BEZNext Performance Assurance for Big Data World

Dr. Boris Zibitsker, CEO
BEZNext
www.beznex.com
Agenda

- Introduction
- Challenges
- What is the role of Performance Assurance
  - Performance Engineering
  - Dynamic Performance Management and Workload Management optimization
  - Strategic Capacity Planning
- Use Cases
- Conclusion
Introduction

- BEZNext expertise - modeling and performance optimization
- We offer Performance Assurance Software and Services for Big Data, Data Warehouses and Cloud computing
  - Optimize applications design
  - Proactive performance management
  - Strategic capacity planning
- Based in the Chicago area
- Proven track record of assisting many customers in different industries in optimizing business and IT decisions
Challenges

• More than 85% of Big Data projects fail to meet expectations according to a Gartner study
• Complexity, interdependence and Growth
• Difficult to predict the outcome of different changes and be proactive
Challenges
How proactively plan and manage implementation of Analytics Platforms: Teradata Vantage platform

How new applications using both Teradata DBMS and ML Engines incorporating Coprocessors, Kubernetes and QueryGrid will perform?
Challenges
How proactively plan and manage implementation of Analytics Platforms: *IBM Big SQL Sandbox*
Challenges
How proactively plan and manage implementation of Analytics Platforms: *The Microsoft Analytics Platform System (APS)*

The combination of an MPP relational database with MPP Hadoop that address the top trends driving the adoption of Big Data

---

**Meeting today’s Big Data analytics requirements**

**The modern data warehouse**

**Enterprise-ready Big Data**

**Optimized performance with MPP technology and In-Memory Columnstore**

---

Performance Assurance Solutions for the Big Data World
What is the role of Performance Assurance

- BEZNext offers Performance Assurance software and services helping customers:
  - Set realistic expectations and Service Level Goals (SLGs)
  - Implement proactive measures throughout the applications life cycle to meet SLGs continuously and cost effectively
Performance Assurance
Performance Engineering Use Cases
How will new application perform in production environment?
How will new application perform in production environment?

Determine proactive measures necessary to meet SLGs

- Workload and volume of data growth affect workloads’ queueing and software delay time
- Response time of new application in production has:
  - Different Response Time, Service Time, Queueing time and Delay Time for new application
- Response time of production workloads and it’s Queuing Time and Delay Time are changed
- Move of Teradata workloads in Intellicloud affects Service Time, Queueing Time and Delay Time for all workloads
- Change of the Workload Management Rules (TASM or YARN) affect the Queueing time and Delay Time of each workload
SLGs will not be met after New Application implementation

Predict Impact of New Application Implementation
Workload Management

• Concurrency
• Priority
• Resource Allocation

• Big Data Clusters
  • YARN, Kubernetes
• Teradata
  • TASM
• IBM Big SQL Sandbox
• The Microsoft Analytics Platform System (APS)
• Oracle
• Dell
Workload Management in YARN

- YARN Capacity, Fair, and FIFO schedulers Rules
- Control tasks, execution and resource allocation
- The resources are divided by LOB or departments and their actual projects
- Incorporation elasticity into the YARN rules
- If resources are available a project that has a need for additional resources can allocate them

The ResourceManager has two main components: Scheduler and ApplicationsManager.
Workload Management in Kubernetes

Performance Assurance Solutions for the Big Data World
Reducing Concurrency will reduce contention but increase waiting time for the tread

Performance Assurance Solutions for the Big Data World

16
Change of Workload Management rules and ML Algorithm will not be Sufficient to meet SLGs
Determine the Minimum Hardware Upgrade Required to Meet SLGs

Predicting the minimum hardware upgrade required to meet SLGs
Verification of Results

Compare Actual Results with Expected

Testing System

Production Environment

BEZNext
Performance Assurance
Dynamic Performance Management and Workload Management Optimization Use Cases
Determine most frequent anomalies and the most frequent root causes for Teradata and Big Data workloads

- Determine most frequent and severe anomalies and root causes
- Determine seasonal peaks and recommend changes for Workload Management rules
- Apply modeling to evaluate Performance Management options
- Automate results verification

Distribution of Anomalies for each workload

**Workloads with Most Frequent Anomalies**

<table>
<thead>
<tr>
<th>Workload</th>
<th>Total Anomalies</th>
</tr>
</thead>
<tbody>
<tr>
<td>LED</td>
<td>123</td>
</tr>
<tr>
<td>Committee</td>
<td>234</td>
</tr>
<tr>
<td>DR</td>
<td>345</td>
</tr>
<tr>
<td>FDR</td>
<td>456</td>
</tr>
</tbody>
</table>

