A Behind the Scenes Look at the IBM Smart Analytics System 9700

Session Number 1353A

Gary Crupi, IBM

IBM Software
Information On Demand 2011

Topics for Today

- A short look beyond the typical importance of “Real Data Insights”
- What is the IBM Smart Analytics System solution for System z?
- How is the solution developed and validated?
- Metrics / Results / Lessons Learned

A quick survey of the typical importance of information slides

Drowning in a sea of data

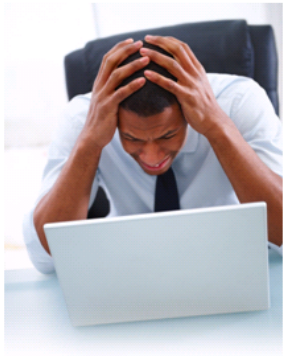
1 in 3

Business leaders **frequently make decisions** based on information they don't trust, or don't have

1 in 2

Business leaders say **they don't have access to the information they need** to do their jobs

Today, many business users are not getting to the information they need, when they need it



60%+ of CEOs need to do a better job capturing and understanding information rapidly in order to make swift business decisions

47% of users don't have confidence in their information

59% of users say that they miss information that might be of value to their jobs because they can not find it

27% of managers time is spend searching for information

50% of the information they obtain has no value to them

Instrumented, Interconnected, Intelligent

Managing the business with metrics

How are we doing?	*Why are we on/off track?*	*What should we do next?*
DASHBOARDING	REPORTING, QUERY	ANALYSIS

Tracking key performance indicators (KPIs) to quickly identify changes and their root causes in the business, creating the agility to response to market opportunities and challenges

Only 14% of organizations can answer questions at the time of decisions

Aberdeen Mar 2011

Analytics-driven Organizations Can…

Detect patterns

…rapidly map symptoms to cures

- *Clinical research studies*
- *Identify best practices*
- *Reduced operating costs*

When is System z interesting for date warehousing and business intelligence / analytics?

- **Start with an existing investment in System z infrastructure**
  - HW/SW/Applications, Processes and People

- **Add the need to see value in one or more of the following areas:**
  - Proximity of the operational production data for security, low latency or application integration
  - Quality of Service provided by System z
    - Continuous Availability, DR Ready, Secure, Responsive Mixed Workload Management, Instrumentation
  - Unique opportunity to mitigate risk and deliver the right information at the right time in a granular, incremental fashion
    - And integrating Dev/Test/QA in a single server

- **With requirements for one of several use cases:**
  - Enterprise Data Warehouse
  - Data distribution hub
  - High speed OLAP data mart
  - Data Mart Consolidation
  - Operational data store for online analytics
  - Competitor data archiving

Top 5 System z customer trends I am seeing

1. **Operational Data Store (ODS) / Near Real-Time ODS / Evolution to ODS Mix**

2. **Consolidation / Migration of Competitive “One Offs”**

3. **Enhancing and Modernizing legacy decision support / reporting systems**

4. **Considering a move to System z from outdated technologies**

5. **Reporting systems outgrowing core OLTP / including “sneaky growth”**

What is the IBM Smart Analytics System solution for System z?

The IBM Smart Analytics System 9600 – June 2010

- Is an **integrated hardware, software and services offering** that enables customers to quickly and cost effectively capitalize on game changing analytics across an enterprise
- Delivers an **expanding portfolio of easy to deploy business analytics**, that seamlessly integrate into operational fabric of a business.
- Enables a centralized view of the business, with an highly available, **advanced workload manager** that can easily prioritize critical queries within a large pool of queries.
- Drives reductions in costs with a **highly available infrastructure**, causing customers to reevaluate the mainframe.

IBM Smart Analytics System 9700 / 9710 Goals

1. **Refresh the entire stack and create THE Transactional Analytics Workhorse**
2. **Hybrid ready to provide a SINGLE transaction and deep analytics server**
3. **Provide on platform, data mining tools**
4. **Uphold the standard of delivering an integrated, tuned and validated stack**
5. **Provide a starter set at an attractive price**
   - Help customers to start small and grow incrementally

Announcements

- **IBM Smart Analytics System 9700 & 9710** 12 Oct
  - *The Evolution of the IBM Smart Analytics System 9600*

- **IBM DB2 Analytics Accelerator-** ***Powered by Netezza*** 12 Oct
  - *The next version of the IBM Smart Analytics Optimizer*

IBM Smart Analytics System 9700 / 9710

Mixed Workloads for Next Generation Business Analytics

What is it?

The next generation of the IBM Smart Analytics System 9700; an integrated solution of hardware, software and services that enables customers to rapidly deploy cost effective game changing analytics across their business. Offered in **standard and foundational configurations**.

The IBM Smart Analytics System 9710 is z114 based and offered at an entry level cost in a **foundational configuration (can customize to a standard configuration).**

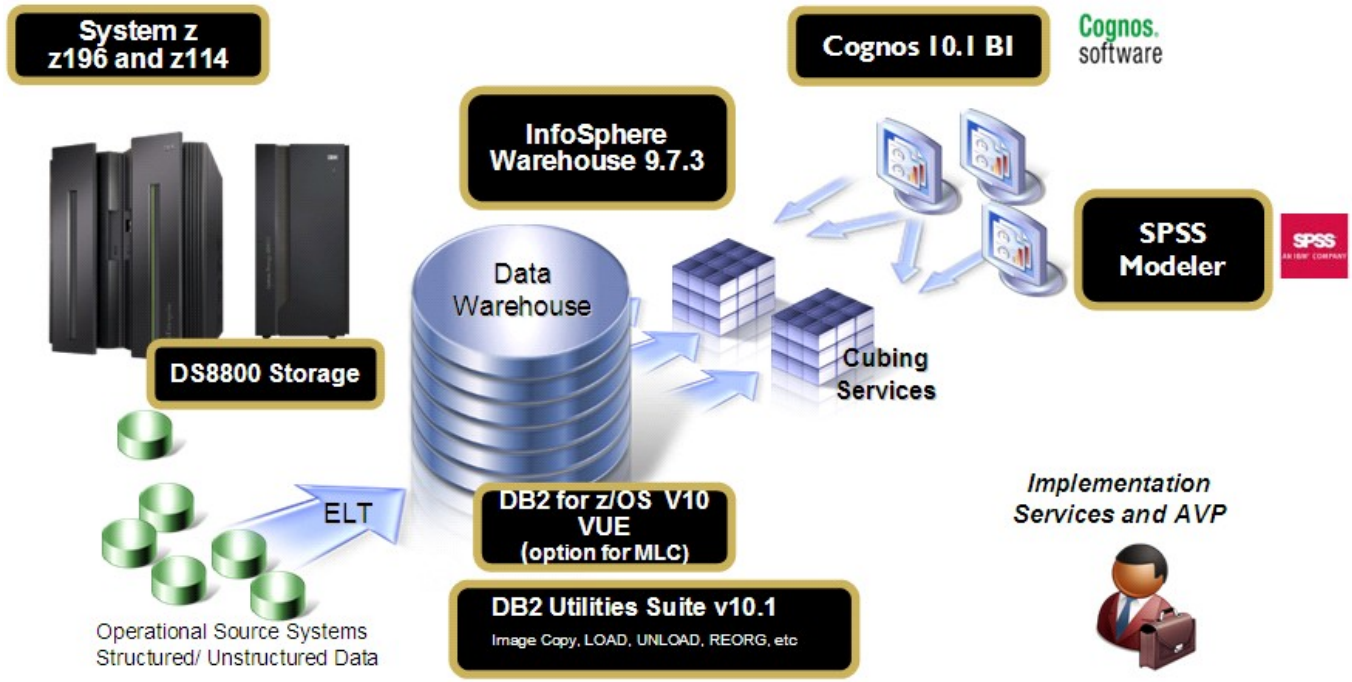
How is it different?

