IBM

IBM INFORMATION INTEGRATION
& GOVERNANCE SYMPOSIUM 2012

Delivering Trusted Information for Smarter Business Decisions

High Speed Analytics

Andrew Conacher – Netezza Specialist

May 2012

Do you want to work harder or smarter?

You are already smart. You are some of the way there.

Your environment needs to be smart too!

What others are doing - where *you* can be

500+ Billion market basket records

Leveraged the same BI applications

70 times more queries on 5 times the amount of data

100,000 queries a day on Netezza systems.

Reduced Storage by almost 100TB

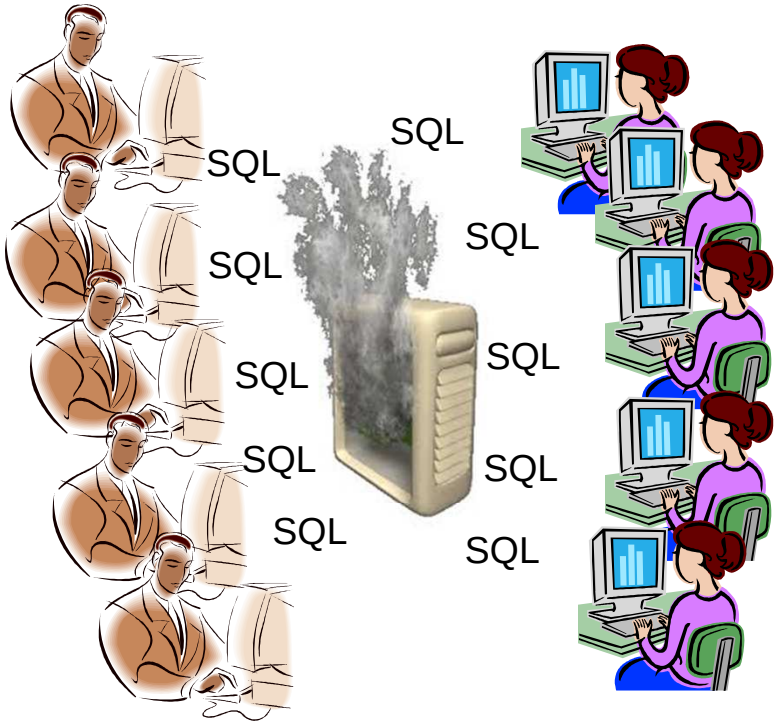
No longer needed aggregates with Netezza