**Most Frequent Root Causes**

<table>
<thead>
<tr>
<th>Root Cause</th>
<th>Total Anomalies</th>
</tr>
</thead>
<tbody>
<tr>
<td>fail</td>
<td>123</td>
</tr>
<tr>
<td>bug</td>
<td>234</td>
</tr>
<tr>
<td>crash</td>
<td>345</td>
</tr>
<tr>
<td>error</td>
<td>456</td>
</tr>
</tbody>
</table>

All rights Reserved
Determine seasonal peaks for each workload and recommend how to change Workload Management Rules (TASM or YARN) to meet SLGs for all workloads

- Repeatable and predictable Anomalies – Seasonal Peaks
- For expected seasonal peaks the resource allocation rules in YARN, TASM or Kubernetes can be changed proactively

<table>
<thead>
<tr>
<th>Workload Name</th>
<th>Parameter Name</th>
<th>Peak Type</th>
<th>Peak Start Date</th>
<th>Duration</th>
<th>Avg Amplitude</th>
<th>Standard Deviation</th>
<th>Min Value</th>
<th>Max Value</th>
<th>95 Percentile</th>
<th>Peak Length STD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acorn1</td>
<td>Response Time</td>
<td>Daily</td>
<td>01/02 00:00</td>
<td>3</td>
<td>72,494</td>
<td>65,221</td>
<td>1,263</td>
<td>205,899</td>
<td>11,246</td>
<td>0.82</td>
</tr>
<tr>
<td>Acorn2</td>
<td>Response Time</td>
<td>Daily</td>
<td>01/02 17:00</td>
<td>2</td>
<td>0</td>
<td>10,562</td>
<td>5,184</td>
<td>20,319</td>
<td>53,736</td>
<td>1.00</td>
</tr>
<tr>
<td>Admin2</td>
<td>Response Time</td>
<td>Daily</td>
<td>01/02 19:00</td>
<td>3</td>
<td>41,674</td>
<td>17,703</td>
<td>4,305</td>
<td>60,988</td>
<td>11,246</td>
<td>1.48</td>
</tr>
<tr>
<td>Dev1</td>
<td></td>
<td>R</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>123,010</td>
<td>0.00</td>
</tr>
<tr>
<td>HR4</td>
<td></td>
<td>R</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>8,575</td>
<td>0.50</td>
</tr>
<tr>
<td>Load0</td>
<td></td>
<td>R</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>289,686</td>
<td>0.00</td>
</tr>
<tr>
<td>Load1</td>
<td></td>
<td>R</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>514,247</td>
<td>0.00</td>
</tr>
<tr>
<td>Load2</td>
<td></td>
<td>R</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>514,247</td>
<td>0.00</td>
</tr>
<tr>
<td>Load3</td>
<td></td>
<td>R</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2,209,780</td>
<td>0.00</td>
</tr>
<tr>
<td>Load4</td>
<td></td>
<td>R</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>10,354</td>
<td>0.00</td>
</tr>
<tr>
<td>Load5</td>
<td></td>
<td>R</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>8,822</td>
<td>0.82</td>
</tr>
<tr>
<td>Other</td>
<td></td>
<td>R</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>300,900</td>
<td>0.00</td>
</tr>
<tr>
<td>Post1</td>
<td></td>
<td>R</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>90,131</td>
<td>0.00</td>
</tr>
<tr>
<td>QA0</td>
<td></td>
<td>R</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>142,853</td>
<td>0.50</td>
</tr>
<tr>
<td>QA2</td>
<td></td>
<td>R</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>39,461</td>
<td>0.00</td>
</tr>
<tr>
<td>Tech/Support2</td>
<td></td>
<td>R</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>514,247</td>
<td>0.00</td>
</tr>
</tbody>
</table>
Old seasonal peak and corresponding Workload Management rules

Historical Data

Old Seasonal Peak

TASM or YARN Rules of Resource Allocation

Workload Management Rules allocate sufficient resources

All rights Reserved
Determine changes in seasonal peak

Old Seasonal Peak

Historical Data

Current Seasonal Peak

Changes of Seasonal Peak

TASM or YARN Rules of Resource Allocation

Current Workload Management Rules do not allocate right resources on time
Adjusting YARN and Kubernetes rules according to changes of workloads’ seasonality

Old Seasonal Peak

Historical Data

Current Seasonal Peak

Current changes in Seasonal Peak

Recommended change for Workload Management Rules

TASM or YARN Rules of Resource Allocation
Analysis Reliability of Big Data Cluster nodes

