Time Series Analytics

Rob Armstrong
Sr. Technologist
Key Takeaways

• Integrate new data sources and include new user communities for broader insights

• Simplified and optimized capabilities to accelerate time based analytics

• Leverage integrated analytics with the scale, optimization, and management of Teradata
Scalability across manufacturing plants, oil fields, fleets of vehicles, power grids, ...

Business users, Operational Technology engineers & data scientists

Joining sensor data to corporate contextual data

Solving complex analysis previously unsolvable

Optimized & parallel performance
Enabling IoT Analytics

The operationalization of real time Operational Technology data from sensors/IoT/RFID and other sources:

Deep Sea Oil & Gas Production...

Operational Failures:
Life Threatening Conditions & Lost Revenues

Preventing equipment failure:
• Know Safety Tolerances are Reached
• Safer Work Environment

Predicting Changes to Sub-Surface Conditions with Time Series Data:
• Ensures Production Targets are Met
• Keeps Cash Flow On-Plan
Other Time Based Use Cases

- Securities/stock trades
- Commodity prices tracking
- Autonomous trading algorithms
- Geospatial device tracking
- RFID/bar code packages
- Gaming – role play, gambling
- Mobile/web application event streams
- Predict staffing requirements

- Merchandise forecasting
- Infrastructure monitoring/audits
- Dev/Ops events
- Employee productivity and security
- Period based Sales analysis (Hourly sales)
- Transportation delays
- Vendor commitment attainment
Making Business Analytics Easier…

• Without time series capabilities
  – Complex SQL to align time intervals
  – Every query = full table scan
  – Complex analytics take hours

• With time series capabilities
  – Optimized data storage
  – Highly efficient queries plans
  – Reduce time to insight and action
• Dramatic reduction in data preparation time
  – It’s all ETL then ad hoc SQL

• Easiest way to join multi variate sensor streams
  – Align different grains of time
  – When you have to compare the shape of two curves
  – Reach across UDA with QueryGrid

• Teradata class scalability
  – Much higher accuracy
  – Turn development into deployment

... and Data Scientists Happier
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
### Time Aware Aggregate Example

```sql
SELECT $TD_TIMECODE_RANGE, $TD_GROUP_BY_TIME, SENSORID, AVG(TEMPERATURE) FROM BUOYS
WHERE TIMECODE BETWEEN TIMESTAMP '2017-08-11 01:00:00' AND TIMESTAMP '2017-08-11 03:00:00'
GROUP BY TIME( MINUTES(30) AND SENSORID) USING TIMECODE(TD_TIMECODE)
ORDER BY SENSORID, $TD_GROUP_BY_TIME;
```

<table>
<thead>
<tr>
<th>Timecode-Range</th>
<th>Group by 30 minutes</th>
<th>Sensor ID</th>
<th>Temperature</th>
</tr>
</thead>
<tbody>
<tr>
<td>'2017-08-11 01:00:00', '2017-08-11 01:30:00'</td>
<td>1</td>
<td>22</td>
<td>63.5</td>
</tr>
<tr>
<td>'2017-08-11 01:30:00', '2017-08-11 02:00:00'</td>
<td>2</td>
<td>22</td>
<td>64.6</td>
</tr>
<tr>
<td>'2017-08-11 02:00:00', '2017-08-11 02:30:00'</td>
<td>3</td>
<td>22</td>
<td>65.0</td>
</tr>
<tr>
<td>'2017-08-11 02:30:00', '2017-08-11 03:00:00'</td>
<td>4</td>
<td>22</td>
<td>65.1</td>
</tr>
<tr>
<td>'2017-08-11 01:00:00', '2017-08-11 01:30:00'</td>
<td>1</td>
<td>23</td>
<td>66.4</td>
</tr>
<tr>
<td>'2017-08-11 01:30:00', '2017-08-11 02:00:00'</td>
<td>2</td>
<td>23</td>
<td>65.1</td>
</tr>
<tr>
<td>'2017-08-11 02:00:00', '2017-08-11 02:30:00'</td>
<td>3</td>
<td>23</td>
<td>64.9</td>
</tr>
<tr>
<td>'2017-08-11 02:30:00', '2017-08-11 03:00:00'</td>
<td>4</td>
<td>23</td>
<td>65.1</td>
</tr>
</tbody>
</table>
Time Aware Aggregation Functions – GROUP BY TIME

<table>
<thead>
<tr>
<th><strong>Existing Aggregate Functions</strong></th>
<th><strong>New Aggregate Functions</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Average</td>
<td>Count</td>
</tr>
<tr>
<td>Describe</td>
<td>Kurtosis</td>
</tr>
<tr>
<td>Maximum</td>
<td>Minimum</td>
</tr>
<tr>
<td>Percentile</td>
<td>Rank</td>
</tr>
<tr>
<td>Skew</td>
<td>Sum</td>
</tr>
<tr>
<td>Std. population deviation</td>
<td>Std. sample deviation</td>
</tr>
<tr>
<td>Population variance</td>
<td>Sample variance</td>
</tr>
</tbody>
</table>

If not in the list above, then function is not time aware and cannot be used with the GROUP BY TIME clause.

