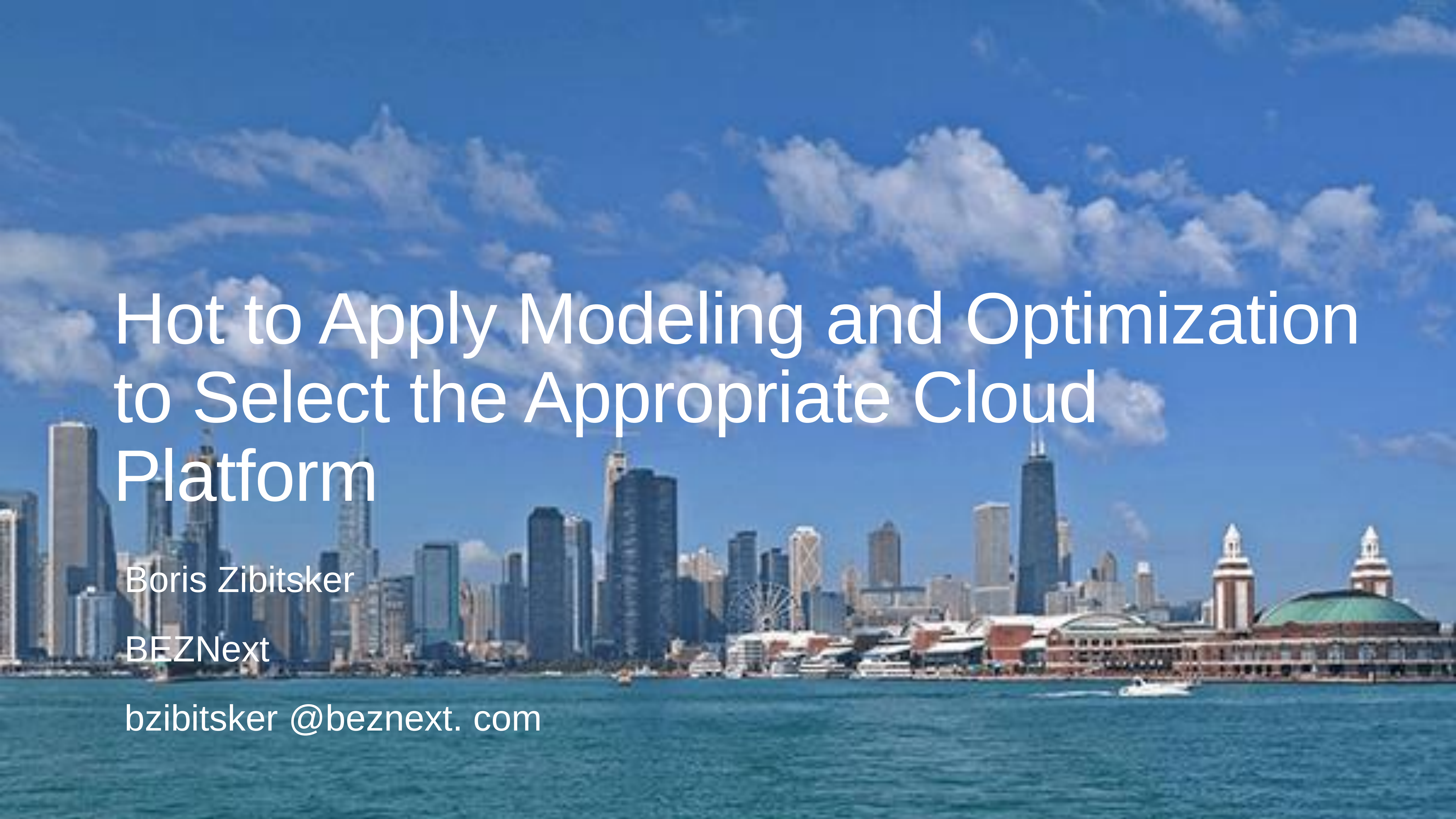


Hot to Apply Modeling and Optimization to Select the Appropriate Cloud Platform

Boris Zibitsker

BEZNext

bzibitsker @beznext. com

About Speaker

Boris Zibitsker

Dr. Boris Zibitsker is a CEO of BEZNext. His focus is on the development of performance assurance, performance engineering, dynamic performance management and long-term capacity planning software tools for big data, data warehouse and cloud applications.

He is a member of SPEC Big Data Research Group.

Boris consults with many Fortune 500 companies, and he manages Capstone projects for graduate students in MS in Analytics at University of Chicago.

Boris a Honorable Doctor of BGUIR and during last 5 years he was a co-chairman of Big Data Advanced Analytics Conference.

BEZNext

Optimizing Business and IT

TOP
Cloud Based Planning
Solution Providers
2020
Recognized by
CIO
APPLICATIONS

Abstract

Organizations want to take advantage of the flexibility and scalability of Cloud platforms. By migrating to the Cloud, they hope to develop and implement new applications faster with lower cost. Amazon AWS, Microsoft Azure, Google, IBM, Oracle and others Cloud providers support different DBMS like Snowflake, Redshift, Teradata Vantage, and others. These platforms have different architecture, mechanism of allocation and management of resources, and sophistication of DBMS optimizers which affect performance, scalability and cost. As a result, the response time, CPU Service Time and the number of I/Os for the same query, accessing the similar table in the Cloud could be significantly different than On Prem.

In order to select the appropriate Cloud platform, we use a modeling and optimization. First, we perform a Workload Characterization for On Prem Data Warehouse. Each Data Warehouse workload represents a specific line of business and includes activity of many users generating concurrently simple and complex queries accessing data from different tables. Each workload has different demand for resources and different Response Time and Throughput Service Level Goals.

- In this paper we will review results of the workload characterization for On Prem Data Warehouse environment.
- Secondly, we must collect measurement data for standard TPC-DS benchmark tests performed in AWS Vantage, Redshift and Snowflake Cloud platform for different sizes of the data sets and different number of concurrent users.
- During third step we use the results of the workload characterization and measurement data collected during the benchmark to modify BEZNext On Prem Closed Queueing model to model individual Clouds.
- And finally, during the fourth step we use the Model to take into consideration differences in concurrency, priorities and resource allocation to different workloads. BEZNext Capacity Planning optimization algorithms incorporate Graduate search mechanism to find the AWS instance type and minimum number of instances which will be required to meet SLGs for each of the workloads. Publicly available information about the cost of the different AWS instances is used to predict the cost of supporting workloads in the Cloud month by month during next 12 months.

