IBM SOFTWARE CONFERENCE 2011
Powered by Impact & Innovate 2011

IBM Netezza Analytics

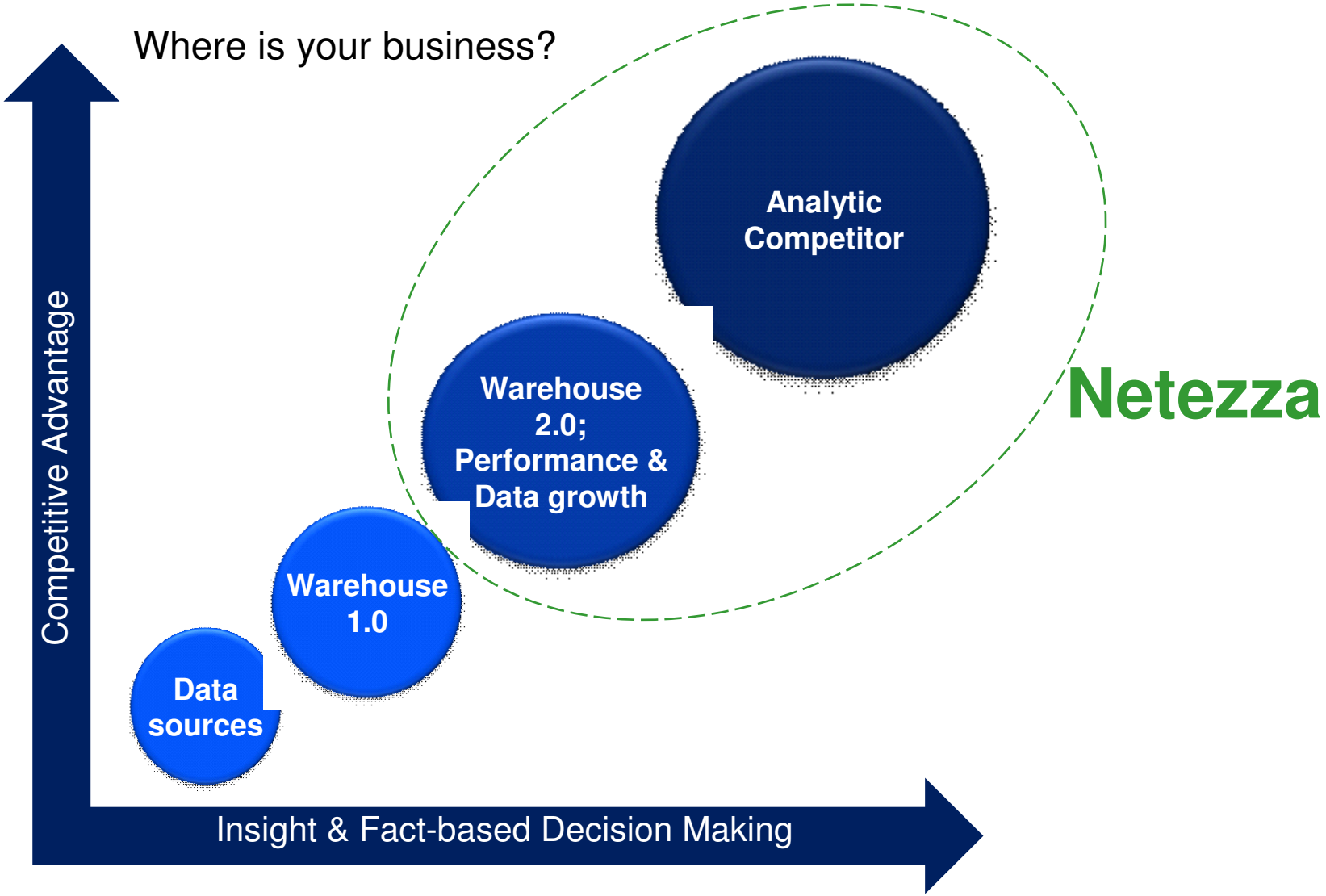
From Data Warehousing to Insight – Enterprise Analytics

