# BEZNext Performance Assurance for Big Data World

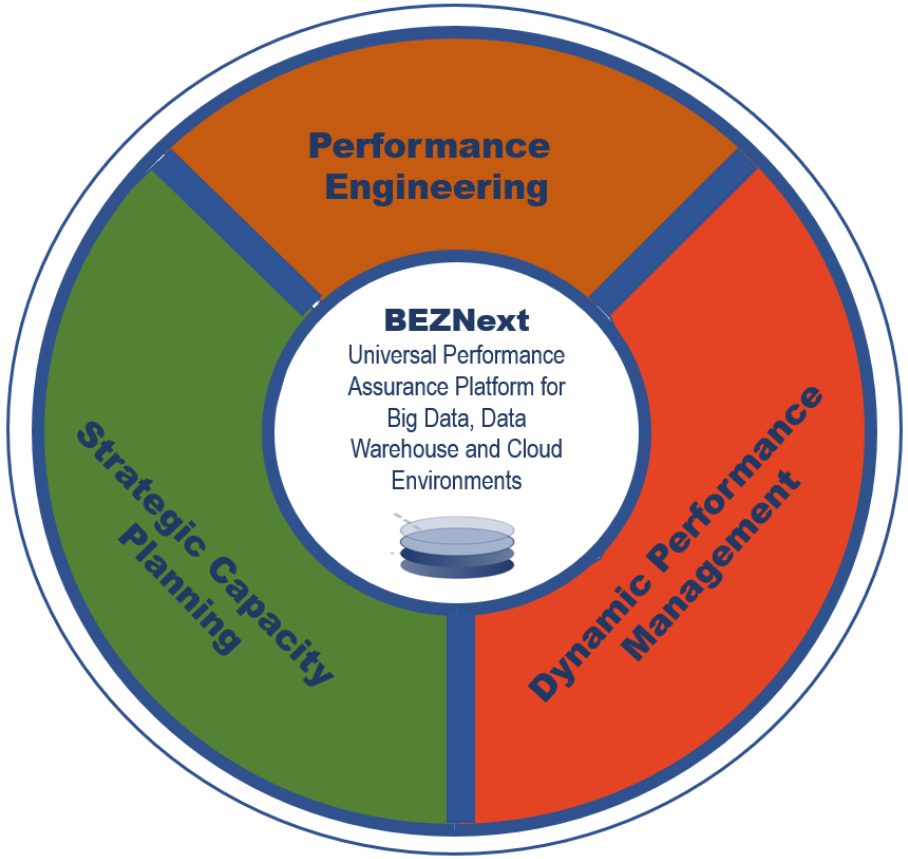

**Dr. Boris Zibitsker, CEO**
**BEZNext**
**www.beznext.com**

# 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

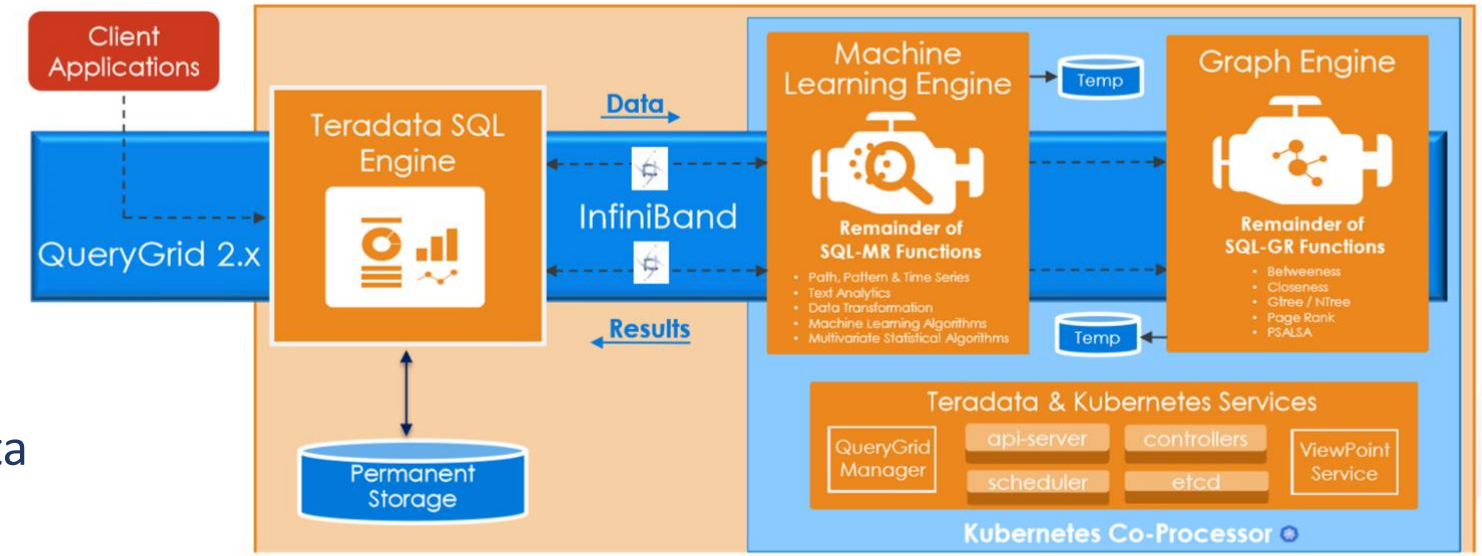


# Challenges

- 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*

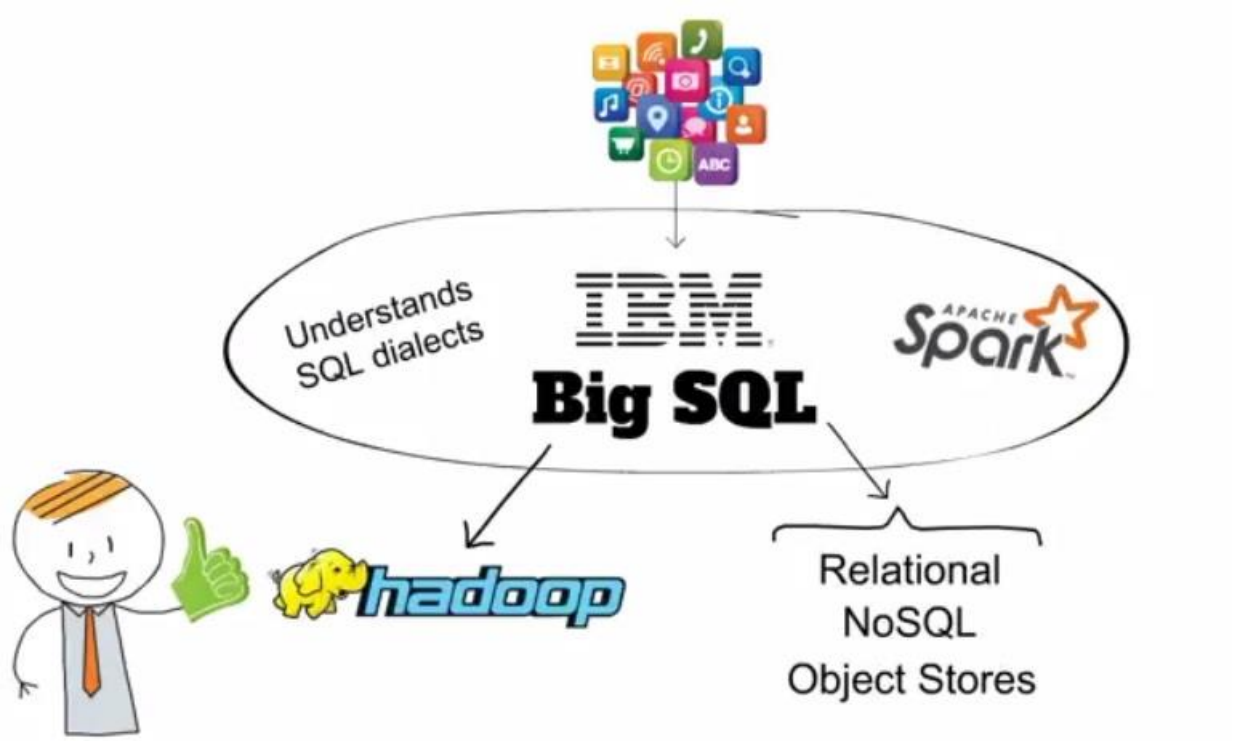



Source: Teradata

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*



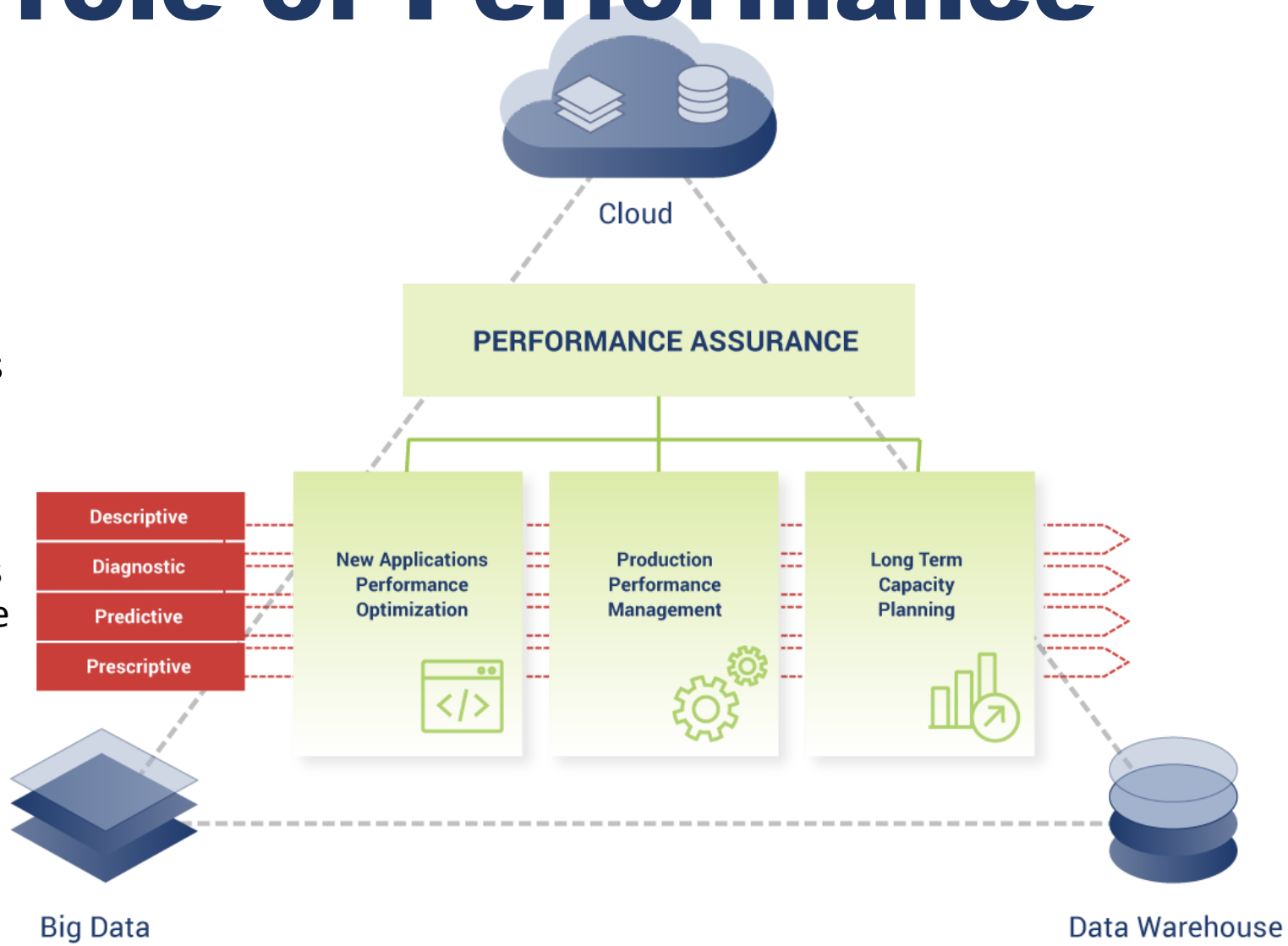
# Challenges

## 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

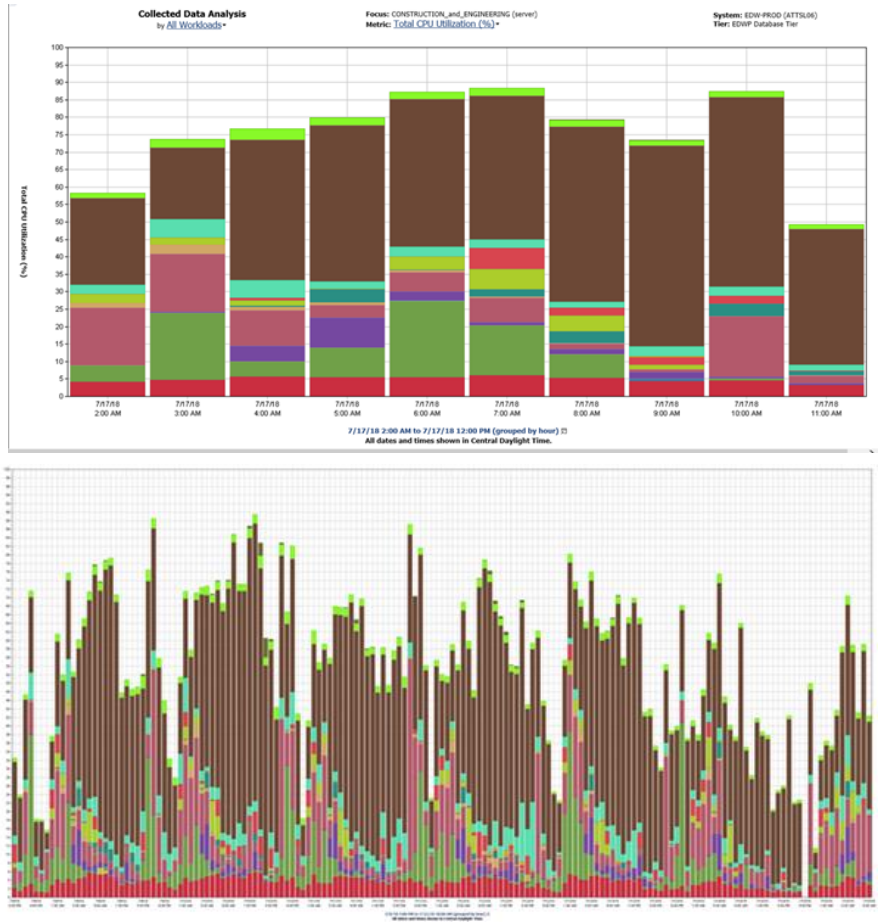
