Teradata Analytics Platform

Sri Raghavan
Senior Global Product Marketing Manager
Agenda

- Driving Business Value with Teradata Analytics
- Evolving from a Database to an Analytics Platform
- Drilling down into Technology
  - Analytic Engines
  - Tools
  - Languages
  - Data types, formats and sources
- Futures
Customer Experience
Finance Transformation
Product Innovation
Risk Mitigation
Asset Optimization
Operational Excellence
New closure event patterns detected the following day

Customer account closure and its impact on satisfaction requires constant model tweaking. Highly manual.

Multi-Channel data consolidation of events of interest.

• Sessionize
• nPath
• Dynamic Time Warping Analysis

Developed a constantly evolving toolkit to detect new closure patterns

Reduce account closure incidence

Multi-Channel Account Closure Detection: Global Bank

CHALLENGE

SOLUTION

ANALYTICS

OUTCOME
Account Closure Analysis Outcome

Account Closure Paths
Customer Experience
Finance Transformation
Product Innovation
Risk Mitigation
Asset Optimization
Operational Excellence
New channel activities are constantly measured throughout the journey.

Customers interact through many channels but the impact of each channel in a customer's journey is not clearly known.

**Multi-Channel data consolidation of events of interest.**

- Sessionize
- nPath
- Attribution
- Customer Satisfaction Index
- Text Analytics

Delivered a repeatable set of experiences that ensures that customers attain their objectives at the least cost in the shortest amount of time.
Advanced Customer Experience Analysis

SQL Engine + Graph Engine + Machine Learning Engine = Best in Class CX Analytics

Statistical Analysis

Multi Channel Behavioral Path Analysis

Influencer Analysis

Sentiment Analysis
Product Innovation (IoT): Major Manufacturer

Sensor data continuously ingested and analyzed for pattern changes and performance triggers.

**CHALLENGE**

Need to understand equipment usage and behavior to prevent down time or unplanned outages across many support groups.

**SOLUTION**

Multi equipment sensor and performance data.

- Log Parsers
- Generalized Linear Model
- Weibull analysis
- Naive Bayes

**ANALYTICS**

Isolated key indicators of performance bottlenecks that helped remove soon to be defective equipment from circulation for proactive maintenance.

**OUTCOME**

Detect parts failure.
Product Innovation (IoT): Event paths leading to shutdown

Shutdown event and sensor paths

versus

Stable running event and sensor paths
Evolution from a Database to an Analytics Platform
## Teradata Everywhere

<table>
<thead>
<tr>
<th>FROM</th>
<th>TO</th>
</tr>
</thead>
<tbody>
<tr>
<td>Structured data warehouse</td>
<td>Analyze <strong>Anything</strong></td>
</tr>
<tr>
<td>On-premises, proprietary hardware</td>
<td>Deploy <strong>Anywhere</strong></td>
</tr>
<tr>
<td>Limited purchasing options</td>
<td>Buy <strong>Any Way</strong></td>
</tr>
<tr>
<td>Inflexible, lock-in</td>
<td>Move <strong>Anytime</strong></td>
</tr>
</tbody>
</table>
Ben, the data scientist, wants the freedom to choose the best tools to create business outcomes through data.

Carole, the CIO, wants to de-risk IT investments while operating at scale within SLAs and leveraging existing data and analytics investments to augment/grow current work.
Teradata Analytics Platform — Bridging the Divide

**Data Scientist**
- Quickly get to actual analysis
- Access preferred tools and languages
- Integrate multi-format data
- Make analytics easy to repeat

**CIO**
- Accelerate time to deliver environments
- Reduce operational costs
- Support end user demands
- Provide leading-edge security
- Leveraging existing data and analytics investment
Teradata Analytics Platform

Analytic Tools

SQL

Analytic Languages

Cross-Engine Analytic Orchestration

SQL Engine

Machine Learning Engine

Graph Engine

High Speed Fabric

Persistent Storage

Data Storage

QueryGrid (Data Store Access)

QueryGrid (Engine Access)
Our Vision: Teradata Analytics Platform

Use your favorite analytic tools

Leverage the most suitable languages

Access a larger variety and volume of data, including new sources

Use the best analytic functions & leading analytic engines
The Best Analytic Functions and Engines
Preferred Tools and Languages
Support for Multiple Data Types

Teradata Analytics Platform
What is an Analytic Engine?

• **Definition**
  – A comprehensive framework that includes all the software components that are well integrated to deliver advanced analytics functionality that can be implemented by a well defined set of user personas

• **Components of an Analytic Engine**
  – Advanced Analytics functions (e.g., Machine Learning, Graph, Time Series)
  – Access points to data storage that can ingest multiple data types
  – Integration into visualization and analytic workflow tools
  – Built in management and monitoring tools
  – Highly scalable and performant environment with established thresholds

• **Advantages of an Analytic Engine**
  – Self-contained (containerized) analytic compute environment that can be separated from data storage
  – Analytic engines can be tailored for access and use by specific personas (e.g., DS, Business Analyst)
Teradata Analytics Platform

Analytic Tools
- Studio
- AppCenter
- Jupyter
- R Studio

Analytic Languages
- SQL
- Python
- R

Cross-Engine Analytic Orchestration
- SQL Engine
- Machine Learning Engine
- Graph Engine

High Speed Fabric

Data Storage
- Persistent Storage

Kubernetes Managed Nodes. Requires additional hardware.
Implemented Through Containerization

• Integrating Docker containers in Teradata Analytics Platform

• Containers enable
  – Add new analytic engines faster
  – Minimized downtime for customers
  – Automatic scale out of engines
  – Improved SLA through compartmenting end-users workload

• Type of Containers: Analytics, UDA, languages, open source, commercial products and custom engines

* Anticipated future capabilities
What is the Teradata SQL Engine?