- *Secure, Available Business Analytics*
  - Rapidly delivers analytic information to decision makers at the time of decision.
  - New environment for the availability, reliability and scalability necessary to stay aligned with the operational systems
- *Simplified administration*
  - Appliance-like delivery
  - Faster deployment at lower cost.
  - Leverages customers existing disaster recovery, and backup processes
- *Proven Operational Characteristics*
  - Extends the qualities of service of System z.
  - Reduces risk through extending System z manageability and security across the entire system.
- *High Value Operational BI*
  - Cost effective way to drive daily operational decisions

Delivering business results in days, not months

IBM Smart Analytics System 9700 / 9710*

Standard Configuration



DB2 for z/OS and Utilities

- Range Partitioned Data
- CPU Parallelism
- Data and Index Compression
- DB2 Data Sharing

InfoSphere Warehouse for DB2 for z/OS

- SQW and Cubing Services
- Runs on Linux on System z targeting DB2 for z/OS
- DBMS provided separately

IBM Cognos BI for Linux on System z v10.1 and SPSS Modeler

- Consolidated and high server utilization rates

System z196 (z114 custom for standard cfg)

- Hardware gold standard
- Reduced costs via specialty processors
- Single-Server reduced complexity, failure points, and latency
- Enterprise Class Storage (DS8800)

z/OS

- Workload Management
- Storage Management System
- Hardware-Assisted compression, I/O and availability

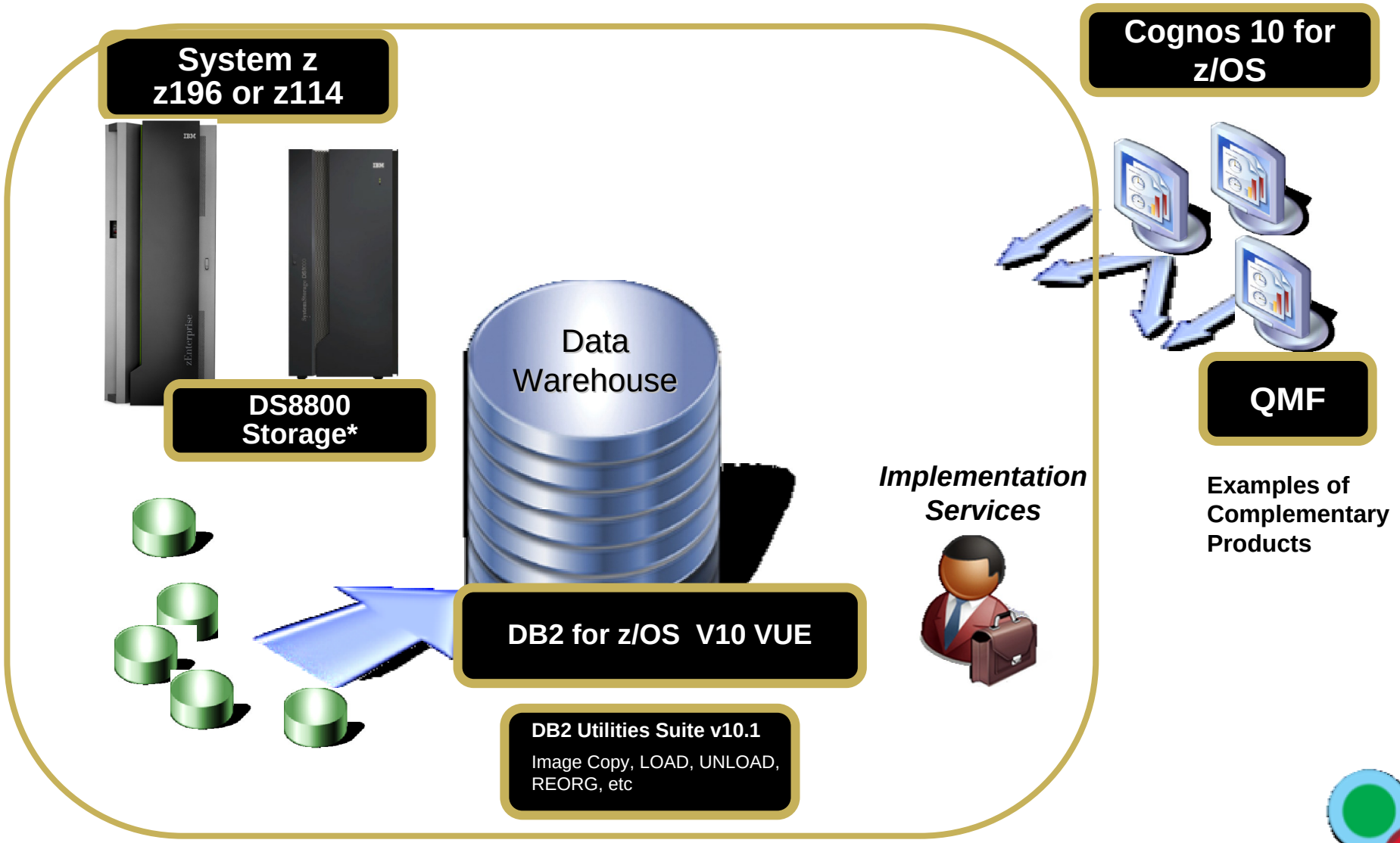
Implementation Services

- Rapid deployment and time to value
- Ensure customer success

*Standard and Foundation Configuration available for the 9700, Foundation Configuration only for 9710 (build on additional components as needed)