# What is the role of Performance Assurance

- 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

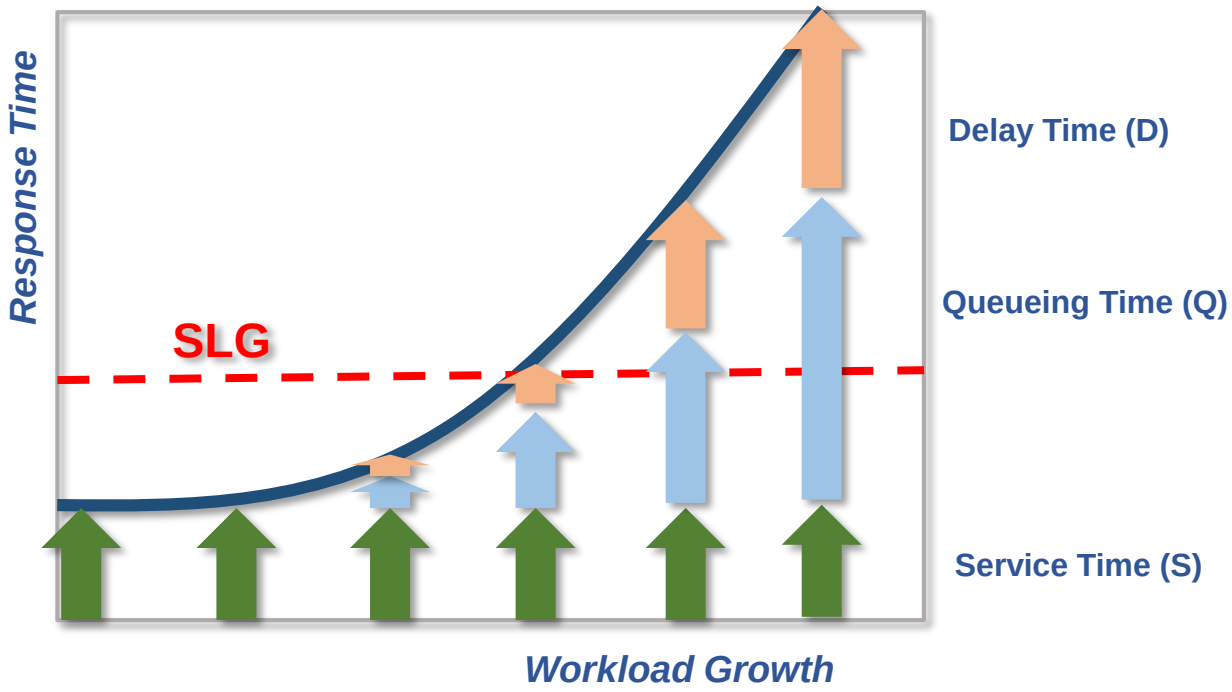
# How will new application perform in production environment?



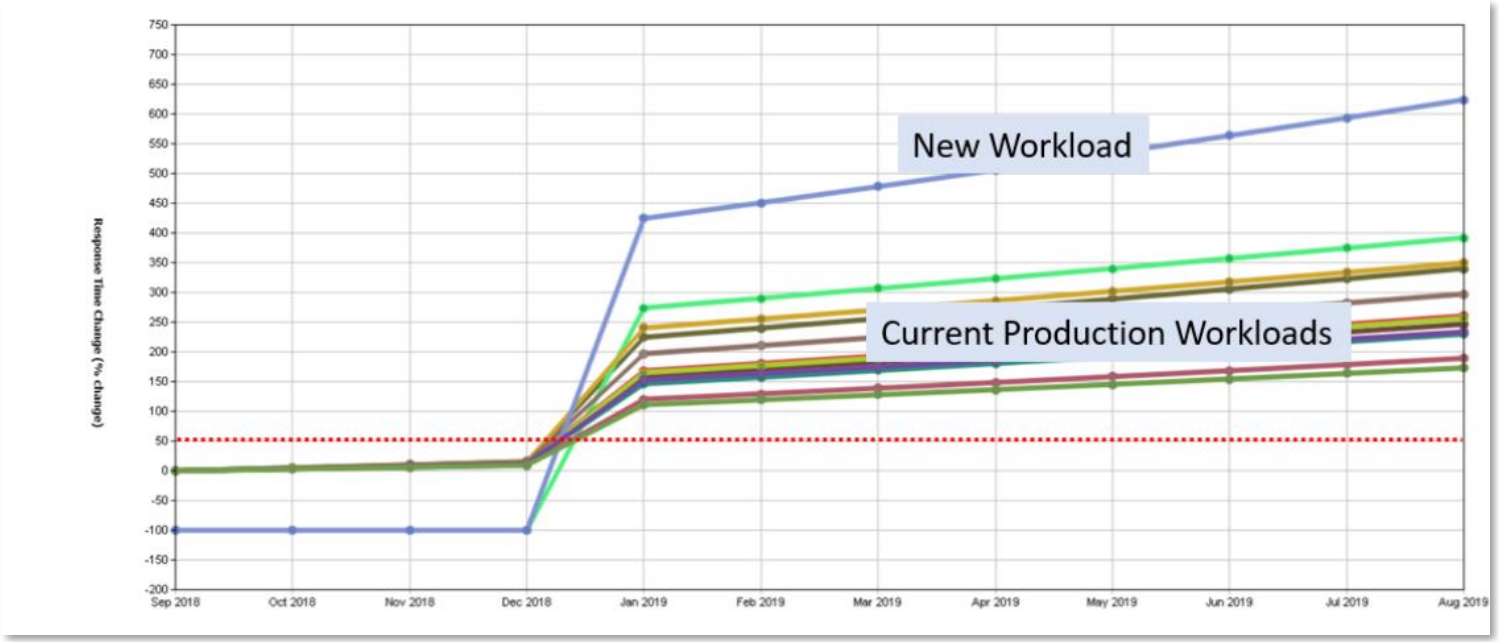

# How will new application perform in production environment?

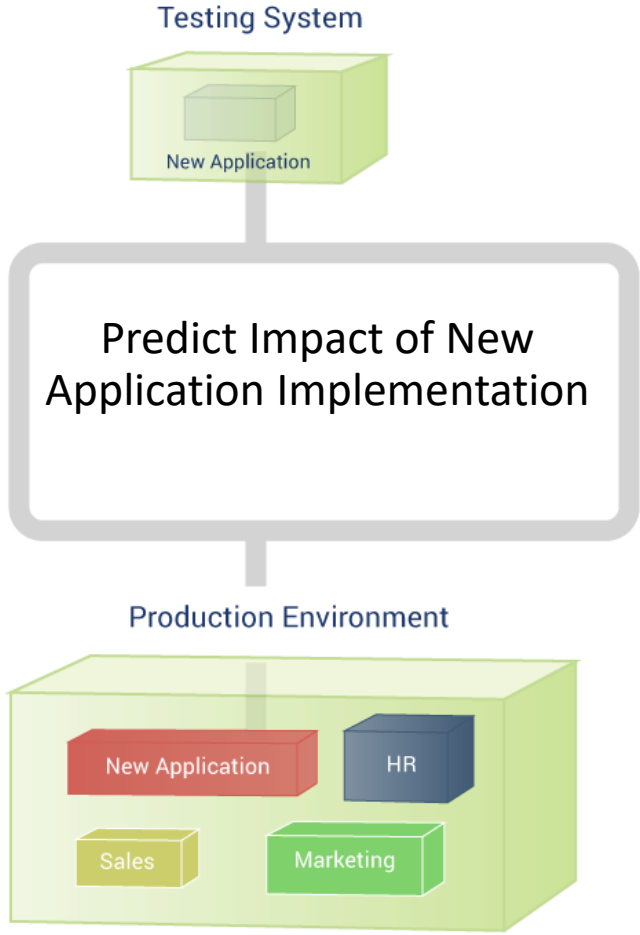
## Determine proactive measures necessary to meet SLGs

- 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



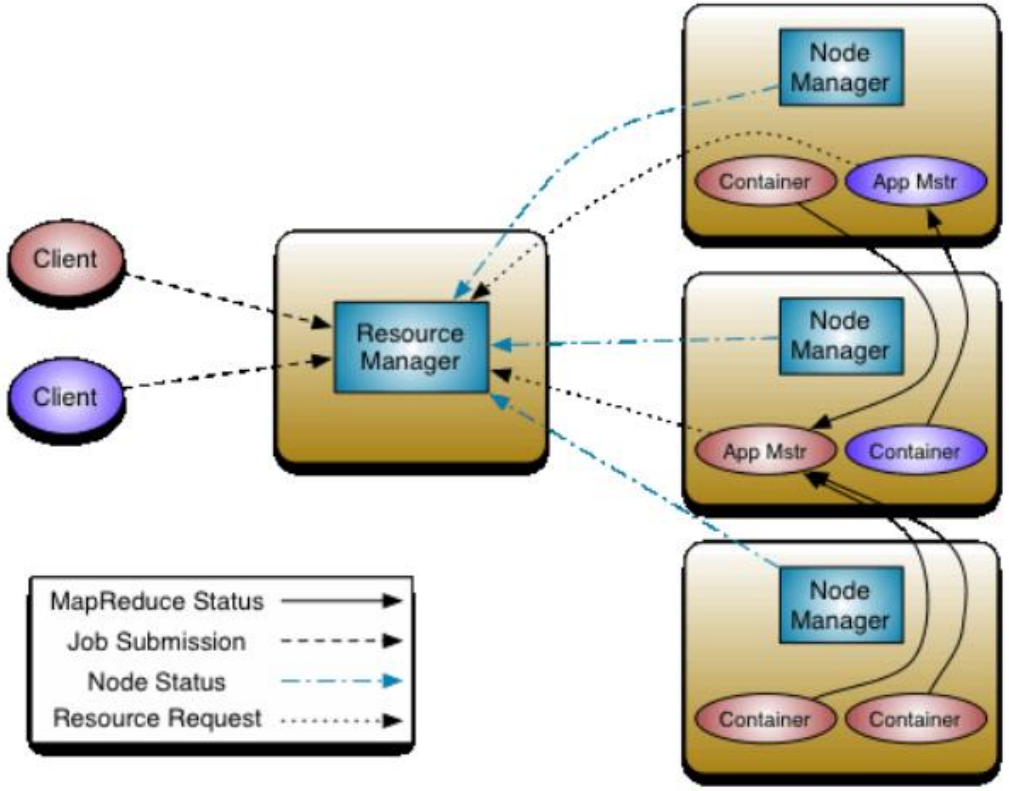



# Workload Management

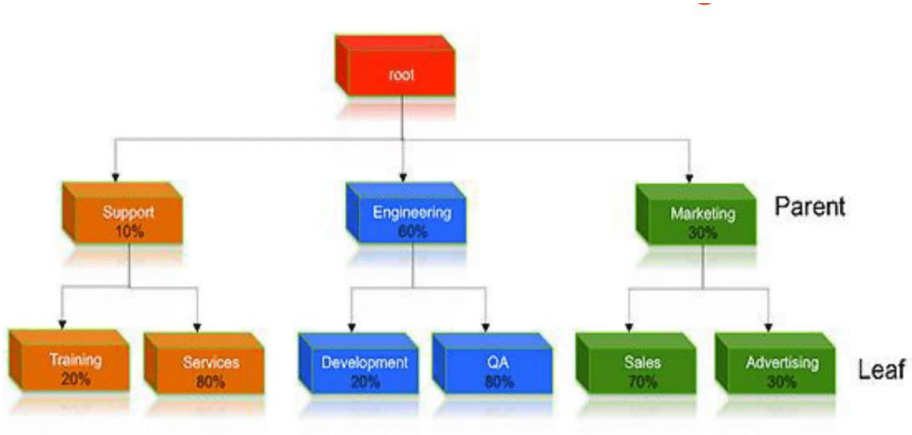
- Concurrency
- Priority
- Resource Allocation

