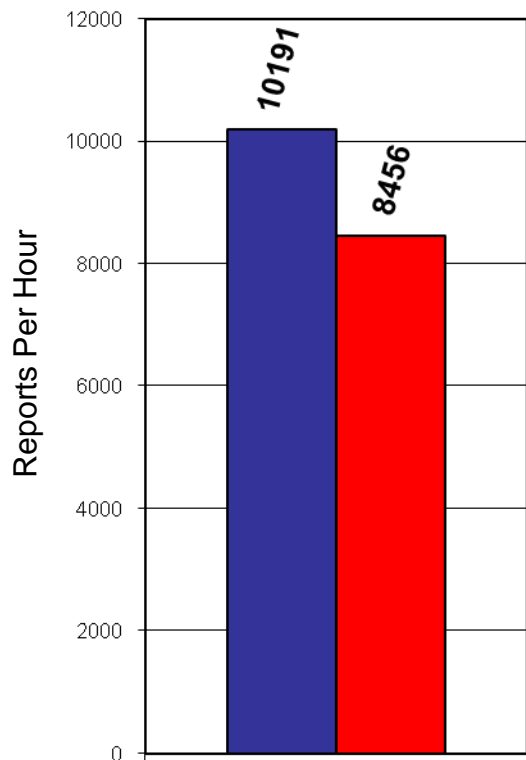


# DB2 For z/OS Is Optimized For Operational BI Queries

- z/OS Workload Management (WLM) optimizes resource sharing between OLTP and Analytical Queries
  - Minimize impact on OLTP performance
- Achieves near-linear scaling and high availability with parallel sysplex clustering
- Cost Based Optimizer provides best access path and query execution plan
  - Provides Hash access path for even faster access to a single row of a table
- DB2 10 supports up to 10x more concurrent users, and up to 20,000 concurrent connections per subsystem

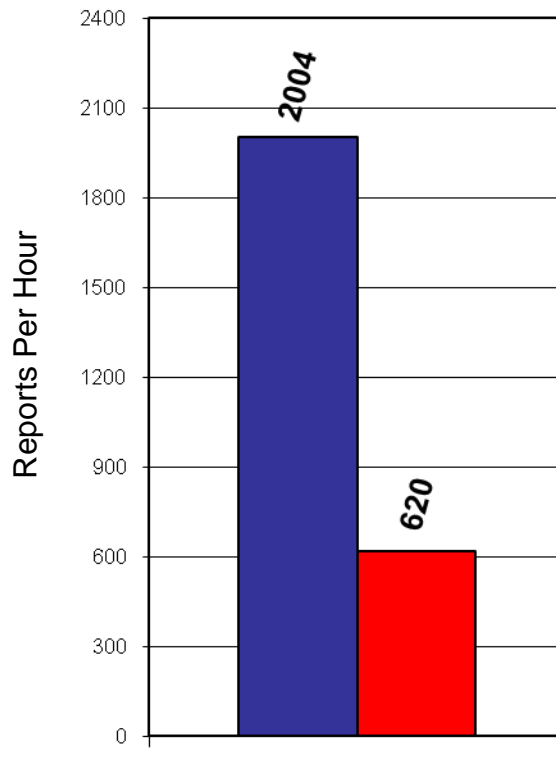
# A Large US Financial Institution – Performance Comparison Of Operational BI

1.2X More  
Small Reports



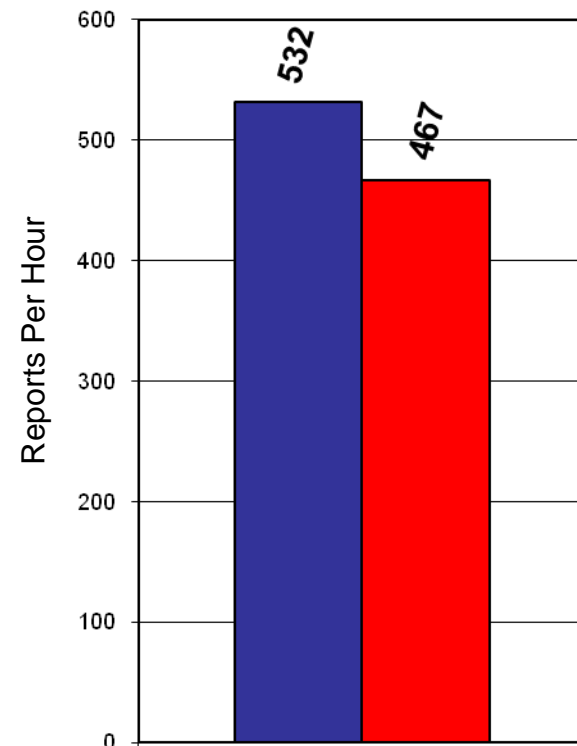
Reports Per Hour  
at 750GB data size  
(Higher is Better)

