

**TERADATA**

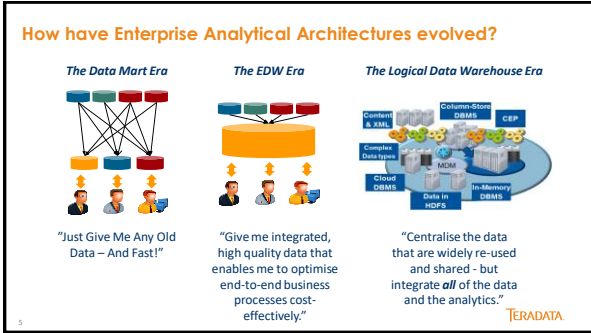
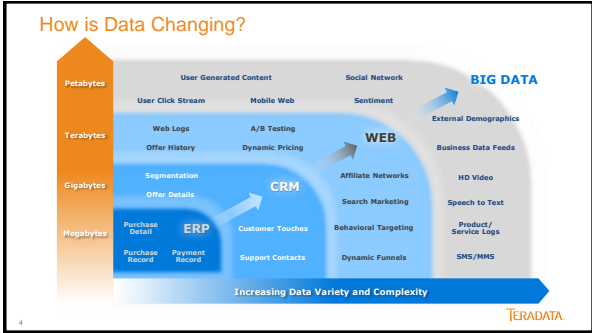
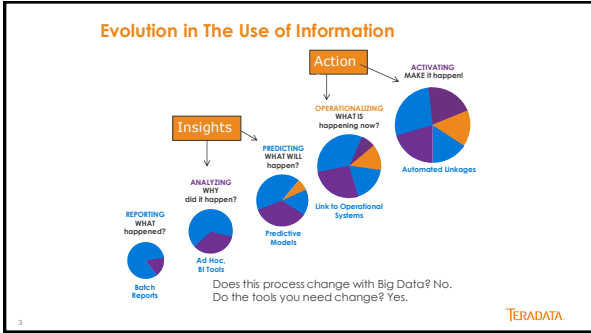
**What is the Analytical Ecosystem?  
And why is it making my life more complicated?**

Brian Grant

**Agenda**

- The Changing Environment
- What are the new Technologies?
- How does it all fit together?
- Conclusions

TERADATA



**Agenda**

- The Changing Environment
- What are the new Technologies?
- How does it all fit together?
- Conclusions

TERADATA

### What is a Database?

- High performance data access
  - Random and sequential
  - An access language + API
- High availability
  - Recovery following errors
- Data model for business applications
  - Isolates schema from application
  - Relationships enforced between data attributes
  - Logical and physical data designs
- ACID properties
- Shared resource, concurrency
- Data controlled by database
  - Data types
  - Secure access controls

**Atomicity**  
apply all changes or none

**Consistency**  
rollback on errors

**Isolation**  
one update at a time

**Durability**  
transactions survive crashes

TERADATA

### So what is Hadoop?

- Open source Data Store Optimized to handle
  - Massive amounts of data
  - Variety of data (Structured/Unstructured/Semi-structured)
  - Contributors: Yahoo!, IBM, Google
- Growing list of supporting tools
- Great Performance for certain Use Cases
- Reliability through replication

TERADATA

### Hadoop vs RDBMS: Complementary Capabilities

**Observations...**

- Technology changing
- Different perspectives

**Focus really needs to be...**

- Capabilities
- Client Skill Sets
- Requirements

TERADATA

### Usage Patterns

#### Data Warehouse extended by Data Lake

Usage Pattern	Example
Best at...	Secure data, Integrated data, Quality data, Scrubbed data
Better at...	Enriched data, Historical data
Good at...	Descriptive analytics, Predictive analytics
Better at...	Machine learning, Document mgmt, Discovery analysis, Streaming data
Best at...	Unstructured data, Raw data, Schema evolution, Natural Language

TERADATA

### How am I supposed to remember all of that?

Remember "Schema on Read"

```

    graph LR
        subgraph Schema-on-Read
            C1[Collect Data] --> W1[Write Data]
            W1 --> AS1[Apply Schema]
            W1 --> AS2[Apply Schema]
            W1 --> AS3[Apply Schema]
            AS1 --> A1[Analyze]
            AS2 --> A2[Analyze]
            AS3 --> A3[Analyze]
        end
        subgraph Schema-on-Write
            C2[Collect Data] --> AS4[Apply Schema]
            AS4 --> W2[Write Data]
            W2 --> A4[Analyze]
            W2 --> A5[Analyze]
            W2 --> A6[Analyze]
        end
    
```

TERADATA

### 1950: Alan Turing defines the imitation game

"Can machines think?"

"And how would we know?"

TERADATA

### New Analytical Tools

Many people associate "Machine Learning" with "Deep Learning"...

Deep Learning is one "supervised" approach to Machine Learning.