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

# Workload Management in YARN

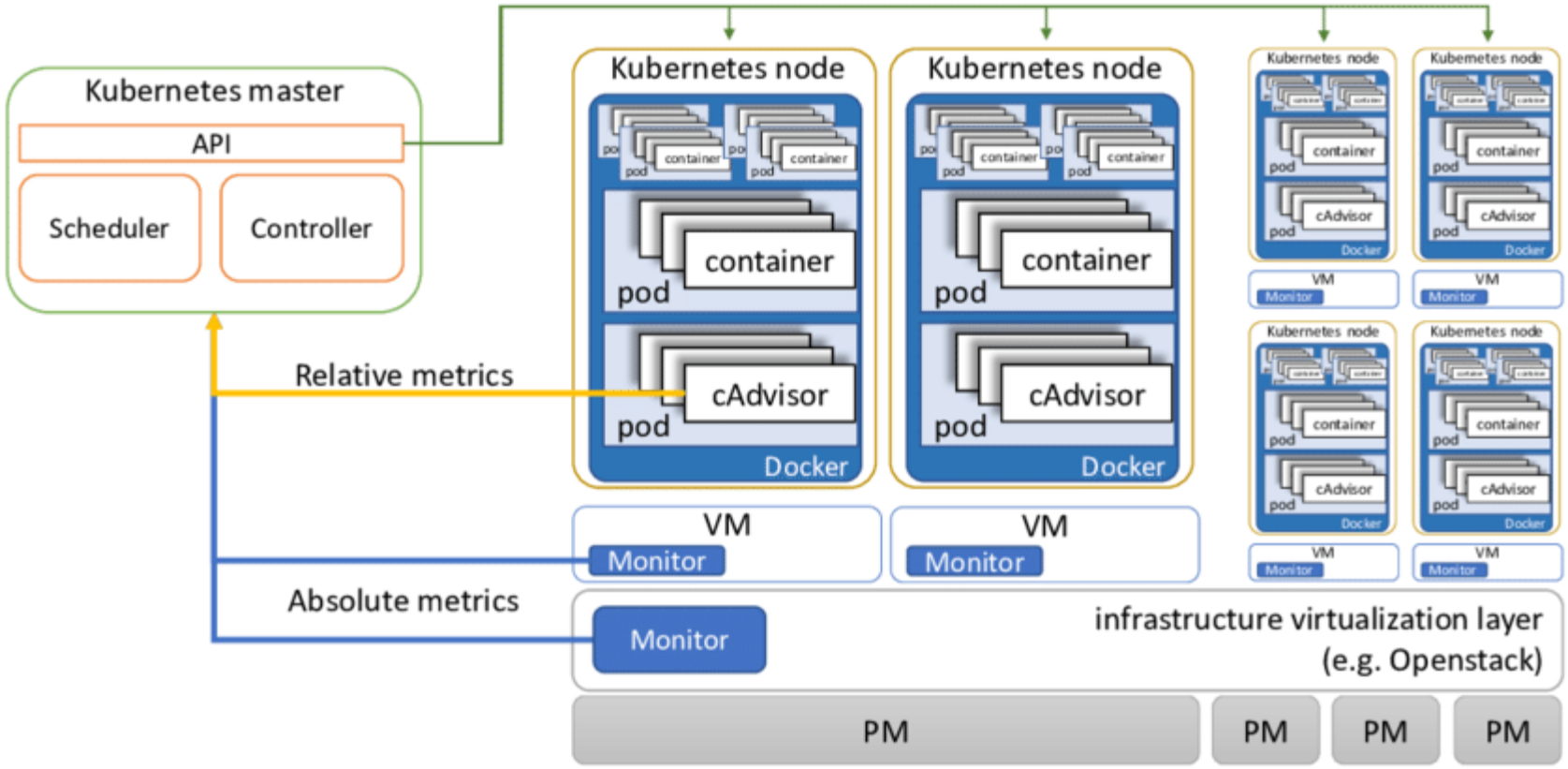



The ResourceManager has two main components: Scheduler and ApplicationsManager.

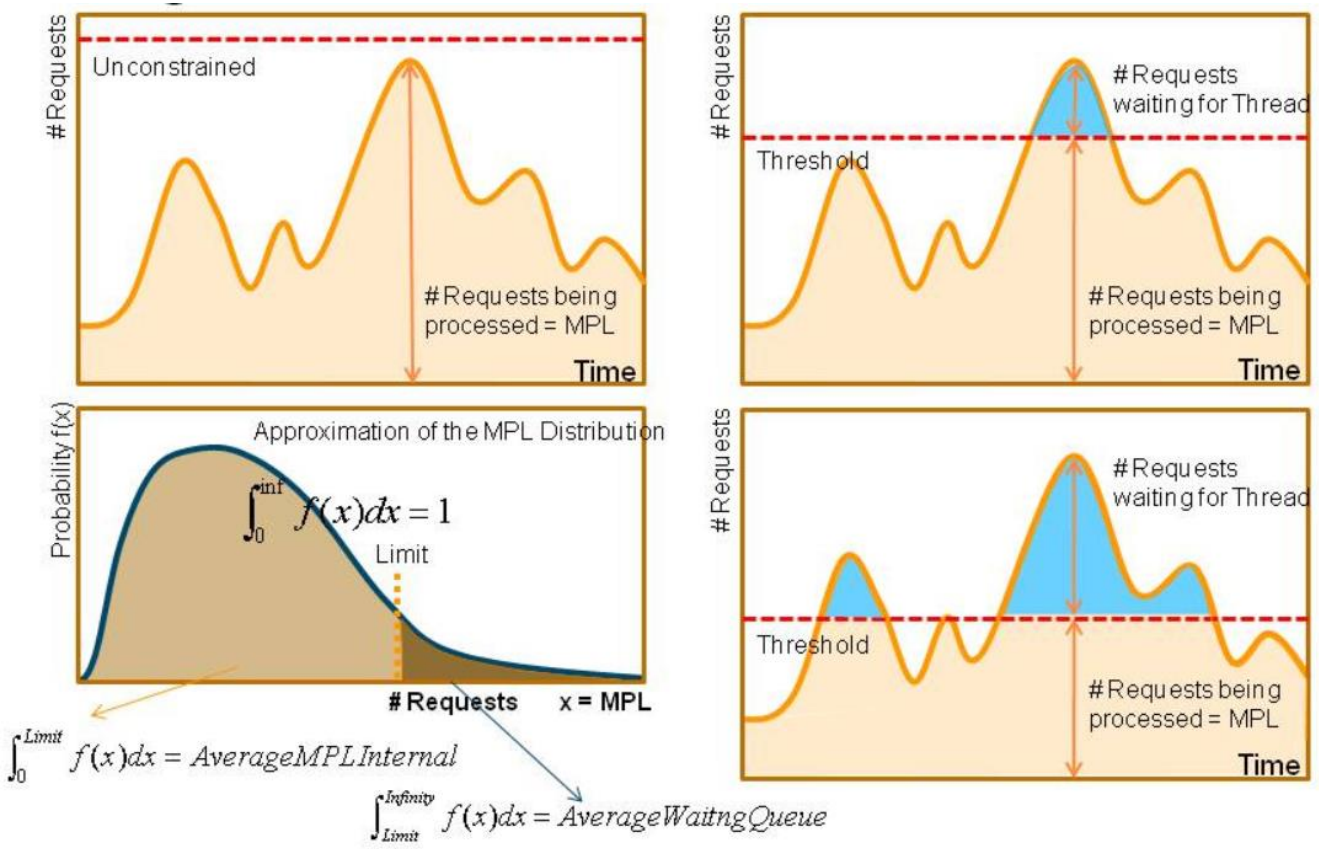



- YARN Capacity, Fair, and FIFO schedulers Rules
- Control tasks, execution and resource allocation
- The resources are divided by LOB or departments and their actual projects
- Incorporation elasticity into the YARN rules
- If resources are available a project that has a need for additional resources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

