Teradata Active EDW 6650 & 6680

Jim Dietz
Enterprise Product Marketing
4/8/11

Platform Introduction & Overview
Agenda

- New Teradata Active EDW platforms
- Teradata Active EDW 6680
- Teradata Active EDW 6650
- Teradata SSD technology & performance
- Active EDW platform solutions
## Teradata Purpose-Built Platform Family

**NEW Platform Models Released @ PARTNERS October, 2010**

<table>
<thead>
<tr>
<th>Data Mart Appliance</th>
<th>Extreme Data Appliance</th>
<th>Data Warehouse Appliance</th>
<th>Extreme Performance Appliance</th>
<th>Active Enterprise Data Warehouse</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Purpose</strong></td>
<td>Test/Development or Smaller Data Marts</td>
<td>Analytics on Extreme Data Volumes from New Data Types</td>
<td>Data Warehouse or Departmental Data Marts</td>
<td>Extreme Performance for Operational Analytics</td>
</tr>
<tr>
<td><strong>Scalability</strong></td>
<td>Up to 11TB</td>
<td>Up to 186PB</td>
<td>Up to 343TB</td>
<td>Up to 18TB</td>
</tr>
<tr>
<td><strong>Sub Segment</strong></td>
<td>Departmental Analytics, Entry level EDW</td>
<td>Analytical Archive, Deep Dive Analytics</td>
<td>Strategic Intelligence, Decision Support System, Fast Scan</td>
<td>Operational Intelligence, Lower Volume, High Performance</td>
</tr>
<tr>
<td><strong>Current Model #</strong></td>
<td>560</td>
<td>1600 1650 (Q1)</td>
<td>2650</td>
<td>4600</td>
</tr>
</tbody>
</table>

*Teradata* raising intelligence
## Teradata Purpose-Built Platform Family

<table>
<thead>
<tr>
<th></th>
<th>Data Mart Appliance</th>
<th>Extreme Data Appliance</th>
<th>Data Warehouse Appliance</th>
<th>Extreme Performance Appliance</th>
<th>Active Enterprise Data Warehouse</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Purpose</strong></td>
<td>Test/Development or Smaller Data Marts</td>
<td>Analytics on Extreme Data Volumes from New Data Types</td>
<td>Data Warehouse or Departmental Data Marts</td>
<td>Extreme Performance for Operational Analytics</td>
<td>Enterprise Scale for both Strategic and Operational Intelligence EDW/ADW</td>
</tr>
<tr>
<td><strong>Scalability</strong></td>
<td>Up to 11TB</td>
<td>Up to 186PB</td>
<td>Up to 343TB</td>
<td>Up to 18TB</td>
<td>Up to 92PB</td>
</tr>
<tr>
<td><strong>Sub Segment</strong></td>
<td>Departmental Analytics, Entry level EDW</td>
<td>Analytical Archive, Deep Dive Analytics</td>
<td>Strategic Intelligence Decision Systems</td>
<td>Operational Intelligence</td>
<td>Active Workloads, Real Time Update, Tactical and Strategic response times</td>
</tr>
<tr>
<td><strong>Current Model #</strong></td>
<td>560</td>
<td>1600 1650 (Q1)</td>
<td>2650</td>
<td>4600</td>
<td>6650/6680</td>
</tr>
</tbody>
</table>

Public Release 4/11/11
Active EDW – Responding to the Market

The Need for Speed

1. Effective use of hot data is crucial to business
   > Near real time operational

2. Massive growth in use of advanced analytics
   > High demands on system resources

3. Reduce energy and floor space while delivering more performance
   > Data centers are limited

4. Big Data demands flexible solutions for growth in both
   > Performance
   > Data capacity
Hot Data Use Case Examples

New Opportunities for Data Warehousing

• BI dashboards
  > 7 X 24 real time operations
• Financial peak periods
  > Month end, quarter end
• Cyber Security
  > Short and long term threats
• Operational reporting
  > Claims processed
  > Inventory instant status
• Machine generated data
  > Rapid response
  > Fast analytics

SSD & Mixed Storage Potential Applications
Steep Growth in Advanced Analytics

• According to a 2009 survey*
  > 38% practicing advanced analytics
  > 85% practicing it by 2012
• Advanced analytics consumes processing horsepower and information
  > For data preparation, scoring, and regression analysis
  > Chew up a huge slice of CPU and I/O bandwidth.
• Two forms of advanced analytics
  > query-based analytics (complex SQL define recent business events)
  > predictive analytics (data mining and statistical methods anticipate future events).

