Business Analytics on zEnterprise

High Performance Analytics & Integrated Attached Co-processors

Carl Parris, parris@us.ibm.com
STSM – System z Performance, Design, Strategy

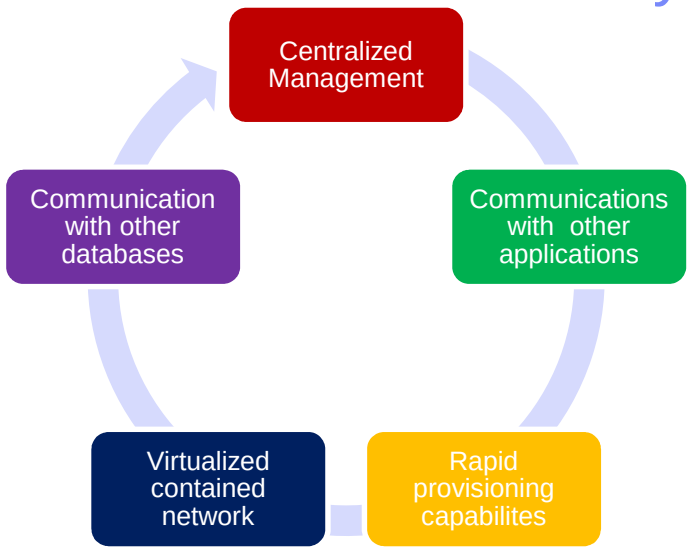
Bill Reeder, breeder@us.ibm.com
WW IT Optimization and Cloud System z Sales Leader

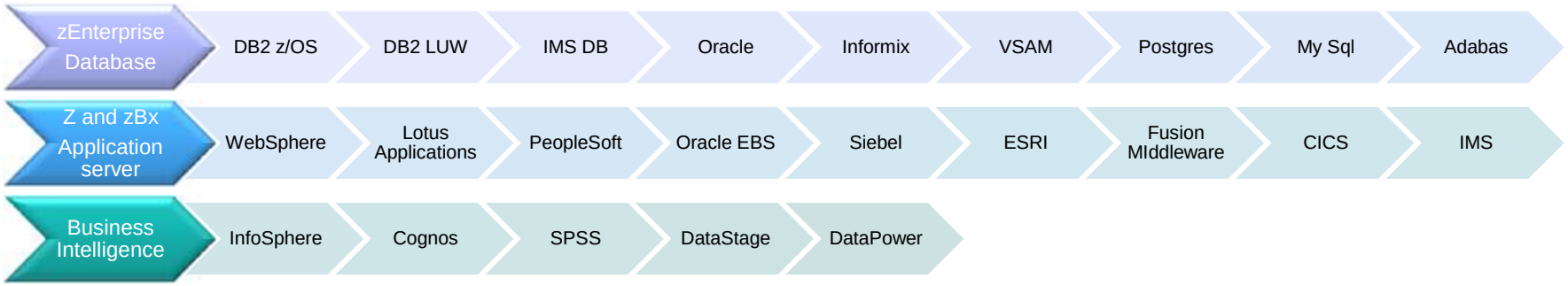
IBM Systems & Technology Group
February 2011

© 2011 IBM Corporation

zEnterprise Solutions Can Support and Integrate Data Like No Other Platform, Providing a Foundation for Other Analytic and Application Capability

- The only platform that can run nine commercial databases, supported at the same time
- Better align and synchronize data, for data integrity. Use the internal architecture to consolidate database communications
- Leverage internal networking between databases and applications
- Centralize management across entire enterprise





- Consolidation of databases
- Tighter integration of data to applications
- Business intelligence close to the data

© 2011 IBM Corporation

These workloads have recognizable patterns

Core Applications

Database (z)
- DB2® for z/OS®, IMS™

Application (z)
- CICS®
- COBOL
- WebSphere®

Database (z)
- DB2 for z/OS
- Oracle on Linux for z

Application (z)
- WebSphere

SAP

Database (z)
- DB2 for z/OS

Application (z)
- Linux® for z

Database (z)
- DB2 for z/OS

Application (x86)
- Linux for x86

Database (z)
- DB2 for z/OS

Application (Power)
- AIX®

Multi-Tier Web Serving

Database (z)
- DB2 for z/OS

Application (z)
- WebSphere

Application (x86)
- WebSphere
- Apache / Tomcat

Database (z)
- DB2 for z/OS

Application (Power / UNIX)
- WebSphere
- JBoss

Database (z)
- DB2 for z/OS, IMS

Transaction Processing (z)
- CICS, MQ

Application (Power /UNIX)
- WebSphere
- JBoss
- WebLogic

Presentation (x86)
- WebSphere
- Windows

Database (z)
- DB2 for z/OS or IMS

Application (Power /UNIX)
- WebSphere
- JBoss

Presentation (x86)
- WebSphere
- Apache / Tomcat

Data Warehouse & Analytics

Master Data Management

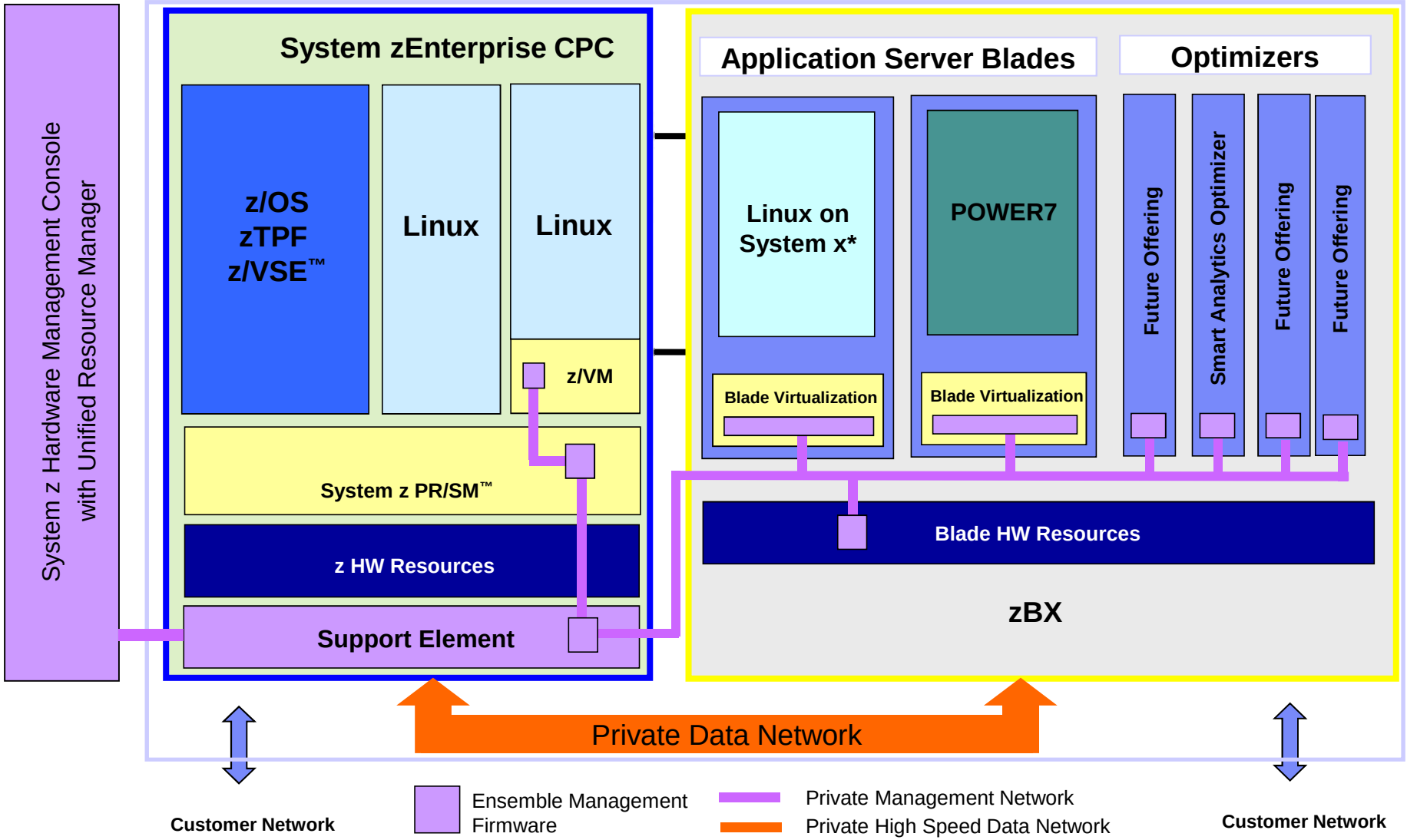
Database (z)
- § DB2 for z/OS

Application (z)
- § WebSphere MDM (AIX, Linux on z)

Analytics
- § System z/OS
  - § DB2
  - § Cognos® (Soon!)
  - § SAS
- § Linux for System z
  - § Cognos
  - § SPSS
  - § InfoSphere™ Warehouse

© 2011 IBM Corporation

zEnterprise with a System z Blade Extension (zBX)



*All statements regarding IBM's plans, directions, and intent are subject to change or withdrawal without notice. Any reliance on these Statements of General Direction is at the relying party's sole risk and will not create liability or obligation for IBM.

