Managing and Analyzing “XL” Data at Market Speed

Steve Hirsch – Chief Data Officer and Chief Architect
XLDB Conference, 10/6/10
NYSE Euronext Profile

- NYSE Euronext (NYX) is the world’s most diverse exchange group, offering a broad and growing array of financial products and services in Europe and the United States that include cash equities, futures, options, exchange-traded products, bonds, market data, and commercial technology solutions.

- With over 8,000 listed issues globally, NYSE Euronext's equities markets -- the New York Stock Exchange, Euronext, NYSE Arca, and NYSE Amex -- represent nearly 40% of the world's cash equities trading volume, the most liquidity of any global exchange group.

- NYSE Euronext also operates NYSE LIFFE, the leading European derivatives business and the world’s second largest derivatives business by value of trading. NYSE Euronext is part of the S&P 500 index and the only exchange operator in the S&P 100 index.

- NYSE Euronext has rich historical roots dating back to 1792 in the United States (NYSE) and to 1602 in Europe (Amsterdam Stock Exchange)
Market Capitalization and Value of Share Trading

In 2008 the United States Gross Domestic Product was $14.347 Trillion

Value of Share Trading $38.7 Trillion
($38.7 \times 10^{12})

Underlying Data Source: World Federation of Exchanges (www.world-exchanges.org)
Financial Markets and the Latency Race to Zero
Trading System Order “Ack” Times are now below 1ms

NYX fastest trading markets transact orders in less than 500 microseconds with the NYSE at 3 milliseconds. To put this in perspective, in 2005 the NYSE was measuring individual trade speeds in seconds.

Drivers – Technology, Algorithmic trading, Co-location, Smaller trade sizes
Data Facts

- NYSE Euronext (NYX) internally generates approximately 4TB of fresh data (3TB Business and 1 TB system logs) on a global basis during the course of a busy day.

- In the U.S., our options and equities markets combined have processed in excess of 15 billion “transactions” (Order, Cancel Order, Route Order, away market price updates) over the course of a single 6.5 hour trading day.

- In mid 2007, our peak message volumes doubled from one day to the next.

- May 7th, 2010 (the day after the “flash crash”) trading session set many new records in our U.S. markets with 684 million equity orders and 4.4 billion MM quotes.

- Data volumes compounding annually from 30-100% depending on market.
OPRA Industry Peak 1 Sec Messages

Max 1 Second on 2010-05-07: 1.9 M
Post Trade Data Platform – Back End

- Primary vendors for our data analytics environment are Netezza (IBM) and Greenplum (EMC)
- 2.5+ Petabytes of Netezza (Twinfin and 10000 series)
- 200 Terabytes of Greenplum (Sun Thumper Cluster and New EMC/GP Appliance)

- All of our business data lands in Netezza (6-12 month retention)
  - Loads of both near real time (<1 minute latency) and EOD batch
- All of our system performance data lands in Greenplum (14 day retention)
  - Loads of both near real time (<30 second latency) and EOD batch

- Appliances Federated by Composite Virtual Data Access Layer
- Flat File Farm 8PB EMC NS960 – growing to 7 years compressed data
  - Used in conjunction with our internally developed On Demand Database tool
On Demand Database – Reducing DM Costs by leveraging the strengths of new platforms – Load speed changed everything

- ODD Provides end users with a high performance facility to load offline compressed data from commoditized file servers to Netezza/GP in a highly transparent way

- ODD keeps track of schema changes over time to allow for reloads of older data

- Flat File Farm presents 7 years of historical data via Clustered file system ready to be reloaded by ODD

- We always maintain a % of our Netezza and GP platforms for use for reloads and ad-hoc analysis

- ODD / FFF is currently capable of pushing 8TB / hour into a single target database
A look inside the NYSE environment...today

NYSE Euronext ETL Grid
Data Processing Factory

End-of-Day
Intra-Day

Data Feeds

FTP

Large Storage Farm

Cluster File System

Input (raw data)

Output (Normalized/Enriched Data)

Fast Loader Services (ODD Services)

End-of-Day Win Client

On-demand request

Fast GigE Network

GRID Server Farm

ETL Workers

Linux

Grid Services

Grid Work Queue

Batch Scheduler

Jobs

Command-Line Interface

End-Users, Developers

Query Services, JDBC, ODBC, BI Tools

New York
Netezza DW Appliances

Chicago
Data Analysis

- **250+ Global Data Users** (Surveillance, Research, Operations, Finance, Capacity Planning)

- **Business Data**
  - Market Surveillance
    - Time-series analysis
    - Pattern matching
  - Trading Pricing Models – predict impact of price changes and product bundles
  - Trading Platform Load Balancing – allocate systems to Trade Engines based on volume
  - Trading Platform Capacity Planning – trend growth at instrument level and predict system requirements
  - Creation of Value Added Data Products

- **System Data**
  - Latency Analysis – determine source of latency in order flow
  - Capacity Planning - trend growth at system level

- **Tools** – in database analytics wherever possible, SAS, R, Business Objects, OBIEE