*TDWI

Both drive the need for performance on data
Teradata Active EDW 6680: Efficient Performance

• What’s new?
  > Mixed storage for data warehouse
    – SSD for I/O performance
    – HDD for capacity
  > Extend Teradata Virtual Storage (TVS) for different drive types
    – Provides automatic data migration
    – Hot data in SSDs & warm/cold data in HDD
  > Single cabinet for each 2 node clique
    – Efficient performance per KW & footprint

• What’s the same as 5650?
  > Nodes, HDD storage, BYNET, etc.
  > Coexistence (with future generations)
### Teradata Active Enterprise Data Warehouse 6680

<table>
<thead>
<tr>
<th><strong>Nodes</strong></th>
<th>Based on Intel six-core Xeon processors</th>
</tr>
</thead>
</table>
| **Storage - Mixed** | SSD: 300GB Enterprise Flash Drive  
HDD: 300GB, 450GB, 600GB drives |
| **Configuration Flexibility** | 4 configurations: 12 to 18 SSD per node and 36 to 48 HDD per node |
| **Data capacity** | From 3.8TB Up to 20PB |
| **Scalability** | Scales up to 4,096 nodes |
| **Availability features** | RAID 1, automatic node failover and recovery,  
performance continuity with hot standby nodes,  
fallback, BAR, Dual Active systems |
| **Software** | Teradata 13.10, TVS, 64-bit SUSE Linux |
| **Workload management** | Full Teradata Active System Manager |
| **System management** | Single operational view of System and Viewpoint |
| **Interconnect** | Teradata BYNET® V4 |

= New
Teradata Active EDW 6650: SSD Ready & “Green”

• What’s new?
  > Solid State Drive (SSD) ready
    - SSD transition platform for investment protection
    - Optional add in late 2011
  > Nodes and storage in same cabinet
    - 1 to 3 nodes + Hot Standby Node in 2 cabinet clique*
    - Flexible configurations save space and energy costs

• What’s the same?
  > All basic elements of previous 5650 model
    - Nodes, BYNET®, software, etc
    - Storage drives and array
  > Coexistence back to 5500 generation

*Clique is the basic Teradata redundancy grouping of nodes
# Teradata Active Enterprise Data Warehouse 6650

<table>
<thead>
<tr>
<th>Feature</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Nodes</strong></td>
<td>Based on Intel six-core Xeon processors</td>
</tr>
<tr>
<td><strong>Storage</strong></td>
<td>300GB or 450GB HDD enterprise-class drives; SSD in future</td>
</tr>
<tr>
<td><strong>Configuration Flexibility</strong></td>
<td>42 HDD per node to 232 HDD per node</td>
</tr>
<tr>
<td><strong>Data capacity</strong></td>
<td>From 7.5TB Up to 92PB</td>
</tr>
<tr>
<td><strong>Scalability</strong></td>
<td>Scales up to 4,096 nodes</td>
</tr>
<tr>
<td><strong>Availability features</strong></td>
<td>RAID 1, automatic node failover and recovery, performance continuity with hot standby nodes, fallback, BAR, Dual Active</td>
</tr>
<tr>
<td><strong>Software</strong></td>
<td>Teradata 12 &amp; higher, SUSE Linux</td>
</tr>
<tr>
<td><strong>Workload management</strong></td>
<td>Full Teradata Active System Manager</td>
</tr>
<tr>
<td><strong>System management</strong></td>
<td>Single operational view and Viewpoint</td>
</tr>
<tr>
<td><strong>Interconnect</strong></td>
<td>Teradata BYNET® V4</td>
</tr>
</tbody>
</table>

"= New"
6650 Upgrade Path to SSD

• Determine data temperature profile
  > Hot data determines SSD size
  > Use Teradata Temperature Assessment tool
• Upgrade current 6650 to SSD when performance is needed
  > Add SSD storage arrays
  > Add CPU capability
  > Upgrade to Teradata 13.10 and TVS
  > Restart system
• Performs as an SSD mixed storage system
Teradata Active EDW Investment Protection

- 6650 supports up to 6 generations of coexistence
  > Enables full performance from all generations of nodes in the system
Mixed Storage:
- SSD
- Teradata Virtual Storage
- Performance Impacts
SSD: Drive Performance Catches Up to CPU
One SSD Provides Throughput of 22 HDDs

The Teradata “Sweet Spot” for drive use

<table>
<thead>
<tr>
<th></th>
<th>Pliant Enterprise SSD</th>
<th>Ratio</th>
<th>Enterprise 15K HDD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Random 80/20 (MB/sec)</td>
<td>&gt;450</td>
<td>22X</td>
<td>20</td>
</tr>
<tr>
<td>Random I/O latency (Sec)</td>
<td>$10^{-6}$</td>
<td>&gt;1000</td>
<td>$10^{-3}$</td>
</tr>
<tr>
<td>I/O per second (4 KB)</td>
<td>$10^5$</td>
<td>150X</td>
<td>$10^2$</td>
</tr>
<tr>
<td>Sequential read BW (MB/s)</td>
<td>&gt;450</td>
<td>3X</td>
<td>&gt;150</td>
</tr>
</tbody>
</table>