© 2011 IBM Corporation

Cloud Service Lifecycle Management

Deploy Service
§ Request Driven Provisioning
§ Management Agents and Best Practices
§ Application / Service On Boarding
§ Self-service interface

Subscribe to Service
§ Request a service
§ “Sign“ Contract

Manage Operation of Service
§ Visualize all aggregated information about situations and affected services
§ Control operations and changes
§ Event handling
§ Automate activities to execute changes
§ Include charge-back

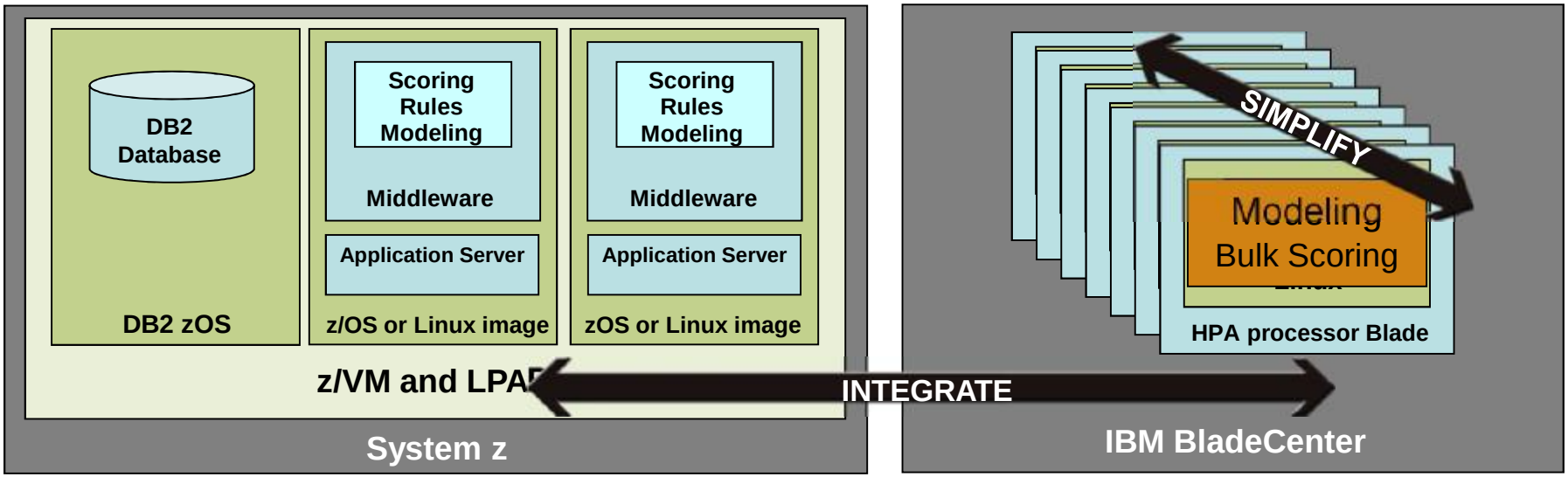
Offer Service
§ Register Services and Resources
§ Add to Service Catalog

Service Creation
§ Scope of Service
§ SLAs
§ Topologies, Best Practices Management Templates

Terminate Service
§ Controlled Clean-up

© 2011 IBM Corporation

Hybrid Schema Mainframe and HPA Accelerator



Why Business Analytics on System z

- **Highest Frequency compute threads in industry z196**
- **Very good floating point performance z196**
- **Large Shared Resource Pool**
  - Single point of resource management
  - Single point of operational control
  - Efficient use of underlying compute resources
  - Manage unpredictable loads between instances
  - Easy/fast provisioning
- **Integration w/Commercial Business Processing**
  - Security
  - Reliability
  - Availability
  - Auditing
  - Monetary Transactions

Why Analytics on HPA Blade

- **Compute thread rich environment**
- **State of the art Vector/SIMD architecture**

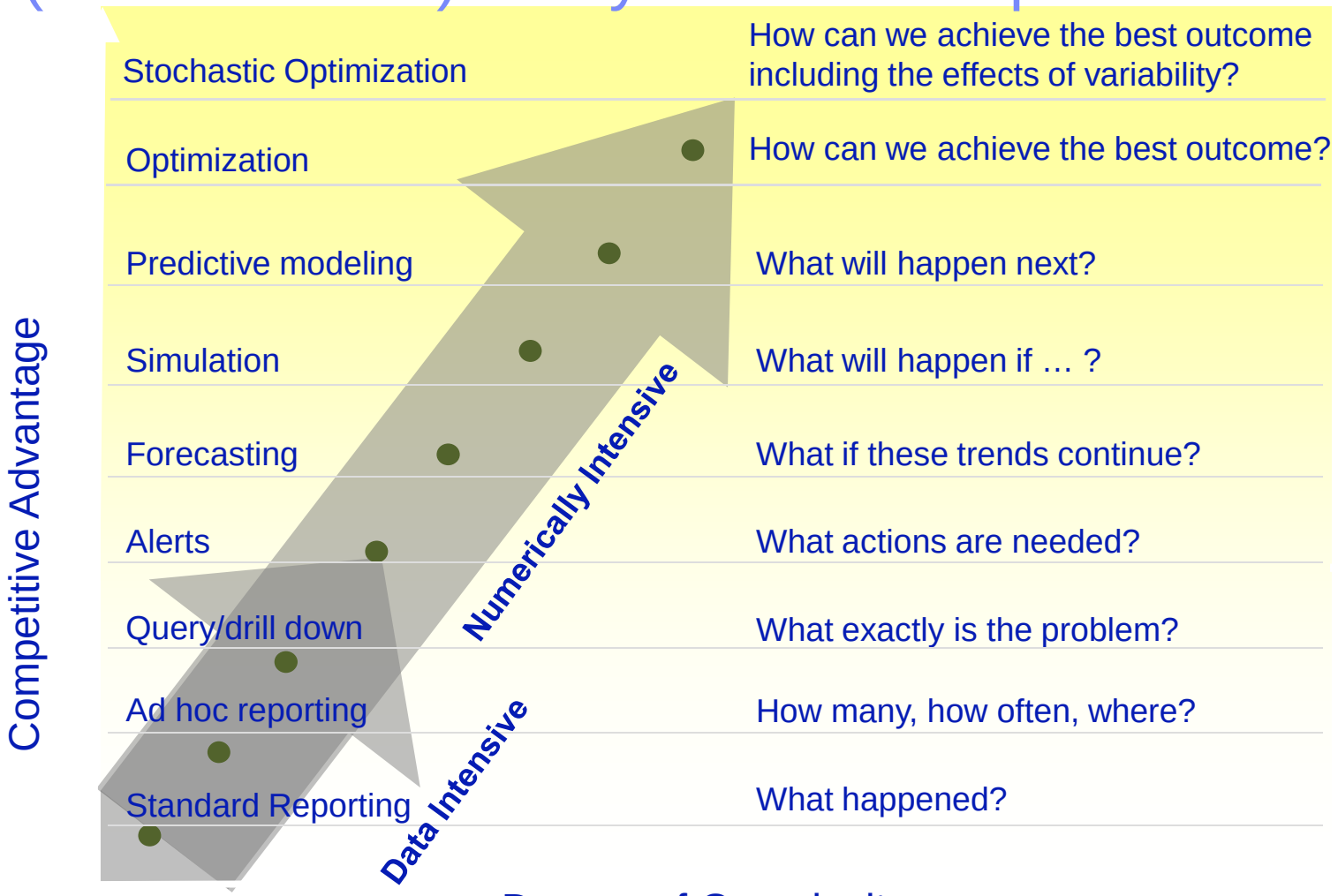
Why Analytics on zGryphon

- **HPC enhanced commercial computing**
- **Single operational domain**
  - **Avoid standalone distributed cluster**
- **Extend strengths of System z**

© 2011 IBM Corporation

zHPC > EdgeHPC > Commercial HPC > Business Analytics

(Mathematical) Analytics Landscape



Increasing prevalence of compute and data intensive parallel algorithms in commercial workloads driven by real time decision making requirements and industry wide limitations to increasing thread speed.