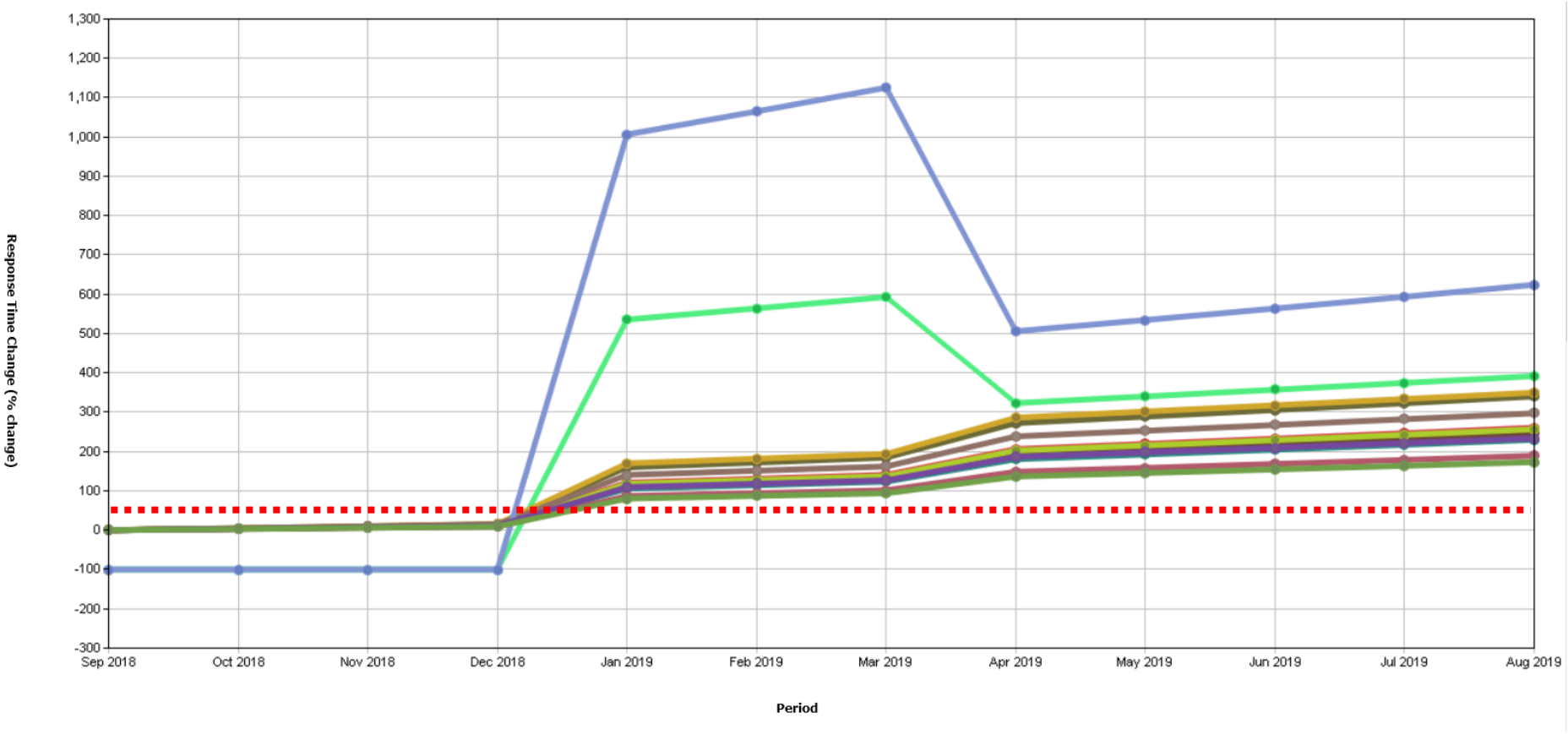
# Workload Management in Kubernetes



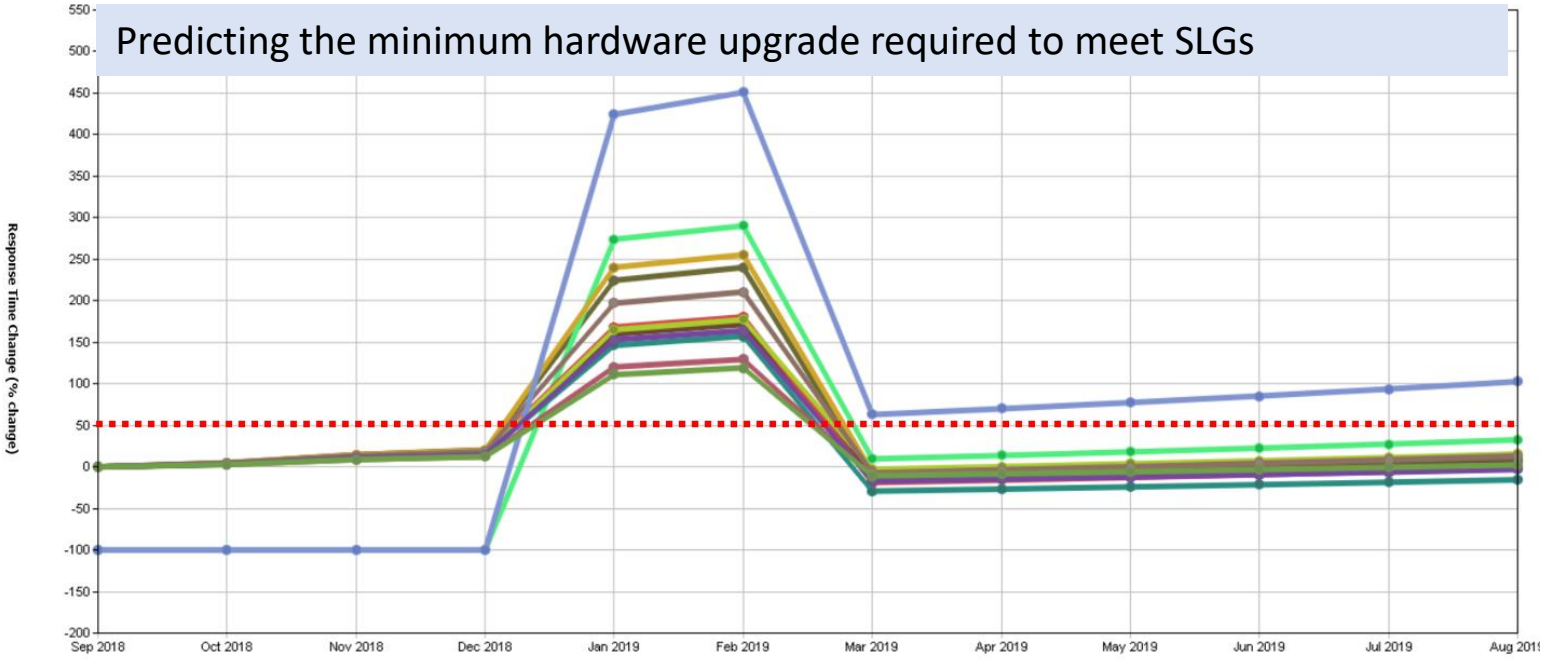
# Reducing Concurrency will reduce contention but increase waiting time for the tread



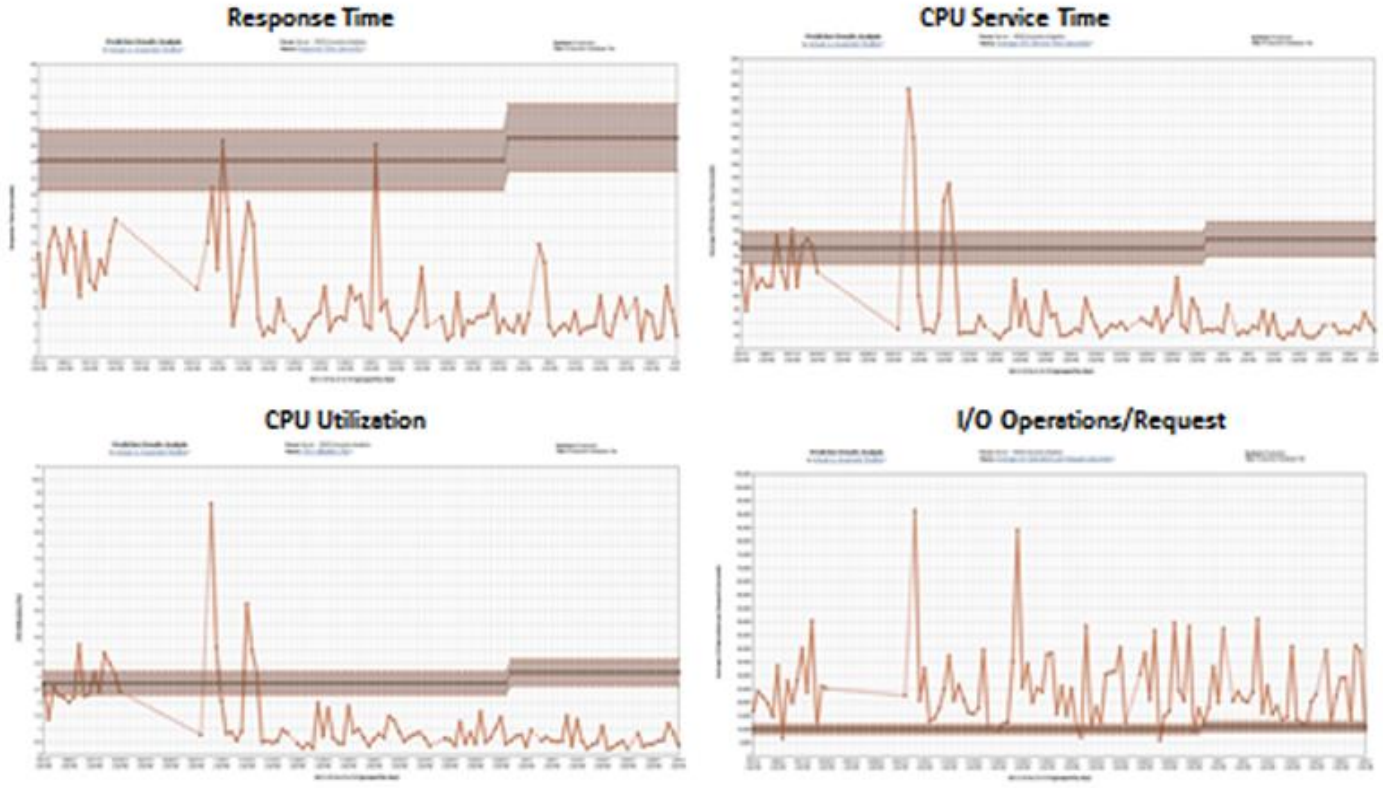

# Change of Workload Management rules and ML Algorithm will not be Sufficient to meet SLGs

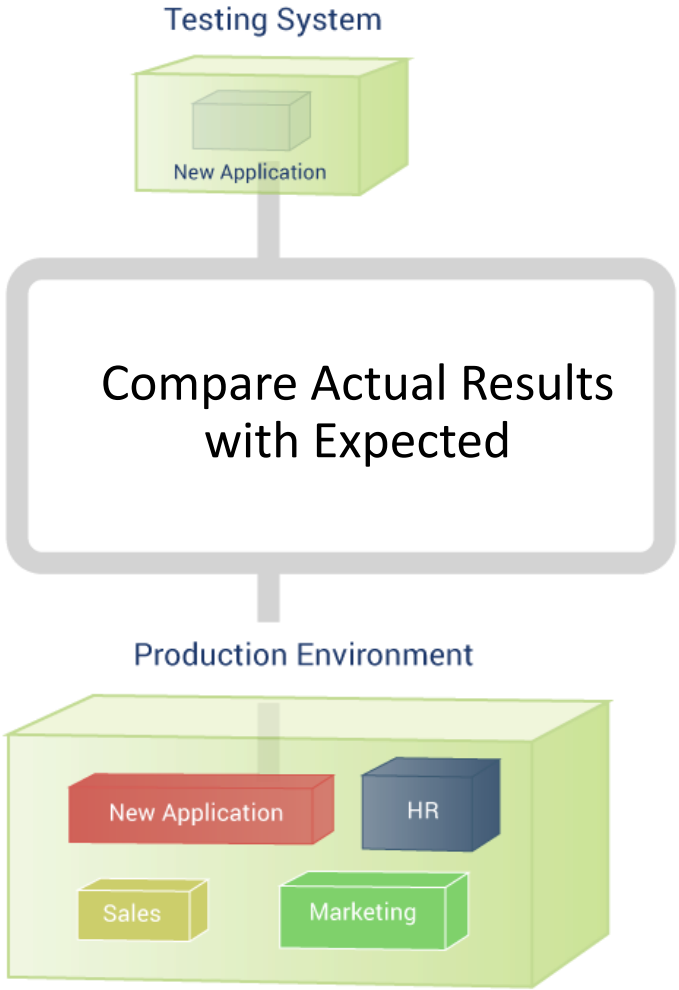


# Determine the Minimum Hardware Upgrade Required to Meet SLGs




# Verification of Results



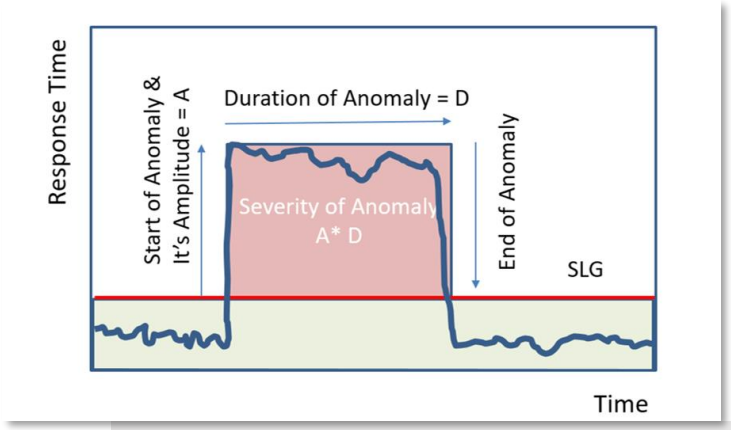



# Performance Assurance

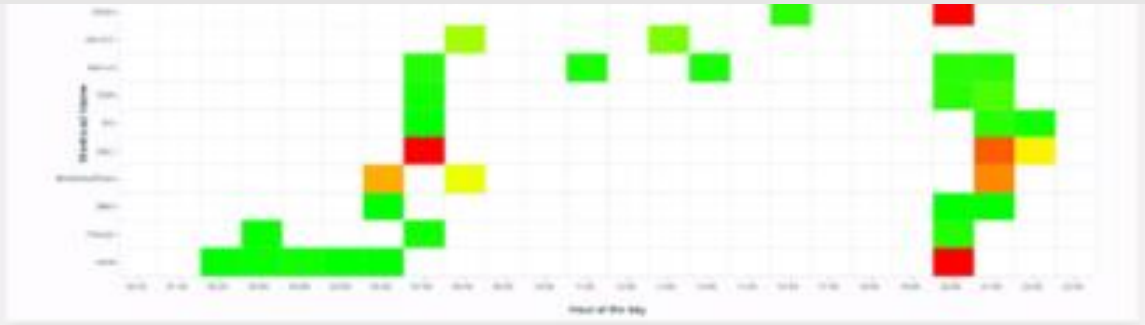
## Dynamic Performance Management and Workload Management Optimization Use Cases

# 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

| Workload | Parameter | # Anomalies | Duration(Hours) | | | | Severity | | | | | Growth Ratio | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | Max | Avg | STD | 98% | Sum | Max | Avg | STD | 98% | Anomalies | Total Severity |
| ADHOC | Response Time [sec] | 1,033 | 5 | 1.18 | 0.52 | 2 | 43.75 | 0.60 | 0.05 | 0.05 | 0.11 | -0.02 | 0.00 |
| BATCH | Response Time [sec] | 936 | 4 | 1.29 | 0.57 | 2 | 81.03 | 1.00 | 0.11 | 0.10 | 0.27 | -0.01 | 0.00 |
| Other | Response Time [sec] | 924 | 5 | 1.21 | 0.54 | 2 | 86.87 | 0.66 | 0.11 | 0.10 | 0.33 | 0.02 | 0.00 |
| FRAUD | Response Time [sec] | 742 | 8 | 1.31 | 0.70 | 3 | 43.63 | 0.76 | 0.08 | 0.09 | 0.24 | -0.04 | 0.00 |
| BATCHCORE | Response Time [sec] | 642 | 7 | 1.39 | 0.82 | 3 | 10.15 | 1.00 | 0.02 | 0.07 | 0.05 | -0.09 | 0.00 |
| UNICAAFF | Response Time [sec] | 608 | 4 | 1.23 | 0.54 | 2 | 15.21 | 0.51 | 0.03 | 0.04 | 0.08 | -0.03 | 0.00 |
| DSR | Response Time [sec] | 595 | 6 | 1.28 | 0.64 | 2 | 39.81 | 0.56 | 0.09 | 0.09 | 0.28 | -0.05 | 0.00 |
| ETL | Response Time [sec] | 577 | 10 | 2.32 | 1.71 | 5 | 68.25 | 1.00 | 0.27 | 0.15 | 0.56 | -2.12 | -0.30 |
| BMG | Response Time [sec] | 560 | 4 | 1.13 | 0.39 | 2 | 74.00 | 0.89 | 0.15 | 0.13 | 0.41 | -0.04 | 0.00 |
| SMART | Response Time [sec] | 560 | 4 | 1.22 | 0.56 | 3 | 38.59 | 0.57 | 0.08 | 0.08 | 0.26 | -0.04 | 0.00 |