Agenda

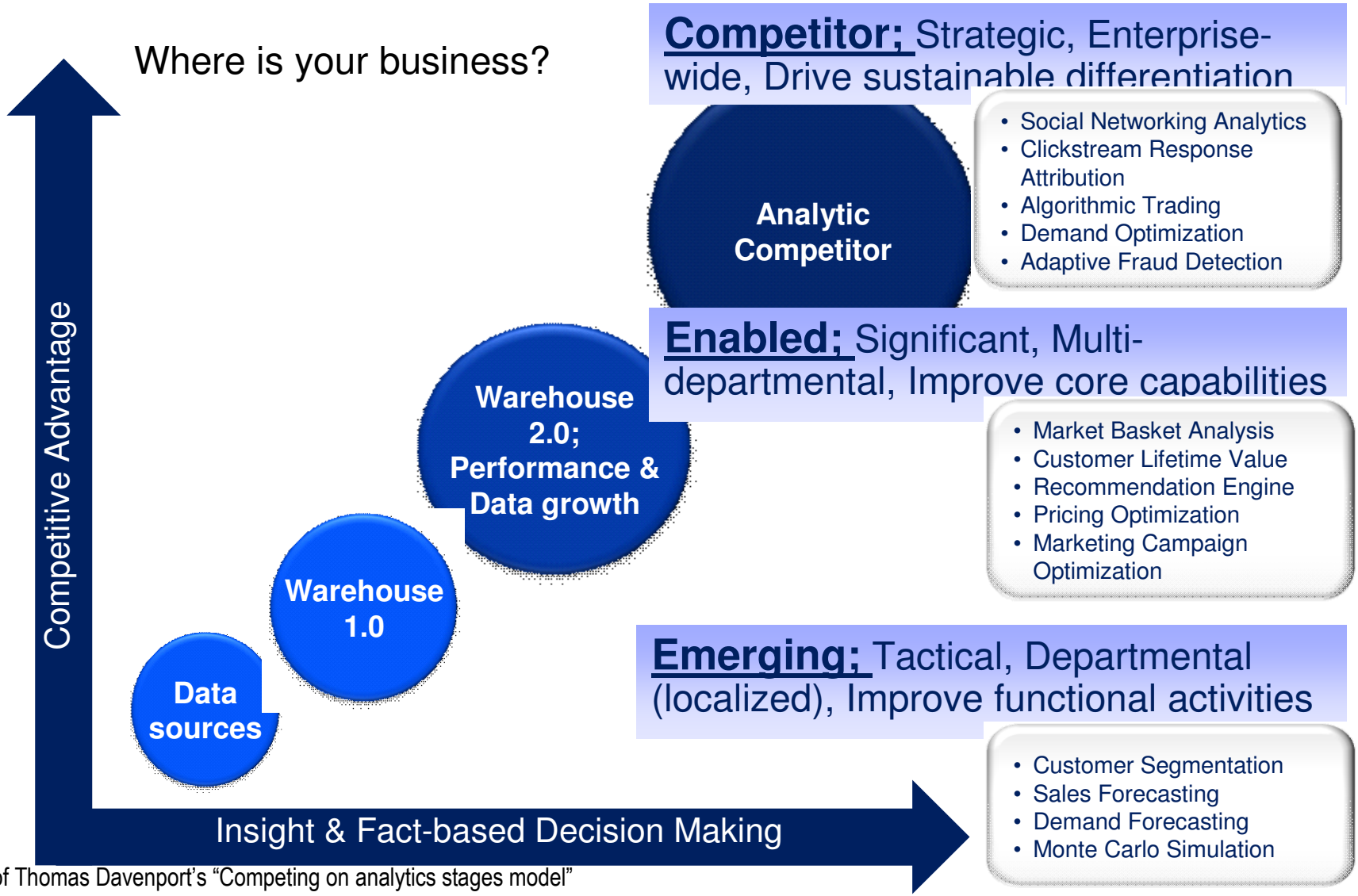
- The journey to insight; path and characteristics
- The Netezza data warehouse appliance
- Analytics today
- How Netezza can help deliver analytic insight

Presentation objective: When I finish this presentation, you would have thought about your businesses current BI position, thought about what it could be, and plan to speak with IBM about your thoughts, and our guidance.

From Data Warehousing to Analytics



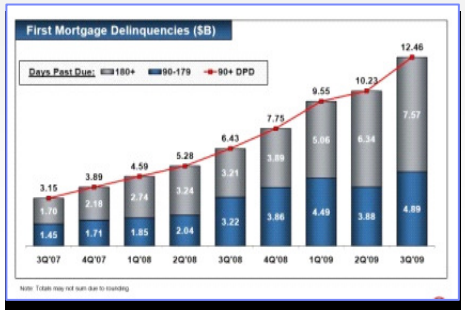
From Data Warehousing to Analytics



*Adaptation of Thomas Davenport's "Competing on analytics stages model"

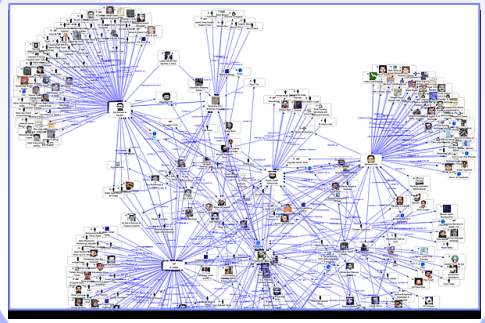
Analytic Benefits Increase as Organizations Move Towards Optimization

BI Reporting and Ad-Hoc Analysis



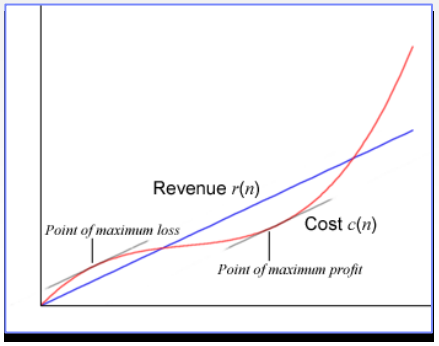
- What happened?
- When and where?
- How much?

Predictive Analytics



- What will happen?
- What will the impact be?