3.2X More  
Medium Reports



Reports Per Hour  
at 750GB data size  
(Higher is Better)

1.1X More  
Large Reports



Reports Per Hour  
at 750GB data size  
(Higher is Better)

# DB2 For z/OS Is Also Optimized For Data Warehouse Queries

- Data Warehouse workloads typically include a mix of simple, intermediate and complex queries
- Data is partitioned to increase parallelism and compressed to increase I/O performance
  - Takes advantage of on-chip compression coprocessors and dedicated I/O subsystem
- Cost Based Optimizer decides best execution plan for each query
  - Simple queries typically assigned to a single processing thread
  - Complex queries may be decomposed into operations that execute in parallel
  - Queries may be automatically rewritten to take advantage of pre-computed partial results in materialized query tables (MQT)
- Result: Optimum Throughput

# Add IBM DB2 Analytics Accelerator (Accelerator) For Even More Optimization

- A workload-optimized, blade-based appliance based on Netezza Technology
  - Storage integrated into the hardware rack
- Deeply integrated with DB2 for z/OS, and transparent to applications
  - Pre-load data from DB2 for z/OS into Accelerator at over 400GB/hr
- Significantly speeds up the response time for a wide variety of complex queries
  - Uses massively parallel processing architecture and data filtering technology at streaming speed
    - Field Programmable Gate Arrays (FPGAs)
- Drives down the costs of data warehousing and business analytics

***Breakthrough technology  
enabling new opportunities***



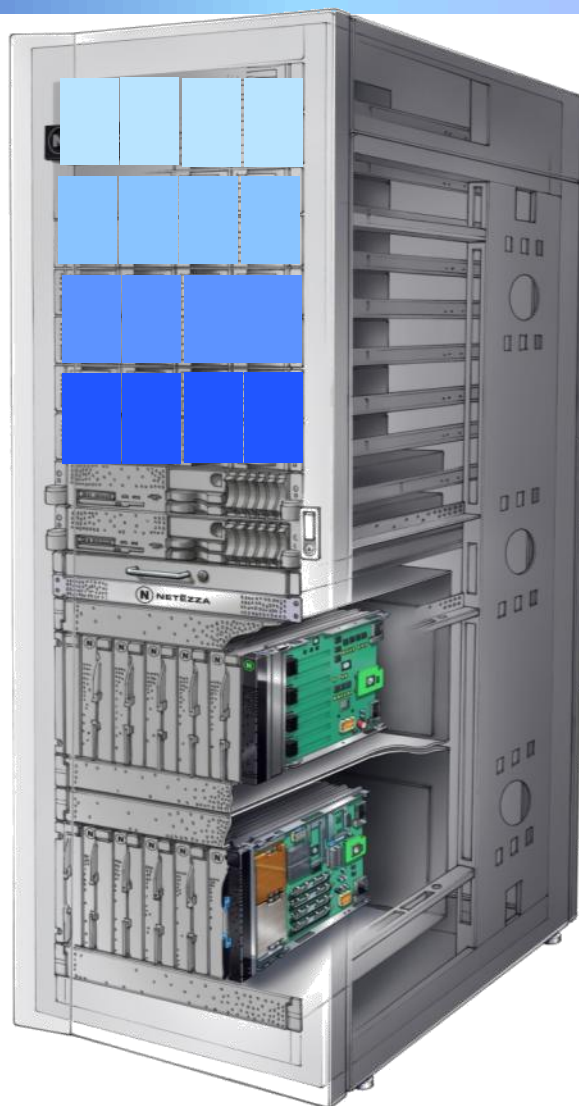


# Accelerator Leverages Massively Parallel Processing To Speed Up Complex Queries

Data partitioned across CPUs and storage

Query distributed to CPU/FPGAs which decompress and filter data in real-time.

