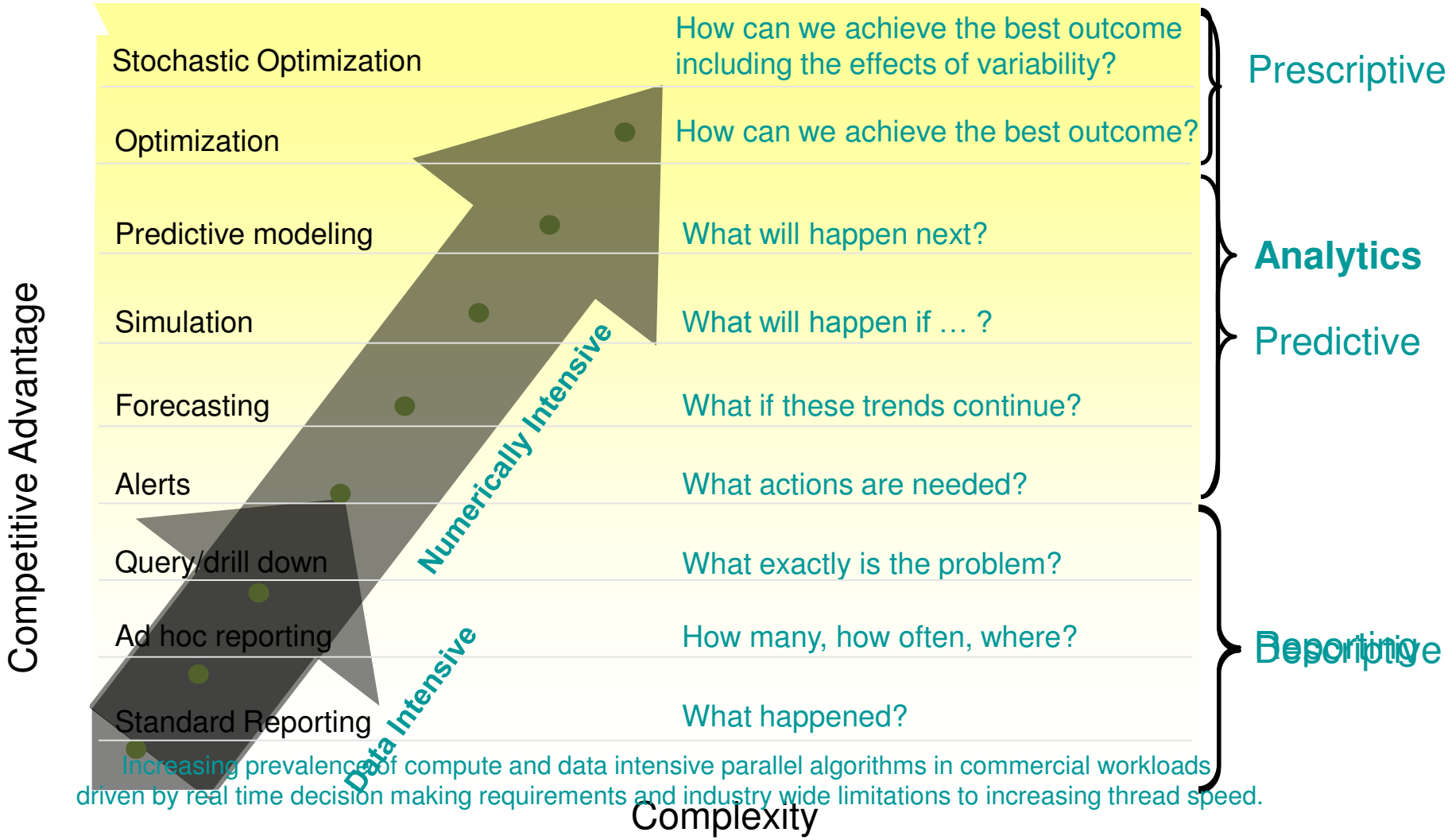
DB2 Architecture for Enhanced Business Analytics

Maryela Weihrauch, IBM Silicon Valley Lab,
Distinguished Engineer, DB2 for z/OS
weihrau@us.ibm.com

ZDB03

zHPC > EdgeHPC > Commercial HPC > Business Analytics
(Mathematical) Analytics Landscape - Today



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

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.