Most Frequent Root Causes

| Root Cause | | | | # Anomalies Caused | Total Severity of Anomalies | Workloads affected |
|---|---|---|---|---|---|---|
| Workload | Parameter | User | Program | | | |
| ETL | Avg CPU Time [sec] | | | 394 | 55.33 | 1 |
| COGNOS_RPT | Avg CPU Time [sec] | | | 359 | 25.80 | 1 |
| DSR | Req/Hour [num] | SUNRISESLOWDASH02 | JDBC15.10.00.14;1.7.0_80 | 276 | 42.42 | 1 |
| BMG | Req/Hour [num] | ABSAPI_NSB | NQSSERVER | 266 | 37.59 | 1 |
| FRAUD | Avg I/O [num] | FELIX | JDBC12.00.00.110;1.8.0_51 | 258 | 34.90 | 1 |
| MonitoringTools | Avg CPU Time [sec] | BEZVISION | JDBC15.00.00.28;1.7.0_17 | 197 | 66.56 | 1 |
| MARS_ONLINE | Req/Hour [num] | | | 187 | 29.49 | 1 |
| EDW_DEV | Avg CPU Time [sec] | | | 167 | 22.23 | 1 |
| Other | Req/Hour [num] | UPSTART | JDBC14.00.00.21;1.8.0_25 | 126 | 8.12 | 1 |
| Admin | Avg I/O [num] | | | 111 | 33.62 | 1 |

Showing 1 to 10 of 210 entries Previous 1 2 3 4 5 ... 21 Next

# 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

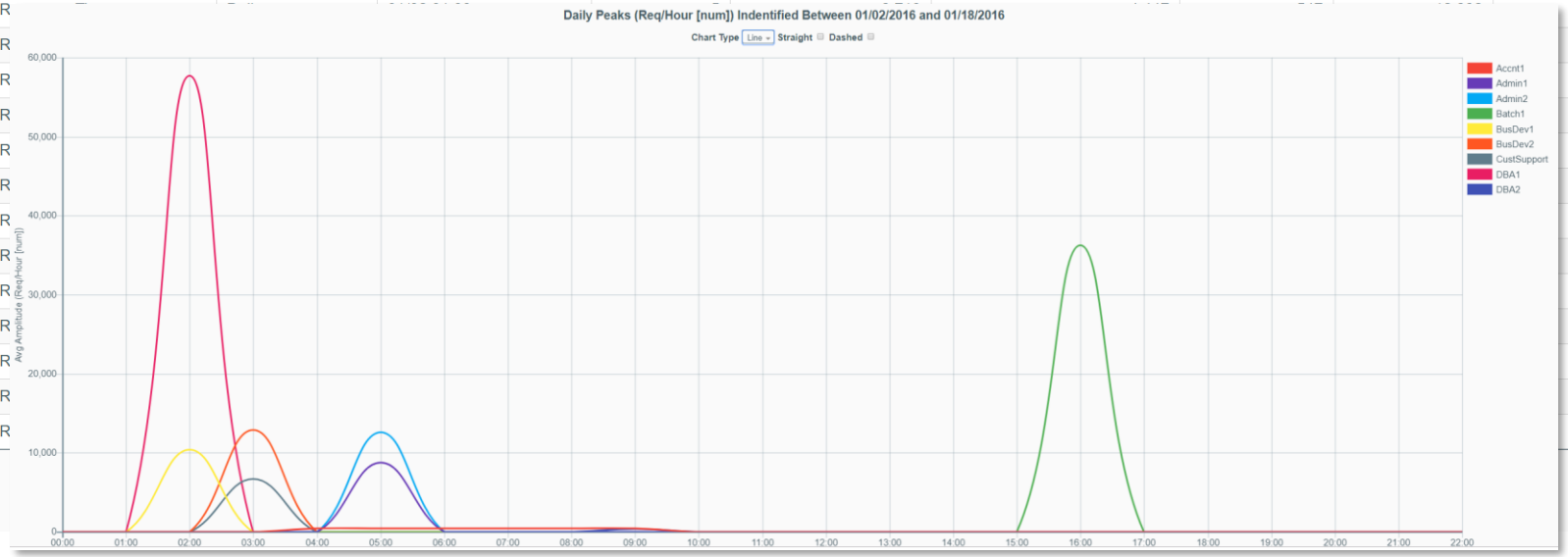
**Seasonal Peaks Determination Results**

**Response Time Seasonal Peaks indentified between 01/02/2016 and 01/21/2016**

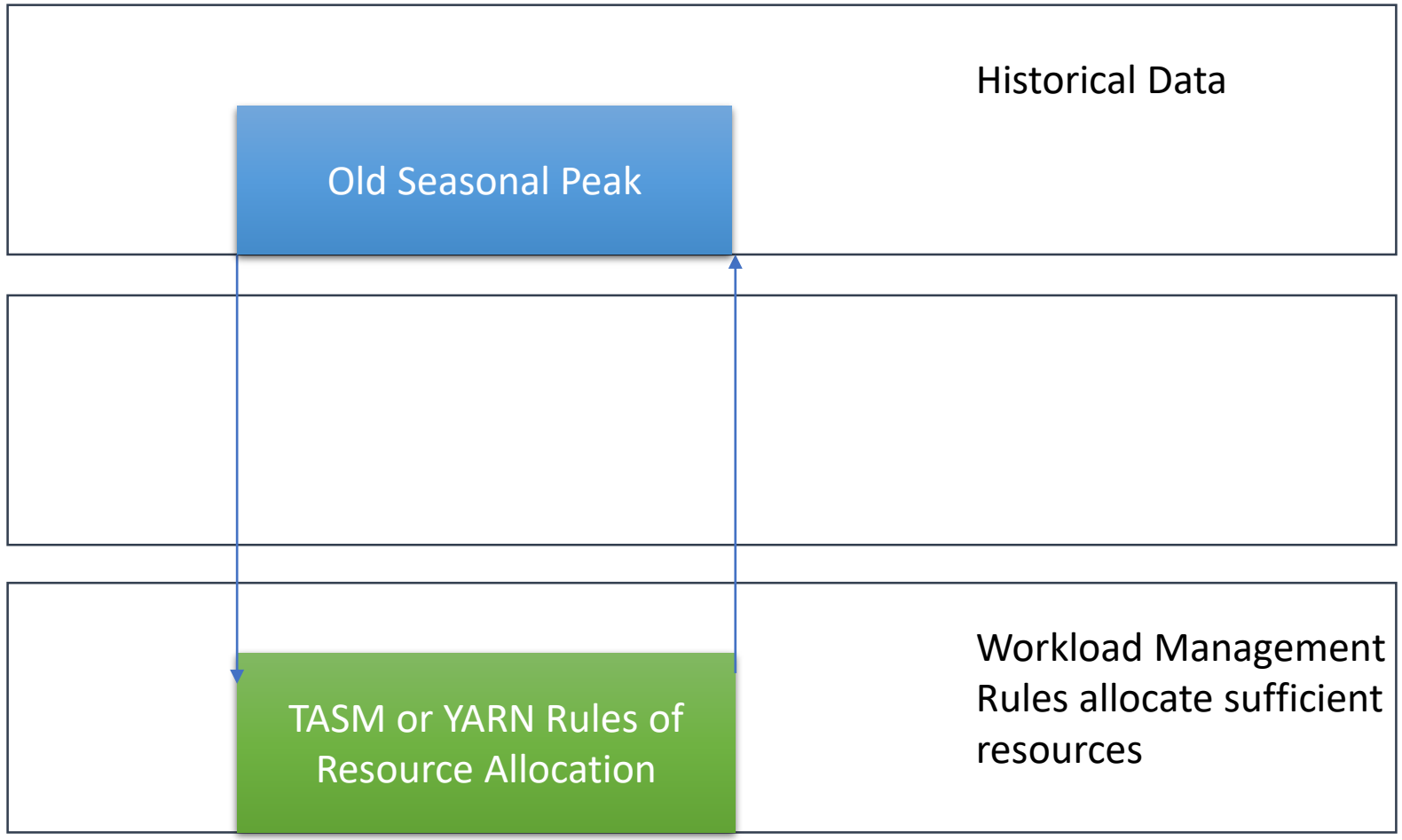
Show 25 entries  Search:

| Workload Name | Parameter Name | Peak Type | Peak Start Date | Duration | Avg Amplitude | Standard Deviation | Min Value | Max Value | 95 Percentile | Peak Length STD |
|---|---|---|---|---|---|---|---|---|---|---|
| Accnt1 | Response Time | Daily | 01/02 06:00 | 3 | 72,494 | 63,221 | 1,293 | 205,899 | 205,899 | 0.82 |
| Accnt2 | Response Time | Daily | 01/02 17:00 | 2 | 0 | 10,562 | 5,194 | 26,319 | 26,319 | 0.50 |
| Admin2 | Response Time | Daily | 01/02 19:00 | 3 | 41,674 | 17,793 | 4,305 | 68,988 | 53,798 | 1.00 |
| Dev1 | R | | | | | | | | 11,246 | 1.48 |
| HR4 | R | | | | | | | | 123,010 | 0.00 |
| Load0 | R | | | | | | | | 8,575 | 0.50 |
| Load1 | R | | | | | | | | 296,968 | 0.00 |
| Load3 | R | | | | | | | | 514,247 | 0.00 |
| Load4 | R | | | | | | | | 514,247 | 0.00 |
| Load7 | R | | | | | | | | 2,209,780 | 0.00 |
| Load8 | R | | | | | | | | 10,354 | 0.00 |
| Other | R | | | | | | | | 8,822 | 0.82 |
| Prod1 | R | | | | | | | | 300,900 | 0.00 |
| QA0 | R | | | | | | | | 90,131 | 0.00 |
| QA2 | R | | | | | | | | 142,853 | 0.50 |
| TechSupport2 | R | | | | | | | | 39,461 | 0.00 |

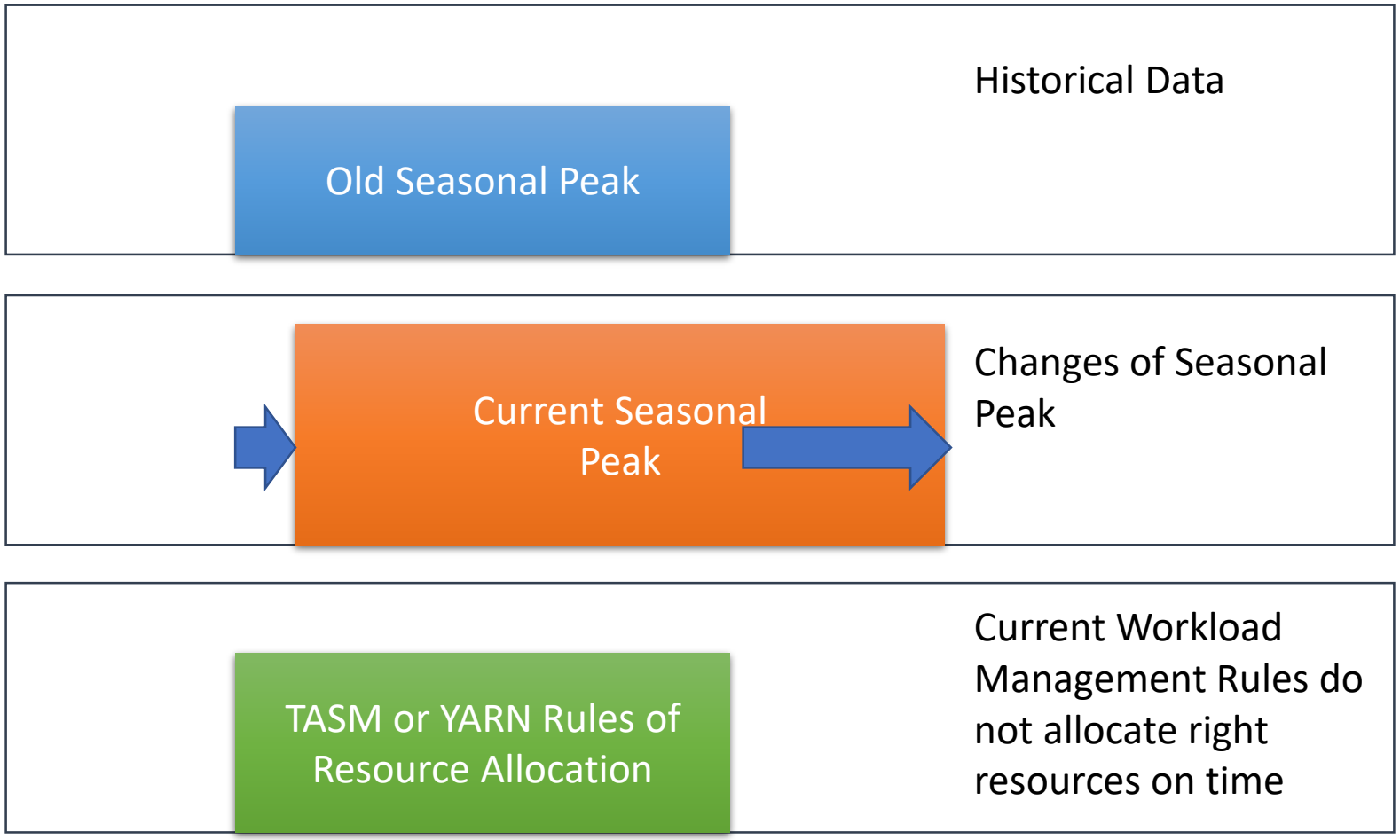
Showing 1 to 16 of 16 entries  Previous 1 Next