Outline

- Introduction to Cloud Selection and Performance Assurance
- Data Collection and Workload Characterization
- Predicting Minimum Cloud Configurations Required to meet SLGs
- Predicting Cost
- Performance Assurance of Data Warehouse Workloads in the Cloud
- Summary

Introduction

BEZNext Performance Assurance Software and Services

Enterprise Wide

Entire Life Cycle

<table>
<tr><th></th><th>Data Warehouse</th><th>Big Data</th><th>Cloud</th></tr>
<tr><td><b><i>Build</i></b>
<b>Performance Engineering during Building New Applications and Infrastructure</b></td><td colspan="3">• How to be sure that new applications will perform well?
• How to predict the impact of new application implementation on performance of existing workloads?
• How to develop proactive recommendations to application developers and operations during DevOps process?
• Does existing platform has enough capacity to support both new and existing applications?
• How to select appropriate On Prem or Cloud platform for new application?</td></tr>
<tr><td><b><i>Manage</i></b>
<b>Dynamic Performance Management and Workload Management Optimization for Production Environment</b></td><td colspan="3">• How to determine most frequent and severe Anomalies and root causes?
• How to determine workloads seasonality and proactively modify Workload Management rules?
• How to determine applications availability and provide actional information to application developers and operations
• How to optimize workload management?</td></tr>
<tr><td><b><i>Grow</i></b>
<b>Strategic Capacity Planning</b></td><td colspan="3">• How to predict the impact of the workload and volume of data growth and determine the most effective measures required to meet Service Level Objectives with lowest cost
• Does existing platform has enough capacity to support both new and existing applications?
• How to select appropriate Cloud platform for Data Warehouse and Big Data applications?</td></tr>
</table>

Criteria of Cloud Platform Selection

MULTIPLE CRITERIA

Performance

Scalability

Cost

Security

Elasticity

Deployment Flexibility

Ecosystem Integration

Database Management

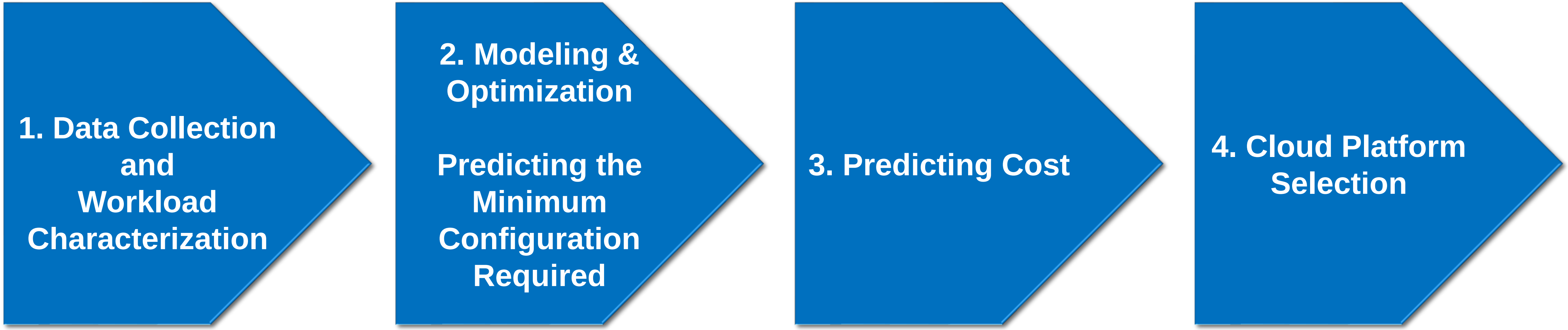
Analytic and Database Functionality

WE WILL FOCUS ON

Performance and

Cost

Major Steps



1. Data Collection and Workload Characterization

2. Modeling & Optimization

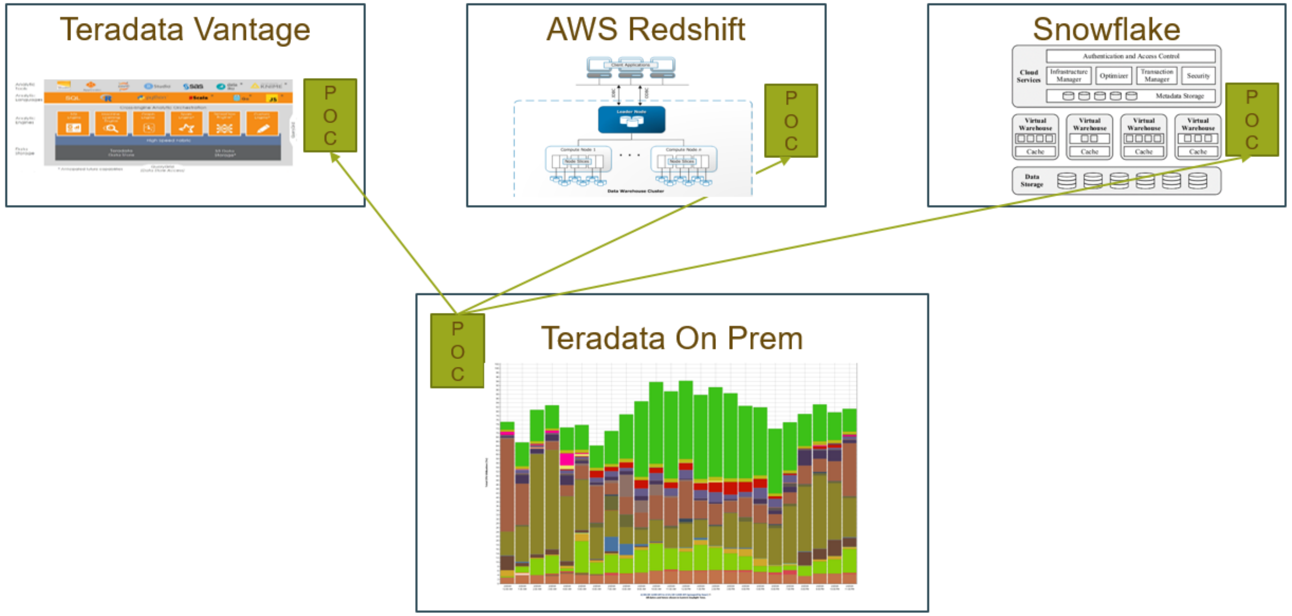
Predicting the Minimum Configuration Required

