In-Database Analytics with R
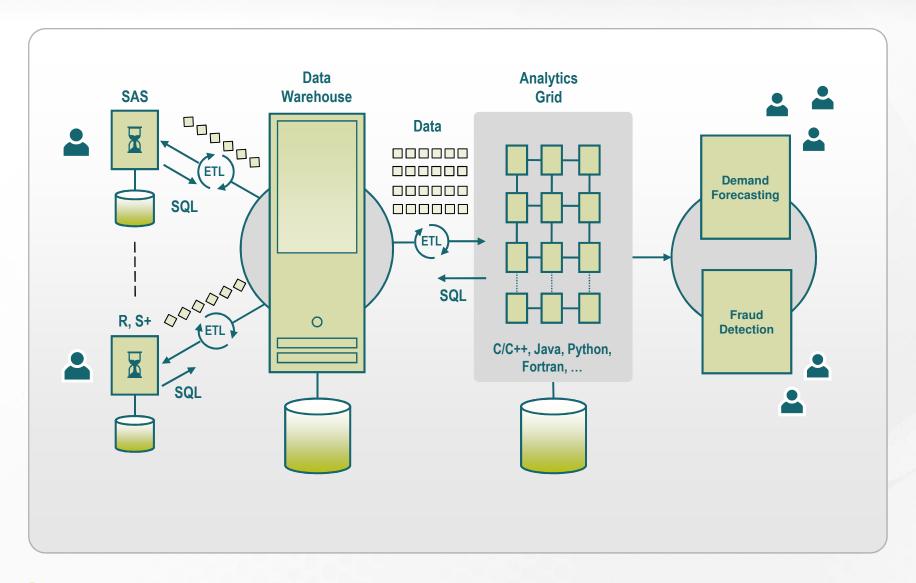
Michele Chambers – Advanced Analytics Product Management, Director

Brian Hess – Advanced Analytics, Director & Principal Mathematician

Agenda

- **What are in-database analytics?**

- **How does in-database analytics processing help you?**

- **Can in-database analytics be used for data mining as well as scoring?**

- **How can you take advantage of a massively parallel architecture to speed up embarrassingly parallel algorithms as well as heroic computations?**
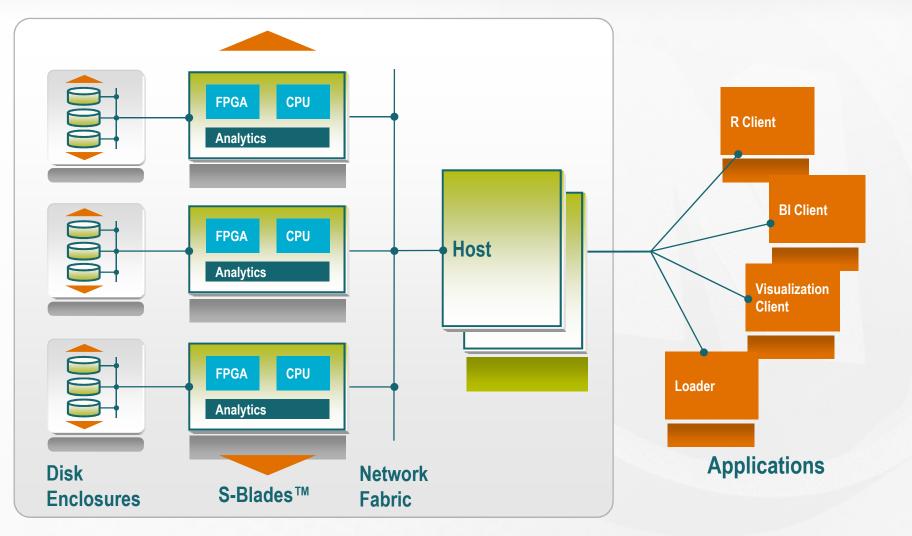
NETEZZA

Advanced Analytics – the Traditional Way

What are in-database analytics?

Embedding of analytics inside the database

so that the computation processing occurs as

close to the data as possible

NETEZZA

What's the Big Deal with In-Database Analytics?

Netezza Appliance

Disk Enclosures · S-Blades™ · Network Fabric

FPGA · CPU · Analytics

Host

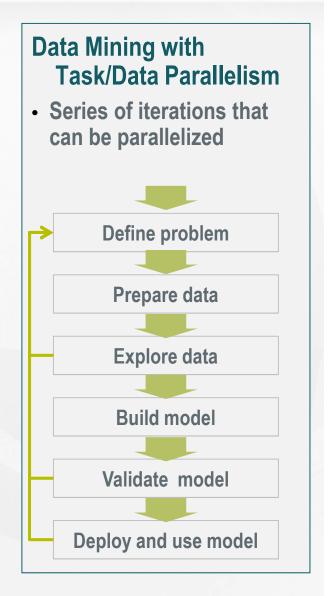R Client · BI Client · Visualization Client · Loader

Applications

NETEZZA

Moving Compute Next to Data As Data Streams By

Task Parallelism

- **Model Simulation / Experimentation**
  - > Concurrent simulation on different data
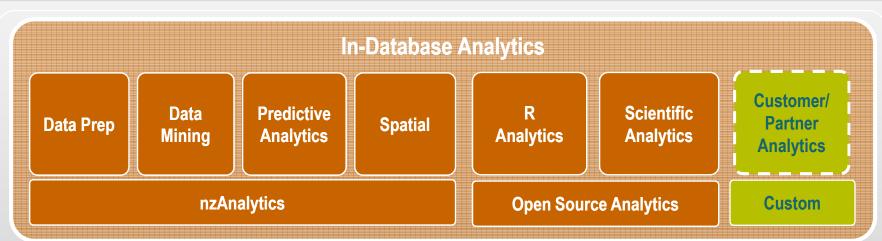 - > 100's different models running against 1M's rows

Monte Carlo Simulation Control Program

Simulation #1 on Core 1	Simulation #X on Core X

Data Parallelism

- **Scoring / Predicting**
  - > Concurrent calculation of a prediction or score on large quantities of data

Scoring Model Control Program

Score calculation on Core 1	Score calculation on Core X

Data Mining with Task/Data Parallelism

- **Series of iterations that can be parallelized**

Define problem
Prepare data
Explore data
Build model
Validate model
Deploy and use model

twinFin™

In-Database Analytics

Data Prep	Data Mining	Predictive Analytics	Spatial		R Analytics	Scientific Analytics	Customer/ Partner Analytics
nzAnalytics					Open Source Analytics		Custom

Parallel Analytic Engines

nzMatrix	nzEngine for Hadoop	nzEngine for R

Software Development Kit
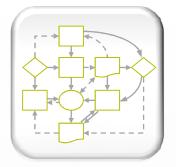
nzAdaptors for C, C++, Java, Python, Fortran	nzPlug-in for Eclipse	nzPackage for R GUI

Streaming Accelerator

Netezza AMPP™ Platform

What does In-Database Analytics Deliver?

Efficient process

Unlimited analytic complexity

Faster turnarounds

Ability to react to market

Ability to experiment

Unlimited/ current data

Reduce expenses

Use R! 2010 - Netezza - In-database Analytics with R

So, What Should I Look For in a Database?

In-database analytics checklist

1. **Data streaming**

2. **Flexible, easy-to-use in-database mechanisms**

3. **Easy, fast, extensible development environment**

4. **Wide availability of tools including open source tools**

5. **Industry accepted standards/tools**

6. **Easy to manage and maintain**

Thank you

Michele Chambers mchambers@netezza.com 508.382.8264

Brian Hess bhess@netezza.com 508.382.8471