Proactive Performance Monitoring with Baselines and Adaptive Thresholds

John Beresniewicz
Technical Staff, Oracle USA
The following is intended to outline our general product direction. It is intended for information purposes only, and may not be incorporated into any contract. It is not a commitment to deliver any material, code, or functionality, and should not be relied upon in making purchasing decisions. The development, release, and timing of any features or functionality described for Oracle’s products remains at the sole discretion of Oracle.
Oracle’s Complete Enterprise Software Stack

Built-in & Integrated Manageability

- Leader in the complete enterprise application stack
- Built-in manageability in every tier
- Integrated manageability across the entire stack
Oracle Enterprise Manager

*Increases Business Efficiency*

- Manage applications top-down, from the business perspective by understanding user experiences and business impact of IT issues

- Manage entire application lifecycle to increase business agility with comprehensive application quality management and compliance solutions

- Reduce operational costs through intelligent diagnostics and automated IT processes
Agenda

• Performance Monitoring
• Understanding Metrics
• Baselines and Adaptive Thresholds
• Enterprise Manager Use Cases
Performance Monitoring
A brief history

• Availability monitoring
  • Simple Boolean (up/down) using ping
  • Notification frameworks constructed

• Performance monitoring
  • Fixed thresholds over system-level counters (V$SYSSTAT)
  • Use existing frameworks

• Vendor metric madness
  • More metrics must be better

• Users complaints are still the primary alerting mechanism
Performance alerting is difficult

- Performance is subjective and variable
  - Better or worse, not best or worst
- Applications vary in performance characteristics
- Workloads vary predictably within system
- Many metrics, few good signals
  - DB Time metrics far superior to counter-based ones
- Metrics lack semantic framework
  - Do alerts point at symptoms, causes, both?
- Setting thresholds manually is labor intensive
  - The M x N problem (M targets and N metrics)
Understanding Metrics
Classifying metrics

- Identify a set of basic metrics

- PERFORMANCE: Time-based metrics
  - KING KONG: Average Active Sessions
  - Response time per Txn, Response time per call

- WORKLOAD TYPE
  - What kind of work is system doing?
  - Typically the “per txn” metrics

- WORKLOAD VOLUME
  - How much demand is being placed on system?
  - Typically the “per sec” metrics

- Triage performance effects by correlating with causes
Demand varies predictably

Autocorrelation of calls per second for email system
Executions per second over a week

- Weekdays show clear hour-of-day pattern
- Weekends different
- What threshold to set?
Scotty, I think we have a problem
Outliers or events?

In stable system, metrics should be statistically stable and rare observations may signal events.

Are these significant?
Baselines and Adaptive Thresholds
Operational requirements

- Set alert thresholds automatically
- Determine thresholds relative to baseline behavior
- Adjust thresholds for expected workload changes
- Adapt thresholds to system evolution
AWR Baselines

• Captured AWR snapshots representing expected performance under common workload
  • Capture can be pre-configured using templates

• SYSTEM_MOVING_WINDOW
  • Trailing N days of data
  • Compare performance against recent history
  • N is settable in days, 3 weeks or 5 weeks are nice settings
  • Out-of-box baseline in RDBMS 11g
Time-grouping

• Captures workload periodicity by grouping data into common diurnal time buckets

• Daily periodicity
  • All hours, Day-Night, Hour-of-Day

• Weekly periodicity
  • All days, Weekday-Weekend, Day-of-Week

• Time-grouping combines daily and weekly periodicities
Metric statistics

• Basic metrics only

• Computed over SYSTEM_MOVING_WINDOW
  • Standard stats: MIN, MAX, AVG, STDDEV
  • Percentiles:
    • Measured: 25, 50 (median), 75, 90, 95, 99
    • Estimated: 99.9, 99.99