3. Predicting Cost

4. Cloud Platform Selection

Data Collection and Workload Characterization On Prem and in the Cloud

Data Collection On Prem and Cloud Platforms



Teradata Vantage

POC

AWS Redshift

POC

Snowflake

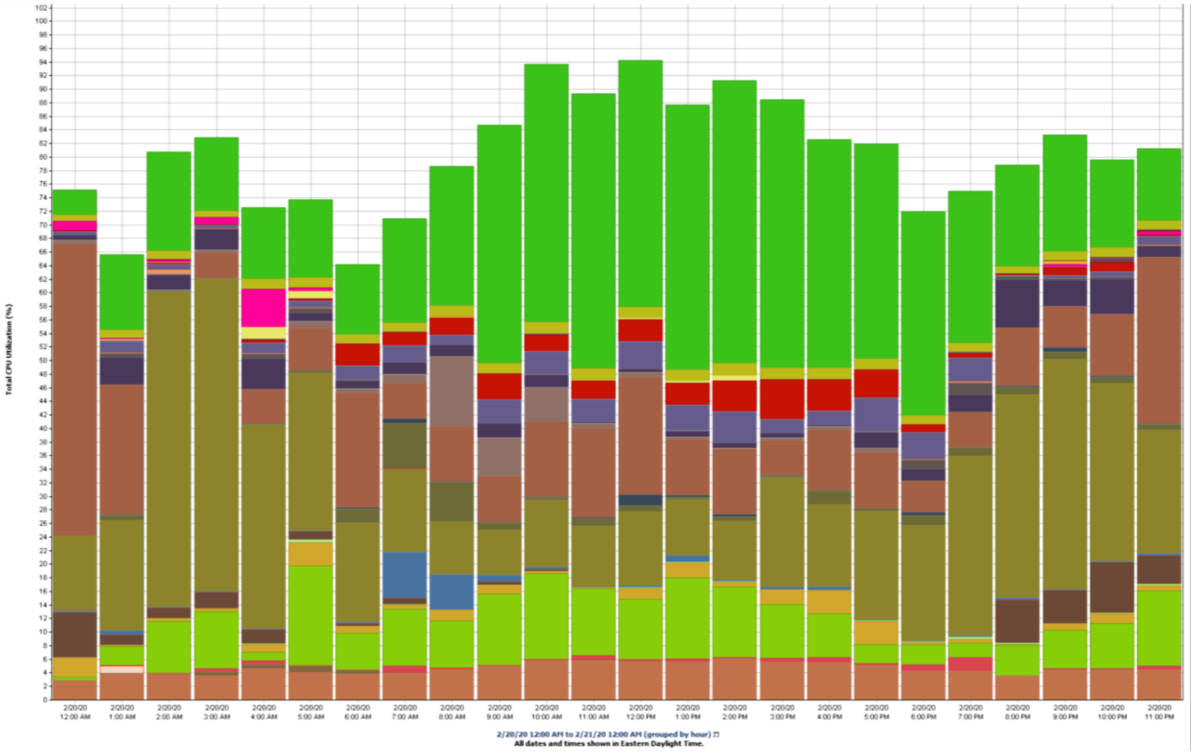
POC

POC

Teradata On Prem

- Production workload On Prem
- POC Benchmark with representative Queries on each Platform
- Standard TPC DS Benchmark on each Platform

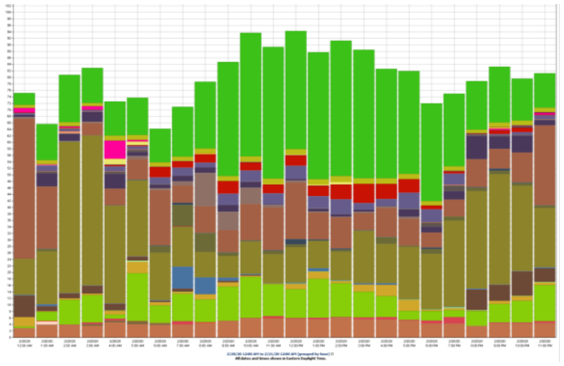
CPU Utilization by Production On Prem Workloads During 24 Hours

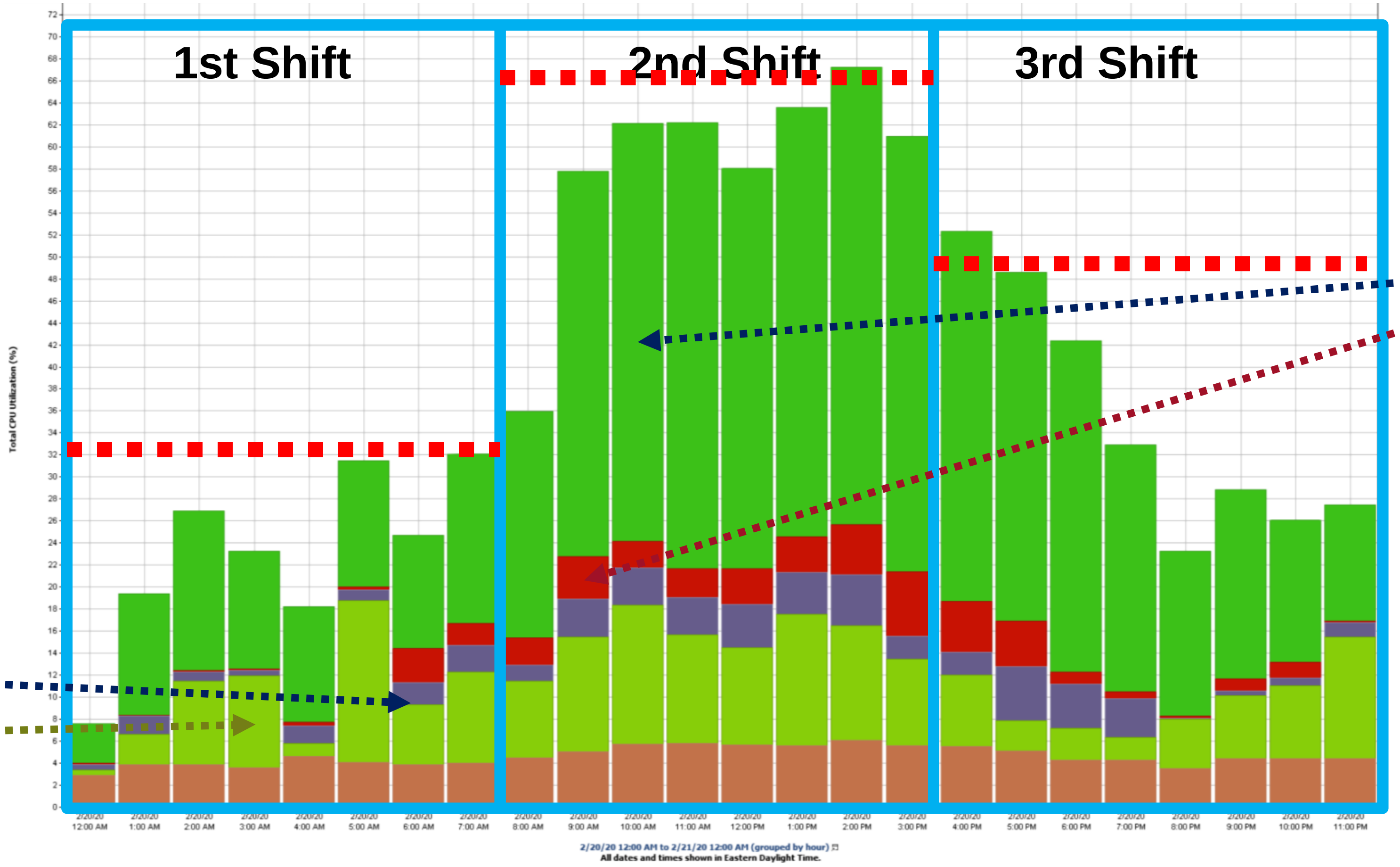


Hourly Profiles for Each Workload are built during Workload Characterization:

- Performance Profile
- Resource Consumption Profile
- Data Usage Profile

CPU Utilization by Workloads Selected for Cloud





Sales

Finance

Marketing

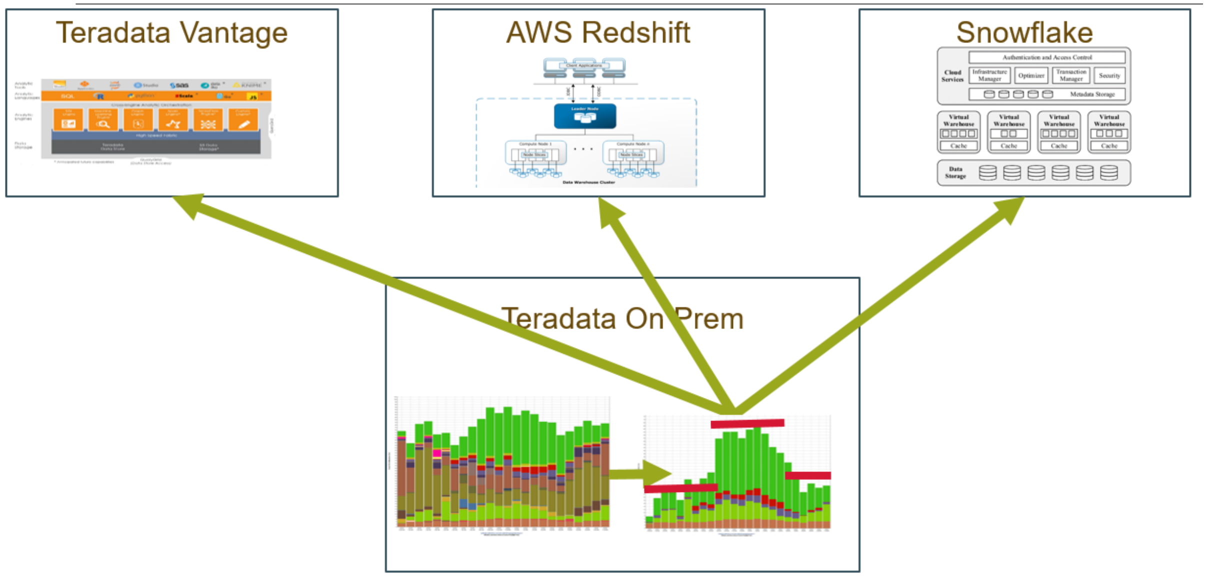
BI

Applying Modeling and Optimization to Determine the Minimum Configuration Required to meet SLG for each Workload

Modeling Approach

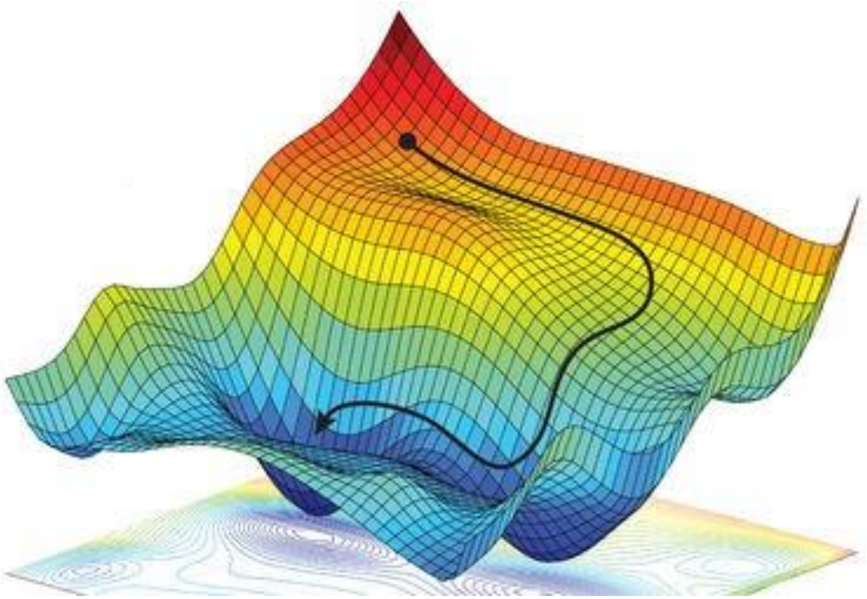
On Prem Queueing Network Model for MPP Data Warehouse reflects Hardware, Software configurations, Workload Management parameters and workload characterization results

Cloud Models are built by modifying parameters of On Prem models to reflect specific architecture of each Cloud DBMS platform and results of the benchmarks



Determining the Minimum Cloud Configuration Capable Meeting SLGs

GRADIENT OPTIMIZATION

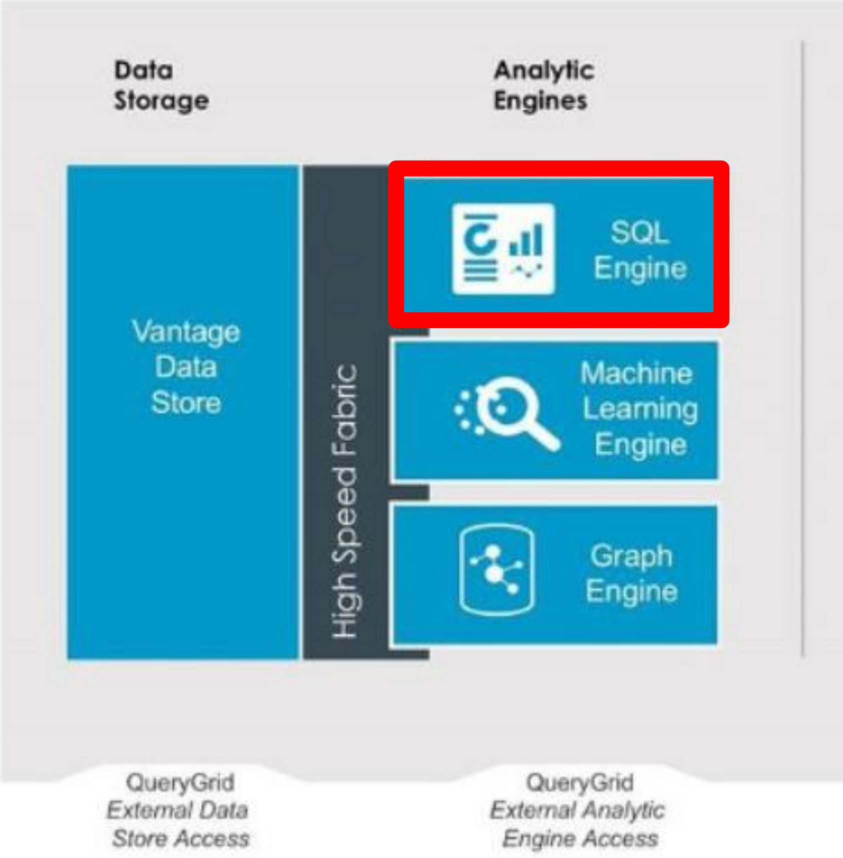


APPROACH

1. Apply Gradient Optimization to Workload Management
2. Use modeling to predict which workload will violate SLGs the most
3. Determine which resource will be the bottleneck
4. Apply Gradient method and iterative modeling to find the minimum hardware configuration required to meet SLGs for all workloads