SMP Host assembles results and returns



Storage

SMP Hosts

CPU/FPGA

*"...when something took 24 hours I could only do so much with it. But when something takes 10 seconds, I may be able to completely rethink the business..." SVP, Nielsen*

# Accelerator Speeds Up Complex Analytics Queries

Query	DB2 (Secs)	DB2 + IDAA (Secs)	Speed Up	Rows Reviewed	Rows Returned
Query 1	9,540	5	1,908x	2,813,571	853,320
Query 2	8,220	5	1,644x	2,813,571	585,780
Query 3	4,560	6	760x	8,260,214	274
Query 4	4,080	5	816x	2,813,571	601,197
Query 5	4,080	70	58x	3,422,765	508
Query 6	3,180	6	530x	4,290,648	165
Query 7	3,120	4	780x	361,521	58,236
Query 8	2,640	2	1,320x	342,529	724
Query 9	2,520	193	13x	4,130,107	137



*“We had this up and running in days with queries that ran over 1000 times faster”, AETNA*

*“We expect ROI in less than 4 months”, Swiss Re*

**Now with  
“trickle”  
feed!**

# Accelerator Speeds Up SAP Business Warehouse Too



- Accelerates SAP Business Warehouse on DB/2 for z/OS
- Dramatic decrease in elapsed time for SAP Business Warehouse ad-hoc reporting

No	Description	Records read	Records returned	DB2 [sec]	IDAA [sec]	Speed Up
1	Simple mass aggregation	17116647	21	117	0.78	150
2	Query #1 + 70% filter	11980812	21	94.2	0.86	110
3	Query #1 + 30% filter	5133708	21	54.8	0.82	67
4	Query #1 + 10% filter	1710293	21	17.6	0.87	20
5	Screwed data, low filtering	10790019	21	96.8	2.47	39
6	Screwed data, high filtering	24	14	7.28	0.83	9
7	Many restrictions	3805941	21	128	7.65	17
8	Navigational attributes	823646	21	17.1	1.27	13
9	Navigational attributes + selective condition	811	21	15.8	1.17	14
10	Open value ranges	2006	21	19.6	3.52	6
11	Hierarchy	1653981	21	17.6	0.97	18
12	Hierarchy + selective condition	55068	21	38.6	0.98	39
13	Restricted key figures on 2 dimensions	1314964	1948	207	7.22	29
14	Query #14 + hierarchy	132564	1499	> 1000	1.27	> 787
15	Calculated key figures (OLAP)	5321586	10	57.8	2.37	24
16	OR linked values	6212609	13	40.5	0.92	44
17	Non uniform data distribution	11016253	13	31.2	0.99	32
18	Selective line item	1724	1706	0.71	1.17	0.6
19	Non-selective line item	115481	68619	33.8	1.36	25
20	All together	3087692	468	87.7	4.42	20

Table 2: Dedicated Query Test on a 18-million-records InfoCube

# BI Day Workload – A Typical Day's Worth Of Analytics Reports

BI Day Fixed Execution Test: Number of reports

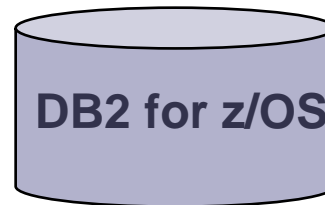
4	3	21	179	99	29,078	84,854	19,600	27,328
Complex Report 1	Complex Report 3	Intermediate Report 9	Intermediate Report 10	Intermediate Report 11	Simple Report 2	Simple Report 4	Simple Report 5	Simple Report 6