'Monday mornings in hell' now a thing of the past

CATALINA MARKETING

How we started out

Many users accessing a general purpose shared everything RDBMS server…

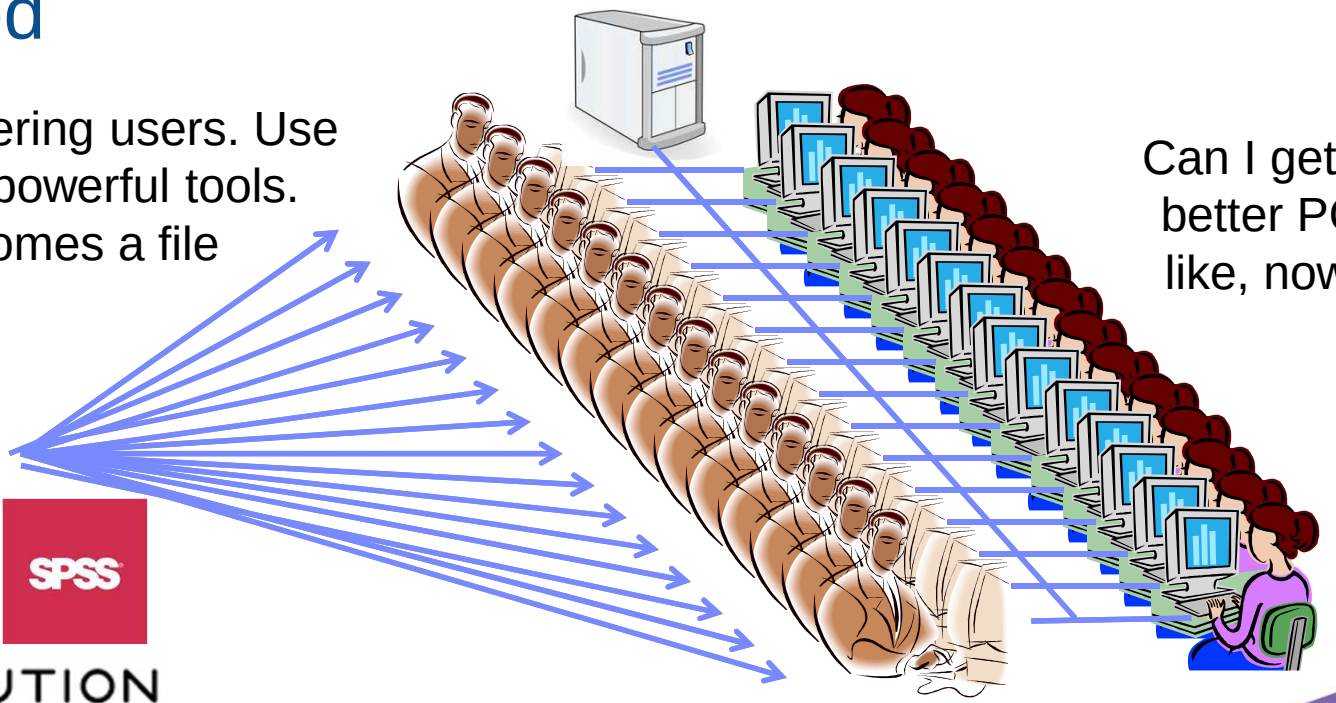


Crossing fingers it won't blow up. Get in early, stay back late to get your work done at quiet times when the server isn't as busy

twitter: Follow @ANZ_IM or mention #IIGS

Analytics at scale disconnected, not integrated

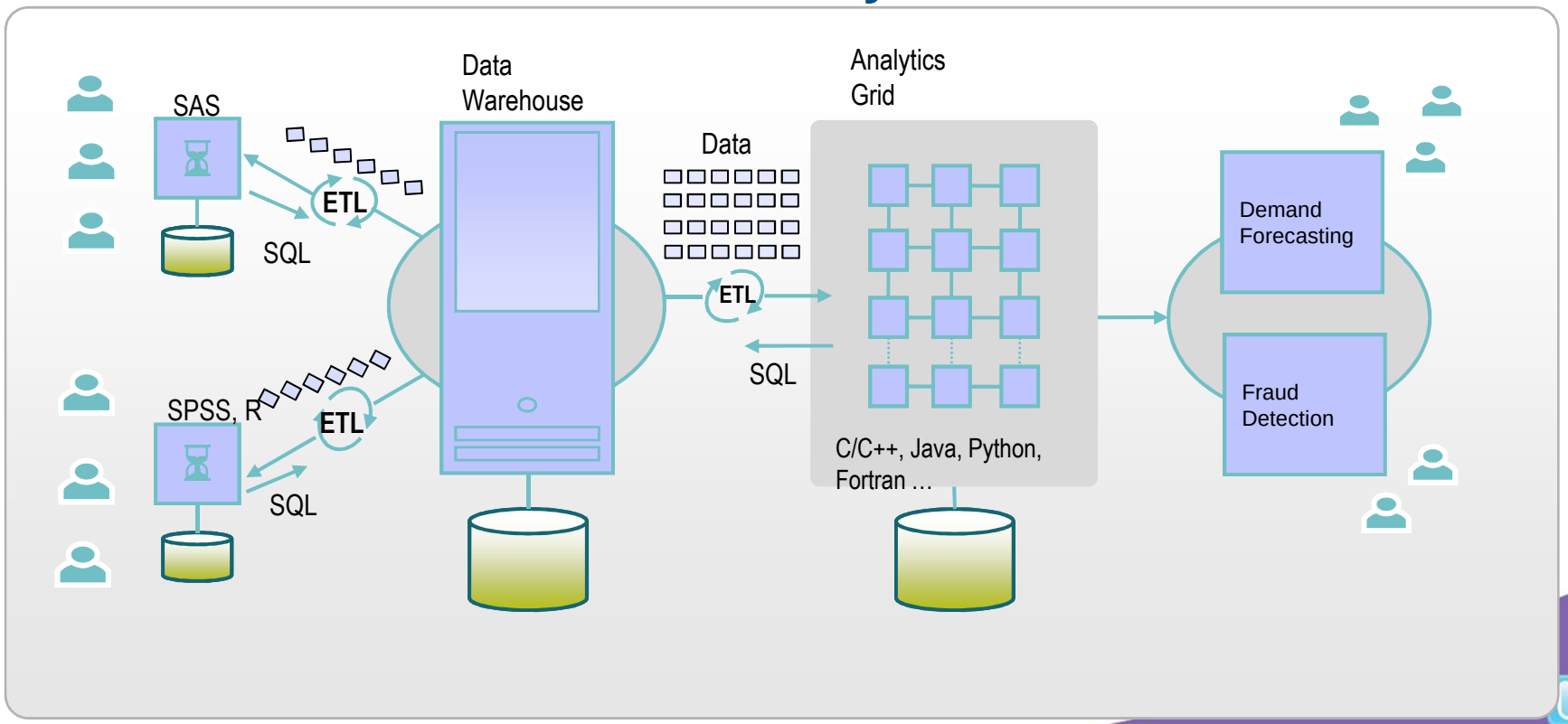
An effort at empowering users. Use Excel, Access and powerful tools. The DB server becomes a file Server.