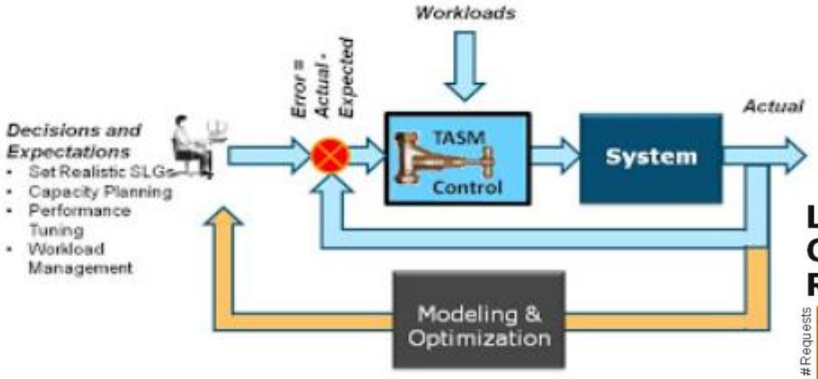
Predicting performance of Data Warehouse workloads in Teradata Vantage environment

- AWS Instance selection
- Limited scalability, but sophisticated optimizer and workload management



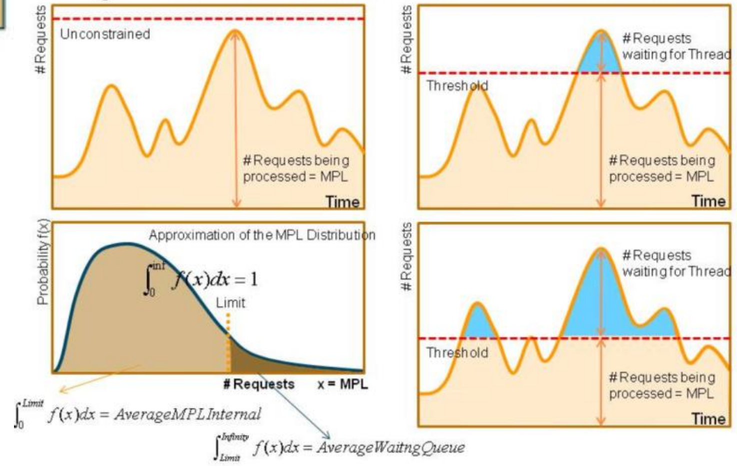
Vantage Workload Management Optimization

OPTIMIZATION OF CAPACITY PLANNING, PERFORMANCE MANAGEMENT AND WORKLOAD MANAGEMENT

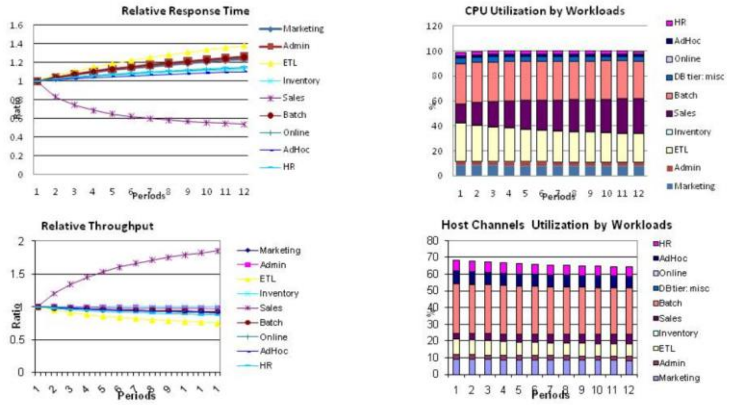


Decisions and Expectations
- Set Realistic SLGs
- Capacity Planning
- Performance Tuning
- Workload Management

LIMIT CONCURRENCY REDUCE CONTENTION BUT INCREASE the # of REQUESTS WAITING for the THREAD



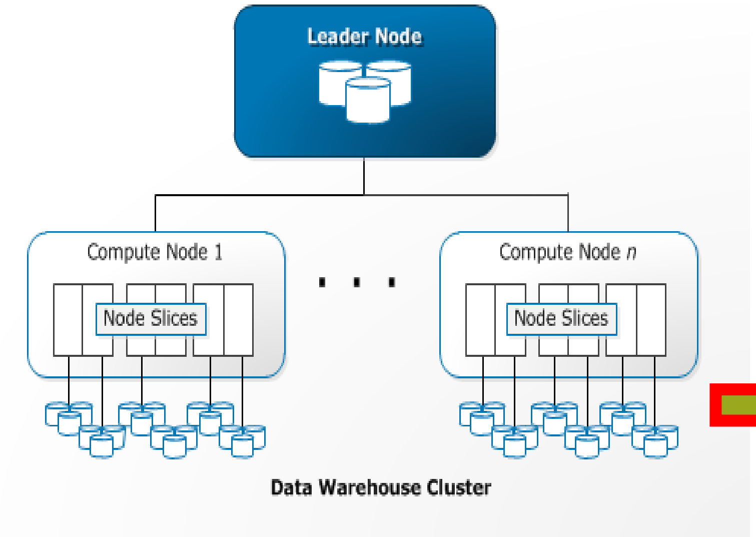
PREDICTED IMPACT OF INCREASING PRIORITY FOR SALES



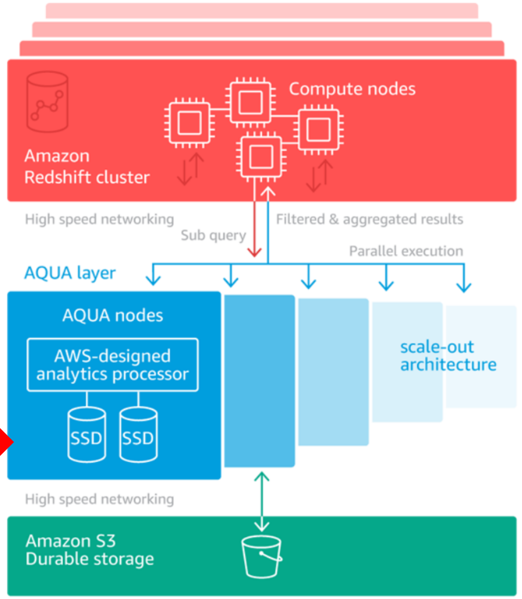
The same approach can be applied to predict the Impact of Changing Classifications and answer questions like: What if Users SQL required less than 1 sec CPU Time will be Processed with Higher Priority…

Predicting performance of Data Warehouse workloads in AWS Redshift environment

- Limited number of nodes/instances in Redshift cluster
- Future release of Redshift will use AQUA to accelerate Redshift queries by running data intensive tasks such as filtering and aggregation, compression and others closer to the storage layer.
- We did not model the impact of AQUA on Redshift performance