4 connections

20 connections

56 connections

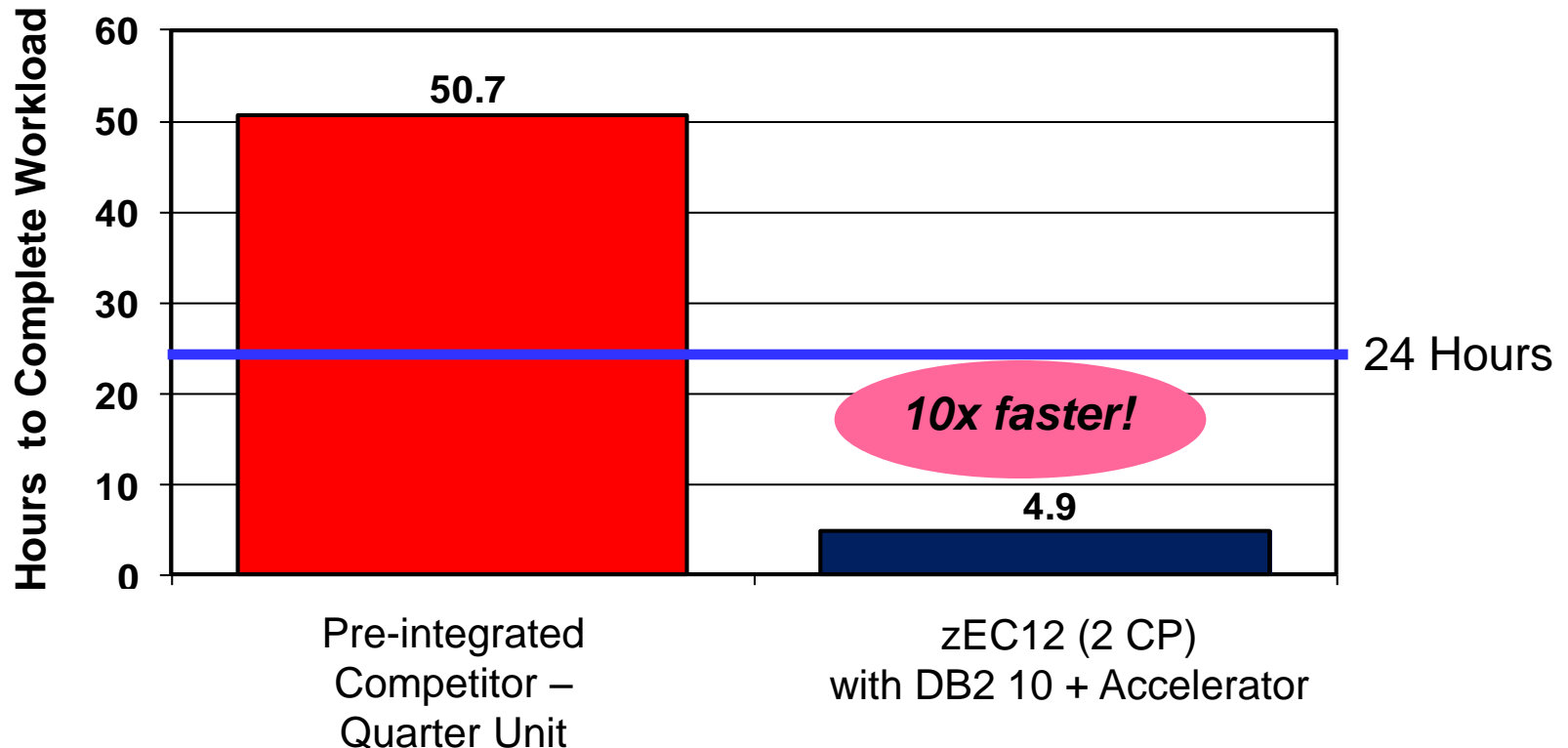
- 80 concurrent connections to DB2 for z/OS
- All short running SQL queries processed by DB2 for z/OS
- Long running SQL queries directed to accelerator



Each report executes one or more SQL queries

# Continuous Improvements For Analytics Workloads – A Typical Work Day

Elapsed Time To Complete BI Day Fixed Execution Test  
(10 TB – 161,166 concurrent executing reports)



Based on BI Day Tests. Lower is better. Performance numbers may vary based on workload profiles.