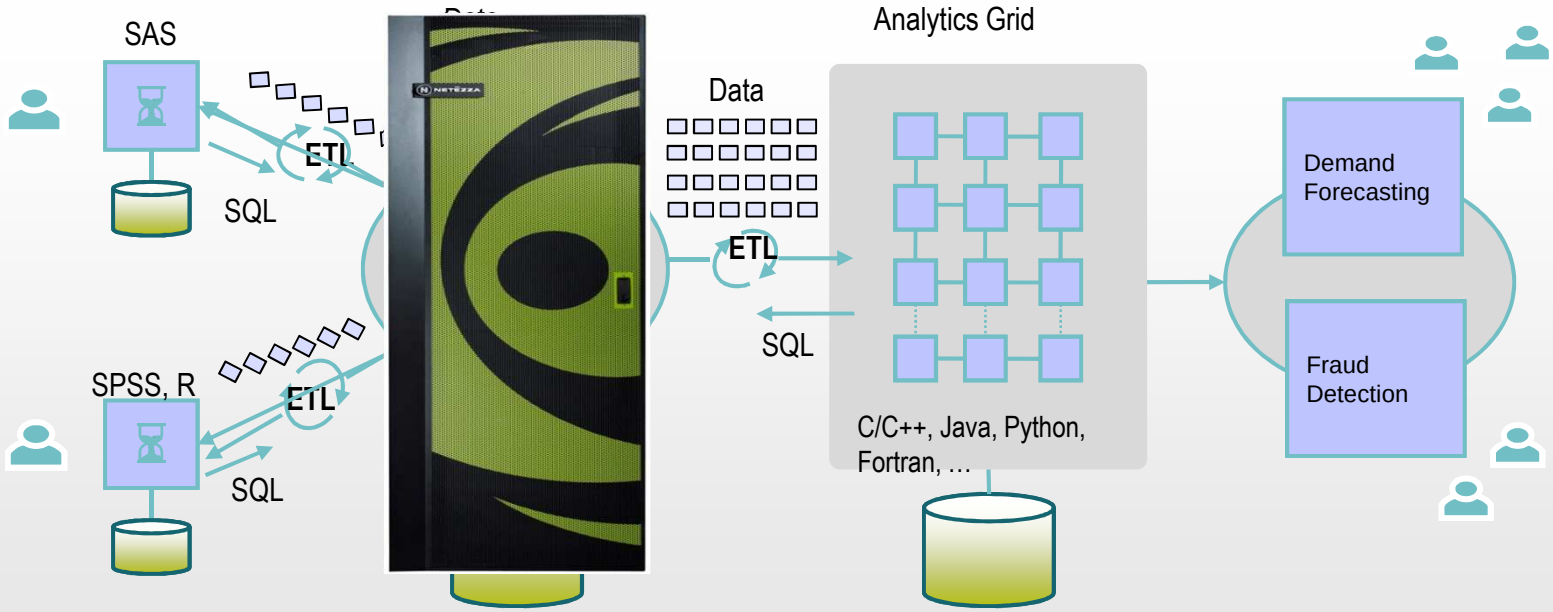
Can I get a better PC, like, now?

but there is gold in your code and formulas!