IBM Smart Analytics System 9700 / 9710

Foundation Configuration



Foundation Configuration

***DS8800 Storage is optional with the z114**

IBM Smart Analytics System 9700 / 9710

A complete software stack

Operating Systems

z/OS Operating System Stack V1.12
z/VM 6.1

D ent

Dashboards / Reporting

Cognos® Business Intelligence V10.1
for Linux® on System z*

Data Mining

SPSS Modeler for Linux on System z V14.2

Database Utilities

DB2 Utilities Suite for z/OS V10.1

Data Movement, Scheduling and Cubing

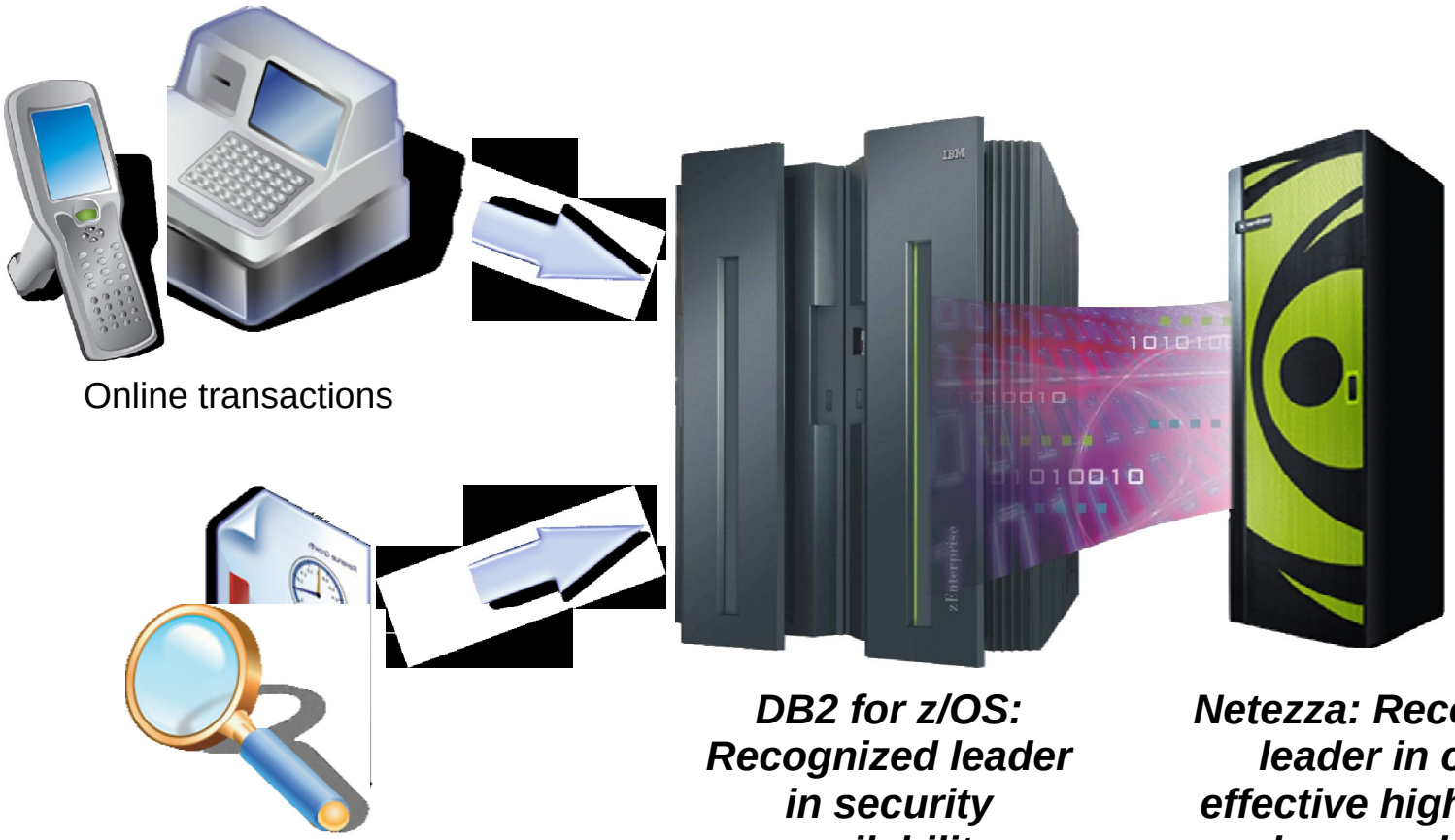
InfoSphere Warehouse on
System z V9.7.3

IBM Smart Analytics System 9700 / 9710 and DB2 Analytics Accelerator

The ultimate hybrid, workload optimized analytics server

- **Applies System z level of security, availability, integrity and reliability (in all the categories the highest in industry) to ALL your data (transactional and analytical)**
- **A true hybrid DBMS: providing applications a single interface to both transactional and analytical data**
- **The fastest performance thanks to specialized query engines: DB2 for transactional and Netezza for analytical access pattern to data**
- **TCO reduction through the effects of consolidation: single security policies, uniform database management, same skills, same tools, same procedures (instead of separate approaches in each of these categories).**
- **Delivering dedicated capacity at a competitive price that matches the value received**

Combining the best transaction system with the best analytics system

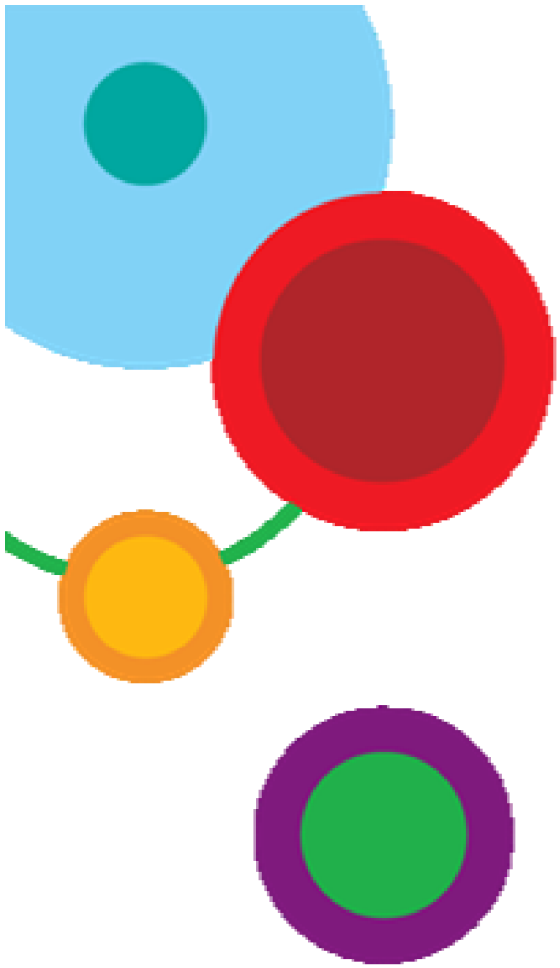


Online transactions

Analytic queries

DB2 for z/OS: Recognized leader in security availability and recoverability for OLTP

Netezza: Recognized leader in cost-effective high speed deep analytics



How is the solution developed and validated?

Let's talk through the phases…

- **Phase 1**
  - System Acquisition and Build
  - Query Suite Contents, Preparation, and Execution
    - Comparison and New Baseline Establishment
  - SPSS Scenario
    - Banking / Modeling and Scoring
- **Phase 2**
  - Workload Simulation – I/O Driven, Enumeration and Elasticity
    - Inputs
      - Scale Factor: 4, 12, 25, 100 TB
      - DB2 Query Concurrency Examples: 10, 25, 50, 100, 300
      - % Data Accessed Examples: .5%, 1%, 5%, 20%, 50%
    - Outputs
      - MIPs, Memory, Storage Capacity, I/O Bandwidth
  - Workload Simulation – Customer Profile – CP and I/O Driven
- **Finalize Configurations**

Development System Build

- System z196 with 80 PUs, 750 GB memory
- DS8800 Storage – Approximately 1,125 TB usable storage
- DB2 10 for z/OS PUT 1104
  - 4, 12, 25, 100 TB databases
  - Split and Order
- Rational Performance Tester

2817-M80 IBM zEnterprise 196

Feature	Description
1129	Model M80 - Air Cooled
1616	16 GB Mem DIMM(5/feat)-A
1876	80-Way Processor CP7
1880	CP7
1901	16GB Memory Capacity Incr
2461	1264 GB Memory
3325	FICON Express8 10KM LX 4 ports
3326	FICON Express8 SX 4 ports
3367	OSA-Express3 1000BASET-EN 4 ports
3371	OSA-Express3 10 GbE SR 2 ports