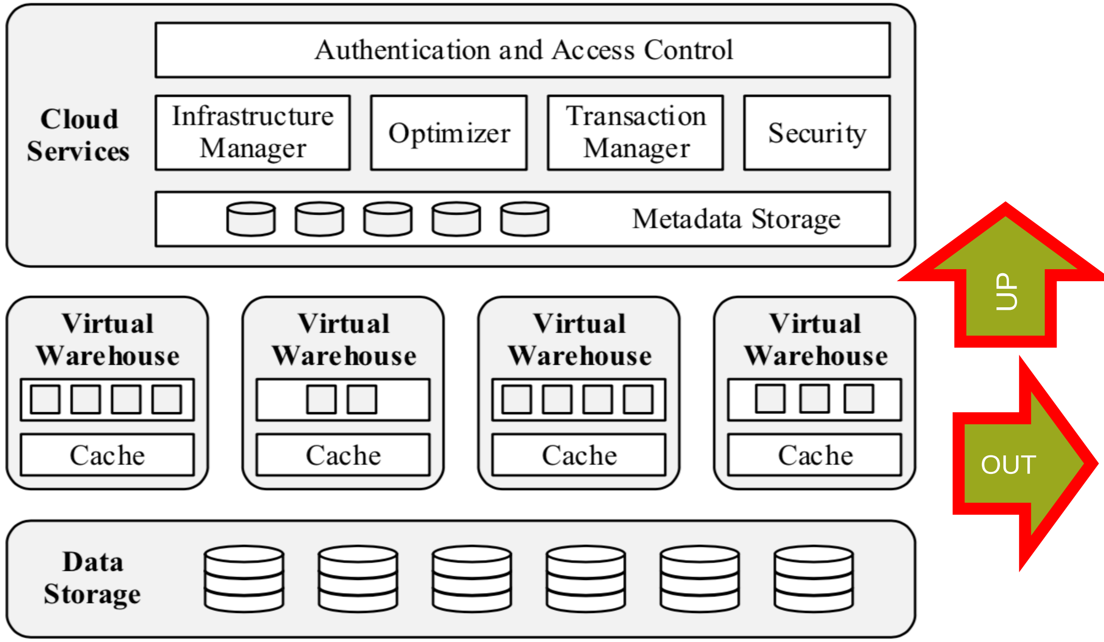
Current Redshift Architecture



Advanced Query Accelerator (AQUA) for Amazon Redshift

Predicting performance of Data Warehouse workloads in Snowflake environment

- Snowflake automatically scales out and scale up
- First scenario - running each workload in dedicated Virtual Warehouse
- Second scenario - running all workloads in one Virtual Warehouse



Examples of Modeling and Optimization Predicting Min Configurations Required to meet SLGs for Each Cloud, 2nd shift during next 10 Month

MIGRATE 4 WORKLOADS TO THE CLOUD

EXPECTED GROWTH IN NUMBER OF USERS - 12% ANNUALLY

EXPECTED GROWTH IN VOLUME OF DATA PROCESSED - 10% ANNUALLY

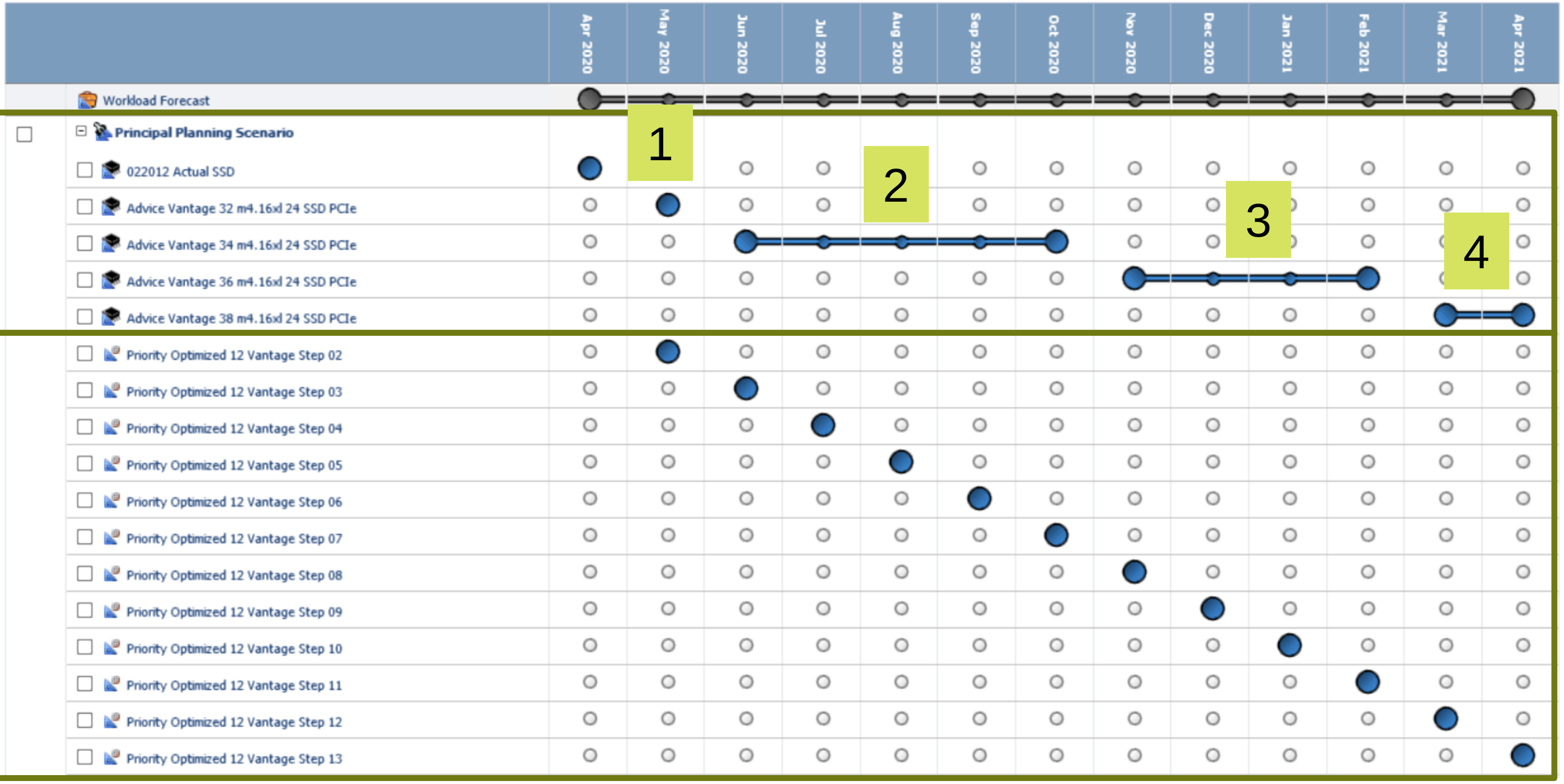
Vantage 2nd Shift / Month

MIGRATE 4 WORKLOADS INTO ONE VIRTUAL WAREHOUSE:

Vantage: Recommended minimum configuration for 2nd shift during next 12 months

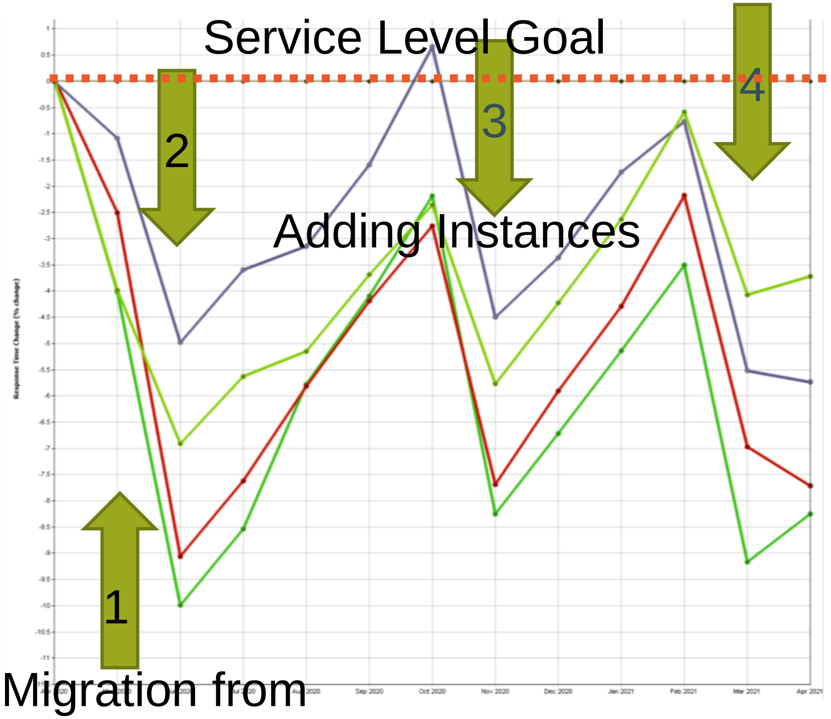
Configuration

Workload Management



	Apr 2020	May 2020	Jun 2020	Jul 2020	Aug 2020	Sep 2020	Oct 2020	Nov 2020	Dec 2020	Jan 2021	Feb 2021	Mar 2021	Apr 2021
Workload Forecast	●	●	●	●	●	●	●	●	●	●	●	●	●
□ ⊟ Principal Planning Scenario													
□ 022012 Actual SSD	●		○	○		○	○	○	○	○	○	○	○
□ Advice Vantage 32 m4.16xl 24 SSD PCIe	○	●	○	○		○	○	○	○		○	○	○
□ Advice Vantage 34 m4.16xl 24 SSD PCIe	○	○	●	●	●	●	●	○	○		○		○
□ Advice Vantage 36 m4.16xl 24 SSD PCIe	○	○	○	○	○	○	○	●	●	●	●		○
□ Advice Vantage 38 m4.16xl 24 SSD PCIe	○	○	○	○	○	○	○	○	○	○	○	●	●
□ Priority Optimized 12 Vantage Step 02	○	●	○	○	○	○	○	○	○	○	○	○	○
□ Priority Optimized 12 Vantage Step 03	○	○	●	○	○	○	○	○	○	○	○	○	○
□ Priority Optimized 12 Vantage Step 04	○	○	○	●	○	○	○	○	○	○	○	○	○
□ Priority Optimized 12 Vantage Step 05	○	○	○	○	●	○	○	○	○	○	○	○	○
□ Priority Optimized 12 Vantage Step 06	○	○	○	○	○	●	○	○	○	○	○	○	○
□ Priority Optimized 12 Vantage Step 07	○	○	○	○	○	○	●	○	○	○	○	○	○
□ Priority Optimized 12 Vantage Step 08	○	○	○	○	○	○	○	●	○	○	○	○	○
□ Priority Optimized 12 Vantage Step 09	○	○	○	○	○	○	○	○	●	○	○	○	○
□ Priority Optimized 12 Vantage Step 10	○	○	○	○	○	○	○	○	○	●	○	○	○
□ Priority Optimized 12 Vantage Step 11	○	○	○	○	○	○	○	○	○	○	●	○	○
□ Priority Optimized 12 Vantage Step 12	○	○	○	○	○	○	○	○	○	○	○	●	○
□ Priority Optimized 12 Vantage Step 13	○	○	○	○	○	○	○	○	○	○	○	○	●

As a result of changing configuration according recommendation and workload priorities, Vantage Response Time will meet SLGs During next 12 months



Service Level Goal

Adding Instances

Migration from
On Prem to Vantage

Optimized Workload Management Priorities

		Sales	Marketing	Finance	BI
Current Priority	Current	24.99	38.54	39.66	16.58
Optimized Priority	Month 1	17.45	29.09	18.25	50.76
Optimized Priority	Month 2	17.26	28.68	18.13	49.73
Optimized Priority	Month 3	13.83	22.92	14.59	39.52
Optimized Priority	Month 4	12.05	21.05	12.14	35.08
Optimized Priority	Month 5	10.84	18.52	11.47	31.38
Optimized Priority	Month 6	10.97	18.44	11.47	31.38
Optimized Priority	Month 7	10.98	19.57	11.36	31.36
Optimized Priority	Month 8	10.96	18.55	11.35	31.36
Optimized Priority	Month 9	11.07	18.7	11.52	31.38
Optimized Priority	Month 10	10.98	18.75	11.22	31.46
Optimized Priority	Month 11	11.02	18.76	11.35	31.07
Optimized Priority	Month 12	11.03	18.76	11.35	31.07