Today's Data Life Cycle "Architecture"

Asynchronous and Distributed

?????

Analyze

OLTP

Operational Systems

Continuous feed

Analytical Foresight

Optimized Business Processes

Customer Support

Sales

Claims Processing

Fraud Management

Underwriting

Marketing

x/p server

Bulk

Data Mining
Segmentation
Prediction
Statistical Analysis

Analytics Server

Data Mover

x/p server

Staging Area

Staging Area

Transformation Server

x/p/z server

ODS (RDBMS)

Batch Process

x/p server

Hourly/daily Batch Process

x/p/z server

Enterprise Data Warehouse (RDBMS)

MIS System
Budgeting
Campaign management
Financial Analysis
Selling Platforms
Customer Profit Analysis
CRM

Multi-Dimensional Analysis

x/p/z server

Departmental

Data Marts

Online Queries & Reporting

Business Objects & Web Intelligence

Report

Cleanse Transform Warehouse

Sample Observations

- **Questionable quality of data in data marts and data warehouses – incomplete, inconsistent scope, …**
  - Data dependency maps for multiple OLTP application sources are large and complex
  - 50-60% of data is replicated between OLTP and ODS/EDW since it's needed by both
- **Moving massive amounts of data between numerous platforms and systems is problematic**
  - Computing resources and time consumption spent to move data instead to process data for business value
  - Data structure changes are complex and slow to propagate through the copy network – business unable to react as fast as it needs on changed requirements
  - Slow to get access to needed data based on changing business requirements
  - Compliance concerns like data usage control, critical business data on personal computers
  - Security exposures, RAS

Business Analytics Life Cycle Architecture on System z

Scoring Rules

Analyze

OLTP

LPAR: z/OS (OLTP)

CICS

OLTP

DB2 for z/OS

Risk Calc.

IMS

WebSphere Classic Federation Server

zLinux

MIS System
Budgeting
Campaign management
Financial Analysis
Selling Platforms
Customer Profit Analysis
CRM

memory

memory

memory

Ad-hoc Queries & Reporting
Business Objects & Web Intelligence

LPAR: z/OS (ODS/EDW)

ODS and EDWH/DM (DB2 for z/OS)

Continuous data feed

Build on request

ELT hourly./daily feed

Departmental

Data Marts

IDAA / Netezza

Bulk

AnalyticsServer

Ongoing/Hourly/daily feed
Under control of DB2

Departmental

Data Marts

Report

Cleanse Transform Warehouse

Business Value of Integrated Architecture

- **Delivers more analytics solutions to the business in shorter time with higher quality**
- **Reduce complexity of data movement results in reduced TCO**
  - Cost to store multiple time
  - Cost to process data movement (HW/SW)
- **Integrated systems to increase data quality and consistency → allows to focus on meaning of the data**
  - quality of data = garbage in -> best algorithm -> garbage out
- **Allows for more focus on reconciliation of data**
- **Common landing area for applications → simplification of consumability**
  - Business has access to all data, cubes and marts for optimization of access

Fit The Solution To The Workload – Access to all Data

Business Analytics / ODS	OLTP
▪ Complex queries against a data warehouse ▪ Read only ▪ Star schema design often used to speed up queries ▪ Different modes of operation	•Transactions against operational data •Reads and writes •Normalized schema design to eliminate redundancy •Multiple user throughput operation

Tightly integrated solution optimized for workloads
Simplification
Single entry point to the true data
High fidelity data
Performance