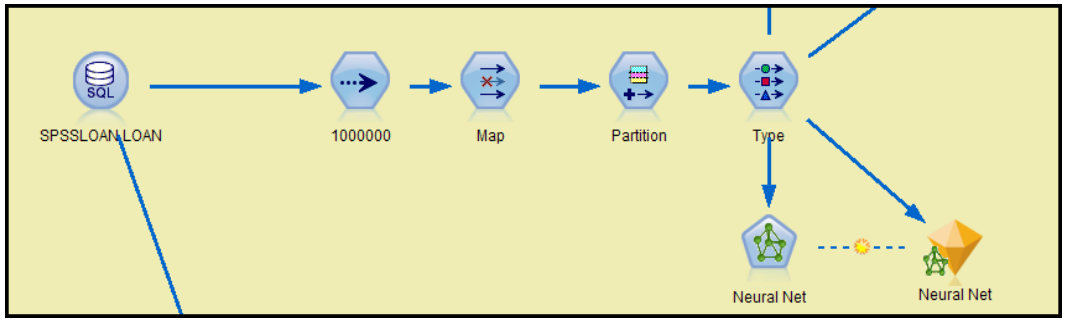
Query Suite

- Comparison to 9600 results (Foundation Queries and Scan Rate) shows:
  - Impact due to new, integrated stack
  - Impact due to hardware upgrades
  - Areas that need to be examined
- New baseline queries for 9700 and beyond
  - Exercise DB2 with varied query complexity and concurrency
- Added Utilities

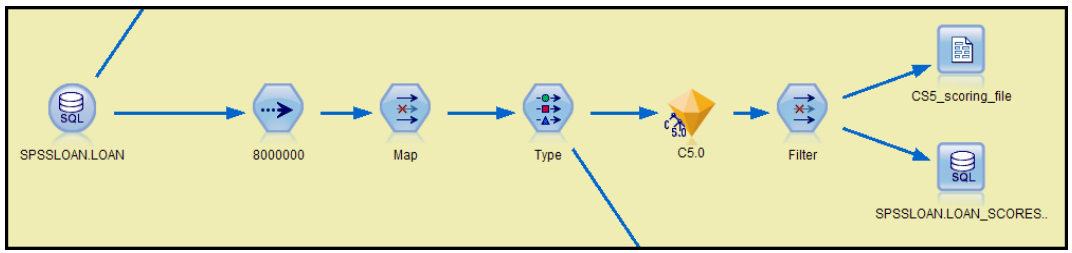
SPSS Modeler

- Modeling and Scoring
  - Customer Clustering
  - Loan Approval
- SQL Pushback Effects
- Scoring to DB2 versus File

Loan Approval Neural Net Model Flow



Loan Approval Neural Net Scoring Flow



Workload Simulation – Elasticity Testing

- Primary Workload Categories
  - Transactional Analytics
  - Mix of Reporting and Ad-Hoc
  - Deep Analytics

	A	B	C	D	E	F
1		% Data Accessed				
2	Concurrency	0.5	1	5	20	50
3	10					
4	25					
5	50					
6	100					
7	300					
8	Host Adapters = 2	Disk Adapters=2	Channels=4		Memory=	
9	Concurrency	% Data Accessed	CP MIPs	zIIP MIPs	I/O Bandwidth	I/O Rate
10	10	0.5				
11	10	1				

- Drive I/O … lots of it with low and high concurrency

- Light and Heavy queries with Low and High % data accessed

Workload Simulation

- Collected customer production SMF data generated by DB2 for z/OS Data Warehouses
- Profiled the customer workloads to understand:
  - CPU Consumption Patterns
  - Concurrency
  - Getpage activity
  - Data volume
  - Acceptable Elapsed Times
- Built modified queries that were used to mimic customer query CPU consumption patterns
  - Trivial <= 1 cp sec
  - Small >1<= 10 cp sec
  - Medium >10<=60 cp sec
  - Large >60 cp sec
- Launched queries from Rational Performance Tester (RPT), through Cognos on Linux for System z targeting DB2 for z/OS with…
  - the right query mix using the right templates
  - the appropriate concurrency levels
- Determine if our configurations are valid and that a “typical” customer workload will fit into our pre-defined solutions

Workload Profiles Requested

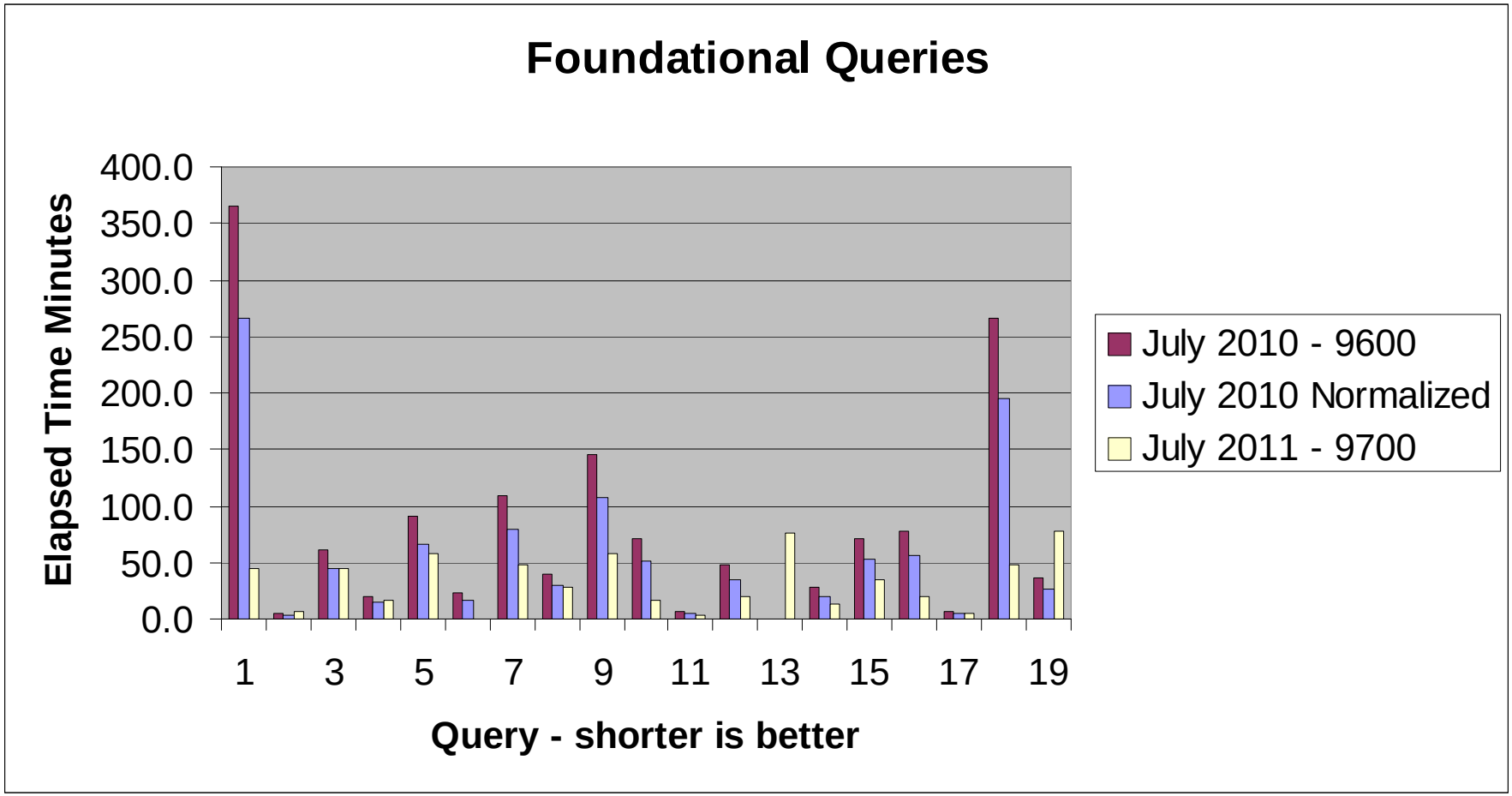
- The IBM Smart Analytics System 9700 / 9710 development team leverages a unique workload simulation process to ensure that our solution is properly configured and validated. The information our customers provide is instrumental in our understanding of resource usage characteristics including processor, data access and I/O. It also helps us advance our solution to better handle current and projected needs.