• Computed over time-groups
  • Automatically determined in 11g

• Computed weekly
  • Saturday 12 midnight Scheduler job
Time-grouped statistics

Baseline Statistics Visualization
Baseline: GMAIL all data NW
Type: Static baseline
Time Period: From Jun 2, 2004 to Aug 12, 2004
Day Partitioning: By Day and Night
Week Partitioning: By Weekdays and Weekend

User Calls (per second) Percentile Values by Baseline Partition
Measured percentiles (.50, .95 and .99) displayed in darker shades. Estimated percentiles (.999 and .9999) displayed in shades of grey.
Adaptive alert thresholds

- Percent of maximum thresholds
  - User input multiplier over time group maximum
  - Good for detecting load peaks

- Significance level thresholds
  - Signal on unusual metric values
    - HIGH (95 pctile)
    - VERY HIGH (99 pctile)
    - SEVERE (99.9 pctile)
    - EXTREME (99.99 pctile)

- Computed and set automatically
  - Thresholds can reset every hour (MMON task)
Enterprise Manager
User Interface
Early 10g visualization: seismograph

Performance Metrics

Normalized Metric Values: JB T1

Workload Volume Metrics

Workload Type Metrics
Enterprise Manager entry points

• DB home page: Related Links

• 10g: Metric Baselines
  • Need to enable metric persistence
  • Static and moving window baselines
  • Time grouping selected by user

• 11g: Baseline Metric Thresholds
  • Out-of-box metric persistence and statistics computation
  • Improved use case based interface
  • Automatic time grouping selection
  • Statistics computed over SYSTEM_MOVING_WINDOW
RDBMS 11g use case goals

• Quickly configure Adaptive Thresholds
• Adjust thresholds in context
• Identify signals for known problem
• Advanced metric analysis
### Baseline Metric Thresholds

#### Category/Name

<table>
<thead>
<tr>
<th>Category/Name</th>
<th>Alerts (Last 7 Days)</th>
<th>AWR Baseline</th>
<th>Threshold Type</th>
<th>Adaptive</th>
<th>Edit Thresholds</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Performance Metrics</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average Active Sessions</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average Synchronous Single-Block Read Latency (ms)</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Response Time (centi-seconds per call)</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Response Time (per transaction)</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Workload Volume Metrics</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Workload Type Metrics</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

#### Related Links

- Threshold Configuration
- Metric Analysis
Quickly configure Adaptive Thresholds

Quick Configuration: Baseline Metric Thresholds
Select one of the following workload profiles to configure a basic set of thresholds that you can expand or change later, if desired.

Workload Profile
- Primarily OLTP (pure transaction processing 24 hours a day)
- Primarily Data Warehousing (query and load intensive)
- Alternating (OLTP during the daytime and batch during the nighttime)

⚠️ By choosing a profile and proceeding with quick configuration, some thresholds that have already been set will be cleared. Refer to the Online Help for complete descriptions of the Quick Configuration effects for each system profile.
Quick configure: OLTP

**OLTP Threshold Settings**

<table>
<thead>
<tr>
<th>Metric Name</th>
<th>AWR Baseline</th>
<th>Threshold Type</th>
<th>Warning Level</th>
<th>Critical Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average Active Sessions</td>
<td>SYSTEM_MOVING_WINDOW</td>
<td>Significance Level</td>
<td>Very High (0.99)</td>
<td>Extreme (0.9999)</td>
</tr>
<tr>
<td>Redo Generated (per second)</td>
<td>SYSTEM_MOVING_WINDOW</td>
<td>Percentage of Maximum</td>
<td>100%</td>
<td>120%</td>
</tr>
<tr>
<td>Response Time (per transaction)</td>
<td>SYSTEM_MOVING_WINDOW</td>
<td>Significance Level</td>
<td>Very High (0.99)</td>
<td>Extreme (0.9999)</td>
</tr>
<tr>
<td>Session Logical Reads (per transaction)</td>
<td>SYSTEM_MOVING_WINDOW</td>
<td>Significance Level</td>
<td>Very High (0.99)</td>
<td>None</td>
</tr>
</tbody>
</table>

**Impact on Existing Thresholds**

⚠️ Applying the OLTP threshold settings will also clear the following settings.