Based on: Competing on Analytics, Davenport and Harris, 2007

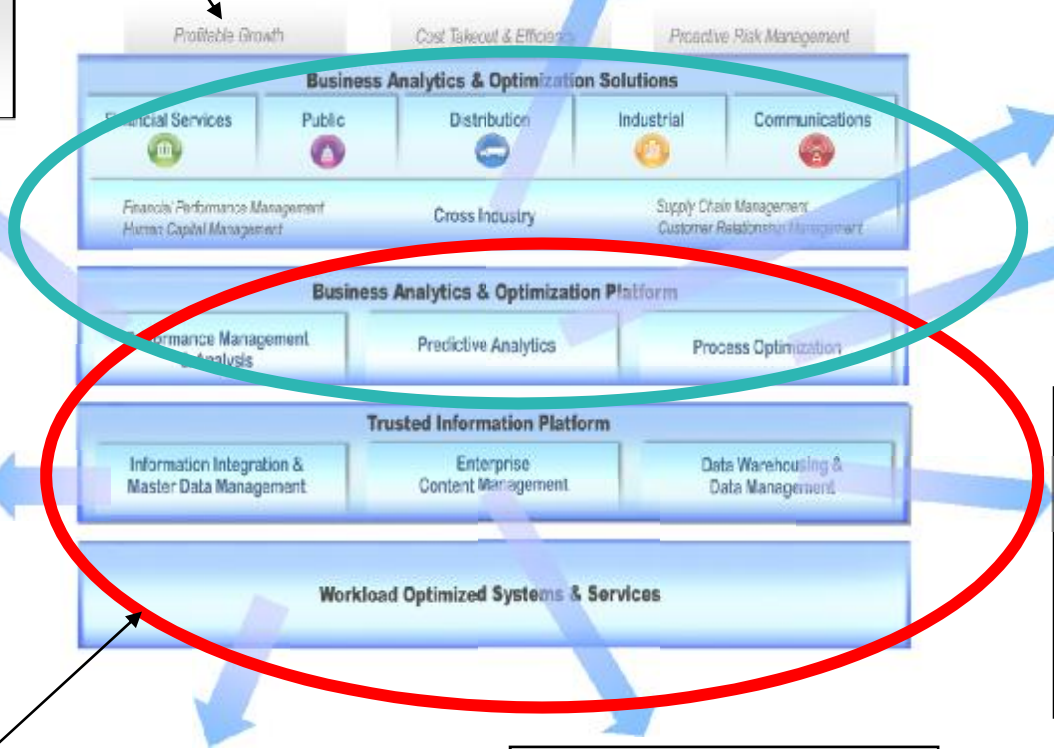
© 2011 IBM Corporation

Market Leading Business Intelligence & Analytics Software

Concentrating on this bit

- Cognos 8 Customer Performance Sales Analytics
- Cognos 8 Workforce Performance
- Cognos 8 Financial Performance Analytics
- Cognos 8 Supply Chain Performance Procurement Analytics
- Entity Analytic Solutions

- Cognos 10 BI
- Cognos Planning
- Cognos TM1



- SPSS
- iLog

- •Filenet BPM
- •iLog

- InfoSphere Information Server
- InfoSphere MDM Server
- InfoSphere MDM Server for PIM
- InfoSphere Foundation Tools
- Telco Data Warehouse & Other Industry Models
- Traceability Server

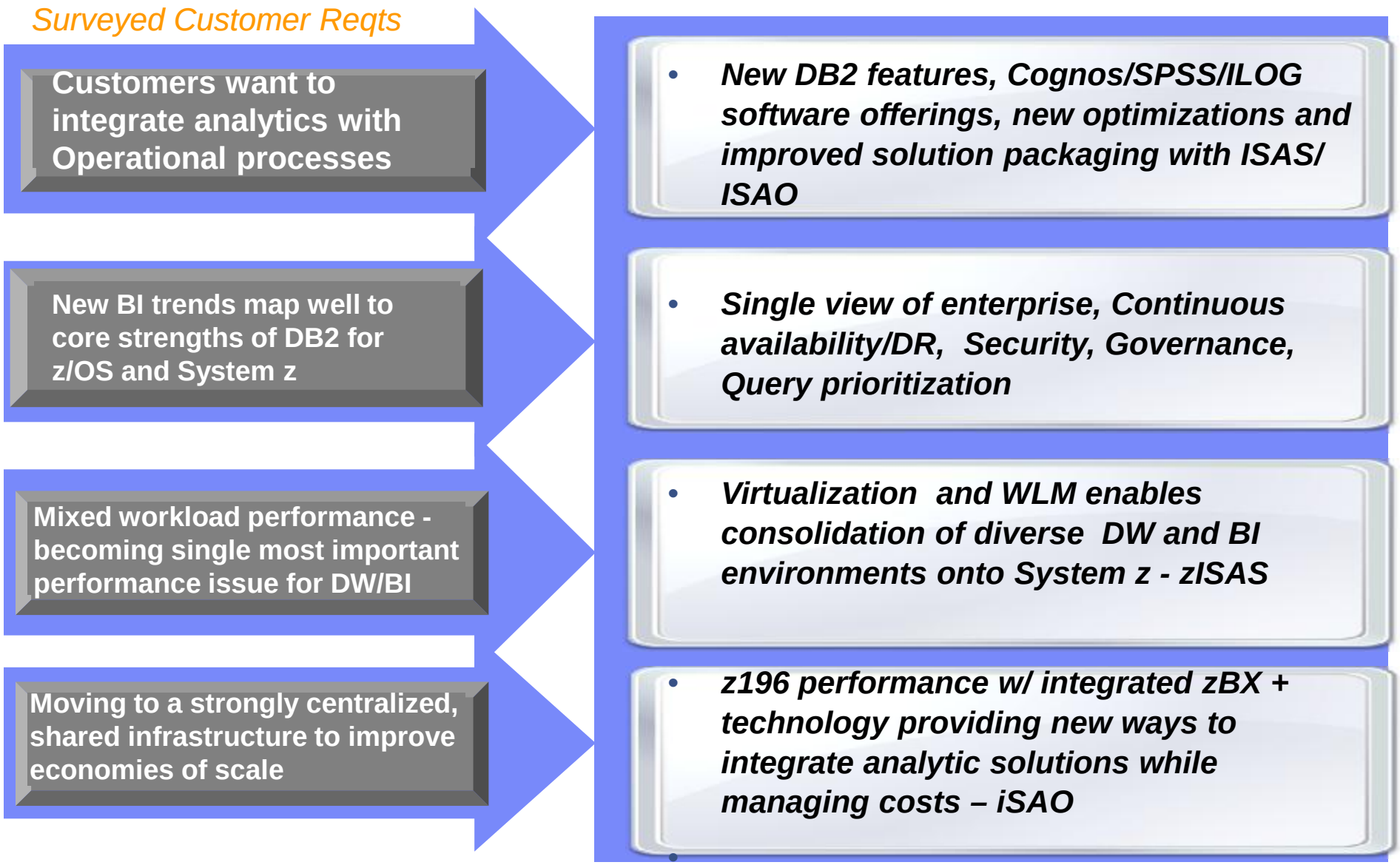
- DB2
- Informix
- IMS
- solidDB
- Optim
- Datastage
- Discovery
- Database tools
- InfoSphere Warehouse
- InfoSphere Streams
- Mashup Hub
- DB2 for z/OS