- What questions will it help customers answer?
  - Do you know what your workload profile looks like?
  - Did you ever wonder what a monster looks like?
  - Do you know how many queries your system services each day?
  - What level of concurrency are you really running?

- How can you participate?
  - Select a system and a 4 – 8 hour interval of interesting activity
  - Ensure the right traces are active
  - Select SMF data
  - Attach data to a PMR
  - Run some basic SQL against the catalog

Metrics / Results / Lessons Learned

Foundational Query Comparison



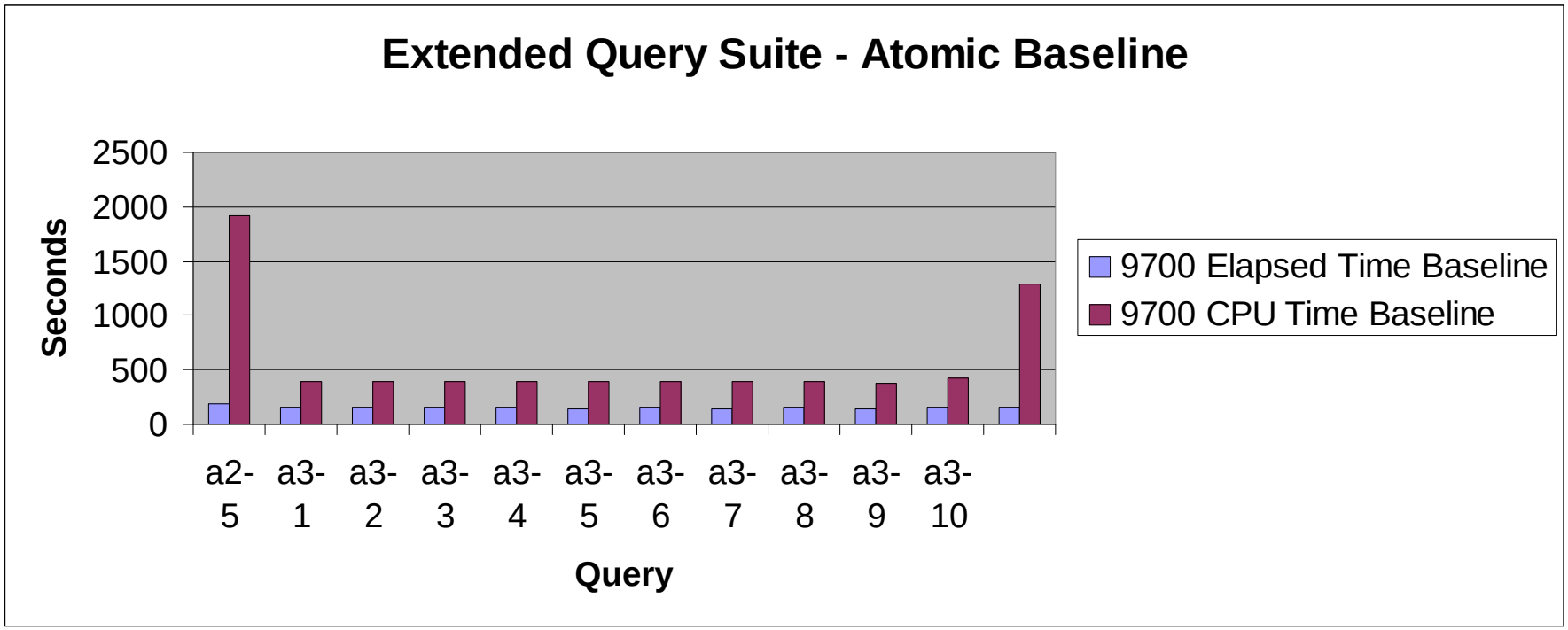
Scan Rate Comparison

DB2's ability to consume data

- Table with 8k page size and 9 columns
- IBM Smart Analytics System 9600 (DB2 9 for z/OS and System z10)
  - Compressed scan achieved 249 mb/sec/cp *
  - Uncompressed scan achieved 322 mb/sec/cp *
- IBM Smart Analytics System 9700 (DB2 10 for z/OS and System z196)
  - Compressed scan – 445 mb/sec/cp
  - Uncompressed scan – 915 mb/sec/cp
  - Extra Credit – Table with 16k page size and 16 columns
    - Compressed scan – 537 mb/sec/cp
    - Uncompressed scan – 1,141 mb/sec/cp