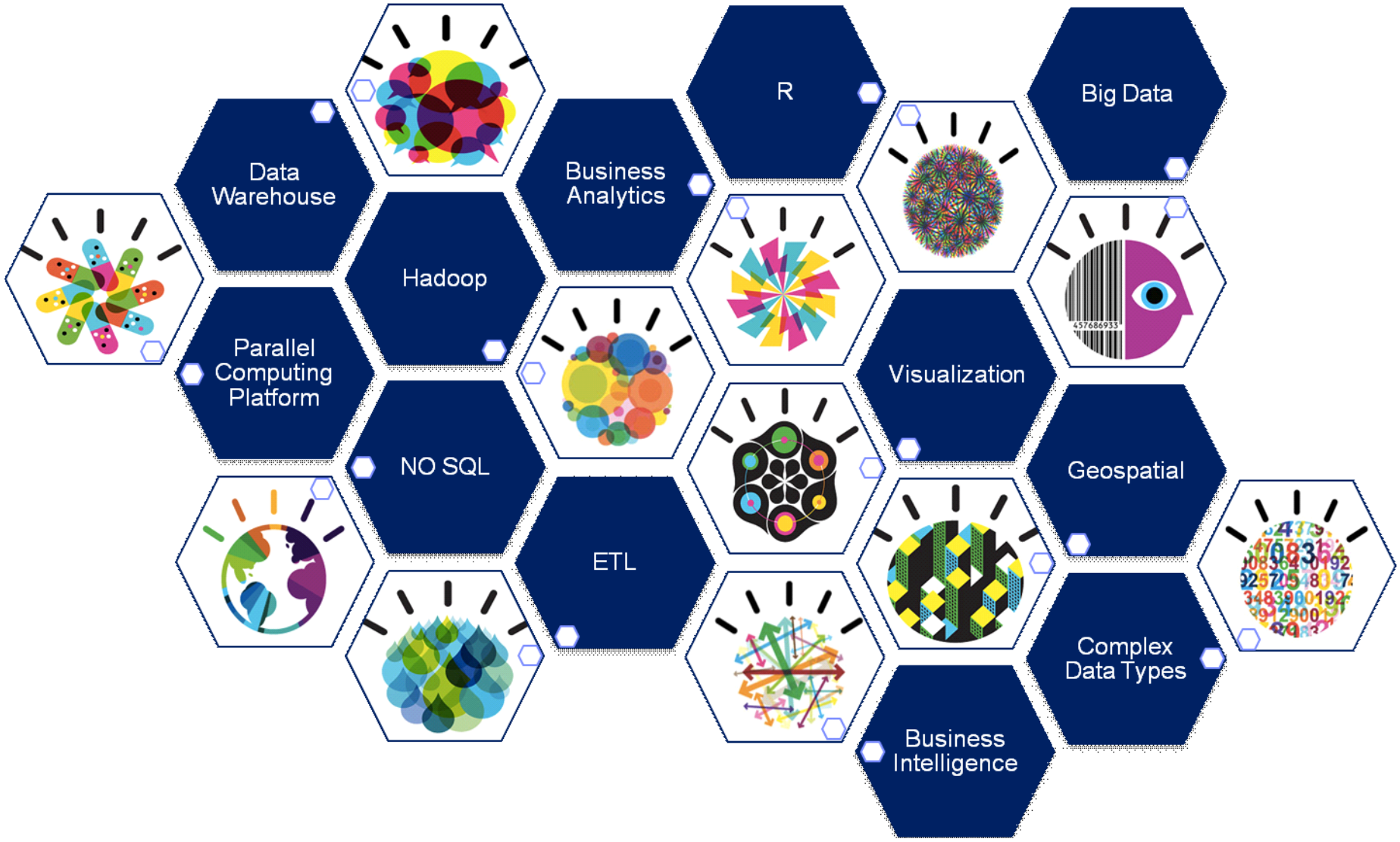
Optimization



- What is the best choice?

More challenging

What is Critical for Enterprises Building a Smarter Planet?



IBM Netezza – A Smarter Appliance for Smarter Customers

- Parallel Computing Platform
- Warehouse for Complex Data Types
- Hadoop & NoSQL
- Business Analytics & R
- Business Intelligence
- ETL & ELT
- Geospatial

TwinFin: The true data warehousing appliance

- *Purpose-built analytics engine*
- *Integrated database, server and storage*
- *Standard interfaces*
- *Low total cost of ownership*

- Speed: 10 -100x faster than traditional system
- Simplicity: Minimal administration and tuning
- Scalability: Peta-scale user data capacity
- Smart: High-performance advanced analytics
- Value: price / performance, price / terabyte