Outages Report

<table>
<thead>
<tr>
<th>Node</th>
<th>Number Of Outages</th>
<th>Down Time(h)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ypf1000</td>
<td>10</td>
<td>12</td>
</tr>
<tr>
<td>ypf1005</td>
<td>10</td>
<td>12</td>
</tr>
<tr>
<td>ypf792</td>
<td>9</td>
<td>12</td>
</tr>
<tr>
<td>ypf1001</td>
<td>9</td>
<td>11</td>
</tr>
<tr>
<td>ypf1003</td>
<td>8</td>
<td>11</td>
</tr>
<tr>
<td>ypd548</td>
<td>5</td>
<td>10</td>
</tr>
<tr>
<td>ypf1002</td>
<td>6</td>
<td>9</td>
</tr>
<tr>
<td>ypf1004</td>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>ypf024</td>
<td>3</td>
<td>6</td>
</tr>
</tbody>
</table>

Showing 1 to 9 of 9 entries
Analysis of Big Data Cluster Resource Utilization
Nodes CPU utilization is unbalanced and varied between 32% and 58%.
Analysis of CPU Time consumed by Users and Applications on each Node of the cluster
Performance Assurance
Strategic Capacity Planning Use Cases
Determining the minimum hardware upgrade required to meet SLGs
Determining proactive changes required to meet SLGs
Organizing continuous proactive capacity management process

- Apply Predictive and Prescriptive Analytics to evaluate options
- Justify proactive capacity management measures necessary to meet SLGs with minimum cost and set expectations
  - Predict the impact of an increase in the number of Users and Volume of Data to determine when SLGs will not be met and justify necessary hardware and software upgrades
  - Predict the impact of new application implementation
  - Predict the impact of anticipated move of workloads and data between Data Warehouse and Big Data Clusters or Cloud environment
- Verify results
  - Automatically compare Actual Results with Expected

Predicting the Impact of the Expected Growth and Changes

Recommended Actions Plan
Predict how workloads’ consolidation and move to Intellicloud will affect performance
What will be an impact of moving workloads to different platform

Big Data → Cloud → Data Warehouse

New Application → Cloud → Data Warehouse → Big Data
Automatic Continuous Performance Assurance Control

Apply ML, AI and QNM Models, Optimization & Automation

Automatic Data Collection
Automatic Workload Aggregation and Characterization
Auto Discovery of Hardware & Software Configuration Changes
Automation of Anomaly and Root Causes Detection

SLGs
Options

Automation of Modeling
Optimization of Proactive Actions

Workload Forecasting
Modeling Scenarios
New Applications

Verifications & Feedback

Design and Development
Dynamic Performance Management
Strategic Capacity Planning

All rights Reserved
# BEZNext Performance Assurance Technology

<table>
<thead>
<tr>
<th>Enterprise IT Systems</th>
<th>Data Collections</th>
<th>Data Preparation</th>
<th>Repository</th>
<th>Advanced Analytics</th>
<th>Functions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Big Data Clusters</td>
<td>Auto Discovery Agent</td>
<td>Data Transformation</td>
<td>Data Lake Performance Repository</td>
<td>Descriptive Analytics</td>
<td>Workload Characterization</td>
</tr>
<tr>
<td>Teradata, Oracle, DB2, MS Data Warehouses</td>
<td>OS / Linux Agent</td>
<td>Workload Aggregation</td>
<td>Diagnostic Analytics</td>
<td>Workload Forecasting</td>
<td></td>
</tr>
<tr>
<td>Clouds</td>
<td>Kafka Agent</td>
<td></td>
<td>Predictive Analytics</td>
<td>Performance Prediction</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Spark Agent</td>
<td></td>
<td>Prescriptive Analytics</td>
<td>Workload Management</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Storm Agent</td>
<td></td>
<td>Control Analytics</td>
<td>Performance Management</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Cassandra Agent</td>
<td></td>
<td></td>
<td>Capacity Planning</td>
<td></td>
</tr>
<tr>
<td></td>
<td>YARN Agent</td>
<td></td>
<td></td>
<td>Verification &amp; Control</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Tez Agent</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Other Agents</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- Big Data Clusters
- Teradata, Oracle, DB2, MS Data Warehouses
- Clouds
Conclusion

• BEZNext offers Performance Assurance Solutions for Big Data, Data Warehouses and Cloud environments

• It helps our customers to succeed in developing, implementing, management and growing Big Data Applications
Thank you

Are any Questions?