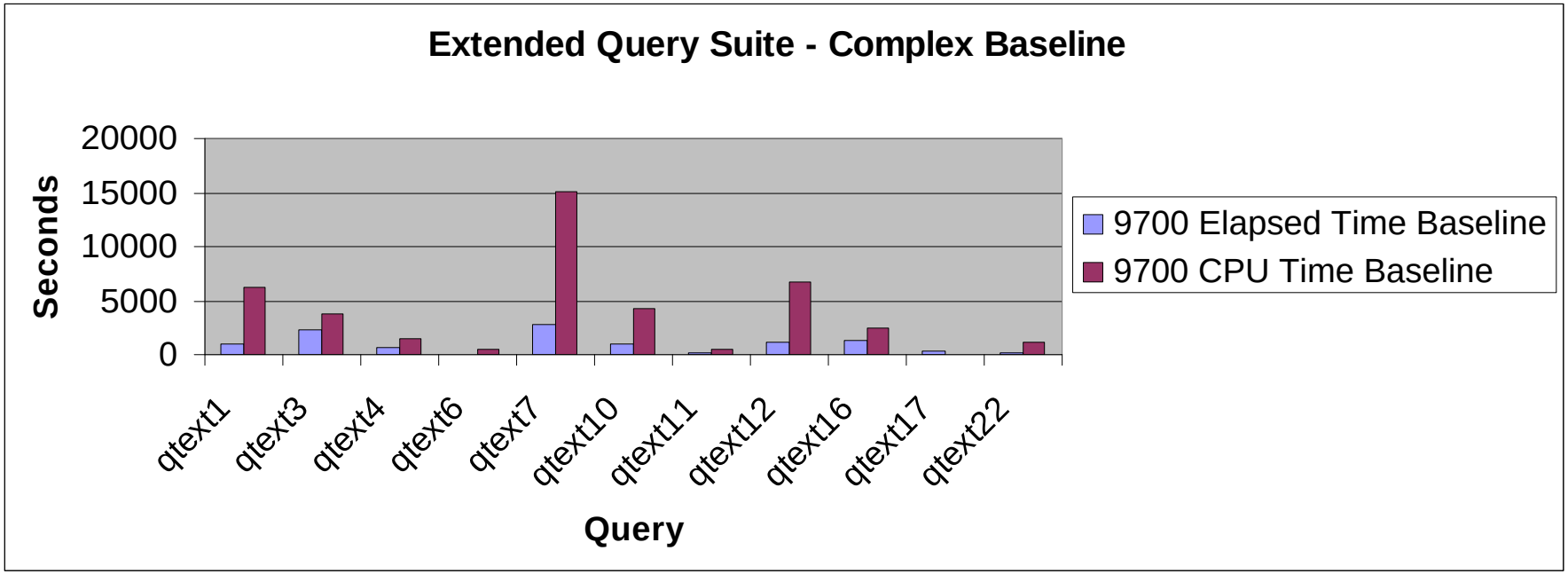
* - Different SQL was used

Extended Query Results – Atomic
No Tuning



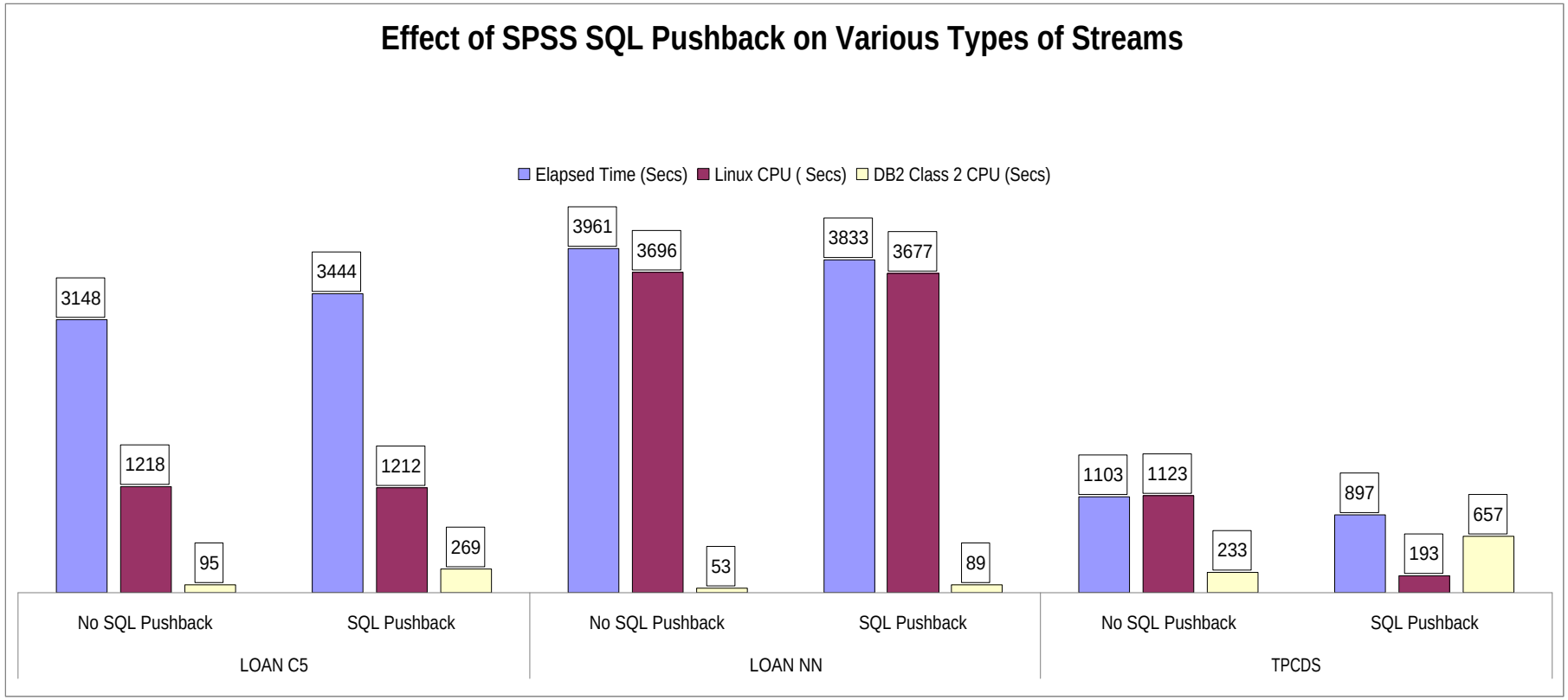
Extended Query Results – Complex

No Tuning



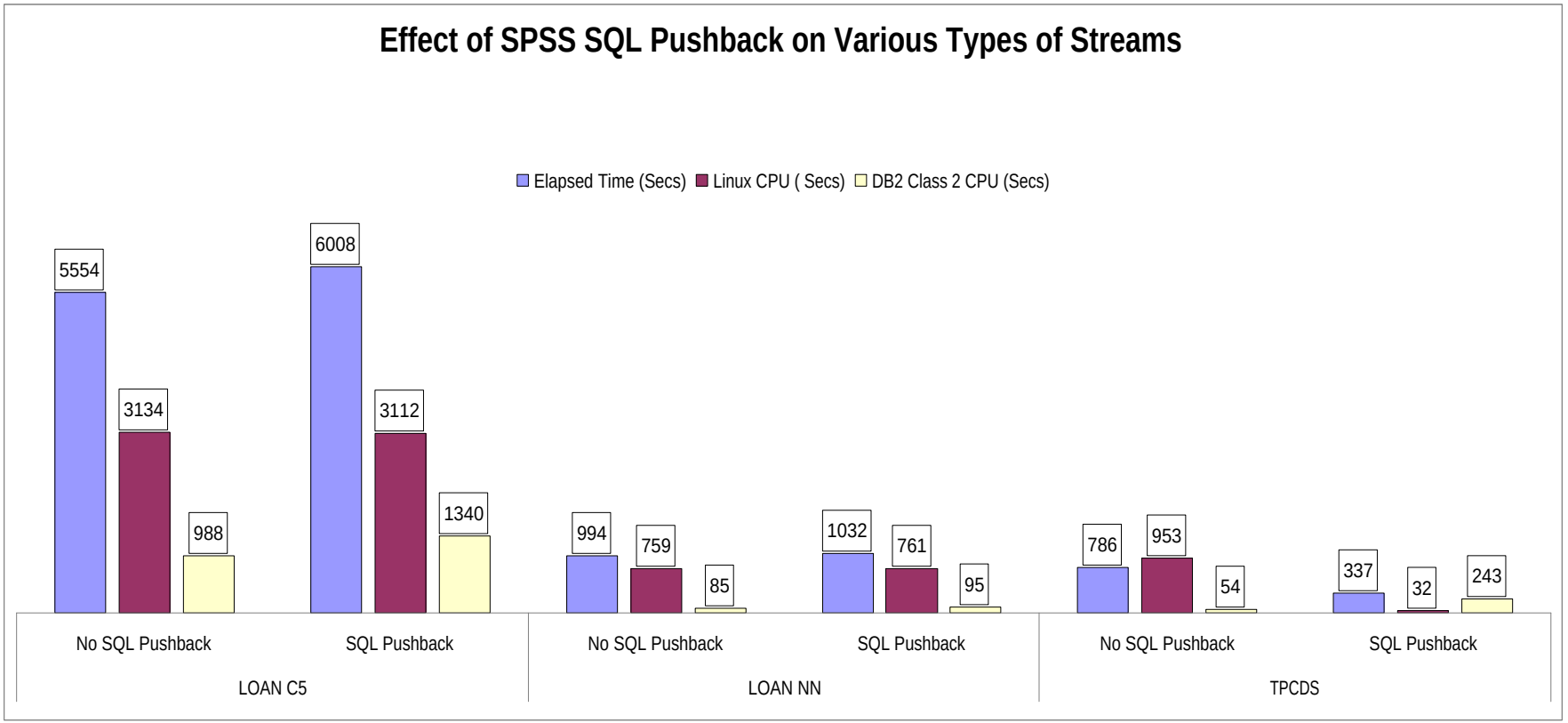
SPSS Results

Modeling