13 TERADATA

### Agenda

The Changing Environment  
 What are the new Technologies?  
**How does it all fit together?**  
 Conclusions

14 TERADATA

### Blended Architecture

Blended architecture is using the right tools and technologies for the right purpose

- IDW **AND** Data Lake
- On Premise **AND** Cloud
- Open Source **AND** Proprietary
- File System **AND** Database
- Agile **AND** Waterfall
- BI **AND** Advanced Analytics
- ETL Tools **AND** Roll Your Own

*The blended architecture era: from "OR" to "AND"*

15 TERADATA

### The Challenge is selecting the right technology

Enough time...  
 Enough money...  
 Work magic with technology...

What is the time to value?  
 What is the cost of re-engineering?  
 What is the optimal solution?

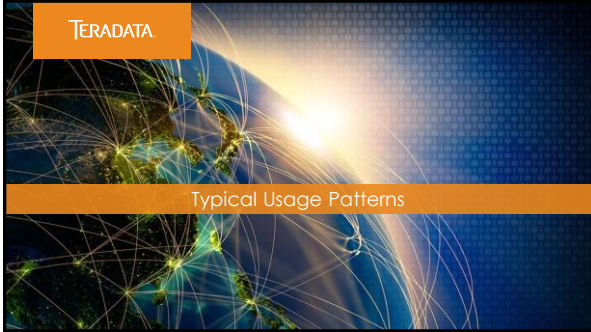
16 TERADATA

17 TERADATA

### Migrate From Projects To Frameworks

Shift focus from a single, tactical exercise to a comprehensive framework

18 TERADATA



### Hadoop for Active Data Archive

Active data archive for better data management

**Situation**  
High performance storage is expensive. A large integrated pharmacy HC provider deals with a variety of data with different business value. All data cannot be store on the same system. Ever expanding data is only adding to this challenge.

**Problem**  
Long terms storage data cannot be queried and it takes a long time for retrieval. No analysis can be performed on the archived data. Losing out on business value from this valuable data.

**Solution**  
Used Hadoop to store all the data coming in from weblogs, medical data, JSON files. Hadoop also serves as an enrichment layer to enhance data for high-end analytics consumption. The complete solution provides easy movement of data from Hadoop, Aster and Teradata.

**Impact**

- Reduced storage costs for data variety
- Perform adhoc analytics on the multiple versions of data
- Retrieve data in minutes (vs. days with tape archives)
- Reduced load and improved performance of DW/Databases

20 **TERADATA**

### Analyze Customer Web Interactions

Capture, Refine, Store ClickStream Data

**Situation**  
Customers interact with public websites of large PC vendor for various purposes --- resulting in huge volumes of raw data. Because of its nature, the data structure and format is not always consistent and because of the volumes, processing the amount of data is difficult.

**Problem**  
Inconsistencies like file errors, corrupted file compressions in the raw omniture data makes the capturing and analysis process error prone. The volume, velocity (Diverse, 1M files) adds to the complexity.

**Solution**  
Teradata Big Analytics solution to provide a landing and staging area for in-coming data at high velocity. Hadoop nodes to curate the data, check for data consistency, and prepare the data for consumption by higher end analytic platforms.

**Impact**

- Reduced data inconsistencies and improved performance
- Capture and curate ALL the data and prepare for analysis
- Perform ad hoc analytics on multi-level interactions
- Improves the marketing campaigns and the customer support process

21 **TERADATA**

### Document Classification

Enabling Real-time Processing of Clinical Records

**Situation**  
Current Mayo Clinic architecture has diverse architectures and varied applications to manage clinical documents for storing, indexing, viewing resulting in high latencies and inefficiencies

**Problem**  
Mayo Clinic's NLP technology has been used to implement various research and clinical use cases which leverage unstructured documents, but these have been constrained in deployment due to RDBMS technology limitations.

**Solution**  
Teradata Appliance for Hadoop platform to design applications to ingest ALL documents types (radiology, operation, ECG/EKG, notes), make data available for free text search while making them simultaneously available for batch processing.

**Impact**

- Enabling near real-time processing of messages and documents
- Real-time querying by end users via Elastic search ( in seconds)
- Combine real-time and batch processing for better results

22 **TERADATA**

### Telematics

Geospatial Analytics for Better Risk Management

**Situation**  
A large diversified insurance provider needed to accurately calculate risk scores and adjust risk premiums for its enterprise fleets based on vehicle data, driver behavior, GPS data, weather data, traffic and DW data. Current custom developed applications limits the effectiveness of these scores.

**Problem**  
Lacks infrastructure and system to handle the huge volumes of real time data. No ad-hoc reporting systems to combine, enrich and analyze the data. Limited storage capacity limits the amount of data that can be captured, refined and stored.

**Solution**  
Used Big Analytics to design a platform to streamline the ingestion process for telematics data from multiple sources, data types, structure, and frequency and combine with other data sources to perform meaningful analytics.

**Impact**

- Quickly analyze data for informed decisions and ad hoc reporting
- Streamlined process to calculate vehicle and fleet scores
- Cost effectively quantify, adjust and manage risk premiums

23 **TERADATA**

### Cyber Analytics

Network Analysis for Advanced security threats

**Situation**  
Current network traffic solutions are not real-time. They either employ a deep packet inspection later or try to analyze one packet at a time in the hope of catching bad apples.