Where other DW use drives
Today’s HDD Based Data Warehouse

Challenge to Balance CPU and Storage I/O

- Ever growing TB per node
  - Higher drive counts
  - Larger drive sizes reduce I/O per TB ratio
    - 300 and 450 GB
  - Performance per data space has eroded
    - 75% lower over 4 years

- How does the Market Respond?
  - Partially fill disk space – “short stroke”
    - Boosts effective performance per data space
  - Commercial concessions

Active EDW 6650

14TB user data
54TB raw disk

The Performance Balance = TPerf

CPU speed
HDD
Node
Storage IOPS

HDD = Hard Disk Drive
IOPS = I/O Operations per Second
Teradata Active EDW with Mixed Storage

- SSD reduces # of drives required
  - Maintains CPU to IOPS balance
  - SSD speed = up to 22X HDD for Teradata I/O usage

- Higher performance per TB
  - Reduced TB per node
  - Maintains same Node TPerf
  - Can boost performance per data space 2X – 4X

- Deliver performance without excess data capacity

Example configuration:
SSD = Solid State Drive
HDD = Hard Disk Drive
All drives = 300GB
IOPS = I/O Operations per Second
SSD Concentrates Performance on Less Data

- Balance node CPU and Storage I/O with SSD impacts data space
  - Same performance per Node
  - Up to 4X more nodes for same data space as HDD only
  - Up to 4X higher overall performance per TB
- Result: Flexible and efficient systems
  - Match both performance and data space needs

![Diagram showing SSD vs HDD usage](image)

- Can use 450GB or 600GB HDD for more data space per node
- 3.5TB per node
- 188 HDDs = \(\approx 14\)TB CDS equal data space
The next step in the Teradata Multi-Temperature Warehouse

- Drive Types are mixed within the system
  - Fast solid state disks (SSD), high speed enterprise drives (HDD)
- Teradata Virtual Storage automatically places data on drive type that corresponds to the data temperature
  - Optimum use of the storage resources – unique in the industry!

Example Of Use:

<table>
<thead>
<tr>
<th>Heat Level</th>
<th>Data Type</th>
<th>Example Data Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hot</td>
<td>Solid State</td>
<td>1-3 Month Sales Data</td>
</tr>
<tr>
<td>Warm</td>
<td>HDD</td>
<td>.5 - 5 Years Sales Data</td>
</tr>
</tbody>
</table>
Mixed Storage and Teradata Virtual Storage

- Optimizes hot data queries by migrating into high performance SSDs
- Cold data in HDD for lower performance reporting and history analytics
- Teradata optimizes use of large main memory in each node as Hot Cache
Temperature of Data is Set by Usage

- ≈25% of EDW data is hot
  > Used most frequently
  > Recent data
  > Last day, week, month

- ≈75% of data is warm/cold
  > Accessed infrequently
  > History – months ago
  > Deep detailed info
Data Warehouse Workload Testing

<table>
<thead>
<tr>
<th>Application</th>
<th>Test Query</th>
<th>1X 6680 node vs. 5650 node</th>
<th>4X 6680 nodes vs. 5650 node</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Loading</td>
<td>MultiLoad/TPT average</td>
<td>1.15X</td>
<td>4.6X</td>
</tr>
<tr>
<td></td>
<td>TPump/TPT stream average</td>
<td>1.12X</td>
<td>4.5X</td>
</tr>
<tr>
<td>Strategic W/L</td>
<td>Multi-Stream DSS average</td>
<td>1.05X</td>
<td>4.2X</td>
</tr>
<tr>
<td>Active Workload</td>
<td>Call center all AMP tactical</td>
<td>1.43X</td>
<td>5.7X</td>
</tr>
</tbody>
</table>

6680: 100% of data on 12 SSD vs. 5650: 100% of data on 188 HDD

Throughput Performance Improvements

Data Space Equal
Faster & Consistent Query Response Times

- Average query times reduced for data on SSD
- SSD eliminates HDD delays
  - Enables more consistent query times
  - Narrows the response time curve

![Graph showing comparison between SSD and HDD response times]

- **SSD Based**
- **HDD Based**

**Number of Queries** vs. **Response Time**

- Max. Response Time Reduced
- Avg. Response Time Reduced
Teradata Active EDW 6650 & 6680 Solutions
Teradata Active EDW: Use 6650 or 6680?