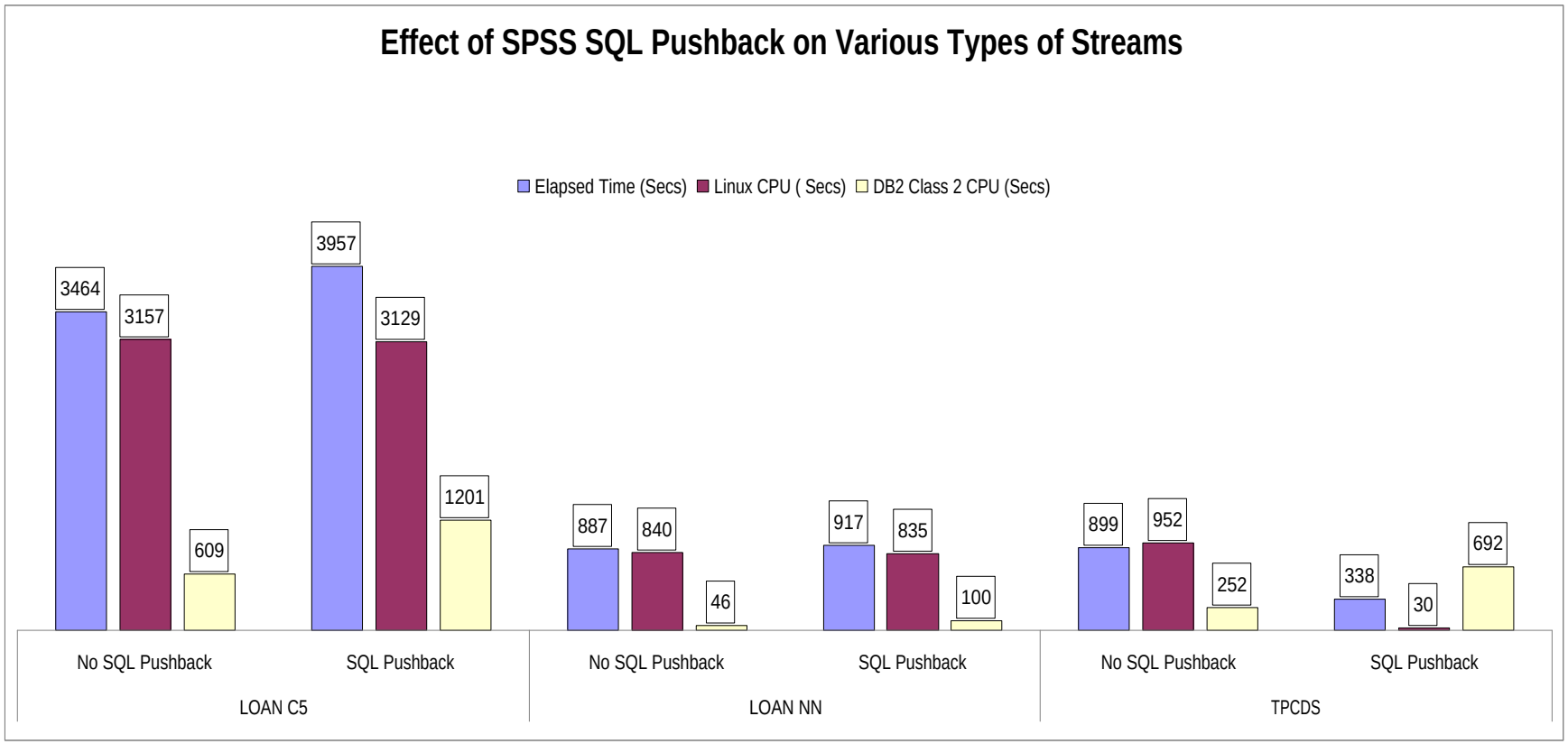
SPSS Results

Scoring to DB2 Table



SPSS Results

Scoring to Linux File

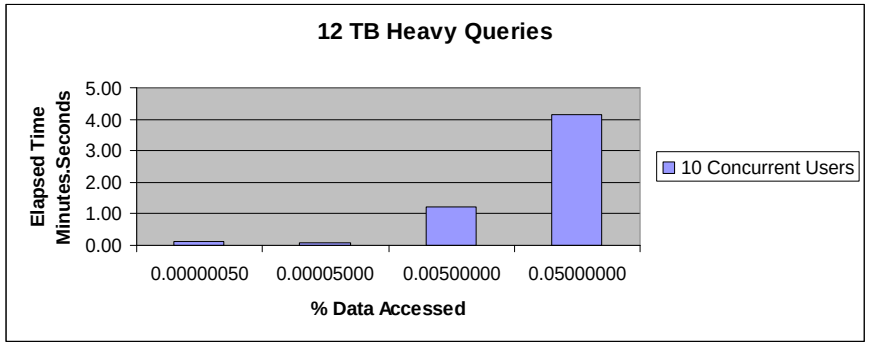


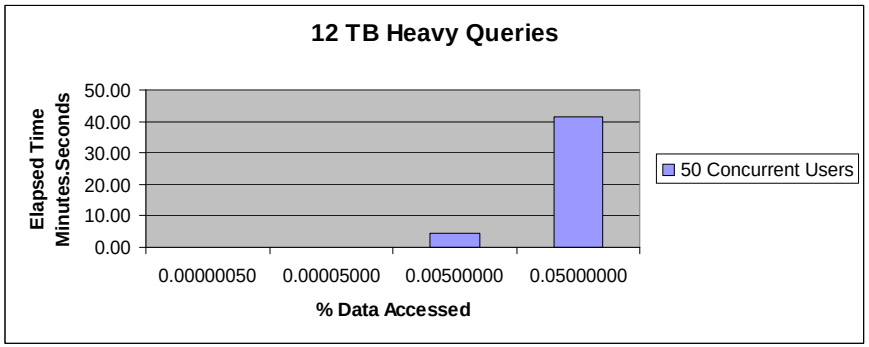
Elasticity Results - % Data Accessed and Concurrency Matters!