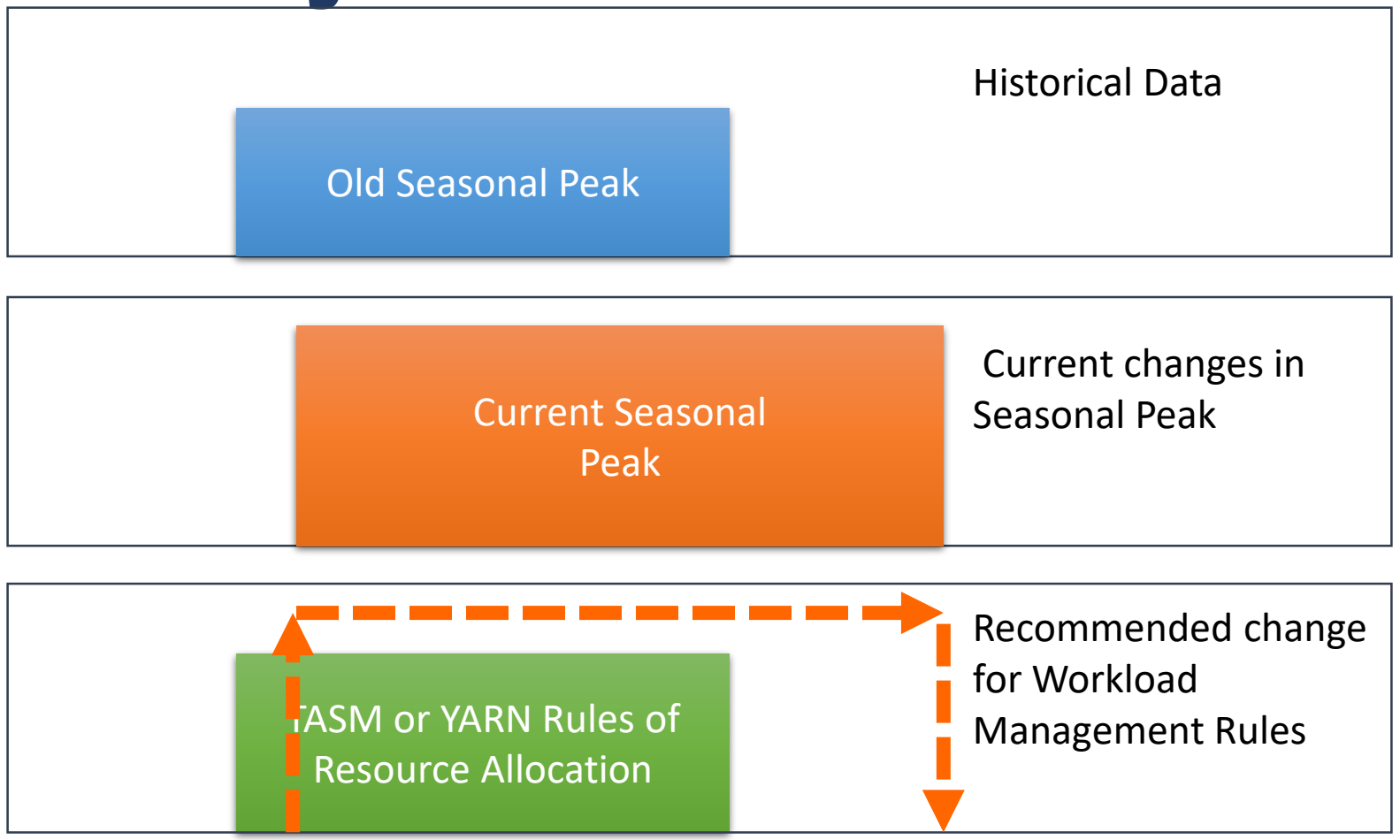
# Old seasonal peak and corresponding Workload Management rules



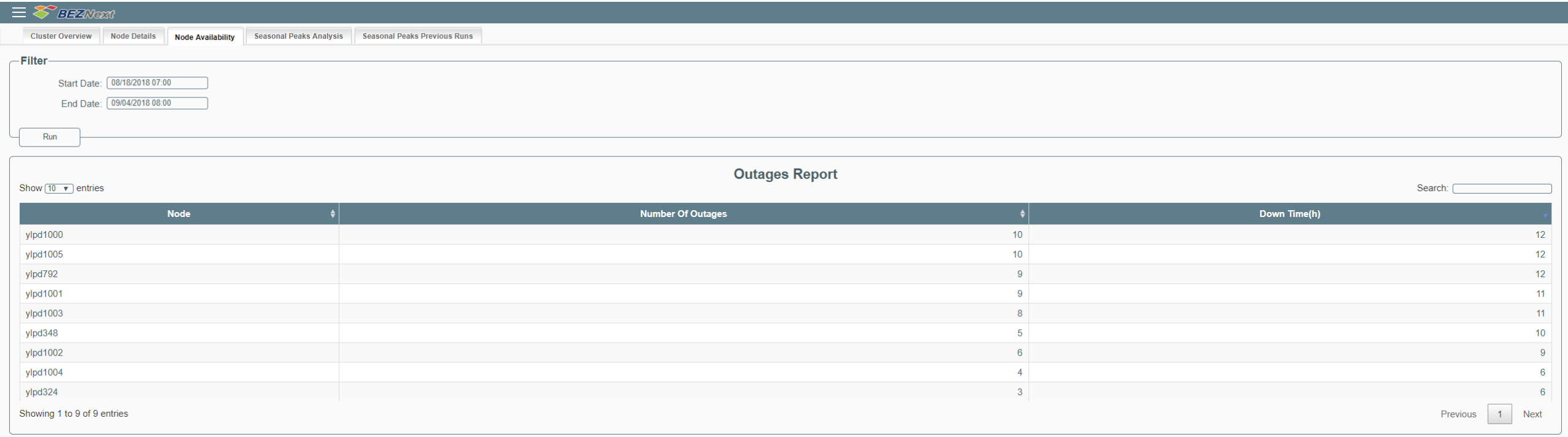
# Determine changes in seasonal peak



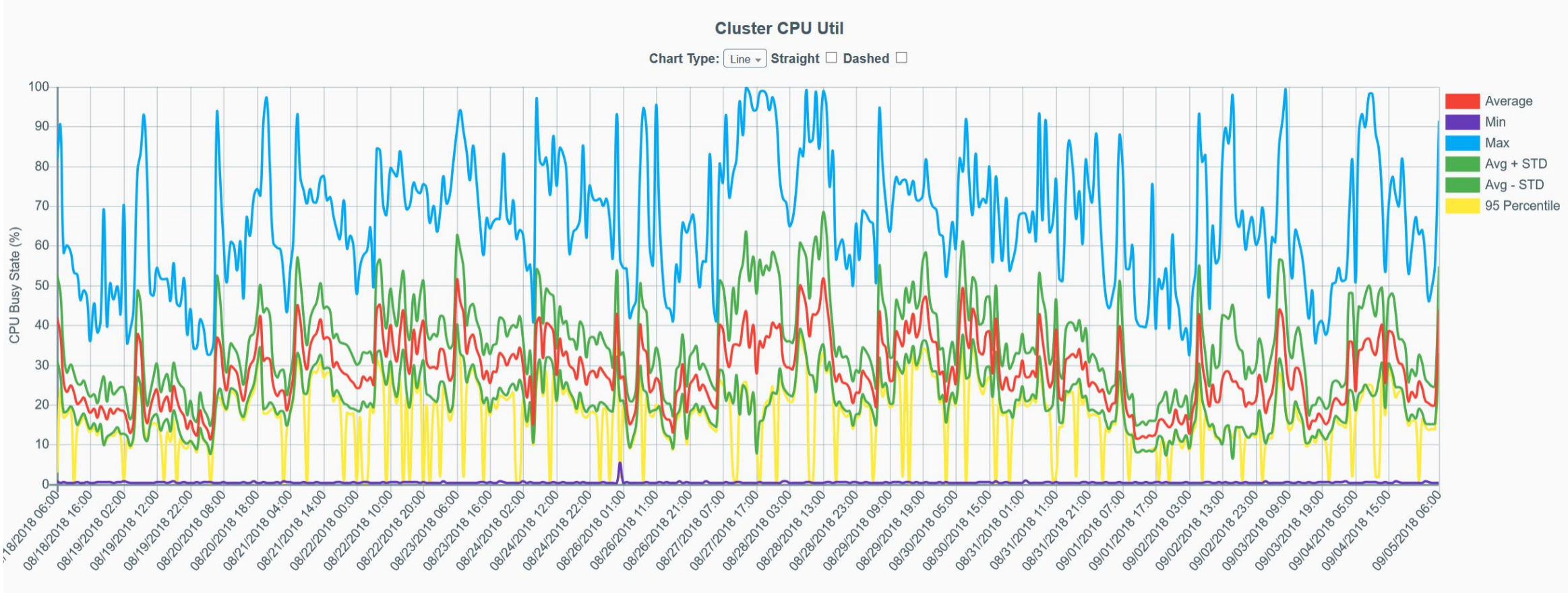
# Adjusting YARN and Kubernetes rules according to changes of workloads' seasonality



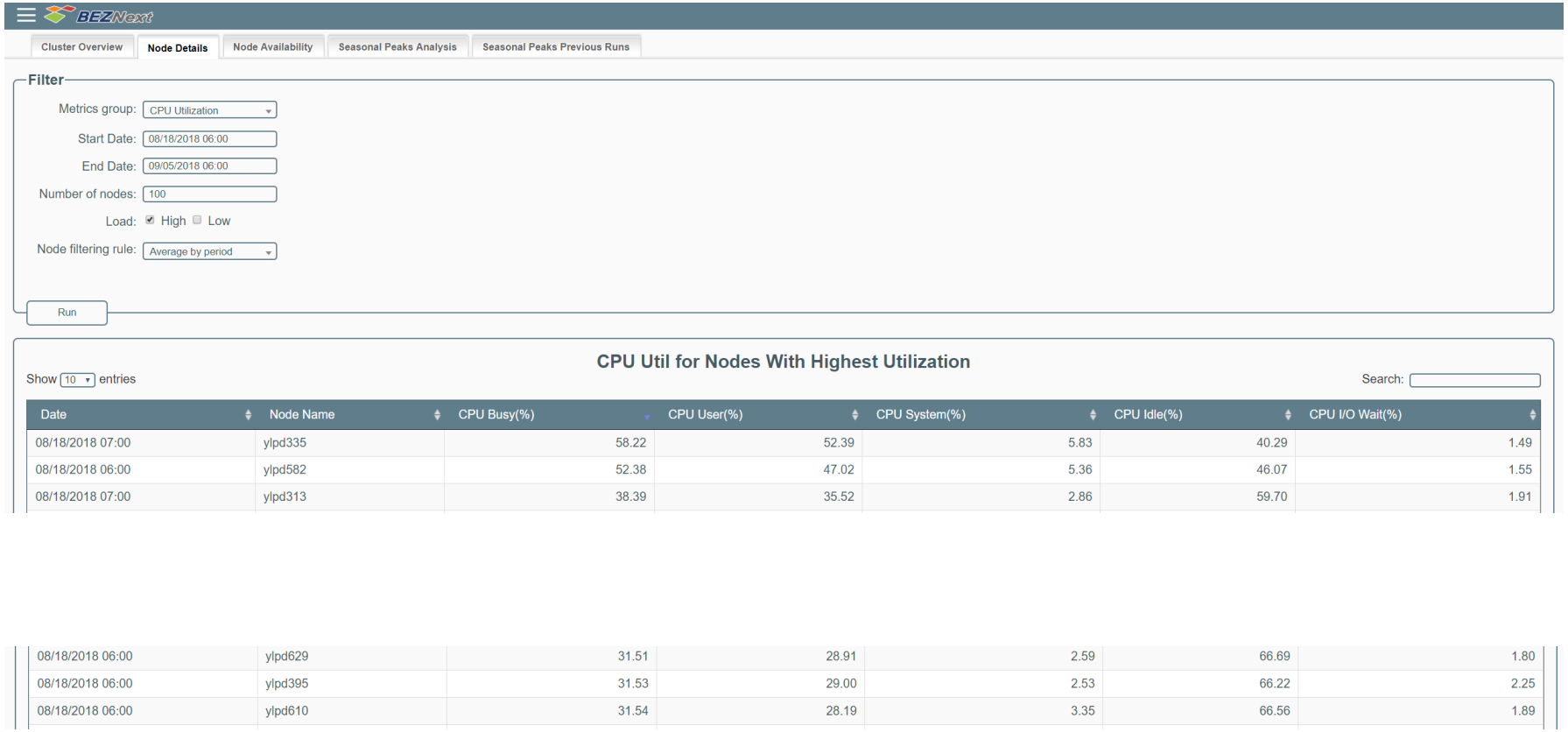
# Analysis Reliability of Big Data Cluster nodes