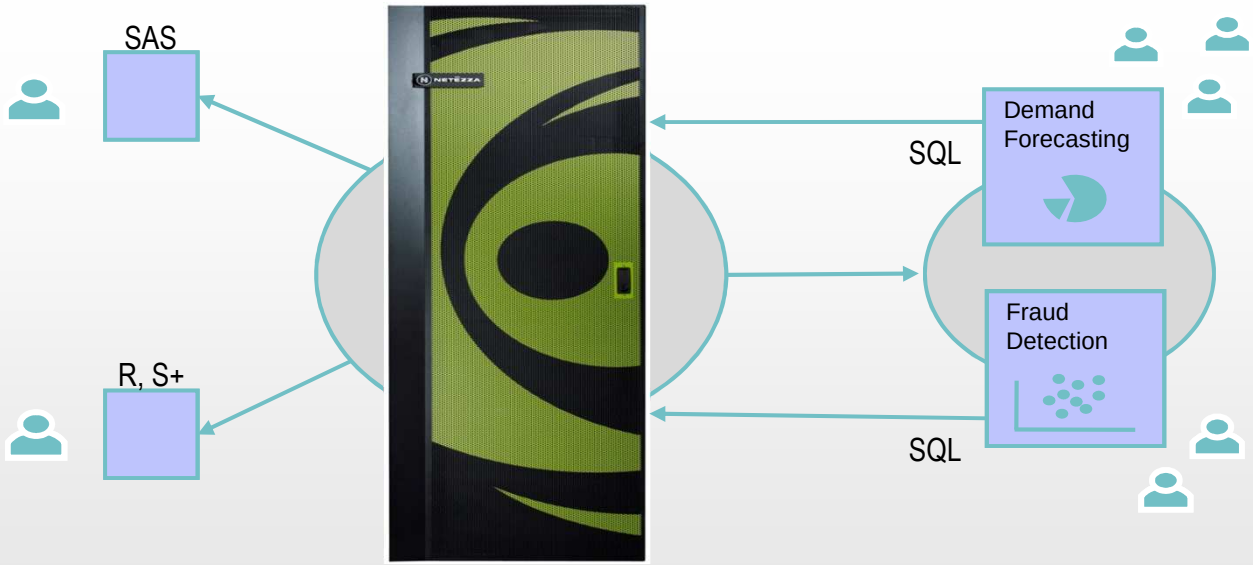
Where to next? Advanced Analytics – the traditional way