This is plumbing

- •Smart Analytics Systems

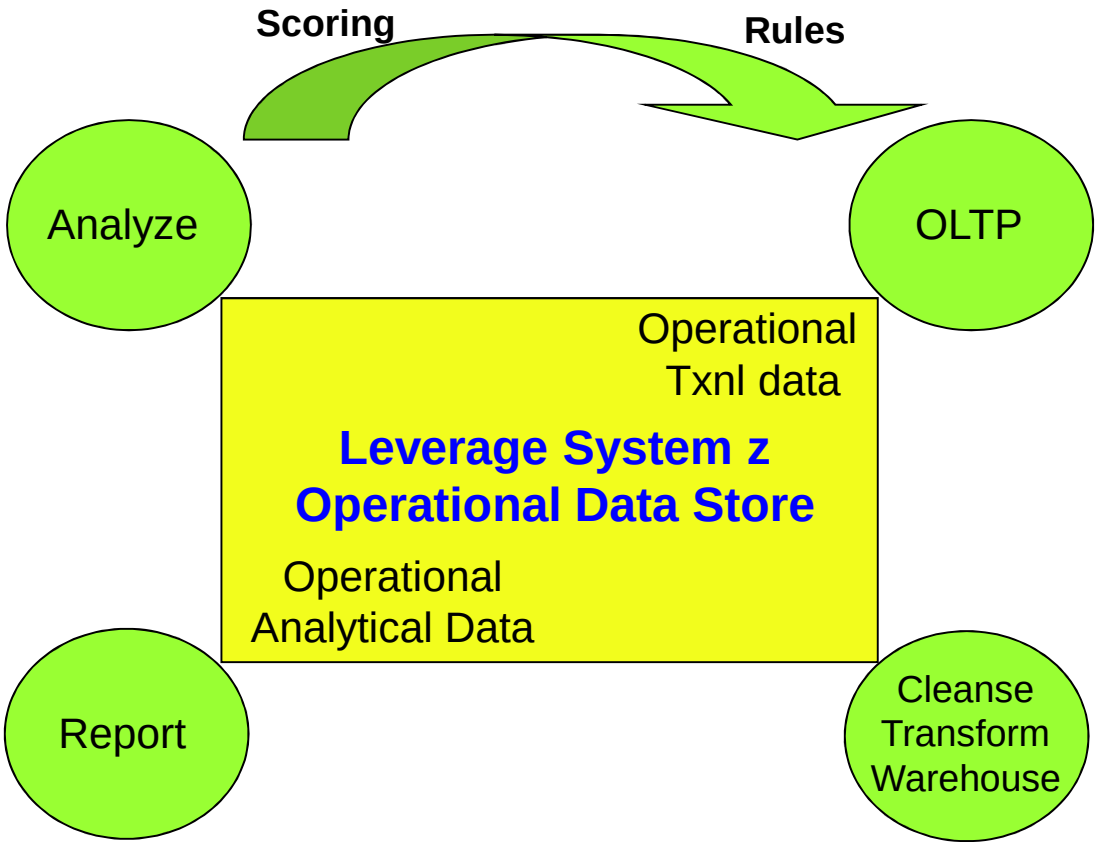
- Filenet P8
- eDiscovery
- Content Manager
- InfoSphere Content Collector
- Records Management
- Content Integrator

© 2011 IBM Corporation



© 2011 IBM Corporation

System z Platform Direction: From Data hub to Analytics hub

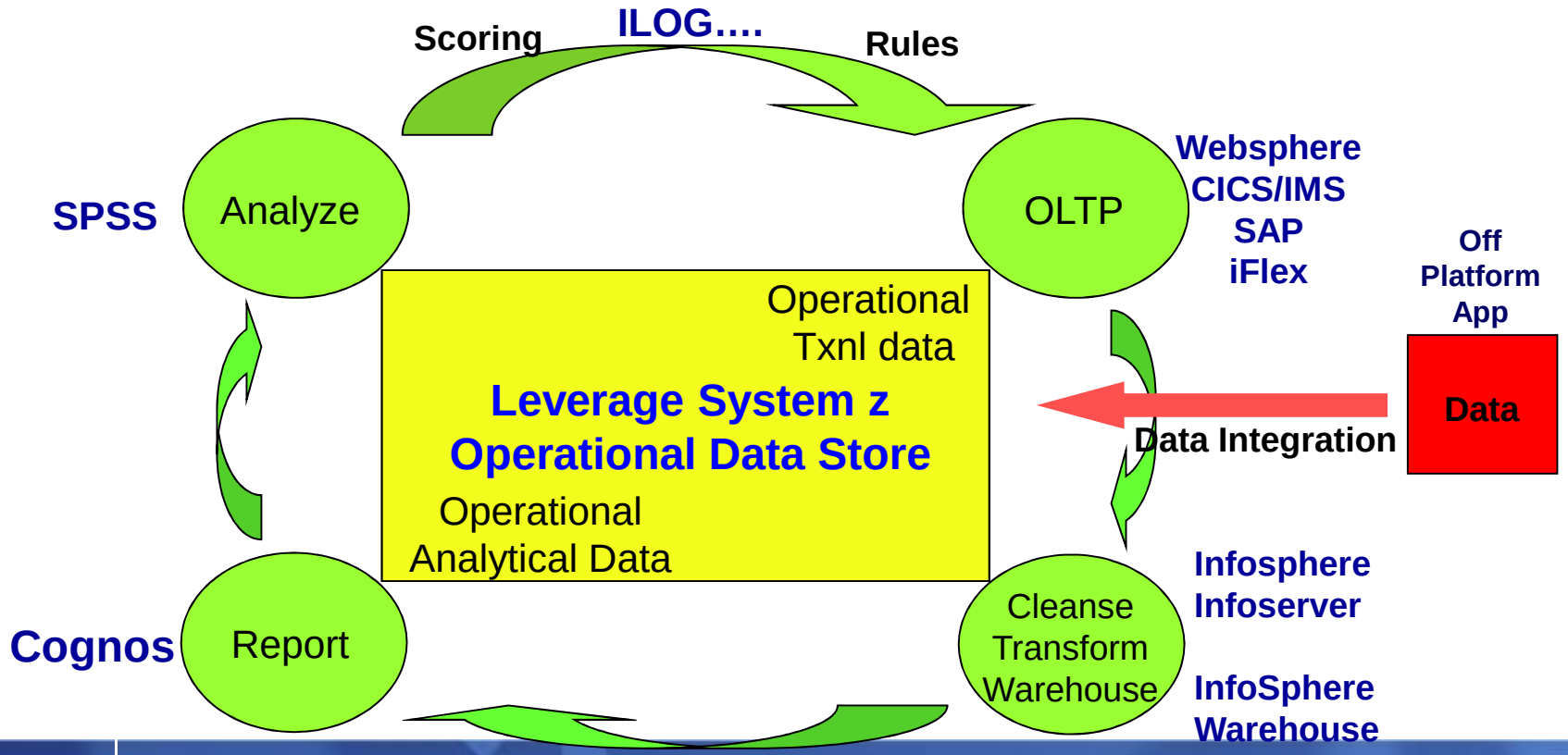


- **Exploit Industry Trends that play to the strengths of System z**
  - **Data Consolidation and creation of "Enterprise Database of Record"**
  - **Operational Business Intelligence with z QOS requirements**
  - **Operational trxs integrated with predictive analytics to provide additional insight**
- **Leverage z Hybrid architecture, accelerators, multi-workload integration (zOS/zLinux)**