Teradata Analytics
Over 1000 in-dbs functions
- Teradata Warehouse Miner
- SAS
- Fuzzy Logix
- Matrix functions
- Time series data types and aggregations

Machine Learning Functions
Popular Machine Learning functions
- Path & Pattern
- Sessionize
- Attribution
- Scoring functions
  - Single Decision Tree
  - Random Forest
  - Naïve Bayes Text Classifier
  - Sparse SVM
  - GLM

Teradata SQL Engine
Analytic (embedded) Functions

Persistent Storage

Industry leading database
- Effortless scalability
- Mission-critical availability
- Query performance
- User management
- Workload management
What are the Machine Learning & Graph Engines?

• The Machine Learning Engines consists of a wide range of analytic capabilities that span analytic genres such as Artificial Intelligence, Statistics, Text & Sentiment determination in addition to Data Preparation and unstructured data Parser functions.

• Examples of Machine Learning Engine analytics include Outcome Categorization, Sentiment Strength Assessment, Customer Path Behavior, Fraud Detection, and more.

• The Graph Engine enables the analysis of relationships and how they influence outcomes in networks of people, products, or processes or really any entities that are connected in a network.

• Examples of Graph Engine analytics include Network Threat Detection, Market Basket Rendering, Social Media Influencer Behavior, and more.
# Machine Learning Engine Functions

## Statistics
- AdaBoost
- Approximate Distinct Count
- Approximate Percentile
- CMAVG
- ConfusionMatrix
- ConfusionMatrixPlot
- Correlation
- CoxPH
- CoxPredict
- CoxSurvFit
- Cross Validation
- Distribution Matching
- EMAVG
- Enhanced Histogram
- Fmeasure
- GLM
- GLMPredict
- Hidden Markov Model
- Histogram
- KNN
- LARS Functions
- LinReg
- LRTEST
- Non-linear Kernel SVM
- Percentile
- Principal Component Analysis
- Random Sample
- ROC Curve
- Sample
- Shapley Value
- SMAVG

## Statistics (cont.)
- Support Vector Machines
- VectorDistance
- VWAP
- WMAVG

## Data Transformation
- Antiselect
- Apache Log Parser
- Fast Fourier Transform
- FellegiSunterTrainer
- FellegiSunterPredict
- IdentityMatch
- IpGeo
- Inverse Fast Fourier Transform
- JSONParser
- FellegiSunterPredict
- IdentityMatch
- MurmurHash
- Number as Categories
- OutlierFilter
- Pack
- Pivot
- PSTParserAFS
- Scale Functions
- StringSimilarity
- Unpack
- Unpivot
- URI Pack
- URI Unpack
- XMLParser
- XMLRelation

## Path, Pattern and Time Series
- Arima
- ArimaPredictor
- Attribution
- Burst
- ChangePointDetection
- Causality Detection
- DTW
- DWT
- DWT2D
- FrequentPaths
- IDWT
- IDWT2D
- Interpolator
- Path_Analyzer
- Path_Generator
- Path_Start
- Path_Summarizer
- SAX
- SAX2
- SeriesSplitter
- Sessionization
- Shapelets
- TimeSeriesOrders
- Unsupervised Shapelets
- VARMAX

## Association
- Basket_Generator
- CFilter
- FP-Growth
- KNN Recommender
- WSR Recommended

## Cluster
- Canopy
- Categorical Clustering
- Gaussian Mixture Model
- KMeans
- KMeansPlot
- Minhash

## Decision Tree
- XGBoost
- Forest_Drive
- Forest_Predict
- Forest_Analyze
- Single_Tree_Drive
- Single_Tree_Predict

## Naïve Bayes
- naiveBayesMap
- naiveBayesReduce
- naiveBayesPredict

## Visualization
- CFilterViz
- NpathViz

## System Functions
- nc_skew
- nc_relationstats

## Text
- Chinese Text Segmentation
- LDA Functions
- Levenstein Distance
- Named Entity Recognition (CRF Model)
- Named Entity Recognition (Max Entropy Model)
- nGram
- PoStagger
- Sentinizer
- Sentiment Extraction Functions
- Text Classifier
- Text_Parser
- TextChunker
- TextTagging
- TF-IDF