Netezza TwinFin

DISK ENCLOSURES

Slice of User Data
Swap and Mirror Partitions
High-Speed Data Streaming

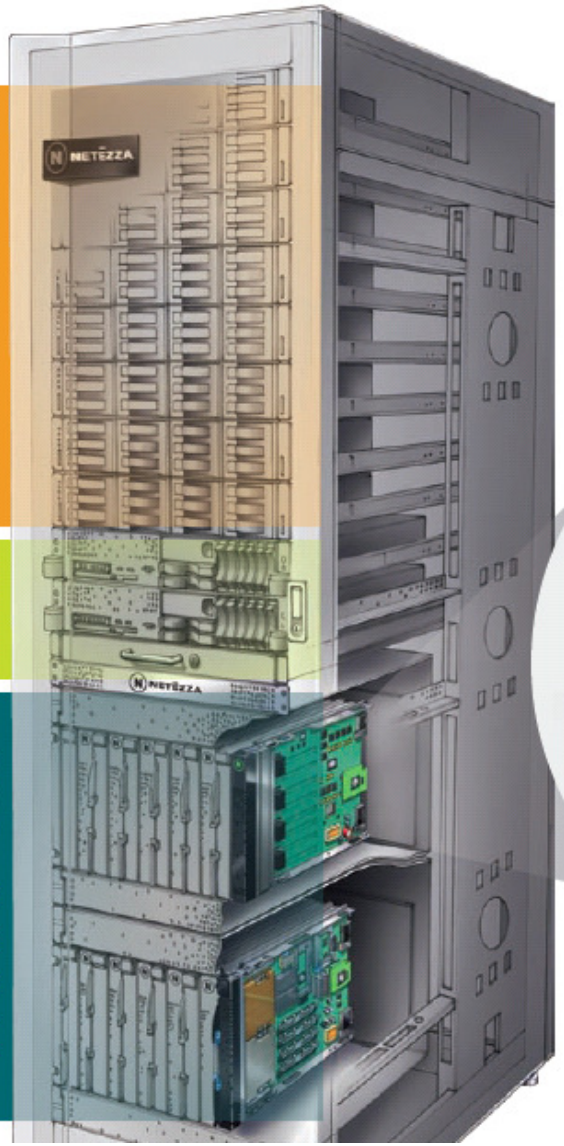
SMP HOSTS

SQL, Query Plan, Optimize, Admin

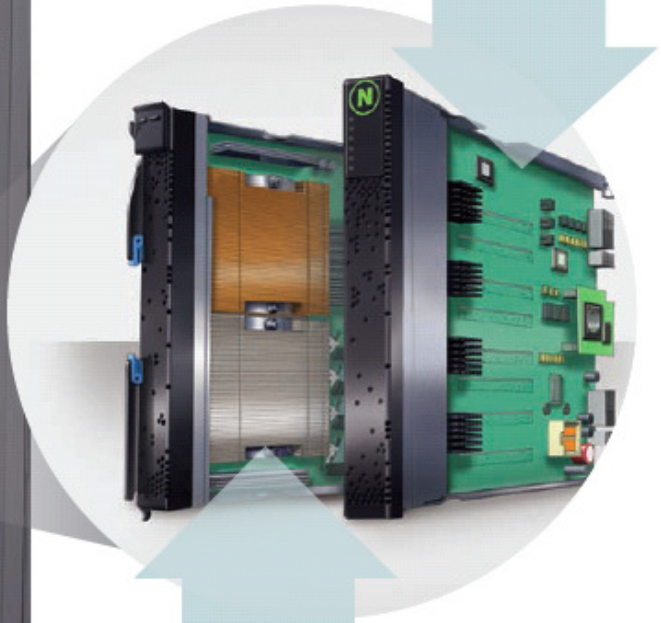
SNIPPET BLADES
(S-BLADES)

Processor & Streaming DB logic

High-performance database engine, complex analytic processing, streaming joins, aggregations, sorts, etc...



TwinFin uses Field Programmable Gate Arrays (or FPGAs) which have been programmed by Netezza specifically to handle **Big Data** very efficiently. These FPGAs filter out extraneous data as fast as it streams off the disk. This removes I/O bottlenecks and frees up downstream components such as the CPU, memory and network from processing unnecessary data, creating a significant turbocharger effect on system performance.



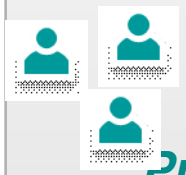
The **Big Math** on TwinFin is performed in powerful multicore CPUs, where database primitives and complex analytics are executed on the filtered data stream. Analytic tasks are run as independent processes operating on data streams on each

Traditional Analytics

The analytics process can be thought of as 2 distinct activities – modelling and prediction.

***Modelling** is the process of mining historical data to identify patterns and relationships of interest.*

Once identified, modellers build a mathematical model that describes a particular behaviour such as propensity to buy or churn, potential fraud, etc.