© 2011 IBM Corporation

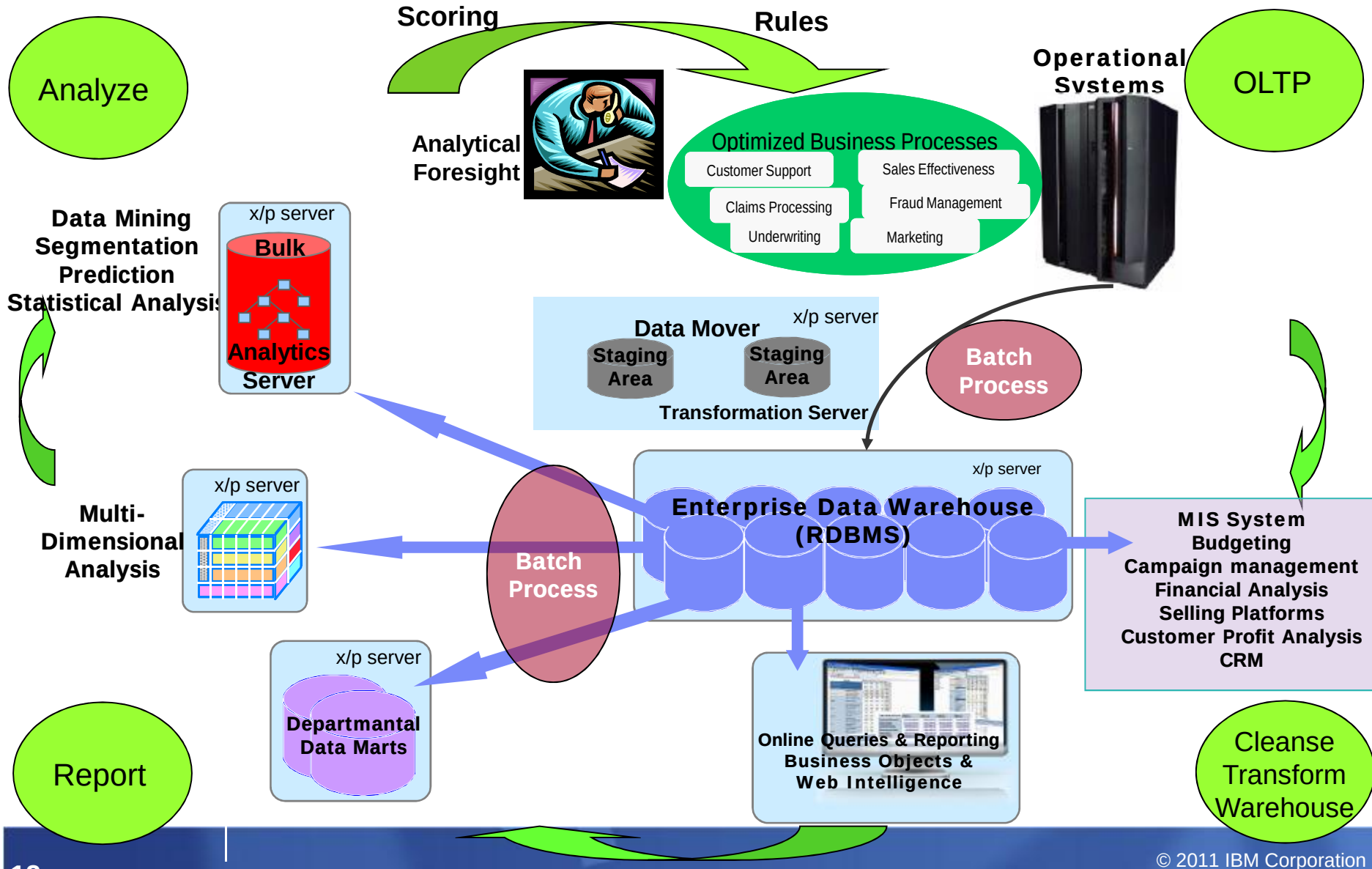
System z Platform Direction: From Data hub to Analytics hub

- Exploit Industry Trends that play to the strengths of System z
  - Data Consolidation and creation of "Enterprise Database of Record"
  - BI/Analytics application consolidation and creation of enterprise single version of truth
  - Operational Business Intelligence with z QOS requirements
  - Operational trxs integrated with predictive analytics to provide additional insight
  - Superior end/end analytics life cycle integration
  - Analytics as a service in an internal or external cloud
- Leverage z Enterprise architecture, accelerators, multi-workload integration (zOS/zLinux)



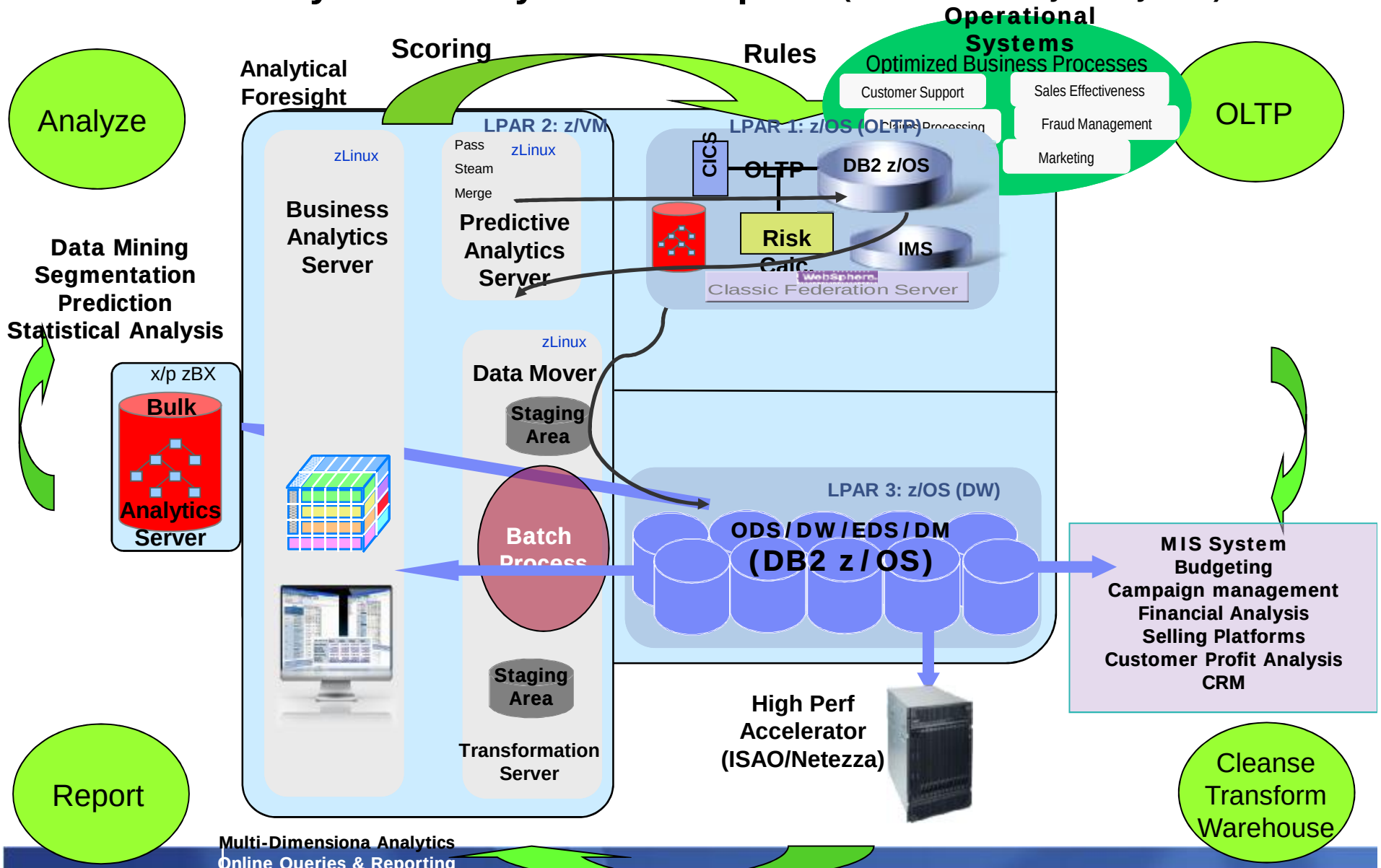
11
© 2011 IBM Corporation

Business Analytics Life Cycle – Async and Distributed



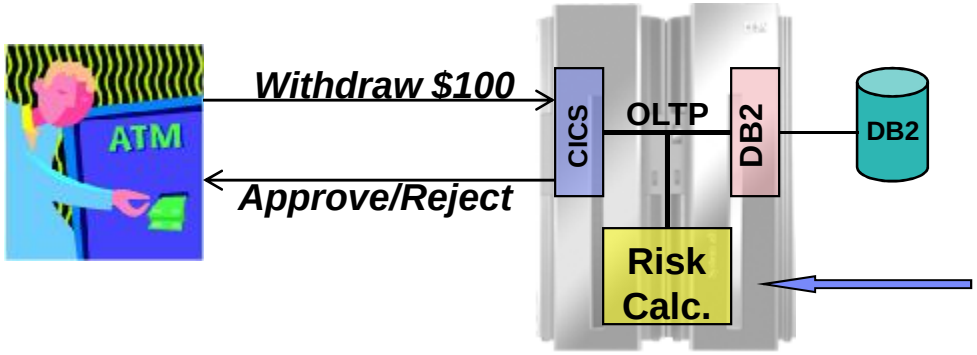
© 2011 IBM Corporation

Business Analytics Life Cycle – zEnteprise (IBM Smart Analytics System)