IBM DB2 Analytics Accelerator DB2 for z/OS ODS members	DB2 for z/OS OLTP members

DB2 Analytics Accelerator

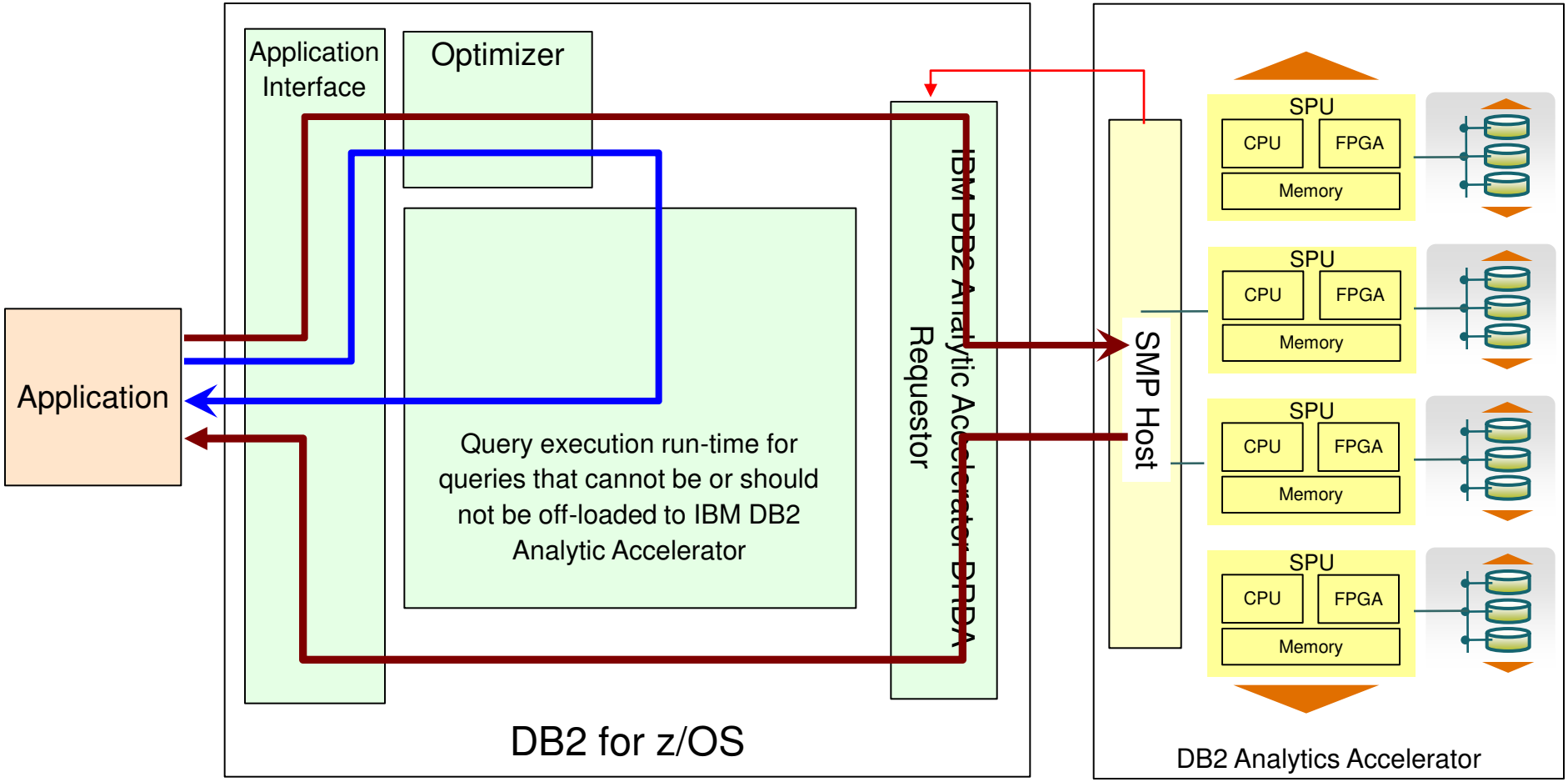
Accelerating decisions to the speed of business

Blending System z and Netezza technologies to deliver unparalleled, mixed workload performance for complex analytic business needs.