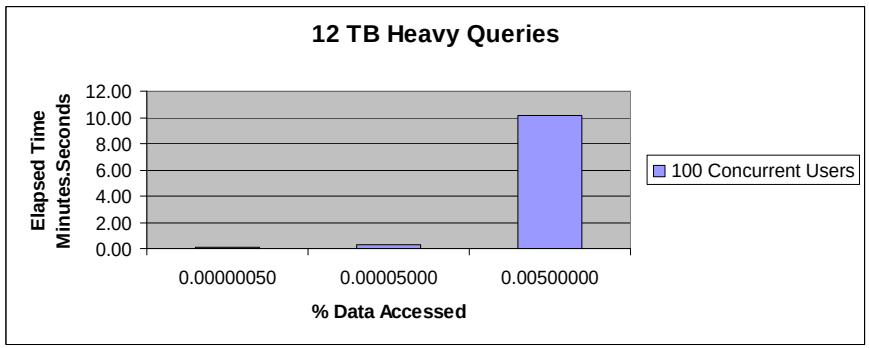
12 TB Database, 12 CPs / 12 zIIPs, MDEG 12 - 24

Heavy Queries with Lower % Data Accessed Values

A - .0000005 : approx 90 orders, 360 line items
B - .00005: approx 9,000 orders, 36,000 line items
C - .005 approx 900,000 orders, 3,600,000 line items
D - .05 approx 9,000,000 orders, 36,000,000 line items
E - .1 approx 18,000,000 orders, 72,000,000 line items







Workload Simulation Results – CPU Based

	Customer A: z10 705 2097 5cps/3ziips The DW data was collected during 4 hrs., of which 1st 2 hrs. the DW workload was capped at 460MIPs, later 2, capped at 1610 MIPs. We used the peak Qry/hr for our simulation	Customer A	Customer A simulation	Customer A simulation 2X
	z/OS CPs	~ 1610 MIPs	1	2
	z/OS zIIPs		1	1
	DB2 Bpools	unknown	11.2GB	11.2GB
DB2 Query Stats	Avg. Trivial Queries/hr.	75	83	159
	Avg. Trivial Query CL2 CP time (s)	0.100	0.004	0.004
	Avg. Trivial Query CL2 Elapsed time (s)	1	0.3	0.6
	Avg. Small Queries/hr.	2	3	6
	Avg. Small Query CL2 CP time (s)	5	5	5
	Avg. Small Query CL2 Elapsed time (s)	58	19	12
	Avg. Medium Queries/hr.	27	28	56
	Avg. Medium Query CP time (s)	14	14	15
	Avg. Medium Query Elapsed time (s)	103	200	221
	Avg. Large Queries/hr.	4	4	8
	Avg. Large Query CP time (s)	655	676	695
	Avg. Large Query Elapsed time (s)	4,290	1,478	1,365
	Max. Large Query Elapsed time (s)			
	Avg. Trivial Query GetPages	2,170	200	200
	Avg. Small Query GetPages	1,030,409	196,000	196,000
	Avg. Medium Query GetPages (normalized to 4K)	220,064	4,757,350	4,757,350
	Avg. Large Query GetPages (normalized to 4K)	217,142,875	19,917,850	19,917,850
RMF	z/OS CPU Utilization (CPs & zIIPs)	unknown	55.2	60.5

Some tidbits…

- Generating 100, 25, 12, and 4 TB databases all on one system takes a lot of time
  - Important reminder about data volume estimates
- Workload based Statistics Advisor works very well – use it for complex workloads
- CP Parallelism is essential (DSNTEP4 FOR FETCH ONLY)
- DFSort – DSA Recommendation
- IFCID 316, 317, 318 Test – no impact
- DB2 Sort – same amount of CPU, but 34% less elapsed time on a REBUILD INDEX
- DM Threshold with high concurrency and levels of parallelism may require sort BP adjustments
- zIIP behavior when batch jobs fall to discretionary

Websites for Additional Information:

IBM Smart Analytics System 9700 Webpage:
http://www.ibm.com/software/data/infosphere/smart-analytics-system/

IBM DB2 Analytics Accelerator Webpage:
http://www.ibm.com/software/data/infosphere/smart-analytics-optimizer-z/

Data Warehousing and Analytics
http://www.ibm.com/software/data/infosphere/data-warehousing/

Data Warehousing and Business Intelligence on System z
http://www.ibm.com/software/data/businessintelligence/systemz/

Terabyte Club for System z BI customers
http://www.ibm.com/software/data/businessintelligence/systemz/terabyte-club.html

Data Governance on System z
http://www.ibm.com/software/data/db2imstools/solutions/compliance.html

Data Warehouse Community in the World of DB2 for z/OS.
http://db2forzos.ning.com/group/datawarehousebusinessintelligenceonsystemz

Some Key Redbooks

- Getting Started with the IBM Smart Analytics System 9600
  - http://www.redbooks.ibm.com/abstracts/sg247902.html?Open
- Enterprise Data Warehousing with DB2 9 for z/OS
  - www.redbooks.ibm.com/abstracts/sg247637.html?Open
- 50 TB Data Warehouse Benchmark on IBM System z
  - www.redbooks.ibm.com/abstracts/sg247674.html?Open
- Housing Transactional and Data Warehouse Workloads on System z
  - www.redbooks.ibm.com/redpieces/abstracts/sg247726.html?Open
- InfoSphere Warehouse: A Robust Infrastructure for Business Intelligence
  - www.redbooks.ibm.com/abstracts/sg247813.html?Open
- DB2 for z/OS: Data Sharing in a Nutshell
  - www.redbooks.ibm.com/abstracts/sg247322.html?Open
- System Programmer's Guide To: Workload Manager
  - www.redbooks.ibm.com/abstracts/sg246472.html?Open

Communities

- **On-line communities, User Groups, Technical Forums, Blogs, Social networks, and more**
  - Find the community that interests you …
    - **Information Management** **ibm.com**/software/data/community
    - **Business Analytics** **ibm.com**/software/analytics/community
    - **Enterprise Content Management** **ibm.com**/software/data/content-management/usernet.html

- **IBM Champions**
  - Recognizing individuals who have made the most outstanding contributions to Information Management, Business Analytics, and Enterprise Content Management communities
    - **ibm.com**/champion

Thank You!
Your Feedback is Important to Us

- Access your personal session survey list and complete via SmartSite
  - Your smart phone or web browser at: iodsmartsite.com
  - Any SmartSite kiosk onsite
  - Each completed session survey increases your chance to win an Apple iPod Touch with daily drawing sponsored by Alliance Tech

Acknowledgements and Disclaimers:

Availability. References in this presentation to IBM products, programs, or services do not imply that they will be available in all countries in which IBM operates.

The workshops, sessions and materials have been prepared by IBM or the session speakers and reflect their own views. They are provided for informational purposes only, and are neither intended to, nor shall have the effect of being, legal or other guidance or advice to any participant. While efforts were made to verify the completeness and accuracy of the information contained in this presentation, it is provided AS-IS without warranty of any kind, express or implied. IBM shall not be responsible for any damages arising out of the use of, or otherwise related to, this presentation or any other materials. Nothing contained in this presentation is intended to, nor shall have the effect of, creating any warranties or representations from IBM or its suppliers or licensors, or altering the terms and conditions of the applicable license agreement governing the use of IBM software.

All customer examples described are presented as illustrations of how those customers have used IBM products and the results they may have achieved. Actual environmental costs and performance characteristics may vary by customer. Nothing contained in these materials is intended to, nor shall have the effect of, stating or implying that any activities undertaken by you will result in any specific sales, revenue growth or other results.