© 2011 IBM Corporation

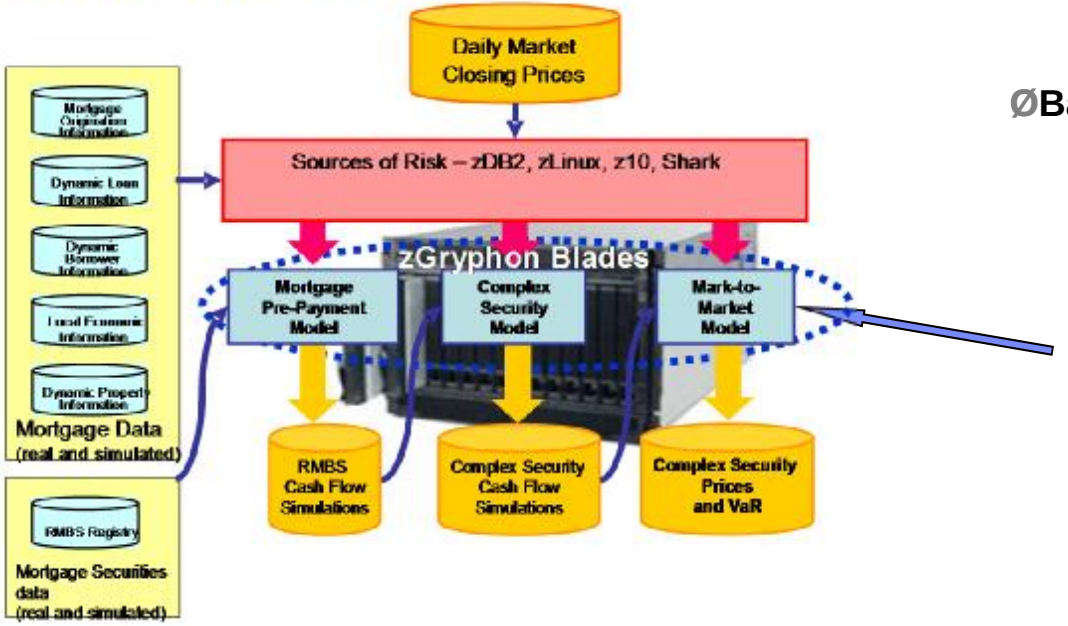
Evolution of OLTP



Ø**Real time 'transactional' analytics**

- **Credit Card Fraud Detection**
  - Ø Compute intensive 'neural network' calculations required off-load to alternative hardware
  - Ø Batch runs overnight – business imperative for real time response. POC w/ ACI/PRM using z/OS and Cell.
    - > Latency costs of offload negated compute advantages of Cell
- **Optimized on-board floating point architecture would re-host this application on z/OS**
  - Ø Eliminate network latency delays
  - Ø Add value to OLTP transaction
  - Ø Huge savings potential the sooner the act of fraud is detected

Bulk Data Analytics



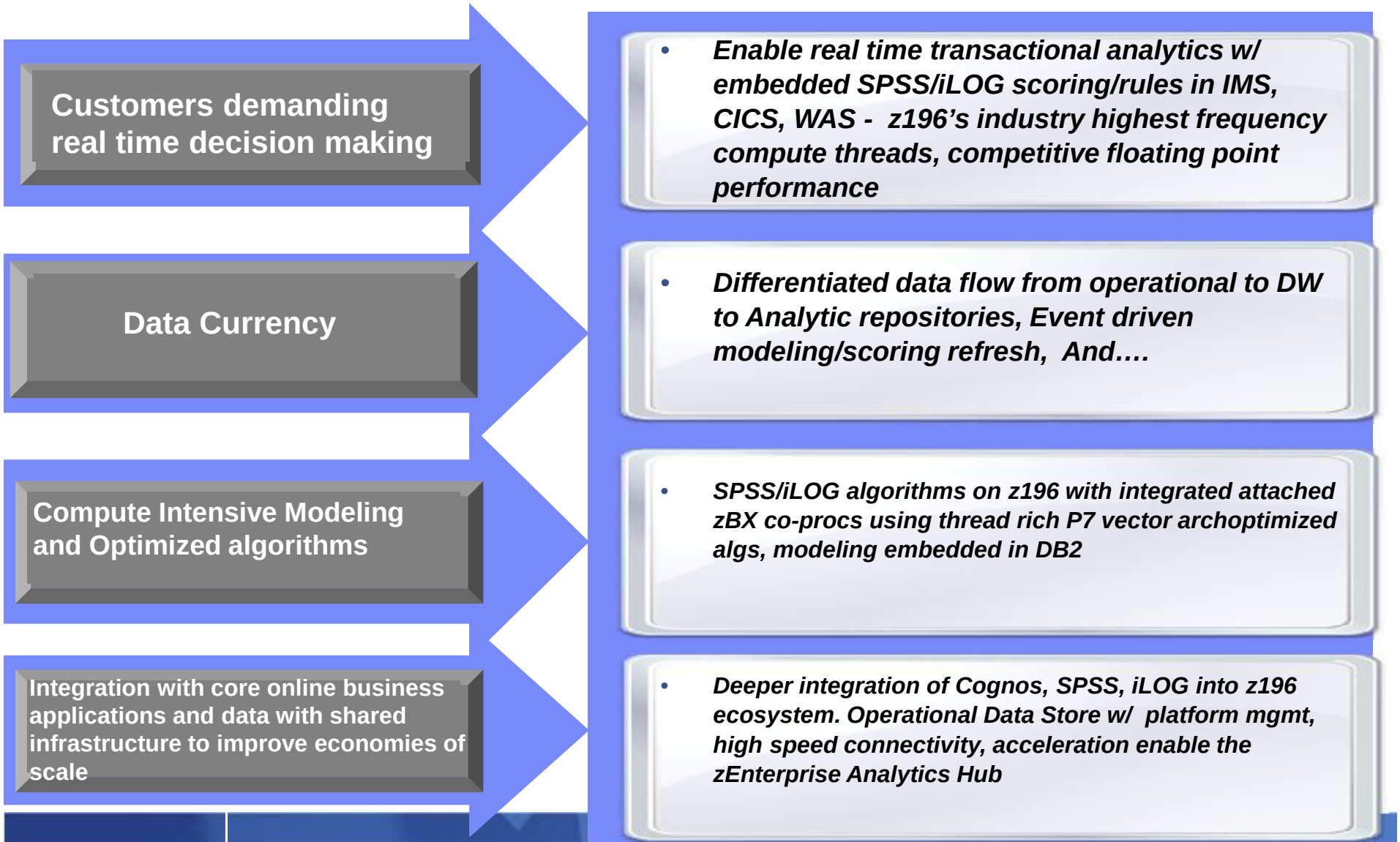
Ø**Batch and near real time**

- **Risk Analysis (IBM Treasury POC)**
- Multiple repositories of operational data
- Sophisticated numerical algorithms
  - > Bayesian probability algorithms
  - > Monte Carlo simulation

- **Batch and near real time good match for host/accelerator offload**
- High performance accelerator HW building block
- High speed bulk data transport
- Efficient data cleansing/transformation engines – ETL
- Value added proprietary data mining algorithms
- Open standard host/accelerator programming model