# zEnterprise Is Optimized For Operational Analytics

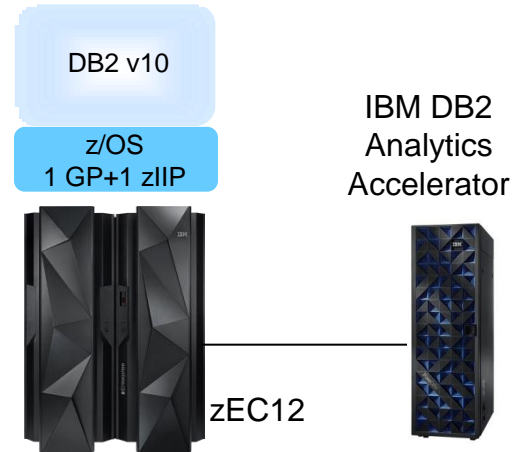
**Standalone  
Pre-integrated  
Competitor  
Quarter Unit**



**Unit Cost (3yr TCA) \$905/RpH**

Workload Time	3,043 mins
Reports per Hour (RpH)	3,178
Competitor ¼ Rack (HW+SW+Storage)	\$2,876,561

**IBM zEnterprise Analytics  
System 9700**



**Unit Cost (3yr TCA) \$71/RpH**

Workload Time	294 mins
Reports per Hour (RpH)	32,891
zEC12 (1 GP + 1 zIIP, HW+SW+50TB Storage) + Accelerator	\$2,337,400

**12x Price Performance!**

Source: Customer Study running 161,166 concurrent reports. Intermediate and complex reports automatically redirected to IBM DB2 Analytics Accelerator for z/OS. Results may vary based on customer workload profiles/characteristics. Note: Indicative 9700 pricing only internal to IBM, quotes to customer require a formal pricing request with configurations.

# Accelerator Trumps Teradata In Execution Of Complex Analytics Queries

## Teradata



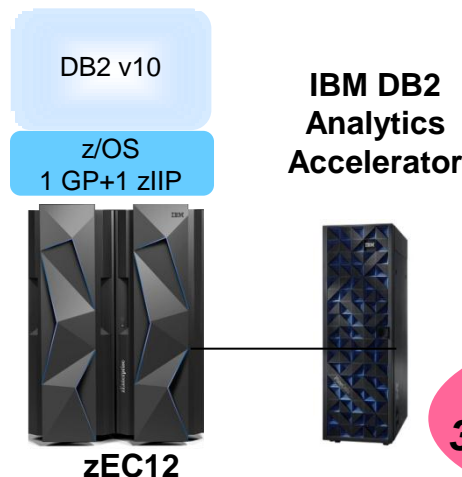
**6650H**

(current generation)

**Unit Cost (3yr TCA) \$330K/QpH**

Workload Time*	1,591 secs
Queries per Hour (QpH)	9.05
Teradata 6650H 1-Node (HW+SW+Storage)	\$2,946,046

## IBM zEnterprise Analytics System 9700



**zEC12**

(current generation)

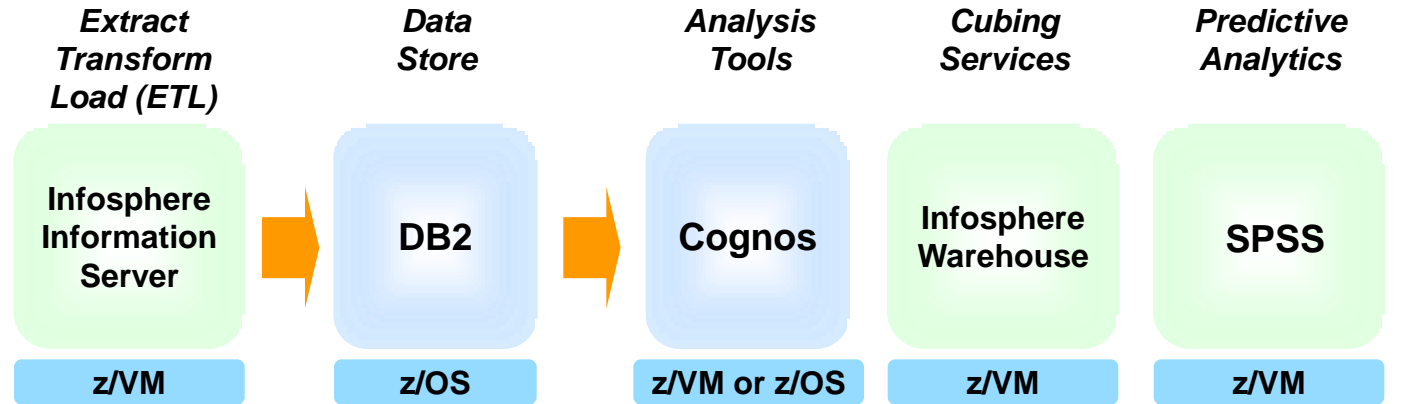
**Unit Cost (3yr TCA) \$10K/QpH**

Workload Time*	60.98 secs
Queries per Hour (QpH)	236
9700 : zEC12 (1 GP + 1 zIIP) + DB2 Analytics Accelerator (HW+SW+Storage)	\$2,337,400