Advanced Analytics with Netezza



Advanced Analytics with Netezza



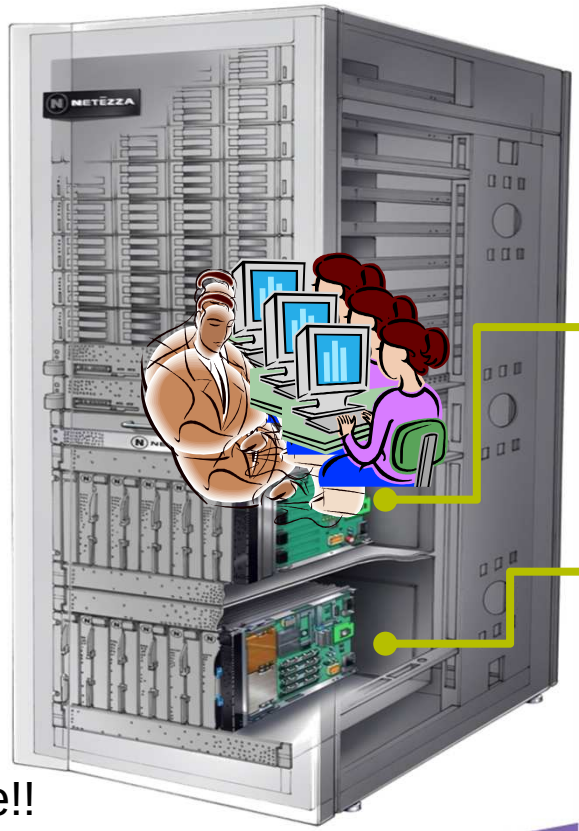
The IBM Netezza 1000 Architecture

Optimized Hardware+Software

Purpose-built for high performance analytics; requires no tuning

True MPP

All processors fully utilised for maximum speed and efficiency