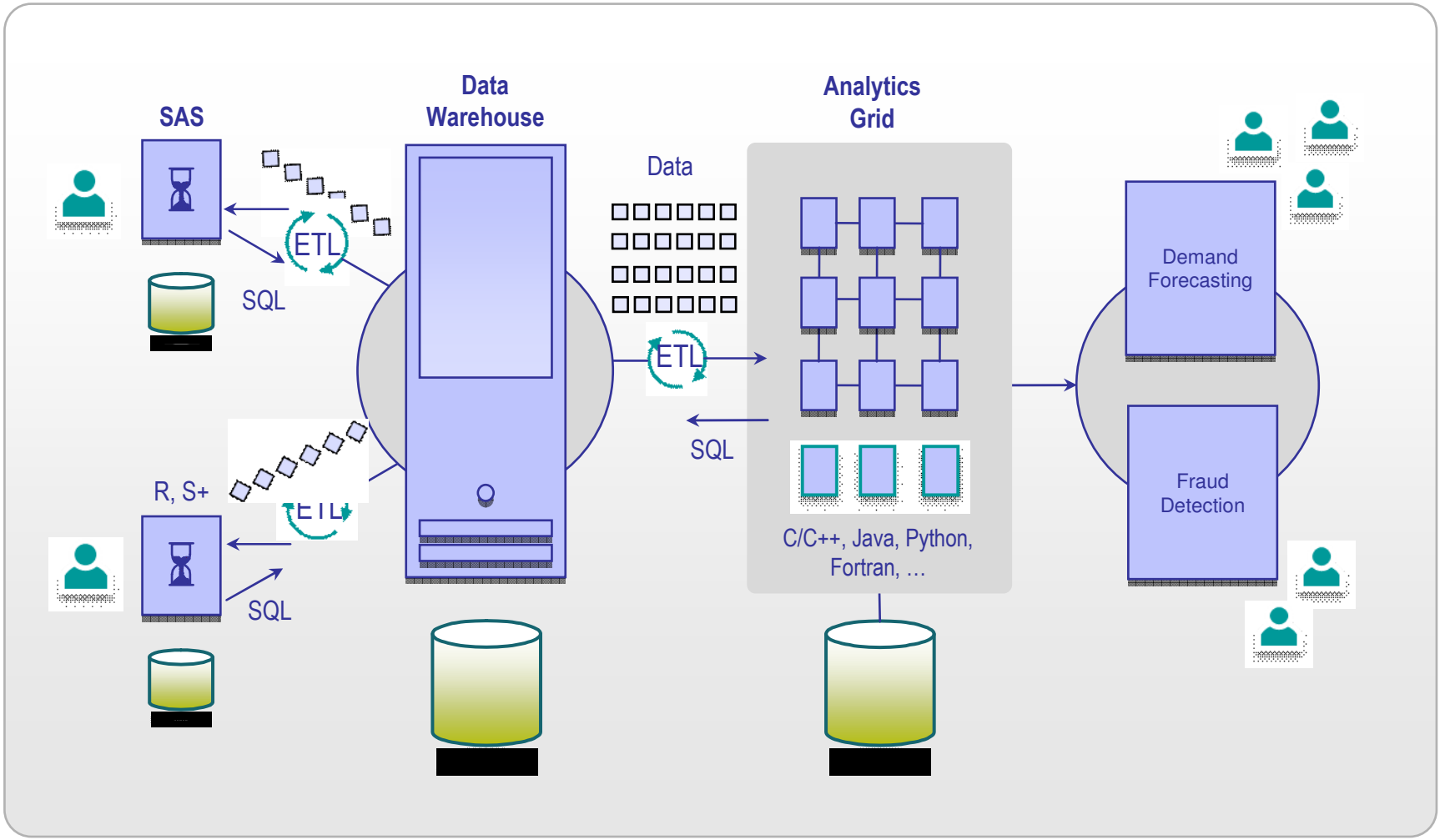


***Prediction** is the process of applying the model to data to predict the event described by the model.*

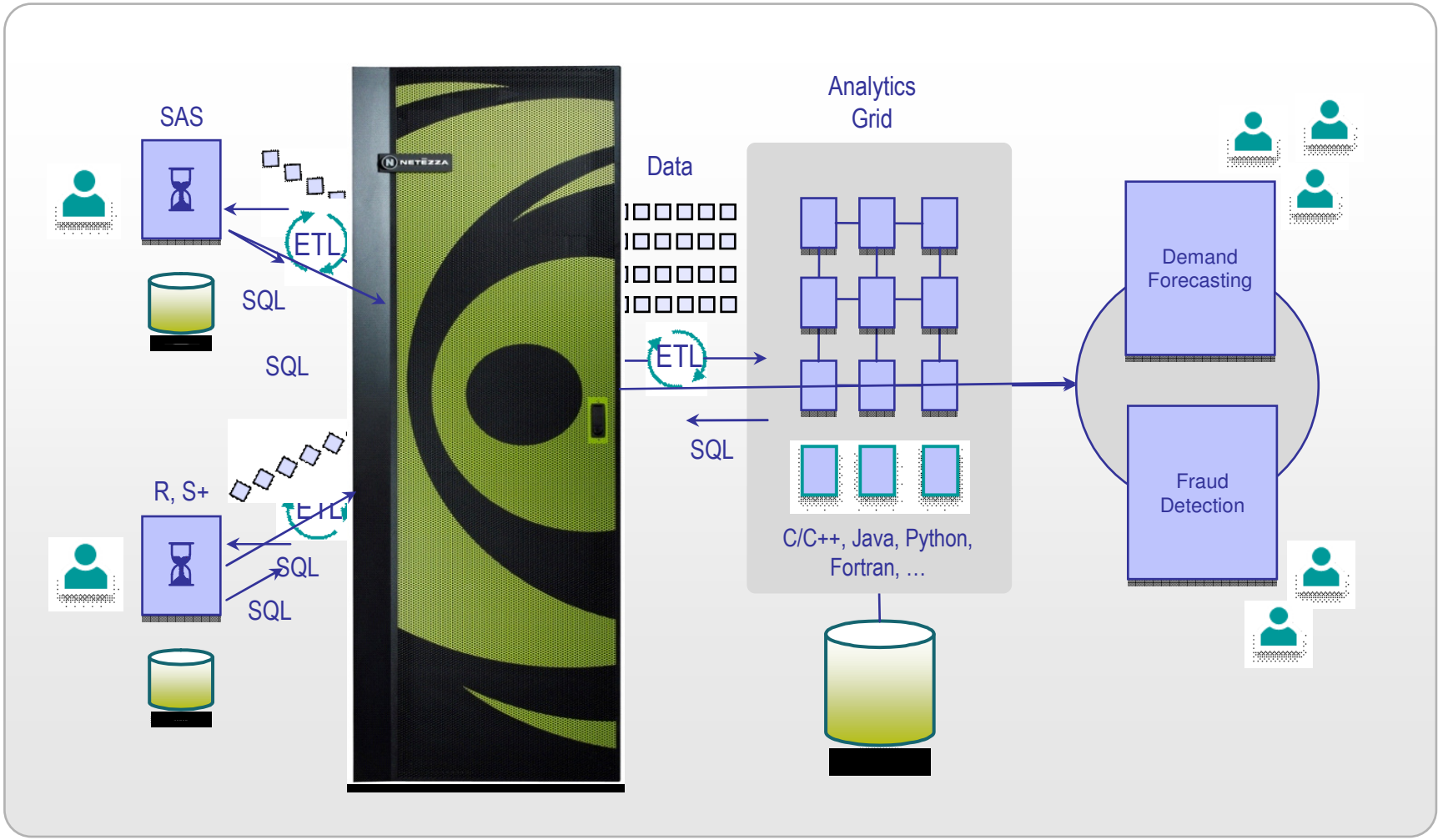
Challenges

- Modellers and statisticians have own workstations or servers where they perform modelling tasks.
- Requires moving and modifying large amounts of data from DW to the analytic tool.
- Forces modelling to be done only on a small sample instead of all the historical data available, lots of IO and data getting distributed in lots of silos across the organization.
- Once a model is built, it is deployed on an analytics server, typically a large SMP server or grid that once again pulls data off the DW to run the predictive model on. The prediction is done off-line and data loaded back onto the DW.