**26x Performance  
33x Price Performance**

Source: Customer Study running 4 complex analytical queries across a 353GB Data Mart, Total User Data shared across 8 older generation Teradata Servers was 15TB. Customer workload was not from any benchmark applications, nor are they based on any benchmark standard. As such, customer applications, differences in the stack deployed, and other systems variations or testing conditions may produce different results and may vary based on actual configuration, applications, specific queries and other variables in a production environment. List Prices used in the comparison. Teradata 6650H price available from <ftp://public.dhe.ibm.com/software/data/sw-library/infosphere/analyst-reports/ITG-ISAS-Exadata-Teradata.pdf>. The 3 year total cost of acquisition includes expected hardware, software, service & support. Prices will vary by country. Results will vary based on actual configuration, applications, specific queries, and other variables in a production environment. Users of this document should verify the applicable data for their specific environment. Contact IBM and see what we can do for you. \*Throughput projected for Teradata 6650H using [TPerf data](http://www.monash.com/uploads/Teradata-Active-EDW-6660-6680.pdf) (<http://www.monash.com/uploads/Teradata-Active-EDW-6660-6680.pdf> : Rated TPerf per 6650H node 121.5) . Throughput for zEnterprise Analytics System 9700 was based on lab measurements for equivalent workload performance improvement on upgrading to DB2 10 for z/OS on zEC12

# Run A Complete Portfolio Of Analytics Software On IBM zEnterprise EC12



## IBM Smart Analytics System –

A comprehensive packaged solution including hardware, OS, and business analytics software



zEC12

Accelerator

*Full function analytics on the same platform*



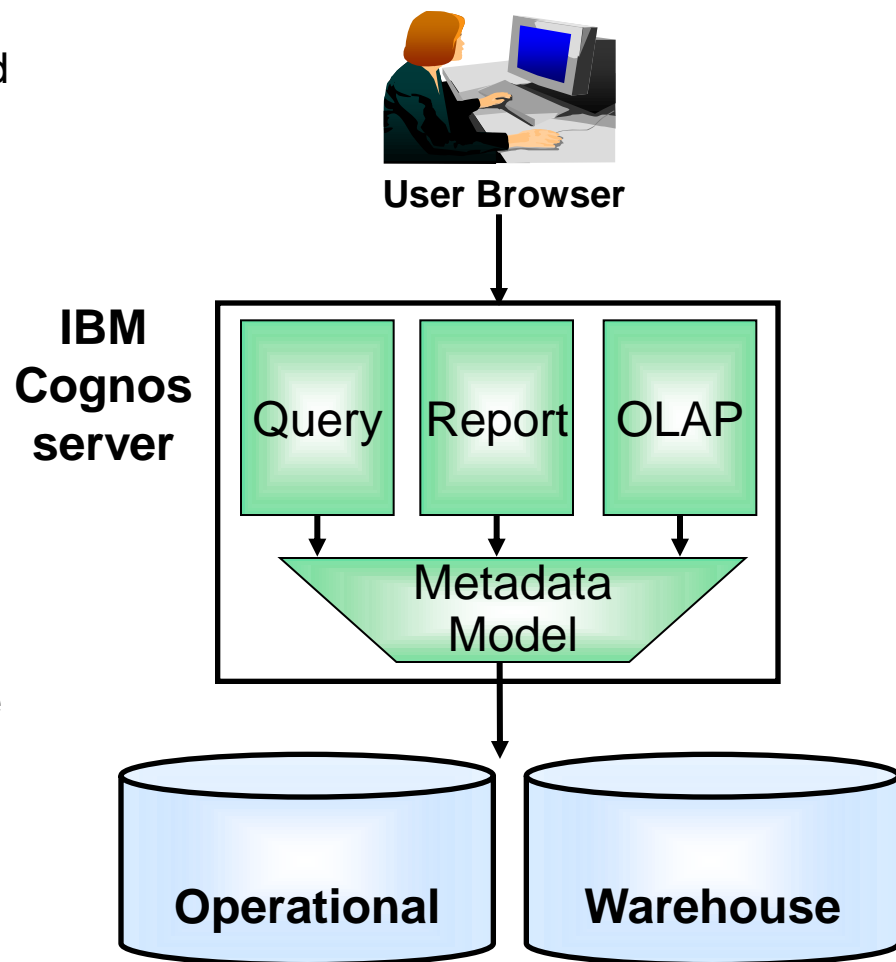
# Cognos Can Generate Reports And Dashboards For Operational BI And Deep Analytics Queries