Streaming Data

Hardware-based query acceleration for blistering fast results

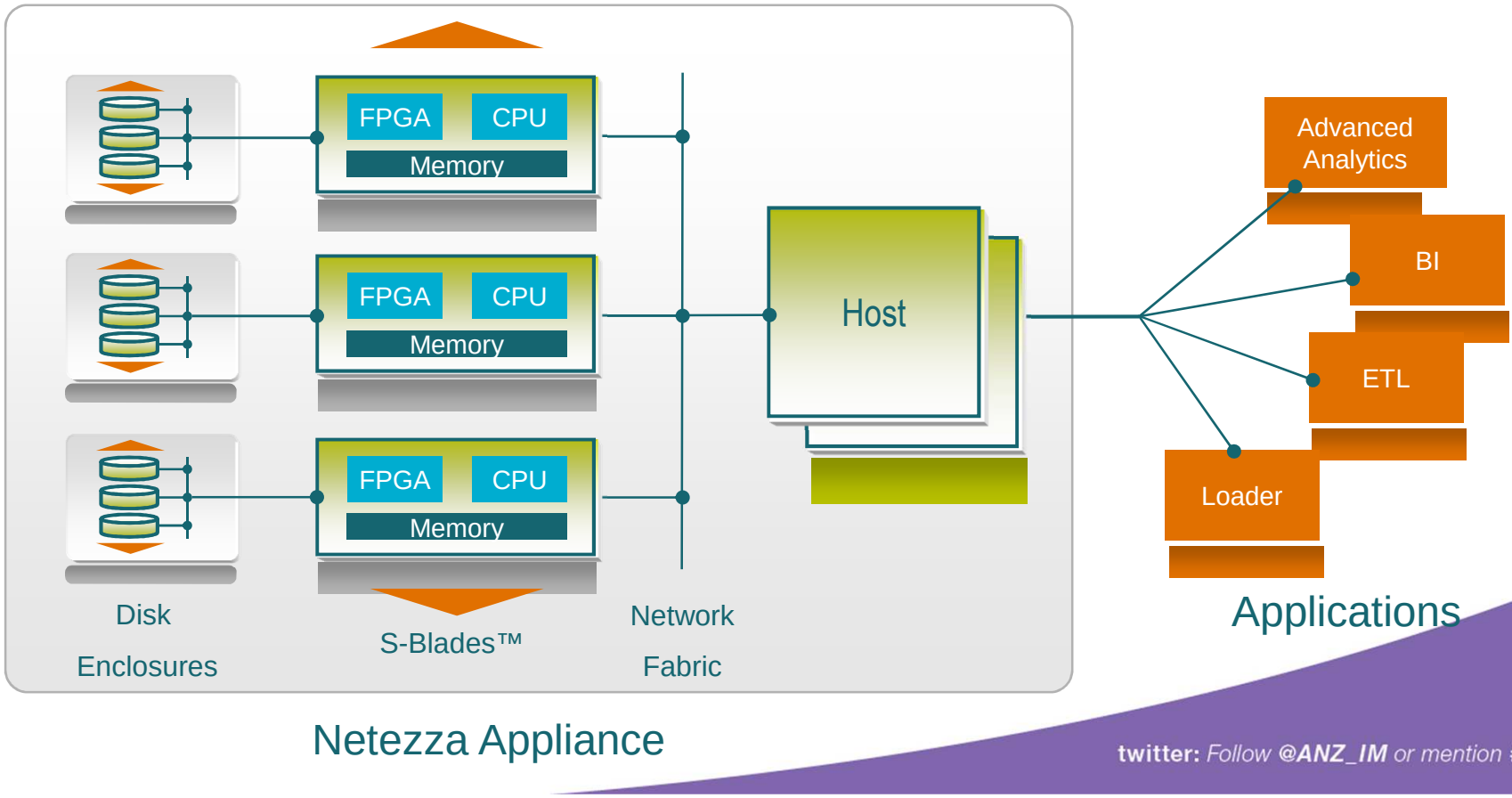
Deep Analytics

Complex analytics executed in-database for deeper insights

96 Cores working for you here!!

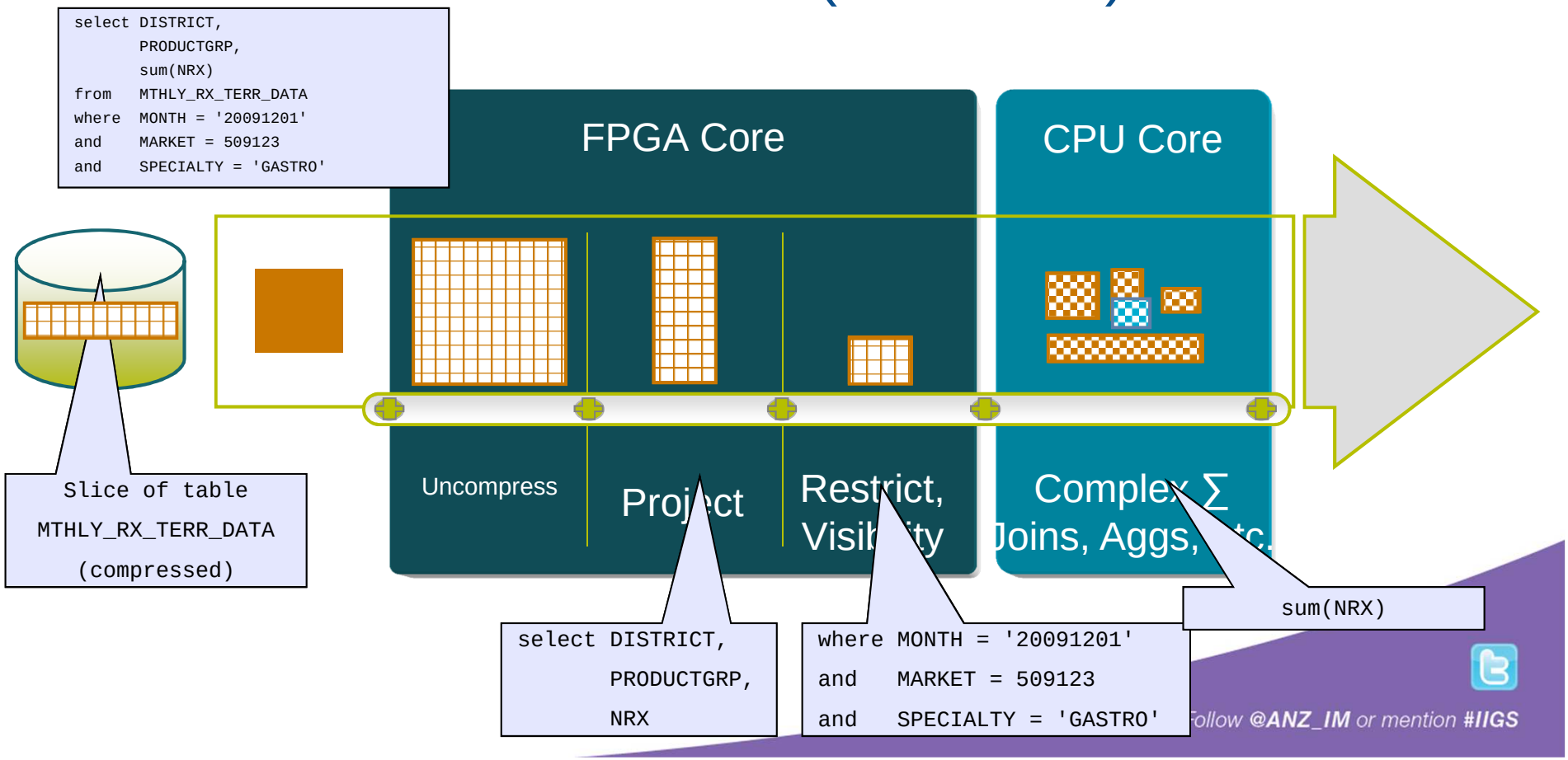
twitter: Follow @ANZ_IM or mention #IIGS