# 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

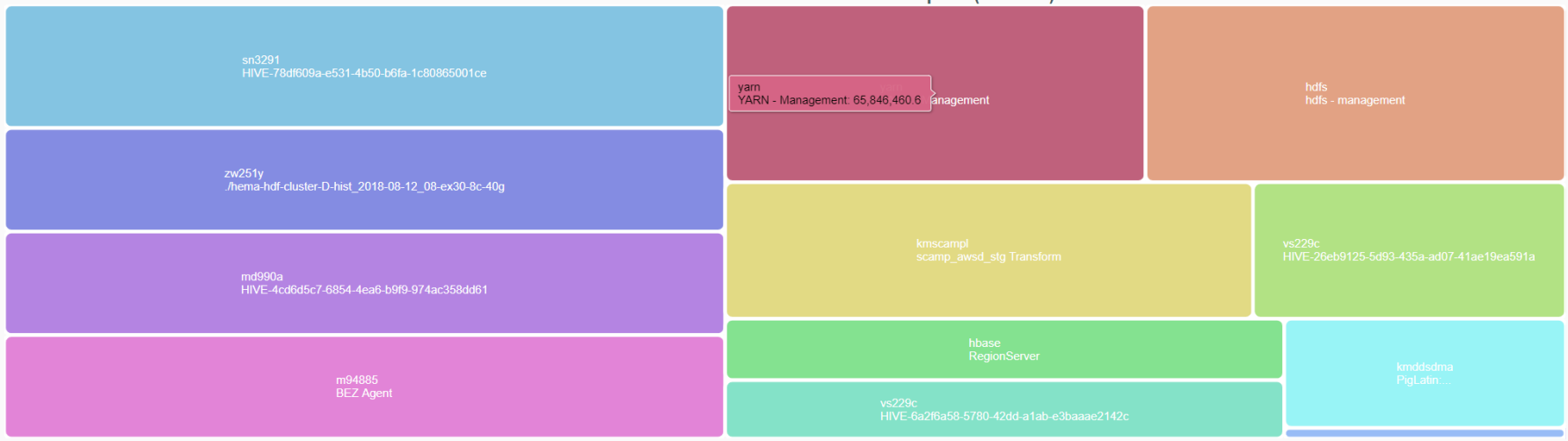
CPU Util by User & Application for Node ylpd313 During 08/18/2018 07:00

Show 10 entries Search:

| User Name | Application Name | CPU Time |
|---|---|---|
| dv203v | HIVE-899f5c61-791e-497a-bd19-d30c2fe04ed4 | 0.00 |
| dv203v | HIVE-1e49f2b0-18f2-4389-bee9-a052dd5457e5 | 0.00 |
| dv203v | HIVE-8f8ebc29-3942-498c-a33f-4532bbe9b823 | 0.00 |
| hbase | RegionServer | 30,297,223.65 |
| hdfs | hdfs - management | 65,846,460.60 |
| jd5648 | HIVE-eecc76a0-b004-45bb-8b18-47e3211c6578 | 0.00 |
| kmddscca | PigLatin:pig01_ddsccnt_daily_util_uverse_ban_details.pig | 0.00 |
| kmddscca | PigLatin:pig02_ddsccnt_daily_util_uverse_program_metadata_live_dvr.pig | 0.00 |
| kmddsdma | PigLatin:pig01_ddsdmsp_scamp_hourly_subs_usage_vol.pig | 27,166,300.52 |
| kmddsdma | PigLatin:pig03_ddsdmsp_scamp_hourly_sms_usage.pig | 2,640,670.42 |

Showing 1 to 10 of 30 entries Previous 1 2 3 Next

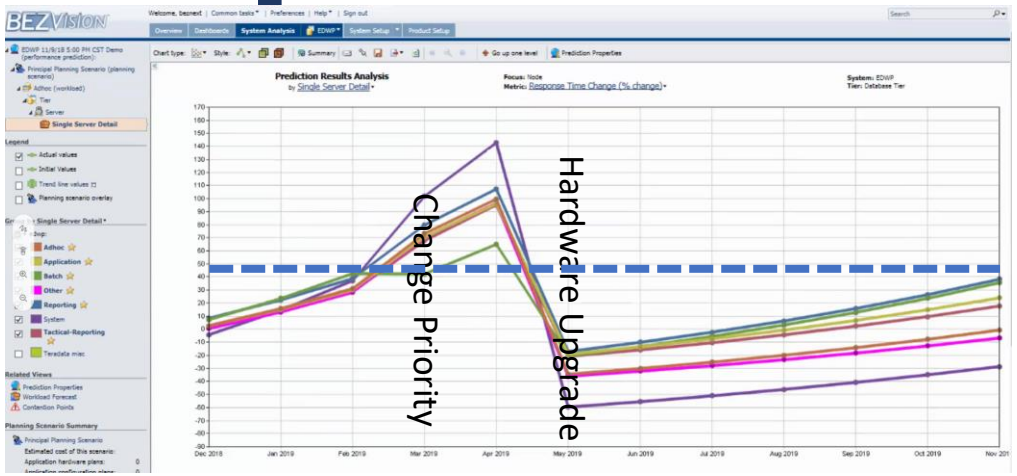
Share of Each Process in Total Resource Consumption (CPU Util)

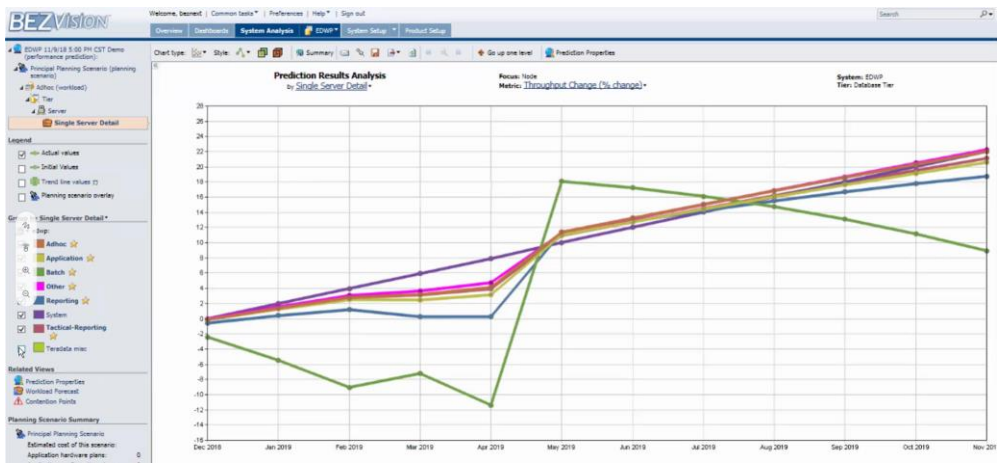


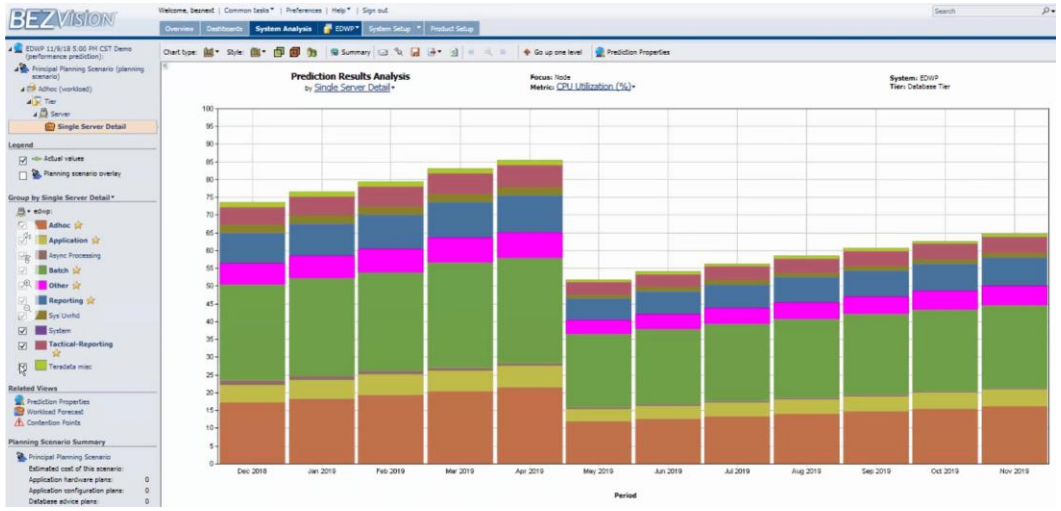
# Performance Assurance
## Strategic Capacity Planning Use Cases

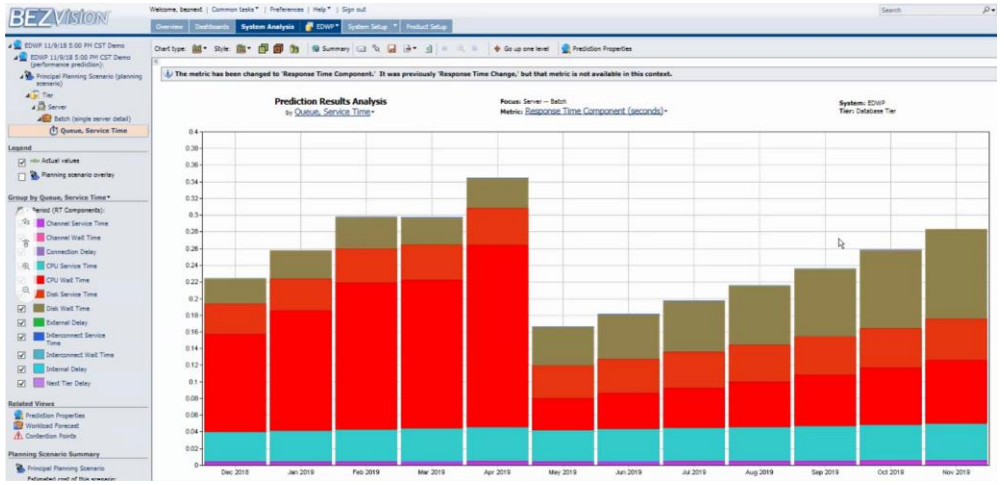
# Determining the minimum hardware upgrade required to meet SLGs