Advanced Analytics – the Traditional Way

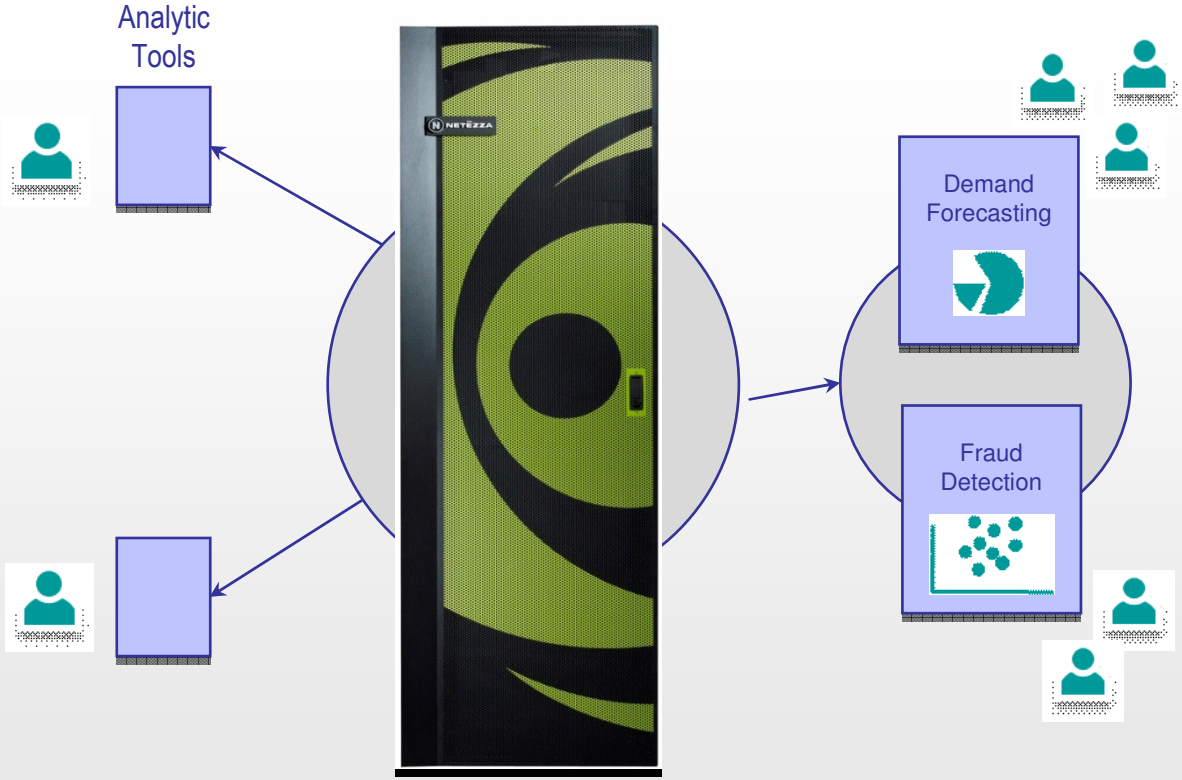


Advanced Analytics with IBM Netezza



Advanced Analytics with IBM Netezza

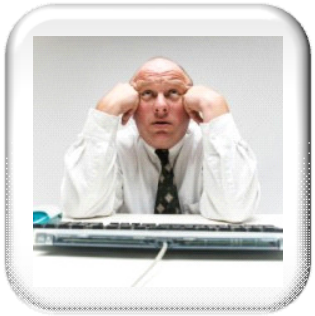
Extremely flexible analytics platform that offers orders of magnitude performance at petascale.



Brings analytics to the data, modellers who can operate on the data directly inside the appliance, ask the most complex questions on all the enterprise data.

Once the model is developed, it is seamless to put it into prediction mode. The prediction and scoring can be done right where the data resides

Challenges to Building & Deploying Analytics at Scale



Inefficient process

Limited analytic sophistication

Inability to course correct

Time consuming – model & moving data

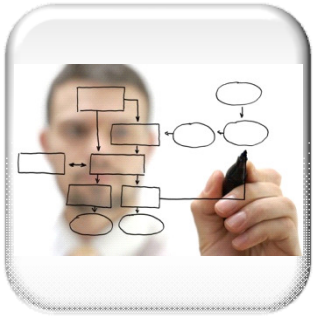
Inability to experiment

Limited & stale data

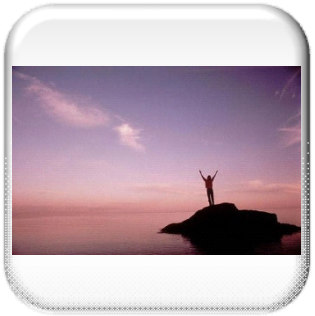
High total cost of ownership

Despite applying time and resources, companies are unable to fully exploit their data resources and ultimately

What does in-database analytics deliver?



Efficient processes



Unlimited analytic possibilities

Ability to react to market

Faster turnaround

Ability to experiment

Unlimited & current data

Low total cost of ownership

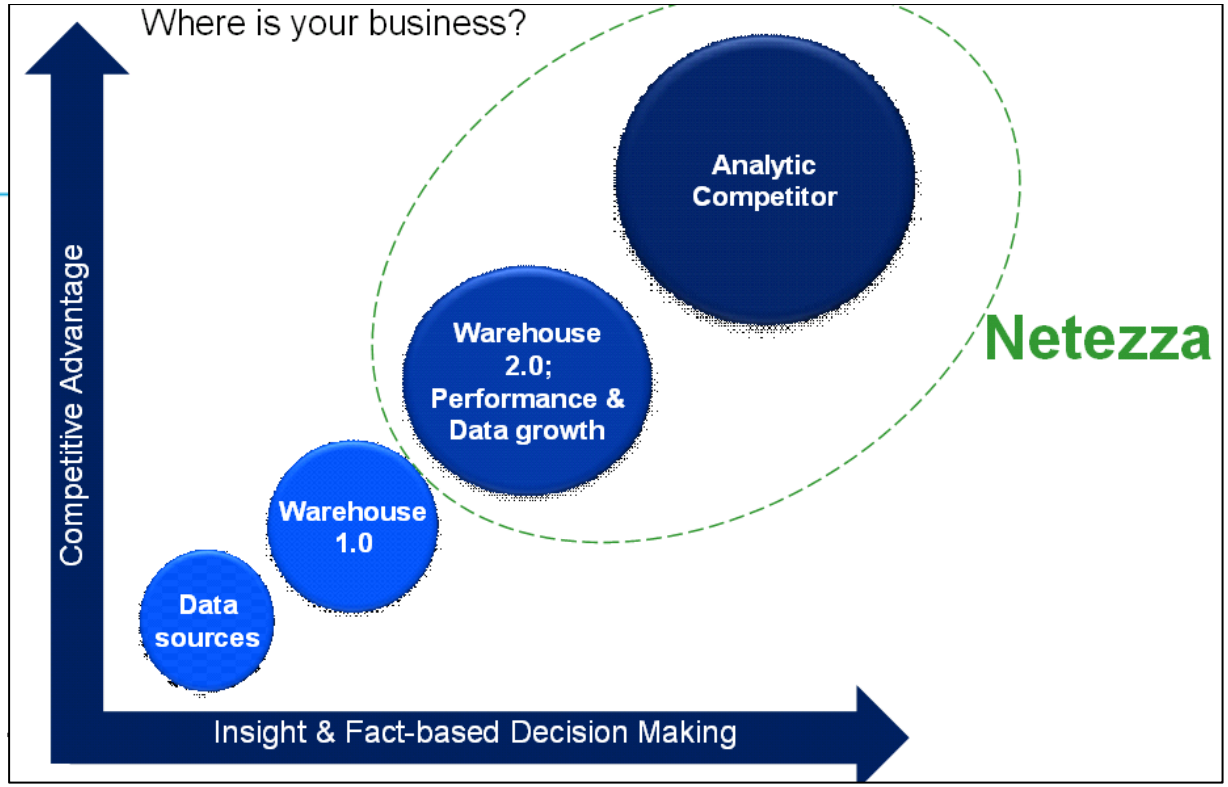
The Simple Appliance Built for Serious Analytics