The Netezza AMPP™ Architecture

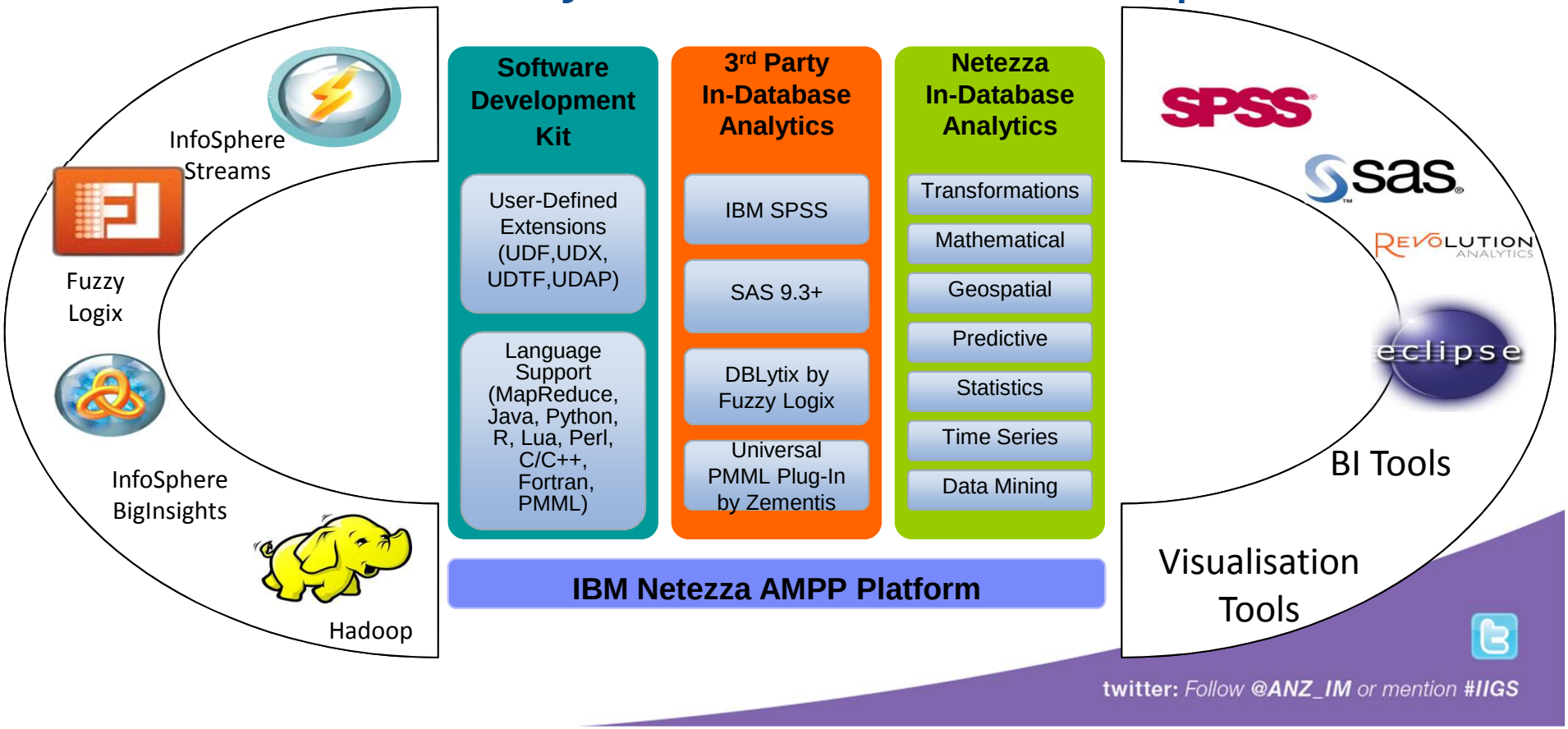


twitter: Follow @ANZ_IM or mention #IIGS

The Netezza Secret Sauce (some of it)