<table>
<thead>
<tr>
<th>Metric Name</th>
<th>AWR Baseline</th>
<th>Threshold Type</th>
<th>Warning Level</th>
<th>Critical Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cumulative Logons (per second)</td>
<td>Fixed Values</td>
<td>100</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Current Open Cursors Count</td>
<td>Fixed Values</td>
<td>1,200</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
## Quick configure: Data Warehouse

### Data Warehousing Threshold Settings

<table>
<thead>
<tr>
<th>Metric Name</th>
<th>AWR Baseline</th>
<th>Threshold Type</th>
<th>Warning Level</th>
<th>Critical Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average Active Sessions</td>
<td>SYSTEM_MOVING_WINDOW</td>
<td>Significance Level</td>
<td>Severe (0.999)</td>
<td>Extreme (0.9999)</td>
</tr>
<tr>
<td>Cumulative Logons (per second)</td>
<td>SYSTEM_MOVING_WINDOW</td>
<td>Percentage of Maximum</td>
<td>120%</td>
<td></td>
</tr>
<tr>
<td>Physical Reads (per second)</td>
<td>SYSTEM_MOVING_WINDOW</td>
<td>Percentage of Maximum</td>
<td>100%</td>
<td></td>
</tr>
<tr>
<td>Response Time (centi-seconds per call)</td>
<td>SYSTEM_MOVING_WINDOW</td>
<td>Significance Level</td>
<td>Severe (0.999)</td>
<td>None</td>
</tr>
</tbody>
</table>

### Impact on Existing Thresholds

Applying the Data Warehousing threshold settings will also clear the following settings.

<table>
<thead>
<tr>
<th>Metric Name</th>
<th>AWR Baseline</th>
<th>Threshold Type</th>
<th>Warning Level</th>
<th>Critical Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>Current Open Cursors Count</td>
<td>Fixed Values</td>
<td>1,200</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Adjust thresholds in context

AWR Baseline

Name: SYSTEM_MOVING_WINDOW

Threshold Settings

- Threshold Type: Significance Level
  - Critical: Severe (0.999)
  - Warning: Very High (0.99)

Occurrences: 2

Average Active Sessions vs. Baseline

Last Updated: September 23, 2008 3:06:16 PM PDT

Related Links

- All Metrics
- AWR Baselines
Adjust thresholds in context

Edit Thresholds: Average Active Sessions

AWR Baseline

Name: SYSTEM_MOVING_WINDOW

Threshold Settings

<table>
<thead>
<tr>
<th>Threshold Type</th>
<th>Percentage of Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Critical</td>
<td>150</td>
</tr>
<tr>
<td>Warning</td>
<td>100</td>
</tr>
<tr>
<td>Occurrences</td>
<td>2</td>
</tr>
</tbody>
</table>

Average Active Sessions vs. Baseline

Week of Sep 20, 2008

TIP: To see a different day, click the image for that day above the chart.

Related Links

All Metrics  AWR Baselines
Identify signals for known problem

Find Metrics Correlating to a Known Problem Time
Specify an approximate time at which a problem occurred and an AWR Baseline to compare against metrics at that time.

Problem Time: 9/21/08 AM
AWR Baseline: SYSTEM_MOVING_WINDOW

Overview
Metric Analysis is used to identify metrics that may have provided a good indication for a known problem time in the recent past.

Use this feature to interactively learn which metrics provide good indicators for performance problems specific to your database. Then, set thresholds for those metrics.

TIP Click on a chart to edit the thresholds for a displayed metric.
Identify signals for known problem

**Tips** Click on a chart to edit the thresholds for a displayed metric.

**Metrics with Strong Correlation to the Problem**
There are no metrics strongly correlated to the problem.

**Metrics with Moderate Correlation to the Problem**

- **Response Time (per transaction)**

- **Average Active Sessions**

- **Response Time (centi-seconds per call)**

- **I/O Requests (per second)**