© 2011 IBM Corporation

System z and the Predictive Business



Predictive Analytics Use Case Scenarios – US Credit Union Example

A. Higher withdrawal limits to increase customer satisfaction

- Many Neighborhood Financial Centers, ATMS, Kiosks do not have service personnel to override withdrawal limits.
- Need real time method of scoring member to determine appropriate limit while limiting risk
- Built a **scoring model and embedded it in credit union's daily transaction processing system** to automatically determine withdrawal limits
- Saved staffing costs, increased customer satisfaction, retention, enabled increased revenue generation with reduced risk

B. Targeted campaigns to improve retention, revenue

- Exported member data from CU's BI system, **applied analytic techniques such as regression to create member profiles** to predict likelihood members will need additional products/services
  - e.g. Home equity line of credit
- Combined member usage characteristics w/ census information (i.e. local home ownership)
  - Filtered out 30-40% of unlikely candidates. Focused on 60-70% most likely to respond
- Increased 'lifted' revenue generated per marketing $$ by 60-100%
- Analysts wrote queries for rules to assist customer service. Recommendations pop-up on monitors during customer calls for relevant offers

C. Grow customer base while risk shrinks

- Attract new customers w/ prior financial problems
- **Used scoring models to control deposit loss**
- Boosted CU bottom line and benefited customers avoiding check cashing services and payday lenders

D. Identify new branch locations

- **Created predictive model** to help identify new branch locations, operate existing branches more profitably, close sites
- **Factor and regression analysis** to identify composite performance based on new customers, deposits, loan distributions

Predictive Analytics enabled getting more mileage of data. Saved over $1M annually, increased revenue and improved member satisfaction

© 2011 IBM Corporation

Analytic Functional Areas

Cross Sell	**Analysis and exploitation of hidden relationships in data about existing customer behavior to predict efficient future activity (purchase of products)**
Direct Marketing	**Analysis of customer characteristics (demographics, responses) to predict the amount of variability and tailoring of a marketing campaign**
Collection Analytics	**Analysis of customer characteristics to predict ability to pay and optimization of resources to facilitate collection.**
Portfolio Prediction	**Analysis of a portfolio of items (patients, products, financials, stores, etc.) to predict (score) a future outcome (survivability, placement, profitability, etc.)**
Customer Retention	**Analysis of a customers past characteristics to predict the likelihood of a customer's future action.**
Risk Analysis	**Quantitative analysis to numerically determine the probabilities of various adverse events and the likely extent of losses if the event occurs**
Fraud Detection	**Analysis of transactions to predict the likelihood of fraud usually based on a score or probability.**

© 2011 IBM Corporation

Mapping industry requirements to analytic functions
Example: FSS (Banking and Insurance)

FSS Analytics Trends	Industry Requirements	Relevant Functional Areas
Core Banking	Customer Insight	Customer Retention, Cross-Sell, Direct Marketing
	Product Recommendations	Customer Retention, Cross-Sell, Direct Marketing
	Fraud Detection and Prevention	Fraud Detection, Risk Analysis, Collection Analytics
	Underwriting	Risk Analysis
Payments	Fraud Detection and Prevention	Fraud Detection, Risk Analysis, Collection Analytics
	Anti Money Laundering	Fraud Detection
	Underwriting	Risk Analysis
Financial Markets	Fraud Detection and Prevention	Fraud Detection, Risk Analysis, Collection Analytics
	Portfolio Analysis	Portfolio Prediction, Risk Analysis
	Product Recommendations	Customer Retention, Cross-Sell, Direct Marketing
Insurance	Cause and Effect Analysis	Portfolio Prediction, Risk Analysis
	Underwriting	Risk Analysis
	Fraud Detection and Prevention	Fraud Detection, Risk Analysis, Collection Analytics

© 2011 IBM Corporation

Mapping Trends and Requirements to Analytical Function Retail Sector

Retail Trends in Analytics	Industry Requirements
Product Optimization and Shelf Assortment	Merchandise Performance
Customer Driven Marketing	Customer Insight/Customer Churn
Fraud Detection and Prevention	Fraud Detection and Prevention
Integrated Forecasting	Merchandise Performance/Customer Insight
Localization and Clustering	Store and Channel Performance
Market Mix Modeling	Promotion Planning
Price Optimization	Merchandise Performance
Product Recommendation	Promotion Planning
Real Estate Optimization	Store and Channel Performance
Supply Chain Analytics	Supply Chain Optimizations
Workforce Efficiency Optimization	Store and Channel Performance

© 2011 IBM Corporation

Mapping Trends and Requirements to Analytical Function
Telco Sector

Telco Trends	Industry Requirements (from Sector Team)
Market Optimization	Customer Churn
	Customer Retention
	Product Cross Sell
	Integrating Telco with retail sales
	Social Networking Models
	Behavioural Analytics
Network Analytics	Cell Tower Energy Management
	Network Traffic Optimization
	Capacity Planning
Revenue Assurance	Circuit Consolidation
	Budget Forecasting

© 2011 IBM Corporation

Mapping Trends and Requirements to Analytical Function Healthcare Sector

Healthcare Trends	Industry Requirements (from Sector Team)
Life Sciences	Gene Pool Analysis
	Drug Discovery
	BioInformatics
Healthcare Payer	Insurance Fraud
	Clinical Cause and Effect
	Medical Record Management analytics
	Network Management analytics
	Employer Group Analytics
Healthcare Provider	Executive Analytics
	Patient Access
	Clinical Resource
	Patient Throughput
	Quality & Compliance

© 2011 IBM Corporation

Mapping Functional Areas to Tasks

Function	Task
Cross Sell	**Association**
Direct Marketing	**Classification, Clustering, Association**
Collection Analytics	**Clustering, Association**
Portfolio Prediction	**Prediction**
Customer Retention	**Classification, Estimation**
Risk Analysis	**Classification, Clustering, Prediction**
Fraud Detection	**Anomaly Detection**

© 2011 IBM Corporation

Mapping Tasks to Techniques/Algorithms

Task	Technique/Algorithm
Association	**Association Rules(Apriori), Decision Trees, Minimum Description Length**
Classification	**Decision Trees, Neural Net, Naïve Bayes, Support Vector Machines**
Clustering	**Clustering, Attribute Analysis, K-Nearest Neighbor**
Estimation	**Logistic, Regression, Discrete Choice Models**
Prediction	**Linear Time Series, Non-linear Time Series, Exponential Smoothing**
Anomaly Detection	**Support Vector Machine**

© 2011 IBM Corporation

SPSS Analytic Components – 1 of 4 Charts

Procedure Family	Procedure	Computation Model Fit
LINEAR	ALM	Automatic linear modeling
LINEAR	ANOVA	Analysis of variance
LINEAR	DISCRIMINANT	Classify cases into groups based on predictor variables
LINEAR	MEANS	Group means and statistics for target variables within categories of predictor variables
LINEAR	ONEWAY	One-way analysis of variance
LINEAR	REGRESSION	Regression
LINEAR	T-TEST	T-tests for one sample, independent samples and pair samples
LINEAR	UNIANOVA	Univariate analysis of variance
LINEAR	GLM	General linear model
LINEAR	2SLS	Two-stage least-squares
LINEAR	WLS	Weighted least-squares
LINEAR	CSGLM	Linear regression for complex samples
NON-LINEAR	GLMM	Generalized Linear Mixed Model
NON-LINEAR	PLUM	Multinomial model for an ordinal target with 5 links
NON-LINEAR	PLS	Partial least squares
NON-LINEAR	COXREG	Cox proportional hazards regression to analysis of survival times
NON-LINEAR	GENLIN	Generalized Linear Model
NON-LINEAR	GENLOG	multinomial & Poisson general loglinear analysis & multinomial logit analysis
NON-LINEAR	HILOGLINEAR	Multinomail hierarchical loglinear models
NON-LINEAR	LOGLINEAR	multinomial & Poisson general loglinear analysis & multinomial logit analysis
NON-LINEAR	MIXED	Linear Mixed Model
NON-LINEAR	VARCOMP	estimates for variances of random effects under a general linear model
NON-LINEAR	CNLR	Constrained nonlinear regression
NON-LINEAR	LOGISTIC REGRESSION	Logistic regression for a binary target
NON-LINEAR	NLR	Nonlinear regression
NON-LINEAR	NOMREG	Multinomial logit model for a polytomous nominal target
NON-LINEAR	PROBIT	Logistic and Probit (binary)
NON-LINEAR	CSCOXREG	Cox proportional hazards regression for complex samples
NON-LINEAR	CSLOGISTIC	Nominal multinomial logistic regression for complex samples
NON-LINEAR	CSORDINAL	Ordinal multinomial regression with 5 links for complex samples
DATA MINING	Bayes Network	Bayes Network
DATA MINING	NaiveBayes	Self Learning
DATA MINING	SVM	SVM (Support Vector Machine)
DATA MINING	MLP	Neural networks
DATA MINING	RBF	Neural networks

