This Symposium brought to you by

www.ttcus.com

Linkedin/Group: Technology Training Corporation

@Techtrain

Technology Training Corporation

www.ttcus.com
BIG DATA ANALYTICS
UNIFIED DATA ARCHITECTURE™

Tasso Argyros
Senior Vice President, Product Strategy & Deployment,
Teradata Labs
Co-Founder, Aster Data
BIG DATA IS A MOVEMENT DEMANDING MORE ANALYTICS ON ALL DATA
From Transactions to Interactions

Transactions

Interactions
What Does it Mean for Government?

• Reduction in Fraud Waste & Abuse

• Further Leveraging Sensor & Machine data analysis

• Optimized Security & protection for citizens & networks
What Does it Mean for Government?

• More effective healthcare

• Better understanding of Citizen Sentiment

• Accountability & Transparency
New Opportunities with Technology

UNIFIED DATA ARCHITECTURE

Data Scientists  Quants  Customers / Partners  Front-Line Workers
Engineers  Business Analysts  Marketing  Operational Systems

LANGUAGES  MATH & STATS  DATA MINING  BUSINESS INTELLIGENCE  APPLICATIONS

AUDIO & VIDEO  IMAGES  TEXT  WEB & SOCIAL  MACHINE LOGS  CRM  SCM  ERP
UNIFIED DATA ARCHITECTURE
Teradata Aster Discovery Platform

DATA SOURCES
- Non-Relational Data
- Multi-Structured Data
- Structured Data
- OLTP DBMS’s

DISCOVERY
- Discovery Platform
  - Structured and multi-structured data
  - Doesn’t require extensive data modeling
  - Doesn’t balance the books
  - Data quality can be good enough
  - No stringent SLAs

DISCOVERY TOOLS
- SQL, MapReduce, Graph, Stats, ML
- Custom Development
- Analytic Applications
  - Fraud patterns
  - Customer behavior
  - Digital marketing optimization

USERS
- Data Scientist
- Business Analyst
- Business Users
Aster SQL-MapReduce
Extend SQL with Integrated MapReduce Engine

PATENTED FRAMEWORK FOR ADVANCED ANALYTICS THAT ARE HARD TO DEFINE IN SQL

- Couples SQL (relational) with MapReduce (SQL-MapReduce) providing a new framework for rich analytics on diverse data (non-relational and relational)

- User code is installed in the cluster, and then it’s invoked on database data from SQL; Execution is automatically parallelized across the cluster

- Includes library of pre-packaged Analytic Modules (50+ currently) to speed analytics development (e.g. time-series, complex pattern/path, affinity, graph, data transformation, text, statistical...)

ASTER MPP DATABASE

- User code is installed in the cluster, and then it’s invoked on database data from SQL; Execution is automatically parallelized across the cluster

ARCHITECTURE FOR DIVERSE, EMBEDDED ANALYTICS PROCESSING

- Supports custom analytics written in a variety of languages (e.g. C, C++, C#, Java, Python, Perl, R); Supports both batch and interactive queries

DELIVERS EASE OF USE AND DEVELOPMENT BY COMBINING SQL AND VISUAL TOOLS

- Makes MapReduce accessible from SQL/SQL-based tools (standard BI tools)

- Visual Aster Development Environment for iterative development with desktop testing
A New Approach

1 SQL Statement to Acquire, Prepare, Analyze & Visualize

Aster Discovery Platform
UNIFIED DATA ARCHITECTURE

Data Scientists  Quants  Customers / Partners  Front-Line Workers
Engineers  Business Analysts  Marketing  Operational Systems

DISCOVERY PLATFORM
INTEGRATED DATA WAREHOUSE

CAPTURE | STORE | REFINE

LANGUAGES  MATH & STATS  DATA MINING  BUSINESS INTELLIGENCE  APPLICATIONS

AUDIO & VIDEO  IMAGES  TEXT  WEB & SOCIAL  MACHINE LOGS  CRM  SCM  ERP
Big Data Management Requirements

• Land/source operational data
  > Only one extract from source system

• History or long term storage
  > Low cost storage

• Preprocess data
  > Sessionize data, remove XML tags

• Transformations
  > Structured and semi-structured data

• Simple Math and Batch Processing
UNIFIED DATA ARCHITECTURE

Data Scientists  Quants  Customers / Partners  Front-Line Workers
Engineers  Business Analysts  Marketing  Operational Systems