- Data Exploration
  - Discover subtle patterns within massive data sets
  - Answer interrelated and complex business questions
- Analytics On-demand
  - Respond quickly to dynamic business conditions and choose the best course of action
- Simple to Use
  - Use your existing analytics infrastructure and skill sets
  - Built-in library of analytic functions
  - Support for popular analytic packages and languages
- High Performance
  - 10-100X complex analytics performance at petascale
  - In-database, parallelized algorithms
  - Asymmetric massively parallel processing (AMPP) architecture

Answer questions that were previously too complex, too time consuming, and required too much data

IBM SOFTWARE CONFERENCE 2011
Powered by Impact & Innovate 2011

Where to from here?



If your business is yet [illegible] competitive tool, then I urge you to start considering your path. Your competitors may very well be.

What's holding you back? Infrastructure, platform, implementation, experience, subject knowledge?

You have a business partner here today, IBM.
We have a lot of experience here in the room to help you.

Thank you

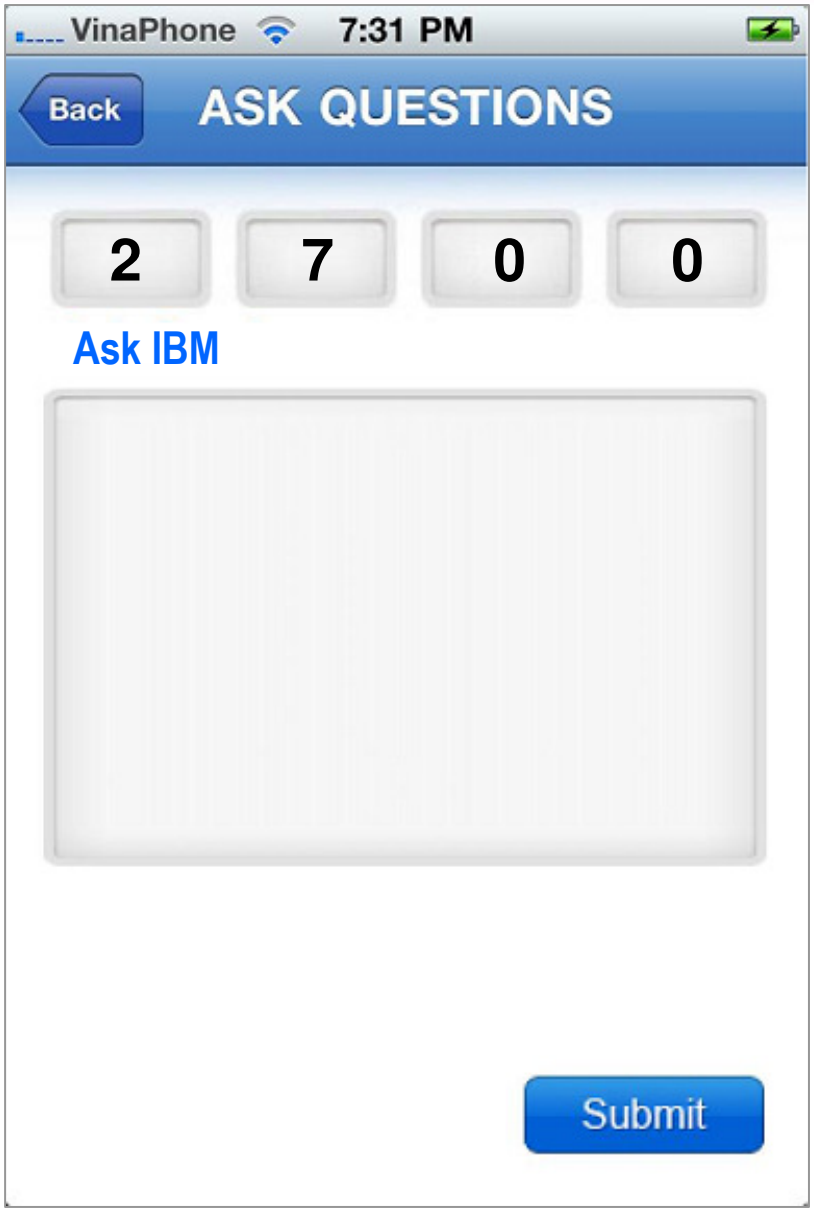
IBM SOFTWARE CONFERENCE 2011
Powered by Impact & Innovate 2011

Question

What isn't one of the common challenges with traditional analytics?

A) Moving of data for analysis
B) Modeling on a subset of data
C) Complex deployment model
D) Analytics isn't strategic

To Ask James a Question:

PIN CODE:

2700