© 2011 IBM Corporation

Categories of Optimization Problems Covered by ILOG Technology

<table>
<tr><td rowspan="3">Mathematical Programming
  </td><td rowspan="2">Continuous Optimization (NP-complete)</td><td>linear programming (LP)
•linear objective function
•linear constraints</td></tr>
<tr><td>quadratic programming (QP)
•quadratic objective function</td></tr>
<tr><td>quadratically constrained programming (QCP)
•quadratic constraints</td></tr>
<tr><td>Discrete Optimization (NP-hard)</td><td>mixed integer programming (MIP)
•one or more non-continuous variables
•includes MILP, MIQP, and MIQCP</td></tr>
<tr><td colspan="2" rowspan="3">Constraint Programming (Combinatorial Optimization)</td><td>Vehicle Routing</td></tr>
<tr><td>Job Scheduling</td></tr>
<tr><td>Custom Search</td></tr>
</table>

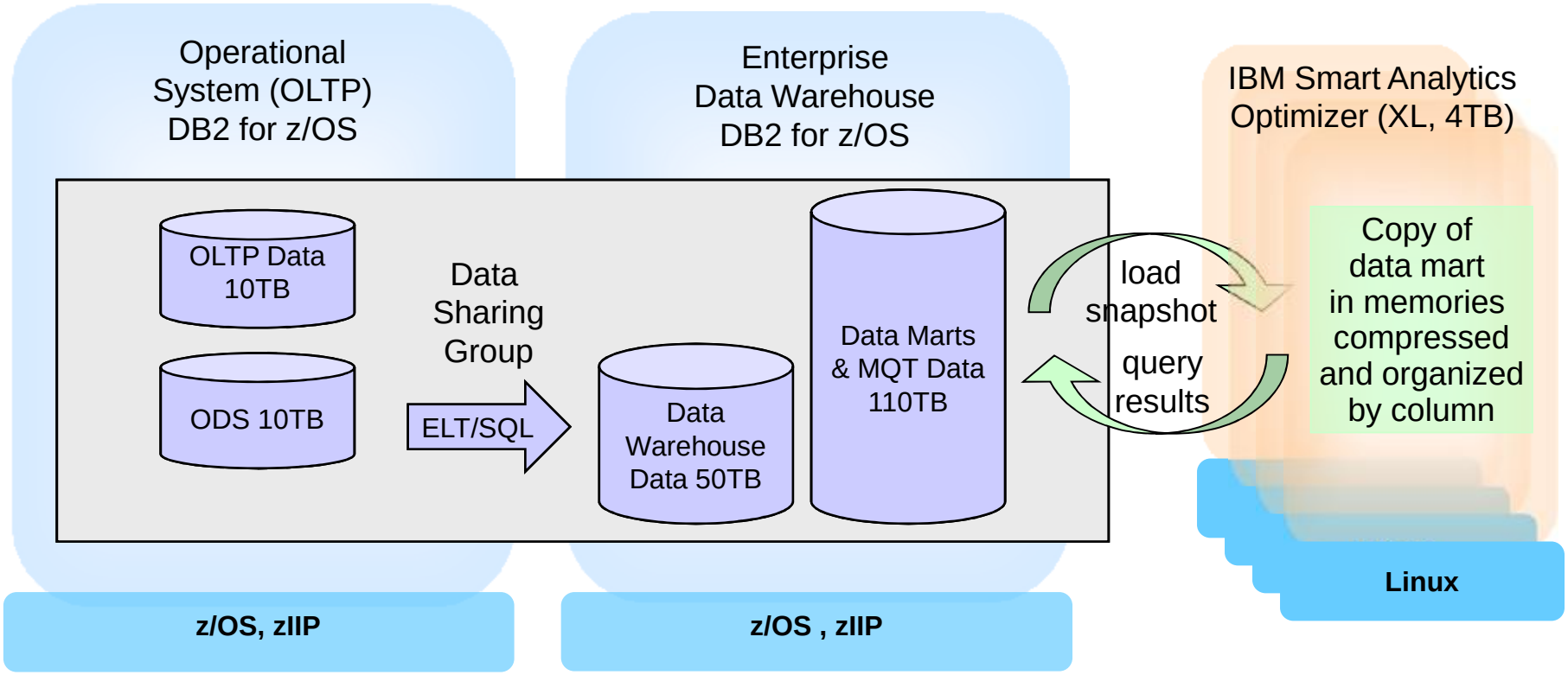
© 2011 IBM Corporation

Major iLOG Algorithms of Mathematical Optimization

- Optimisers
  - Simplex
    - Dual and primal simplex
    - Dual simplex is often the best choice
    - Problems where both dual and primal simplex perform poorly are rare
    - Research literature of running simplex on GPUs exists
  - Barrier
    - Suitable for large, sparse problems
    - The only optimizer for QCP problems
    - Parallel version available
  - Network
    - Suitable for network-flow problems
  - Sifting
    - Suitable for problems with large column/row ratios
    - Extension of simplex
- Search strategies
  - Branch and cut
    - Search tree with nodes being subproblems
    - Parallel version available
  - Dynamic search
    - A variation of branch and cut

© 2011 IBM Corporation

Data Warehousing And OLTP Co-Located On zEnterprise



- Operational data moved to warehouse via ELT
- DB2 for z/OS centrally manages warehouse and data marts

- ISAO accelerates query execution
- Transparent to applications

ODS – Operational Data Store

© 2011 IBM Corporation

Summary

- Business Analytics exploits operational data to try to operate your business better.
- Fully integrated solution: HPC + algorithms + transactions + data => insight

Ø **Cognos, SPSS, ILOG, Infosphere WH with DB2/zOS provide the base for powerful new integrated Business Analytics Solutions with real time OLTP applications**

Ø **Emerging host/accelerator programming models will facilitate the ease of exploiting co-processors without specific accelerator architecture knowledge with cross-vendor portability**

Ø **zEnterprise with integrated attached co-processors provides a unified combination of scalability, aggressive single thread performance and Power based throughput computing threads and vector processing**

© 2011 IBM Corporation

Questions

© 2011 IBM Corporation

SPSS Predictive Analytics Models Available on System z

- SPSS on Linux for System z supports over 30 models,
  - The 8 popular models support database push back for scoring in DB2 z/OS.
  - 5 popular models now available listed below:

1. Logistic regression, Trees (Algorithm names Include CHAID, Quest, C&R Tree)
   - Finance-Used in banking to predict which customers are credit worthy. Which customers should I make a loan to?
   - Finance, Retail, Insurance, Entertainment-Used in marketing departments to determine which customers are going to respond to an offer
   - Insurance-Used in insurance to determine which claims are legit vs. Fraudulent
   - Telecommunication -Predicting customer churn
2. Cluster Analysis (Algorithm names Include K Means, Kohonen, Two Step)
   - Finance, Banking, Insurance -Used in marketing departments across industries to better understand customer segments
   - Customer attrition analysis
3. Market Basket Analysis (Algorithm name" Apriori)
   - Retail -Product assortment planning
4. Time series analysis/forecasting
   - Retail -forecasting catalog sales, forecasting demand, sales planning
5. Cox Regression
   - Retail, Telecommunications -Predicting the time for customer churn
   - Healthcare -determining the efficacy of a drug

© 2011 IBM Corporation