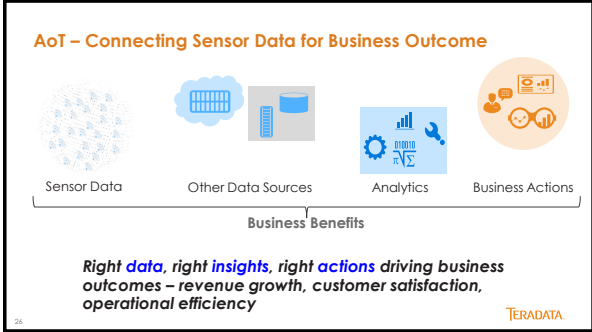
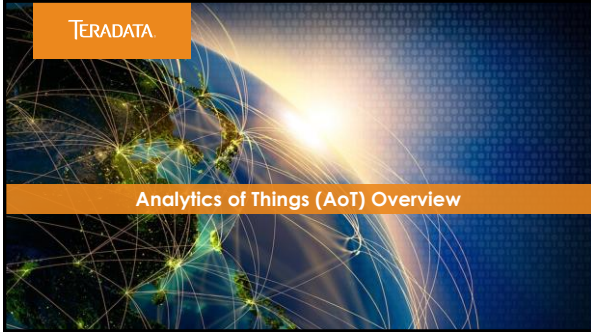
**Problem**  
Ineffective detection techniques with only 0.001% of the corporate traffic which might be infected. Today's API ( advanced persistent threat type viruses ) make it even more difficult to get detected with the current processing architectures

**Solution**  
Teradata Appliance for Hadoop is critical part of the UDA solution architecture that can process enormous volumes of data in near real-time and the ability to point accurately malicious code, predictively preventing malicious attacks

**Impact**

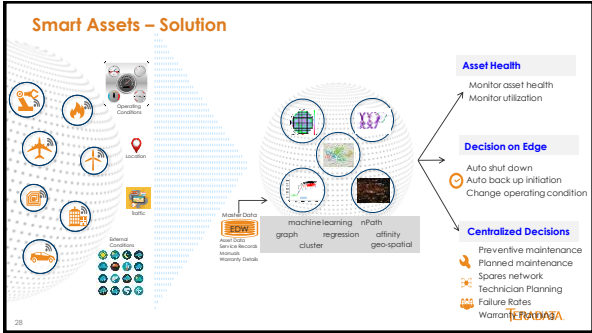
- Improve response times drastically to network traffic events
- Improve effectiveness of controls, fraud detection, and DUP efforts
- Proactively & automatically respond to malicious events or risks

24 **TERADATA**



### Our Most Frequently used Capabilities in AoT

Industrial AoT	Consumer AoT	Smart City
<ul style="list-style-type: none"> <li>Smart Assets</li> <li>Service 360</li> <li>Connected Factory</li> <li>Connected Supply Chain</li> </ul>	<ul style="list-style-type: none"> <li>Health &amp; Fitness</li> <li>Smart Home</li> <li>Connected Car</li> <li>Usage based Insurance</li> </ul>	<ul style="list-style-type: none"> <li>Smart Infrastructure</li> <li>Smart Utilities</li> <li>Smart Citizen</li> <li>Smart Mobility</li> </ul>
<b>IoT Data</b> <ul style="list-style-type: none"> <li>Sensor Data Qualification (AoT accelerators)</li> </ul>		



### Case Study: Predictive model for engine failures on trains

- Business Objective**
  - Avoid unplanned train downtime through prediction model
  - Repair trains before failure occurs, secure uninterrupted process
- Data Challenge**
  - Sensor data from engines and data from maintenance management system
  - Understand patterns to failure (pattern analysis)
  - Understand correlation of errors / sensor readings around a failure
  - Based on this insight, build a predictive model
  - Example: daily pattern of temperature readings mid – low – mid often occurs 3 days prior to an engine problem
- Solution**
  - Data structure to support ingestion, transformation, storage of data for analytics
    - Leverage Teradata analytics capability for predictive modeling
    - Scoring of predictive model to predict failures

**Opportunity to Impact**

- Pre-Dispatch required spare parts in time
- Avoid unplanned downtime, penalty, process interruption
- Save time on failure analysis

### Where might Smart Assets Help You?

- Are you responsible for the uptime of an asset used by end customers?
- Does a downtime affect your production, revenue or customer service?
- Will you benefit from predicting a failure of a remote asset?
- Can you improve your product quality by understanding product behavior in the field?





**Agenda**

- The Changing Environment
- What are the new Technologies?
- How does It all fit together?
- Conclusions

TERADATA

**Blended Architectures are no longer an option; They are a requirement**

- Data Warehouse + Data Lake
- On Premise + Cloud
- RDBMS + HDFS
- Commercial + Open Source

TERADATA



**Questions?**

TERADATA