LANGUAGES  MATH & STATS  DATA MINING  BUSINESS INTELLIGENCE  APPLICATIONS

Big Data Analytics

Big Data Management

TERADATA ASTER
DISCOVERY PLATFORM

TERADATA
DATA WAREHOUSE

CAPTURE | STORE | REFINE

AUDI0 & VIDEO  IMAGES  TEXT  WEB & SOCIAL  MACHINE LOGS  CRM  SCM  ERP
Teradata Unified Data Architecture™ with Streaming

Data Scientists
Engineers

Business Analysts
Customers / Partners

Marketing
Executives

Front-Line Workers
Operational Systems

LANGUAGES | MATH & STATS | DATA MINING | BUSINESS INTELLIGENCE | APPLICATIONS

DISCOVERY PLATFORM

DATA WAREHOUSE

CAPTURE | STORE | REFINE

EVENT PROCESSING

AUDIO & VIDEO | IMAGES | TEXT | WEB & SOCIAL | MACHINE LOGS | CRM | SCM | ERP
BIG DATA EXAMPLES
Use Case #1

**Situation:** Needed a centralized, integrated, and complete data warehouse ecosystem to support Clinical Business Intelligence & Research Informatics.

**Challenge:** The Health Services Data Warehouse (HSDW) requires greater scalability in its primary data store to accommodate the very large data volumes and complex analytics necessary to support the program mission.

**Solution:** Migrating the core data warehouse and analytical mart layers to reside on a multi-rack Teradata Aster implementation. The project will leverage inherent parallelism, key MapReduce functions, and columnar storage as appropriate to support intensive data operations.

**Anticipated Impact:**
- Designed to scale out with commodity hardware
- Prepared to accept In-Database processing of complex statistics;
- Able to support tactical recurring health system reporting as well as research analytics

Air Force Medical Service consists of 60,000 active duty, reserve, civilian, and contract medical and support professionals. The Air Force operates 63 medical facilities in the U.S. and 12 at over-seas locations, with more than 1,700 Air Force medical personnel in 19 countries.

Analytics to support clinical business intelligence
Use Case #2: Network Security
Detecting network threats with MapReduce Analytics

Application Characteristics

- Millions of PCs, files and URLs translate into 100s of billions of potential paths
- Full network detail required for analysis
- *Near-real-time scoring of every network node* required for fast response

**Examples**

- Security event management
- Zero day attack identification
- Data loss prevention
- Data packet inspection

**What this Means for Analytics?**

- Recognize the pattern of activity that the threat represents
- Off-line models not sufficient to keep up with quickly morphing attack behaviors
- Requires a behaviorally-based, “belief propagation” system to automate analysis, threat identification, & rapid response

**Advanced Analytics**

- Constant analysis of all connections and traffic
- Interactive root cause analysis
- Rapid attack identification, validation, and response
Use Case #3: Medical Surveillance
Monitor and identify new and spreading medical threats

Challenge
• Need to identify spreading and emerging medical threats as quickly as possible to prevent spread and loss of life
• Data around an emerging threat can often first be found in social media data (e.g., Facebook, Twitter)
• Amount of data across all sources is tremendous and varied

With Teradata Aster
• Ability to store and analyze massive data volumes without complex data movement
• Ability to combine multiple data sources in one system increases power of analysis
• Flexibility of graph processing engine simplifies detection of complex patterns

Impact
• See relationships among disparate data more quickly
• Identify and communicate implicit connections among events (e.g., similar symptoms diagnosed at different medical facilities within short distance of each other)
Graph Analytics with Teradata Aster

- **Link analysis**: predicting connections (among people, symptoms, etc.) that are likely to be of interest by looking at known connections
- **Influence analysis**: identifying clusters and influencers (e.g., doctors, hospitals) in networks
Summary

• Big Data is about ALL data. Expand from Transactions to also include Interactions & Sensor data.

• Technologies enable Discovery on ALL data while using standard, SQL interfaces

• Significant Big Data use cases involve analyzing micro-interaction, sensor, text data to predict events, behavior
THANK YOU

• For more, visit www.asterdata.com

• Follow me on twitter: @targyros

• Follow our blogs:

  http://www.teradata.com/blogs/