## Location Analysis
- LoadGeometry
- PointInPolygon
- GeometryOverlay

## Deep Learning
- Neural Networks

---

**See the Analytics Foundation Guide for a Complete Review of Analytic Operators**

The Best Analytic Functions and Engines

Preferred Tools and Languages

Support for Multiple Data Types

*Future
Analytic Tools Support

- Teradata Studio or any SQL editor
  - SQL

- AppCenter
  - SQL & Python

- Jupyter Notebooks
  - SQL and R client package
  - AppCenter

- Rstudio
  - Call R functions via R library
Analytic Languages Integration

- **SQL**
  - Native and AppCenter

- **R**
  - R client package, embedded process and AppCenter

- **Python**
  - Embedded process and AppCenter
The Best Analytic Functions and Engines

Preferred Tools and Languages

Support for Multiple Data Types

*Future
Teradata Support for Multiple Data Types, Formats and Sources

Data Formats
- JSON, BSON, AVRO
- CSV, XML, PDF, Voice, Video and Images

Data Types
- Geospatial, Temporal and Time Series
Business Outcomes through Time Series Analysis

- **Optimized delivery and pricing plans** by geography, product type, and other variables
- **Smarter maintenance protocols** to catch equipment degradations well before a full blown crisis
- Monitoring of prescriptions and pharmacological disbursements to **prevent abuse and theft**
- **Optimizes usage patterns** by providing efficiency recommendations and **prevention of unauthorized or illegal access** to power systems
Teradata Database – Time Series Capabilities
Teradata Database 16.20

Agile Analysis enabled by Time-Aware Functions

• Time period aware aggregations
• Work with ANY time component data
• Impute missing values
  • Ignore, removed, update with constant

High-Performance enabled by Primary Time Index (PTI)

• Supports time sensitive decisions
• Fast access through:
  • Hash distribute by time bucket
  • AMP-local processing
  • Sequenced data
Data Sources and Storage

- **Analytic Tools**: AppCenter, jupyter, R Studio
- **Analytic Languages**: SQL, Python, R
- **SQL Engine**: Cross-Engine Analytic Orchestration
- **High Speed Fabric**: SQL Engine, Machine Learning Engine, Graph Engine
- **Persistent Storage**: S3, Others: Cassandra, Kafka, MySQL, MongoDB, SQL Server, PostgreSQL, InfoSphere BigInsights, Hadoop
The Reality

- More analytic tools & techniques
- Proliferation of departmentalized analytic
- Dynamically changing analytic workloads
- Wide range of deployment choices
- Large upfront investment

Constant Change is the norm
It’s extremely difficult to understand what products to buy that create value for me, especially around the data movement products because there are so many to choose and understand.

Global Chip Manufacturer
Teradata’s Comprehensive Software Portfolio that Enables the Orchestration of an Analytic Ecosystem

- Key enabler of the Teradata Everywhere strategy
- Core software required to Ingest, Access, Deploy, and Manage a Unified Data Architecture
- One simple, subscription-based software license bundle
Key Takeaways of Teradata IntelliSphere

• IntelliSphere™ offers capabilities for complete solutions to solve challenges through apps and use cases.

• The Teradata IntelliSphere portfolio is a single purchase that covers all deployments within your organization.

• Built on a modern stack that provides evolving capabilities through continuous innovation.
We know that no single analytic system can meet all customers’ needs.
IntelliSphere Unlocks Several Key Capabilities

All included in one simple subscription-based software license

**Ingest**
Ingest and distribute high volume data streams, with ready-to-run elastic architecture and quick access for business-critical analysis

**Manage**
Management software allows for ad-hoc data movement, as well as ongoing monitoring and control via an operational interface

**Deploy**
Deploy applications and analytic models for easy user access and enterprise collaboration

**Access**
Gain easy access to data stored anywhere, even in a hybrid cloud or heterogeneous technology environment
Modern Stack
Teradata AppCenter

Ecosystem Services
- Dictionary
- Identity
- Notification
- Audit
- Catalog
- Logging

UI - Covalent
Visualization Engine

Container Apps
- Open Data Science
- Teradata Products/Solutions
- Bring Your Own Container

Data Access Layer – Teradata, Aster, Presto, Hive

SDK
- Python
- Java
- R (future)

Platform
- Kubernetes
- Docker Containers

Cloud and On-Prem Deployment
- AWS
- IntellïCloud
- VMWare
- IntellïBase
Built On Key Use Cases

- Data Lab
- Disaster Recovery
- Dual Active/High Availability
- Dual Systems/Workload Balancing
- Self-Service Data Ingest for Analytics
- Borderless Analytic Apps
Teradata Analytics Platform Future Vision

Analytic Tools
- Studio
- AppCenter
- Jupyter
- R Studio
- SAS
- Datiiku
- KNIME

Analytic Languages
- SQL
- R
- Python
- Scala
- Go
- JS

Analytic Engines
- SQL Engine
- Machine Learning Engine
- Graph Engine
- Spark Engine*
- TensorFlow Engine*
- Custom Engine*

Cross-Engine Analytic Orchestration

Data Storage
- Teradata Data Store
- S3 Data Storage*

High Speed Fabric

QueryGrid (Engine Access)

* Anticipated future capabilities

QueryGrid (Data Store Access)