IBM Netezza Analytics: Features and Capabilities



Netezza Analytics implementations

UDF
User Defined Function

Customer	Date	Item
ABC	1-1-2001	1
DEF	1-1-2001	1
GHI	1-2-2001	1

→

UDF
2
4
6

UDA
User Defined Aggregate

Customer	Date	Item
ABC	1-1-2001	1
DEF	1-1-2001	1
GHI	1-2-2001	1

→

UDA
2

UDTF
User Defined Table Function

Customer	Date	Item
ABC	1-1-2001	1
DEF	1-1-2001	1
GHI	1-2-2001	1

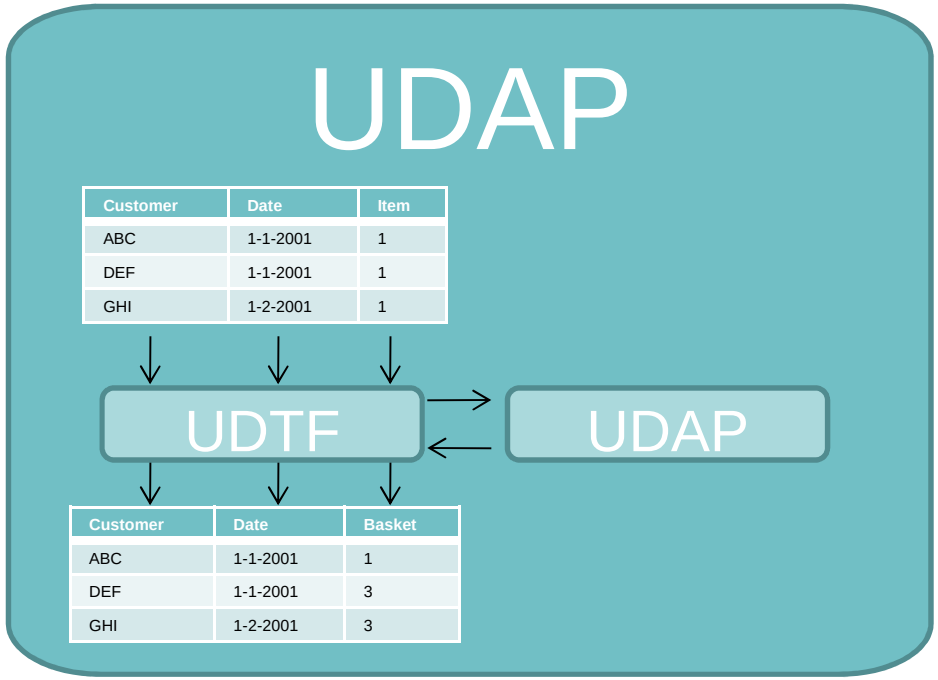
↓

Customer	Date	Basket
ABC	1-1-2001	1
DEF	1-1-2001	3
GHI	1-2-2001	3

User Defined Extensions API consists of C/C++ classes Called via SQL

twitter: Follow @ANZ_IM or mention #IIGS

Newest User Defined Extension
User-Defined Analytic Process



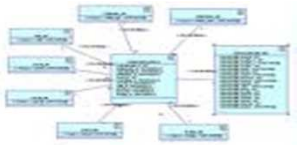
UDAP can be written in C/C++, Java, Python, Fortran, R, LUA, Perl

twitter: Follow @ANZ_IM or mention #IIGS

Typical Netezza/Hadoop Use Cases

Unstructured data → hadoop (Create context (classification, text mining)) → NETEZZA (Analyse)

Semi-structured data → hadoop (Parse, aggregate) → NETEZZA (Analyse, report)



Structured data → NETEZZA (Analyse, report) → hadoop (Active archival, Long running queries)

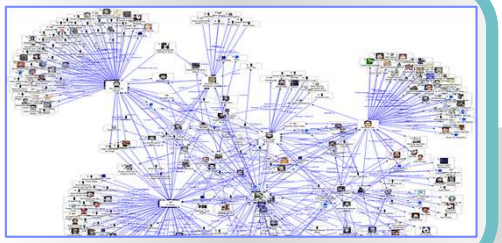
twitter: Follow @ANZ_IM or mention #IIGS

The Analytic Enterprise

BI Reporting and Ad-Hoc Analysis

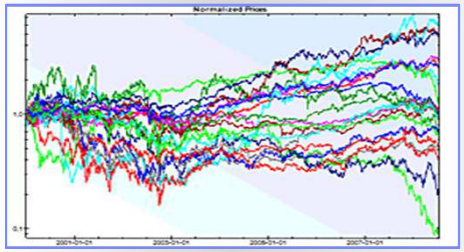
- What happened?
- When and where?
- How much?

Predictive Analytics



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

Optimisation



- What is the best choice?

twitter: *Follow* ***@ANZ_IM*** *or mention* ***#IIGS***

Company Confidential

Netezza at T-Mobile

- Originally a traditional RDBMS Data Warehouse – 40 Terabytes
- Managed by 3 FTEs
- Moved to Netezza
- Scaled from 40TB – 100TB – 500TB – 2 Petabytes today
- Same 3 FTEs
  - Principal Engineer
  - SQL Developer
  - Business conduit
- Used for Network Analytics
- 60 Applications, 1200 users, 400 data feeds
- Loads 17 billion records per day

http://www.youtube.com/watch?v=IBZYJpDT0vQ

T-Mobile: IBM **Netezza** Client Success Video - **YouTube**

www.**youtube**.com/watch?v=IBZYJpDT0vQ
14 Mar 2012 - 3 min - Uploaded by IBMCustomerReference
Network engineering success at scale with IBM **Netezza** Christine Twiford, **T- Mobile's** Manager of Network ...