# Determining proactive changes required to meet SLGs



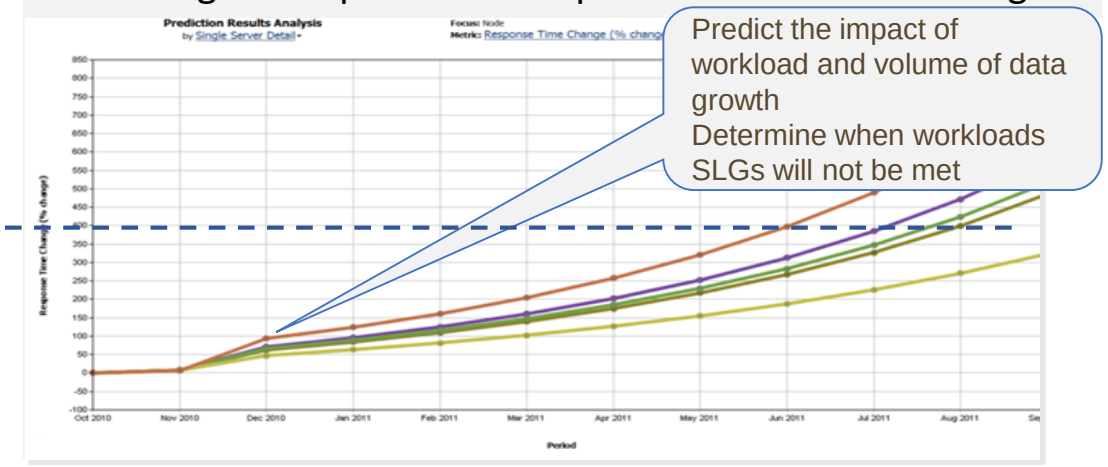




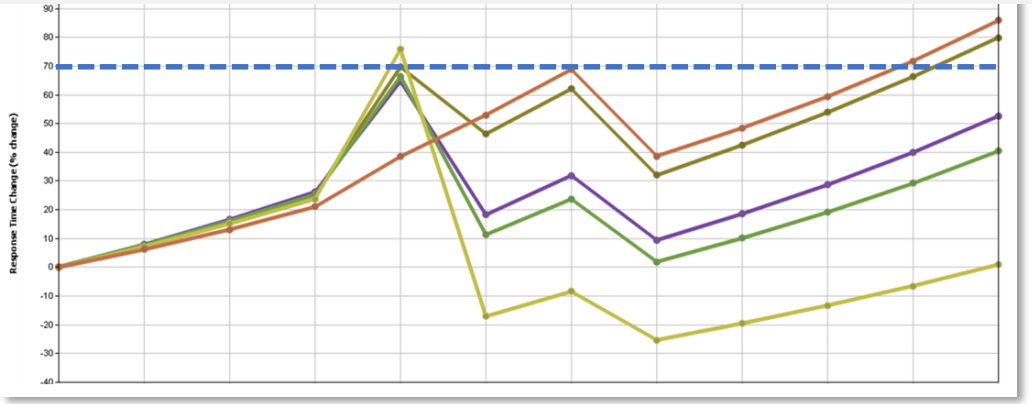
# Organizing continuous proactive capacity management process

- 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

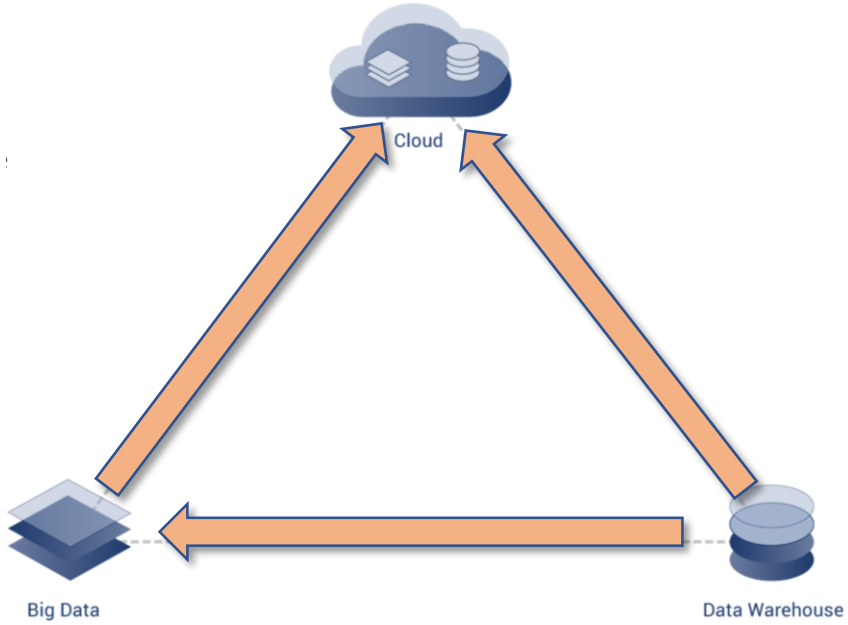


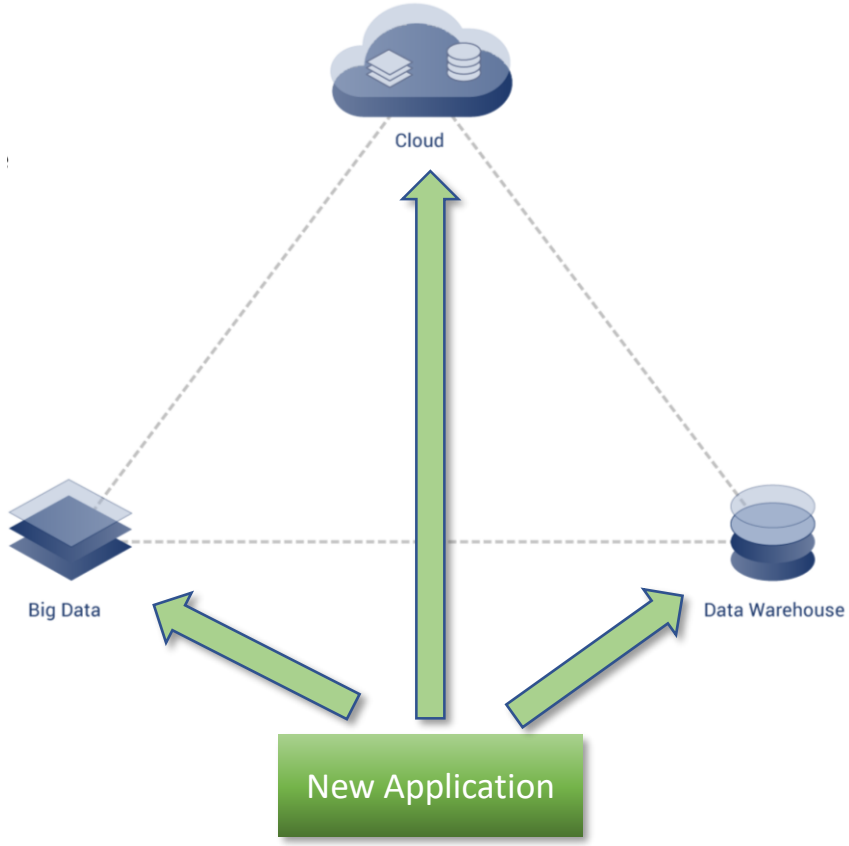

Recommended Actions Plan



# Predict how workloads' consolidation and move to Intellicloud will affects performance

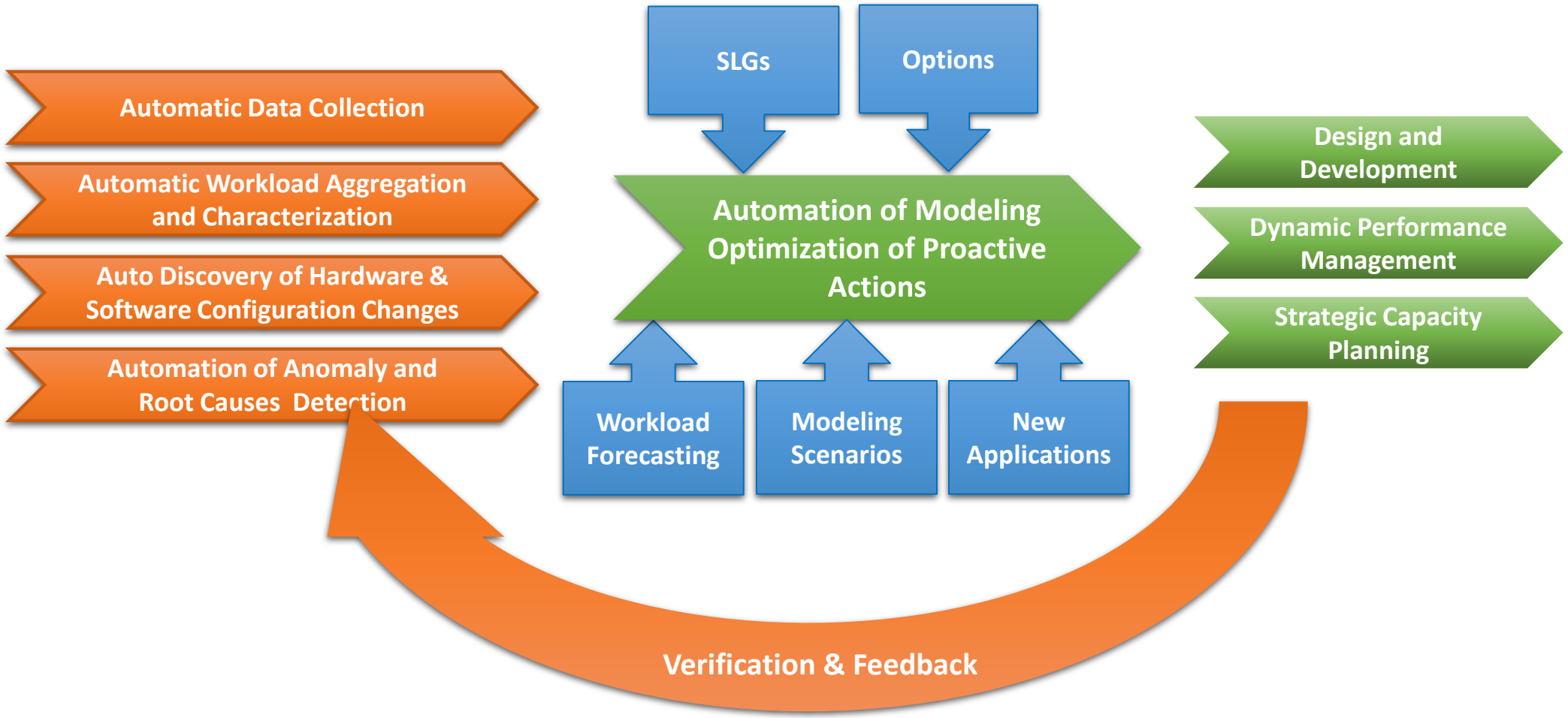
# What will be an impact of moving workloads to different platform



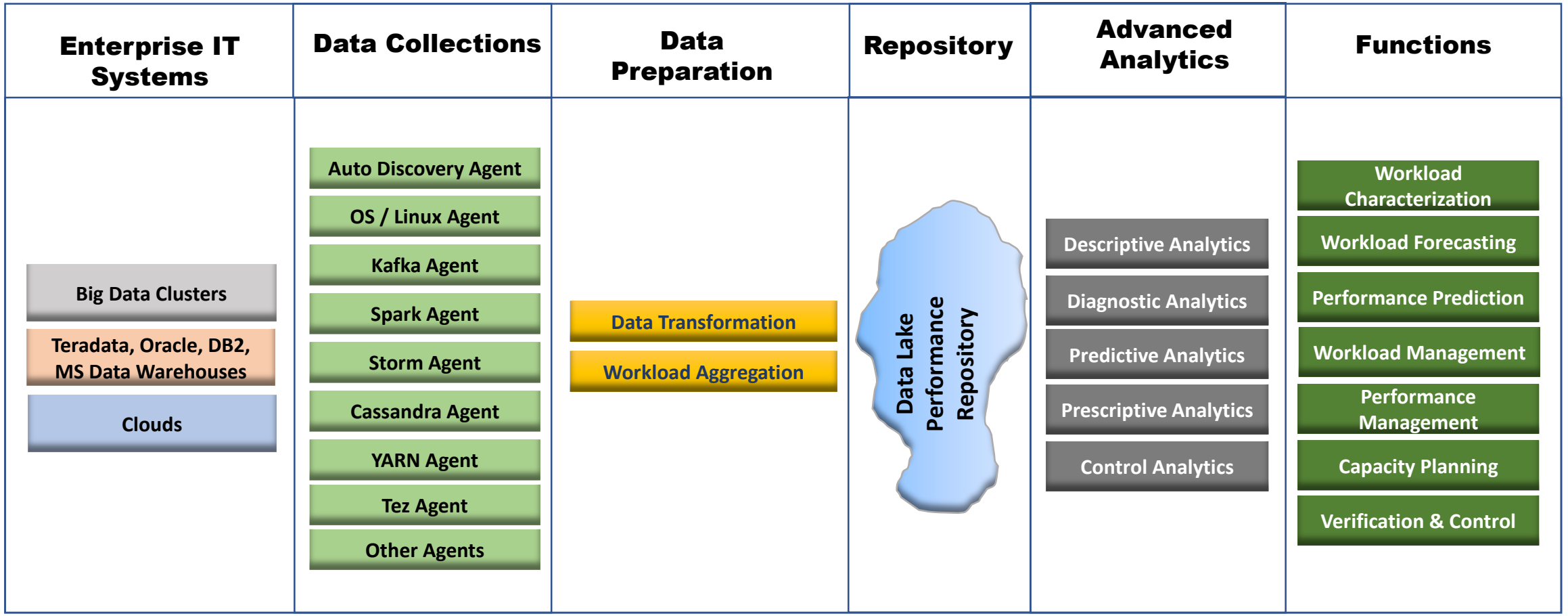
# Automatic Continuous Performance Assurance Control

## Apply ML, AI and QNM Models, Optimization & Automation




BEZNext

# BEZNext Performance Assurance Technology



# 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?