## ■ People-centric

- ⌞ Server based business analytics accessed via browser
- ⌞ Consistent user interface for different analytic activities
- ⌞ Reuse new intelligence assets
- ⌞ Built-in collaboration and social networking
- ⌞ Threaded discussions, activities, and notifications

## ■ Easy to deploy and manage

- ⌞ Implemented in Java, runs on WebSphere
- ⌞ Scales up and out across heterogeneous hardware and operating systems
- ⌞ Runs on Linux on System z



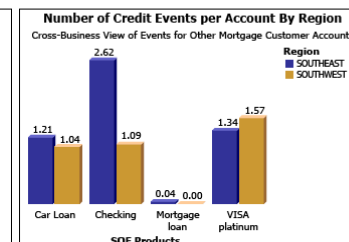
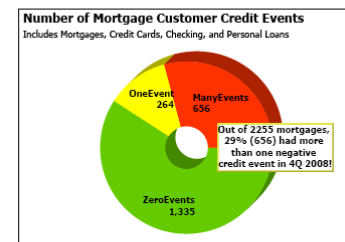
# DEMO: Use Cognos To Identify New Business Insights From The Data Warehouse

1. Reviewing Mortgage data provides false impression of credit risk
2. Report looking at negative credit events (bounced checks, missed payments) of customer across all accounts (Credit Card, Checking, etc...), identifies high-risk mortgages
3. Assess risk of flagged customer using FileNet case management data (Appraisal) linked to the Report



## Identifying At-Risk Mortgages Using Credit Event Data from Across the SOF Business

Many SOF mortgage account holders also hold SOF credit cards, checking accounts, and personal loans. This is a report of negative credit events in non-mortgage accounts belonging to current SOF mortgage holders. A credit event is any non-payment of a balance due. Checking account credit events are Insufficient Fund (ISF) events ("bounced checks").



## 4Q 2008 Mortgage Customer Detail by Region and State

Colors: Credit events numbers are color coded. Accounts with greater than 8 events are shown in **block red**

Links: Customer ID link opens customer's mortgage document folder using FileNet Workplace XT. Authentication required.

Region: **SOUTHEAST**

State: FL

			Checking		VISA platinum		Car Loan		Mortgage loan		Summary	
			Negative Credit Events	Current Balance	Negative Credit Events	Current Balance	Negative Credit Events	Current Balance	Negative Credit Events	Current Balance	Negative Credit Events	Current Balance
TAMPA	<a href="#">7226</a>	Herman Miller	11	\$1,433.86	3	\$865.48	3	\$24,465.55	0	\$232,285.82	17	\$259,050.71
		7928	11	\$1,433.86	3	\$865.48	3	\$24,465.55	0	\$232,285.82	17	\$259,050.71
	<a href="#">7246</a>	Julia P Lamoreaux	4	\$1,251.57	2	\$891.85	2	\$52,120.40	0	\$722,748.89	8	\$777,012.71
		7948	4	\$1,251.57	2	\$891.85	2	\$52,120.40	0	\$722,748.89	8	\$777,012.71
	<a href="#">8044</a>	Kelly O Montcalvo	4	\$1,127.24	2	\$844.82	2	\$74,670.00	0	\$323,366.59	8	\$400,008.65
		8044	4	\$1,127.24	2	\$844.82	2	\$74,670.00	0	\$323,366.59	8	\$400,008.65
	<a href="#">8085</a>	Shad I Davis	4	\$780.11	2	\$830.11	2	\$43,230.00	0	\$919,073.43	8	\$963,913.65