Get more insight from your data timely

- Fast, predictable response times for “right-time” analysis
- Accelerate analytic query response times
- Improve price/performance for analytic workloads
- Minimize the need to create data marts for performance
- Highly secure environment for sensitive data analysis
- Transparent to the application

Query Execution Process Flow



Heartbeat (DB2 Analytics Accelerator availability and performance indicators)

Queries executed without DB2 Analytics Accelerator

Queries executed with DB2 Analytics Accelerator

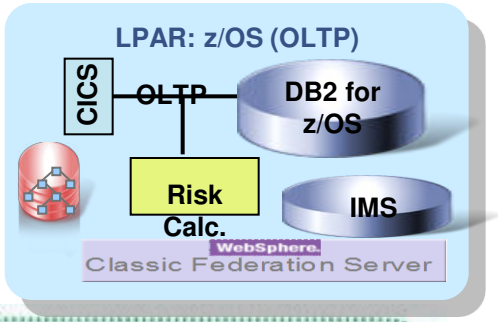
IDAA V3+ Themes

- **Enhancing current capabilities**
  - HPSS enhancements
    - DB2 partition restricted read-only state
    - DB2 Image Copy enhancements
    - Archive SP enhancements
    - Restore functionality
    - RTS last change timestamp column exploitation (DB2 11)
  - Incremental Update enhancements
    - exploiting IFI filtering (DB2 11)
    - Support up to 10 DB2 subsystems
  - WLM support for local applications
  - Enhanced Monitoring (system level)
  - Enable multiple encoding
  - HA: Workload balancing in group
  - Automated NZKit Install
- **Increase offload capability/more query acceleration**
  - Support of static SQL
  - Local multi-row fetch
  - Implicit casting (comparison VARCHAR and numeric)
  - More built-in functions (e.g. BITAND)

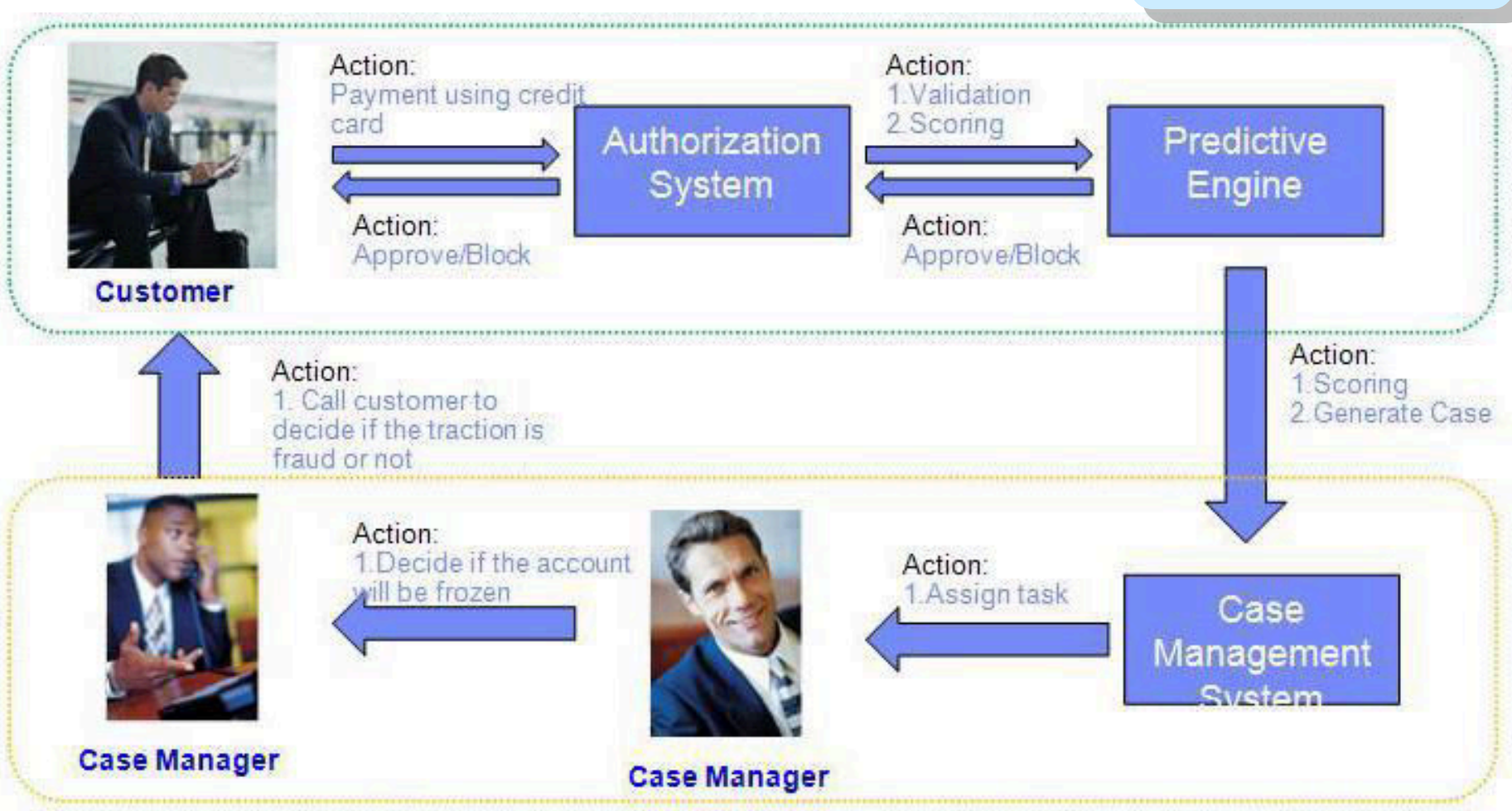
Changing Landscape for Business Analytics on z Based on Well Established zSystem Core Values …