**Growing performance at same pace as data storage**
- Satisfied with performance
- Future high performance needs – add SSD later
- Coexistence with prior generations

**Growing performance at a higher pace than data storage**
- Tighter SLA for Active workloads
- Floor space constrained
- Hot data sets with cold data
- Performance needs are NOW

6650
HDD Only

6680
Mixed Storage

Nodes

HDD

SSD

Nodes

Nodes

HDD

HDD
**Teradata Active EDW 6650 – Matches the 5650**  
**Saves Energy & Floor Space**

### 5650 Clique
- Basic building block
- 4 cabinets
- 2 nodes
- UPS
- UPS
- UPS
- UPS
- 188 drives per node

### 6650 Clique
- Basic Building Block
- 3 cabinets
- 1 node
- UPS
- UPS
- UPS
- 124 drives per node

<table>
<thead>
<tr>
<th></th>
<th>5650</th>
<th>6650</th>
<th>Compare</th>
</tr>
</thead>
<tbody>
<tr>
<td>User data space per clique</td>
<td>29.6TB</td>
<td>29.1TB</td>
<td>equal</td>
</tr>
<tr>
<td>HDD per clique</td>
<td>376</td>
<td>372</td>
<td>equal</td>
</tr>
<tr>
<td>Nodes per clique</td>
<td>2</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>TPerf per clique</td>
<td>238</td>
<td>243</td>
<td>~equal</td>
</tr>
<tr>
<td>Energy &amp; floor $/year</td>
<td>$82K</td>
<td>$67K</td>
<td>-20%</td>
</tr>
</tbody>
</table>

① **Node CPU capacity available for SSD add-in and special workloads**
# 6680 Lowers Data Center Costs

*Where Higher Performance per Data Space Needed*

## 5650 System
- Fill disks only 50% to get 2X performance per capacity
- 2 nodes
- 188 drives per node
- 300GB

## 6680 System
- SSD enables higher performance per data space
- 2 nodes
- 84 HDD (600GB)
- 28 SSD (300GB)

## Comparison Table

<table>
<thead>
<tr>
<th></th>
<th>5650</th>
<th>6680</th>
<th>Compare</th>
</tr>
</thead>
<tbody>
<tr>
<td>User Data Space delivered</td>
<td>29.6TB</td>
<td>15.6TB</td>
<td>Far fewer drives</td>
</tr>
<tr>
<td>User Data Space used</td>
<td>14.8TB</td>
<td>15.6TB</td>
<td>Save wasted space</td>
</tr>
<tr>
<td>Performance per data space*</td>
<td>.96</td>
<td>1.06</td>
<td>1X</td>
</tr>
<tr>
<td>Normalized list price</td>
<td>$1</td>
<td>$.80</td>
<td>.8X</td>
</tr>
</tbody>
</table>

*Normalized TPerf per TB*
### 6680 - Efficient Approach to 2X Performance with Same Data Space

#### 6650 System
- 2 Nodes with Traditional Performance per Data Space
  - 120 drives per node
  - 300GB

#### 6680 System
- 4 Nodes with Same Data = 2X Performance per Data Space
  - 36 HDD/node
  - 12 SSD/node
  - 600GB

<table>
<thead>
<tr>
<th></th>
<th>6650</th>
<th>6680</th>
<th>Compare</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data space</td>
<td>29.1TB</td>
<td>29.2TB</td>
<td>Same</td>
</tr>
<tr>
<td>TPerf per system</td>
<td>243</td>
<td>504</td>
<td>2.1X</td>
</tr>
<tr>
<td>Performance per Data Space*</td>
<td>.49</td>
<td>1.04</td>
<td>2.1X</td>
</tr>
</tbody>
</table>

*Normalized TPerf per TB

27% lower data center costs
6680 - Efficient Approach to 4X Performance with Same Data Space

### 5650 System
Traditional Performance per Data Space

- **2 nodes**
- **UPS**
- **188 HDD per node**

### 6680 System
4X Nodes @ Same Data Space = 4X Performance per Data Space

- **2 nodes**
- **UPS**
- **36 HDD per node**
- **12 SSD per node**

<table>
<thead>
<tr>
<th></th>
<th>5650</th>
<th>6680</th>
<th>Compare</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data space</td>
<td>29TB</td>
<td>30TB</td>
<td><strong>equal</strong></td>
</tr>
<tr>
<td>HDD/SSD capacity</td>
<td>300/NA</td>
<td>300/300</td>
<td><strong>same</strong></td>
</tr>
<tr>
<td>Performance per Data Space*</td>
<td><strong>0.48</strong></td>
<td><strong>2.0</strong></td>
<td><strong>4X</strong></td>
</tr>
</tbody>
</table>

*Normalized TPerf per TB