*At risk customers are identified*

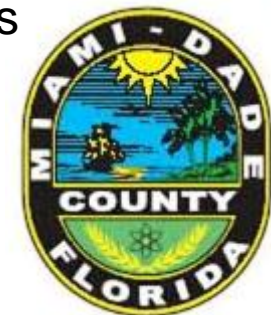
# Miami-Dade County Runs Cognos On Linux On System z

## Requirements:

- New Federal reporting requirements
  - Every new system, every new solution, every new application required a business intelligence component
- Existing distributed-based Cognos 8 system
- Required a standardized enterprise BI solution with:
  - Higher capacity to support expected growth
  - Fewer overall resource requirements (hardware, software, labor, etc.)
- Had available IFL's on System z

## Results:

- Moved Cognos 8 BI deployment from distributed to System z in 11 days
  - Consolidate multiple deployments to a single platform
  - Consolidate multiple disparate data sources
  - Single point for BI administration
  - Offer a complete disaster recovery plan
  - Additional green savings
- Easily meet new requirements for growth and savings



# SPSS Enables Customers To Predict Future Events And Drive Better Business Outcomes

## Capture

Data Collection delivers an accurate view of customer attitudes and opinions

IBM SPSS  
Data Collection



## Predict

Predictive capabilities bring repeatability to ongoing decision making, and drive confidence in your results and decisions

IBM SPSS  
Statistics\*/Modeler\*/Text Analytics



Platform

Pre-built Content



## Act

Unique deployment technologies and methodologies maximize the impact of analytics in your operation

IBM SPSS  
Decision Management  
Collaboration & Deployment \*

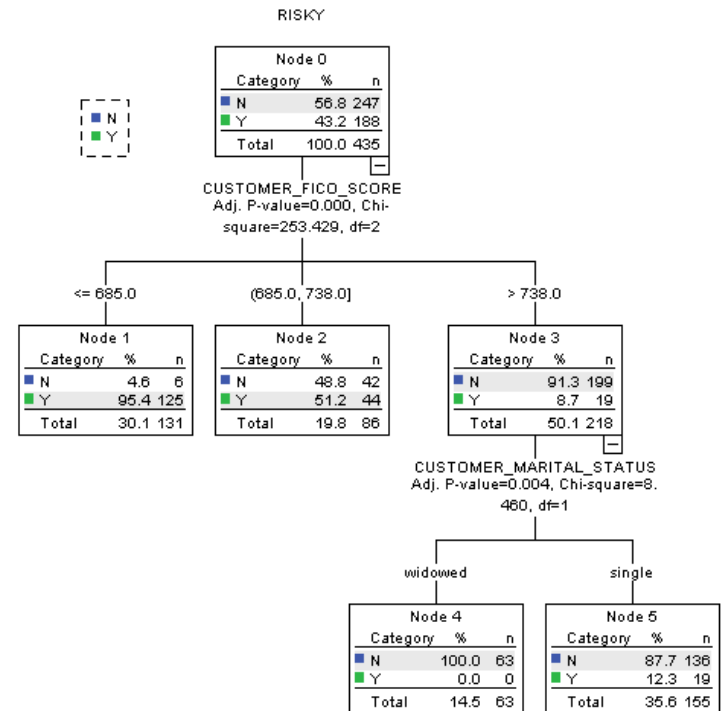


\* Runs on Linux on System z

Cost Effective Business Analytics

# DEMO: Discover Rules For Identifying Risky Customers Using SPSS Statistics

1. Load data from Data Warehouse on DB2 for z/OS into SPSS Statistics
2. Pre-process the data to create new attributes for quantifying negative credit events across different product lines and create a risk flag for mortgage
3. Run Decision Tree to discover rules for characterizing risky customers
4. Evaluate if Herman Miller is classified as “RISKY” by applying the Decision Tree rules



- Credit Limits identified for characterizing risky customers
- Use these credit limits for automated loan approval process

# Run End-To-End BI On zEnterprise To Reduce Costs And Improve Reliability

- 60-70% of operational data resides on System z\*
- zEnterprise offers a fully integrated, optimized solution on one platform
  - From operational data to business analytics
- Consolidating data warehouses on zEnterprise with Accelerator can reduce costs by 75%
- Cognos and SPSS add unmatched predictive intelligence



\* Source <http://www.ibmssystemsmag.com/mainframe/trends/whatsnew/The-Mainframe-at-a-Crossroads/>