- **The z Business Analytics Software Portfolio**
  - Cognos on z/Linux and z/OS
  - SPSS on zLinux
  - SPSS scoring in DB2 10 for z/OS User Defined Functions
  - ILOG BRMS and CPLEX Optimization on z/OS
  - DB2 10 for z/OS and QMF
    - Bi-temporal function – point in time/end of business query
- **IDAA with integrated Netezza technology**
  - Industry proven workload optimized analytics appliance integrated with DB2 for z/OS
- **Efficient data movement (ELT instead of ETL)**
  - Low latency movement of data with Replication Server

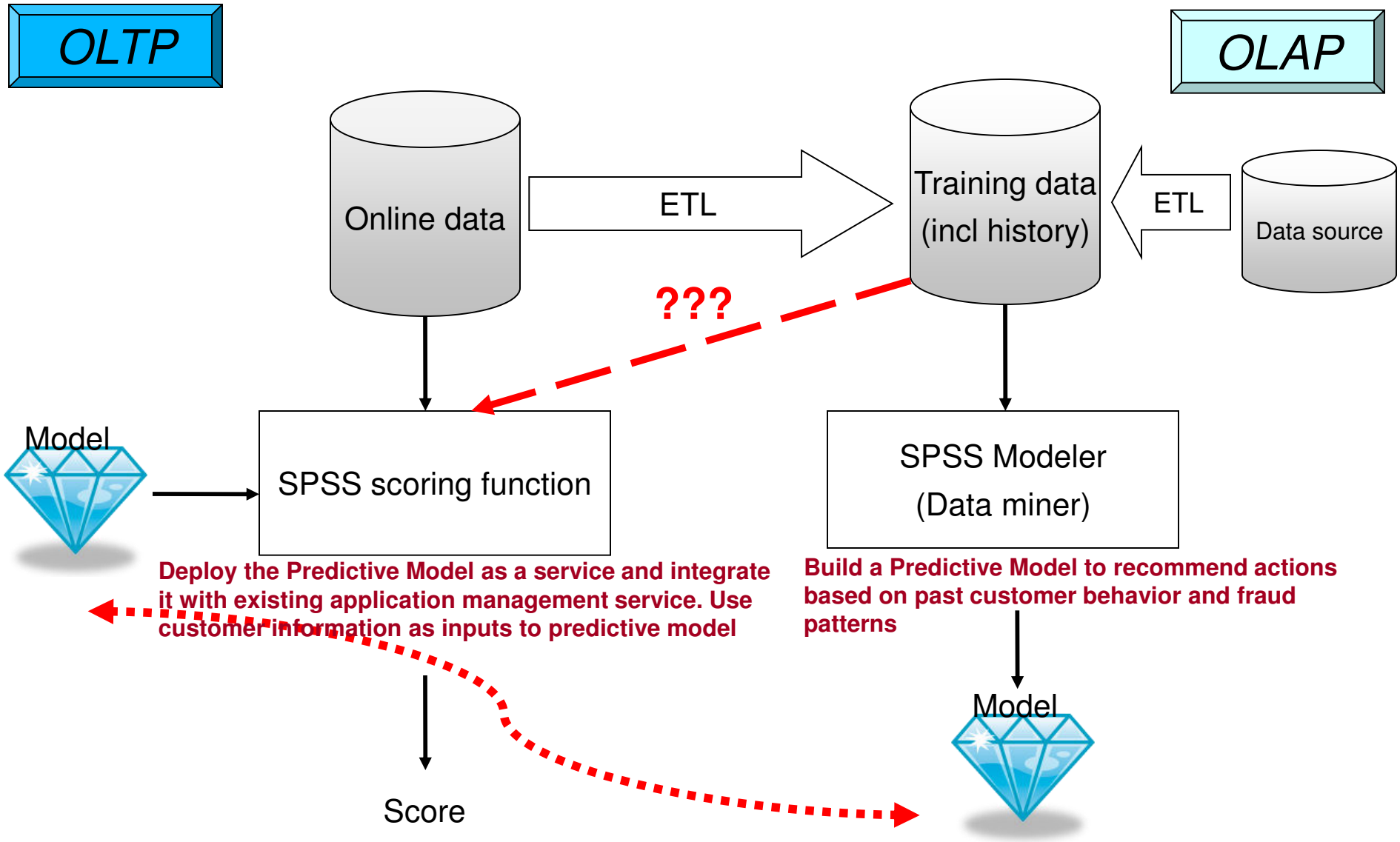
SPSS Real Time Transactional Analytics
Sample scenario: Fraud Prediction and Risk Foresight