These new aggregate functions are only invokable with the GROUP BY TIME clause.
Primary Time Index (PTI)
High Performance Parallelism with Efficient Storage and Access

What are the sensor readings between 4:30pm and 5:30pm?

Time Bucket (customizable)

8AM  10AM  12PM  2PM  4PM  6PM
## Primary Time Index Tables (PTI)

### Storage distribution choice

<table>
<thead>
<tr>
<th>Time interval only</th>
<th>Time + column list</th>
<th>Column list only</th>
</tr>
</thead>
<tbody>
<tr>
<td>PRIMARY TIME INDEX (TIMESTAMP(6), DATE '2016-02-22', HOURS(2))</td>
<td>PRIMARY TIME INDEX (TIMESTAMP(6), DATE '2016-04-19', HOURS(2), COLUMNS(COUNTRYID,CARID))</td>
<td>PRIMARY TIME INDEX (TIMESTAMP(6), DATE '2016-01-01', COLUMNS(SENSORID))</td>
</tr>
</tbody>
</table>

### In-table logical ordering

<table>
<thead>
<tr>
<th>Time code only</th>
<th>Time code + sequence number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time code only</td>
<td>Time code + sequence number</td>
</tr>
</tbody>
</table>
Many SQL Table Designs Include Time

<table>
<thead>
<tr>
<th>Partitioned Primary Index (PPI)</th>
<th>Temporal Tables</th>
<th>Primary Time Index (PTI)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Business</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Multi-dimensional analytics</td>
<td>• Time periods (ranges)</td>
<td>• High volume time stamped data</td>
</tr>
<tr>
<td>• Hierarchical analytics</td>
<td>• Historical relevance</td>
<td>• Time aware analytics</td>
</tr>
<tr>
<td>• Date, character, or numeric levels</td>
<td>• Audit – what was the situation when...</td>
<td>• Sorted data</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Unique algorithms</td>
</tr>
<tr>
<td><strong>Technology</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Multi-level (up to 64)</td>
<td>• Slowly changing dimensions</td>
<td>• Distribution to AMPs by time buckets</td>
</tr>
<tr>
<td>• Does not effect row distribution to the AMPs</td>
<td>• Insert, update, delete</td>
<td>• Updates/deletes rare</td>
</tr>
<tr>
<td>• Data is not ordered</td>
<td>• Normalize and overlap functions</td>
<td>• Insert late arrival data</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Multivariate payload common</td>
</tr>
</tbody>
</table>

All table types can use “GROUP BY TIME”
Integrating Analytics Drive the Most Value

**Assets**
- Location data – *TimeSeries and Geospatial*
- Sensor observations – *Timeseries*

**Open/Public Data**
- Map data – *Geospatial*
- Traffic (Real Time; Historical) – *JSON*
- Emergency data – *JSON*
- Weather data – *JSON / Temporal*

**Business Data (Transactional and Temporal):**
- Requests
- Financial Data
- Historical data
- Supply Chain

**Teradata Analytics Platform**

**Analytics**
- Trip Planning
- Demand Based re-planning (Urgent Request bases)
- Optimal delivery plan
- Capacity prediction planning
- Repair / Replace decisions
- Delay versus Penalty costs

**Data Integration**
Teradata Advantage

Integrated Analytics

- Analyze various data in context with business data
- Use multi-function analytic engine

High-performance enabled by Primary Time Index (PTI) that supports time sensitive decisions

Agile analysis enabled by Time-Aware Aggregate Functions that work with ANY time component data

... on robust Teradata Database: Scalable, highly available, high performance, and secure

... with Teradata Everywhere deployment options. Design for data gravity
Key Takeaways Summary

• Integrate new data sources and include new user communities for broader insights
  – Examples that we shared such as incorporate IoT analytics and extending the tools into the data scientist arena for new insights

• Simplified and optimized capabilities to accelerate time based analytics
  – This would include the new time aware functions and optimized Primary Time Index table options

• Leverage integrated analytics with the scale, optimization, and management of Teradata
  – Examples that we shared such as the integration of Geospatial, Temporal, and Transactional to drive targeted and timely action