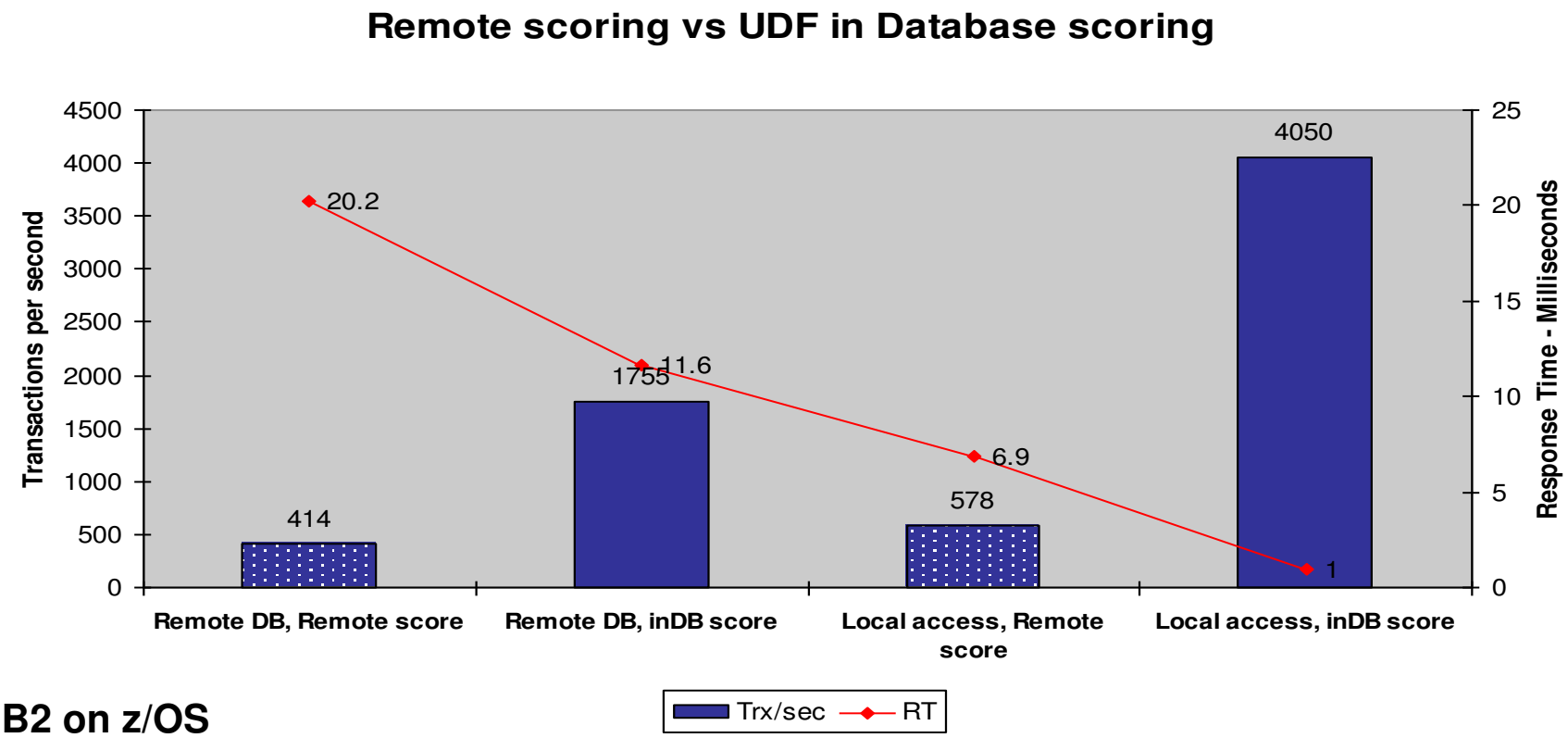
Within a payment transaction, do a real-time decision whether the transaction is a fraud.



Predictive Analytics Processs - OLTP>OLAP>Model>Score>OLTP



Real Time Scoring – Performance Comparison



DB2 on z/OS
z196 LPAR with 2 CPs

SPSS
Linux on z
z196 LPAR with 2 IFL

- Measurements optimized for max throughput on fully utilized system.
- Response times include full transaction with multiple DB accesses

Real-time, data-oriented solutions transform these questions into actionable insights

How do I target and retain my best customers?

- **ex.: churn management**

Work from a single, current view of the customer

Next Best Action ***solutions***

How do I reduce fraud?

- **ex.: real-time fraud identification/prevention**

Drive insights directly into payment systems

Anti-Fraud ***solutions***

How do I manage risk?

- **ex.: operational and financial risk visibility**

Develop a real-time enterprise view of risk-related data

Governance, Risk and Compliance ***solutions***

How do I focus limited resources where they will be most effective?

- **ex.: supply-chain optimization**

Know where everything and everyone is right now

Resource Optimization ***solutions***

Solutions based on System z unique technology make preventing **opportunistic fraud economically viable**

Banking

Anti Money Laundering

Card Fraud

Government

Tax Fraud

Benefits Fraud

Health Care

Provider/ Vendor Fraud

Insurance

Claims Fraud

Workload-optimized for predictive fraud analytics

- Unique ability to score models for real-time updates, at the time of transaction
- Detect a higher percentage of fraudulent claims pre-payment without negatively impacting claims processing efficiency
- Consistent performance due to compute-intensive analytics processing capabilities

Integrated hub for payment and operational data

- Payment processing and predictive fraud analytics are integrated, accessed and analyzed on the same platform
- Maintains current performance and throughput targets
- Reduces movement and proliferation of data copies
- Consistent security and operational processes

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

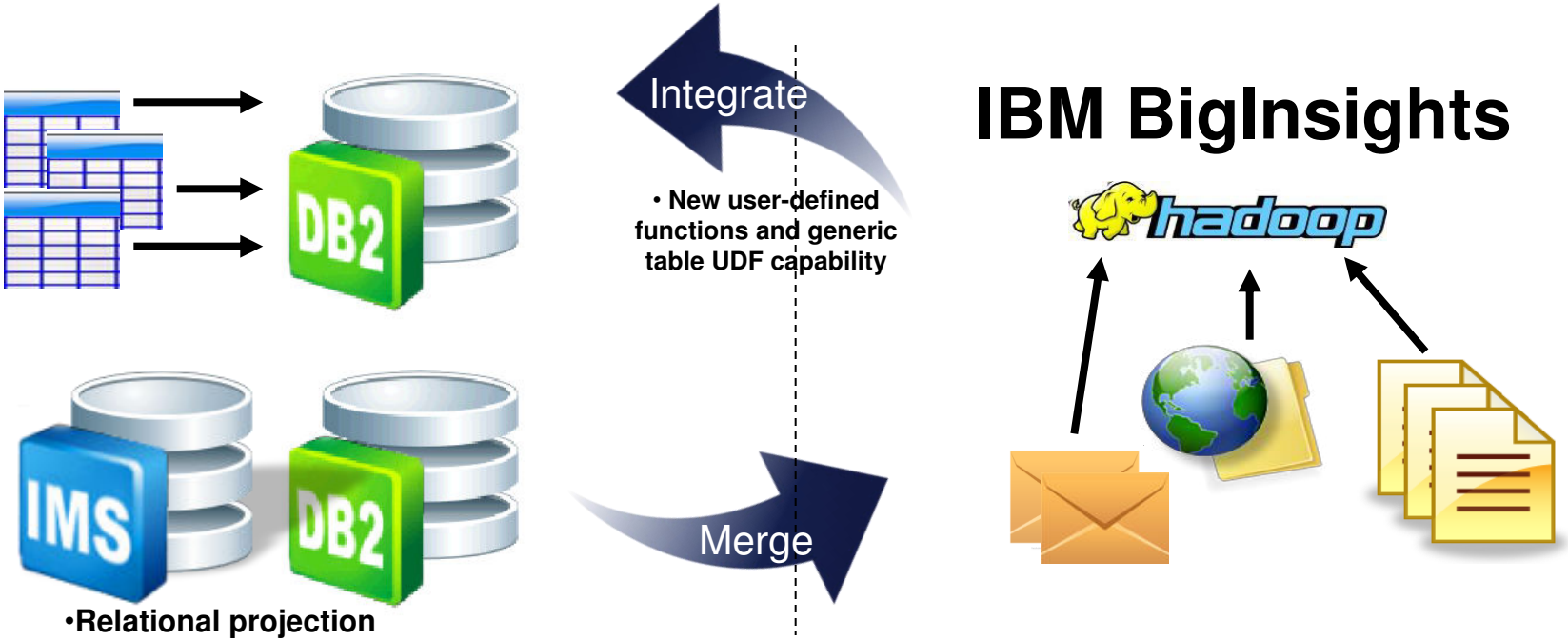
- Analytics and decision engines reside where the DWH / transaction data is
- "Noise" (veractity) surrounds the core business data
  - Social Media, emails, docs, telemetry, voice, video, content
- Expanding our insights – getting closer to the "truth"
  - Lower risk and cost
  - Increased profitability

"Circle of trust" widens

Enhancing Big Data Analytics with IMS and DB2 for z/OS

▪ Much of the world's operational data resides on z/OS	▪ Unstructured data sources are growing fast

- Two significant needs:
  1. Merge this data with trusted OLTP data from zEnterprise data sources
  2. Integrate this data so that insights from Big Data sources can drive business actions
- IMS & DB2 are providing the connectors & the DB capability to allow BigInsights to easily & efficiently access each data source
- DB2 is providing the connectors & the DB capability to allow DB2 apps to easily and efficiently access hadoop data sources



DB2 11 Support for Big Data

- **Goal: integrate DB2 for z/OS with IBM Hadoop based BigInsights Bigdata platform**
  - Enabling traditional applications on DB2 for z/OS to access Big Data analytics.
- **Analytic jobs can be specified using JSON Query Language (Jaql)**
  - Submitted to BigInsights
  - Results stored in Hadoop Distributed File System (HDFS).
- **A table UDF (HDFS_READ) reads the Bigdata analytic result from HDFS, for subsequent use in an SQL query.**
- **Must have a variable shape of HDFS_READ output table**
  - DB2 11 supports generic table UDF, enabling this function

Big Data Use Cases

Big Data Exploration

Find, visualize, understand all big data to improve decision making

Enhanced 360º View of the Customer

Extend existing customer views (MDM, CRM, etc) by incorporating additional internal and external information sources

Security/Intelligence Extension

Lower risk, detect fraud and monitor cyber security in real-time

Operations Analysis

Analyze a variety of machine data for improved business results

Data Warehouse Augmentation

Integrate big data and data warehouse capabilities to increase operational efficiency


Thank You