Redshift 2nd Shift

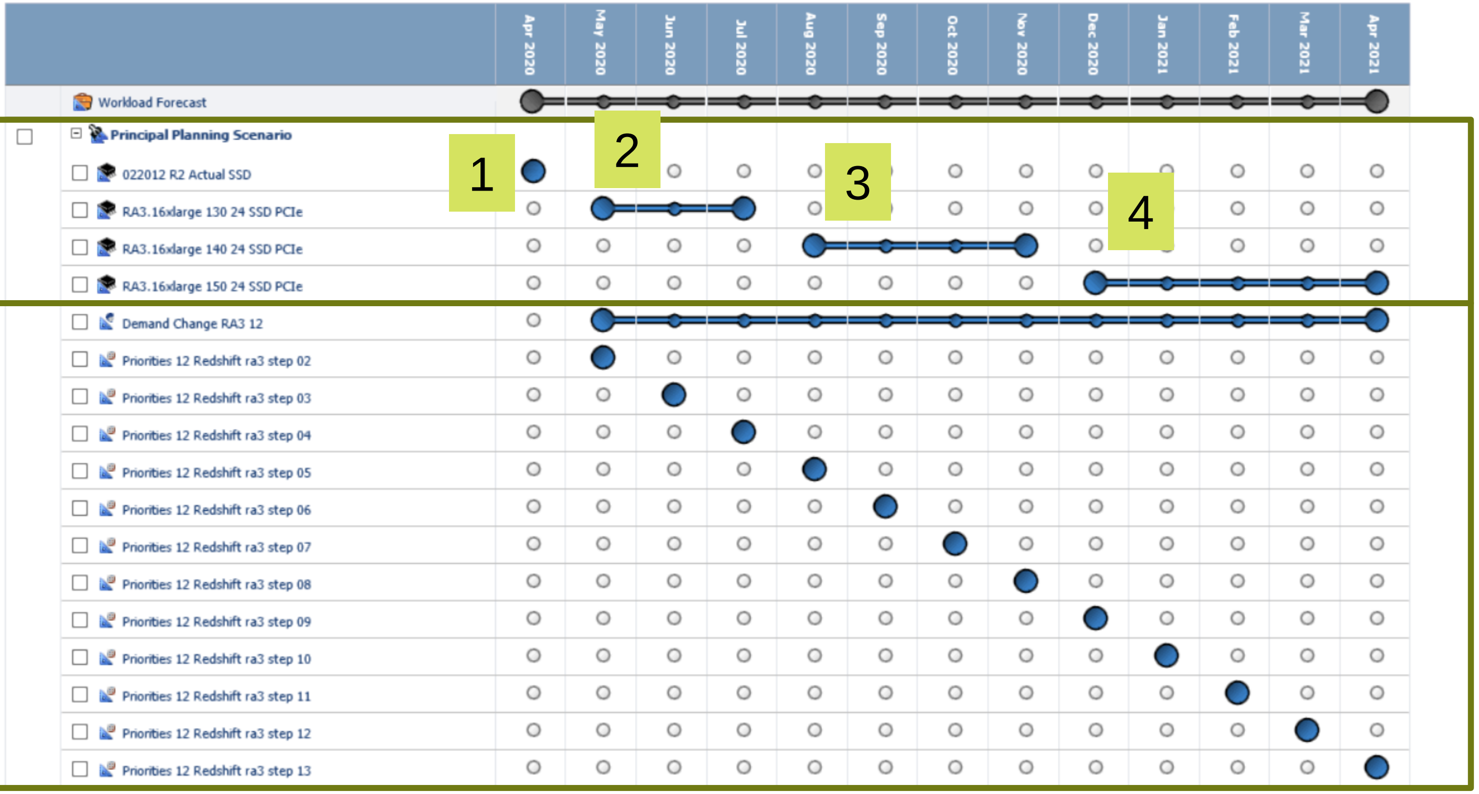
MIGRATE 4 WORKLOADS TO THE REDSHIFT:

CLIENT REPORTING, USER REPORTING, PROGRAM INTEGRITY AND PROVIDER ECONOMICS

As a result of changing configuration according recommendation, Redshift Response Time will meet SLGs during next 12 months

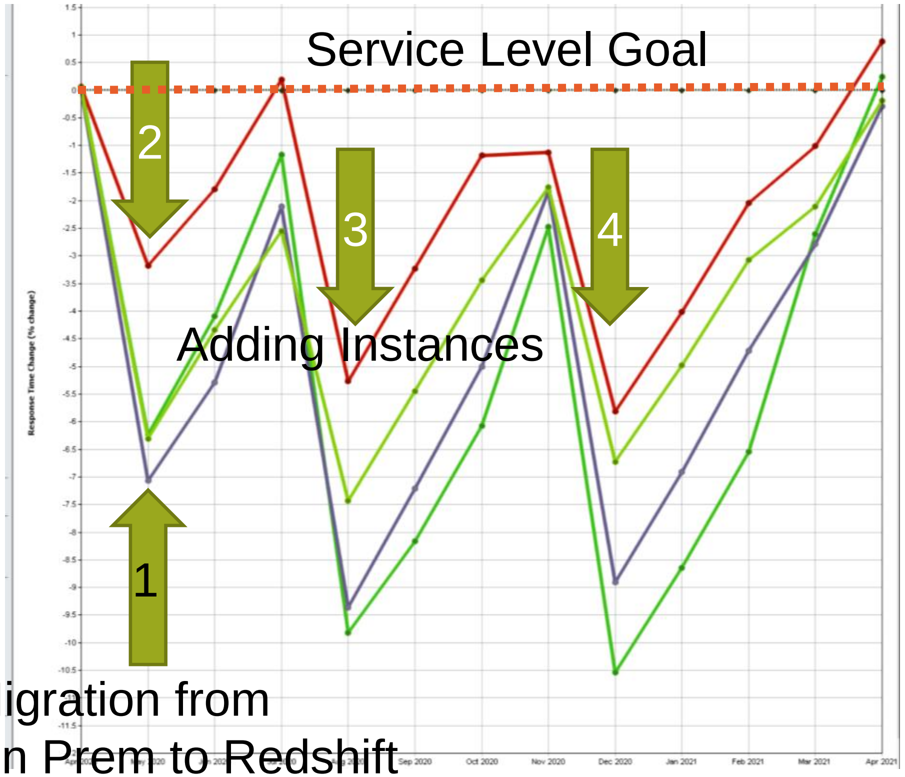
Configuration

Workload Management



	Apr 2020	May 2020	Jun 2020	Jul 2020	Aug 2020	Sep 2020	Oct 2020	Nov 2020	Dec 2020	Jan 2021	Feb 2021	Mar 2021	Apr 2021
Workload Forecast	●	●	●	●	●	●	●	●	●	●	●	●	●
□ ⊟ Principal Planning Scenario													
□ 022012 R2 Actual SSD	●		○	○	○	○	○	○	○	○	○	○	○
□ RA3.16xlarge 130 24 SSD PCIe	○	●	●	●	○	○	○	○	○		○	○	○
□ RA3.16xlarge 140 24 SSD PCIe	○	○	○	○	●	●	●	●	○		○	○	○
□ RA3.16xlarge 150 24 SSD PCIe	○	○	○	○	○	○	○	○	●	●	●	●	●
□ Demand Change RA3 12	○	●	●	●	●	●	●	●	●	●	●	●	●
□ Priorities 12 Redshift ra3 step 02	○	●	○	○	○	○	○	○	○	○	○	○	○
□ Priorities 12 Redshift ra3 step 03	○	○	●	○	○	○	○	○	○	○	○	○	○
□ Priorities 12 Redshift ra3 step 04	○	○	○	●	○	○	○	○	○	○	○	○	○
□ Priorities 12 Redshift ra3 step 05	○	○	○	○	●	○	○	○	○	○	○	○	○
□ Priorities 12 Redshift ra3 step 06	○	○	○	○	○	●	○	○	○	○	○	○	○
□ Priorities 12 Redshift ra3 step 07	○	○	○	○	○	○	●	○	○	○	○	○	○
□ Priorities 12 Redshift ra3 step 08	○	○	○	○	○	○	○	●	○	○	○	○	○
□ Priorities 12 Redshift ra3 step 09	○	○	○	○	○	○	○	○	●	○	○	○	○
□ Priorities 12 Redshift ra3 step 10	○	○	○	○	○	○	○	○	○	●	○	○	○
□ Priorities 12 Redshift ra3 step 11	○	○	○	○	○	○	○	○	○	○	●	○	○
□ Priorities 12 Redshift ra3 step 12	○	○	○	○	○	○	○	○	○	○	○	●	○
□ Priorities 12 Redshift ra3 step 13	○	○	○	○	○	○	○	○	○	○	○	○	●

Predicted Redshift Relative Response Time Change



Service Level Goal

Adding Instances

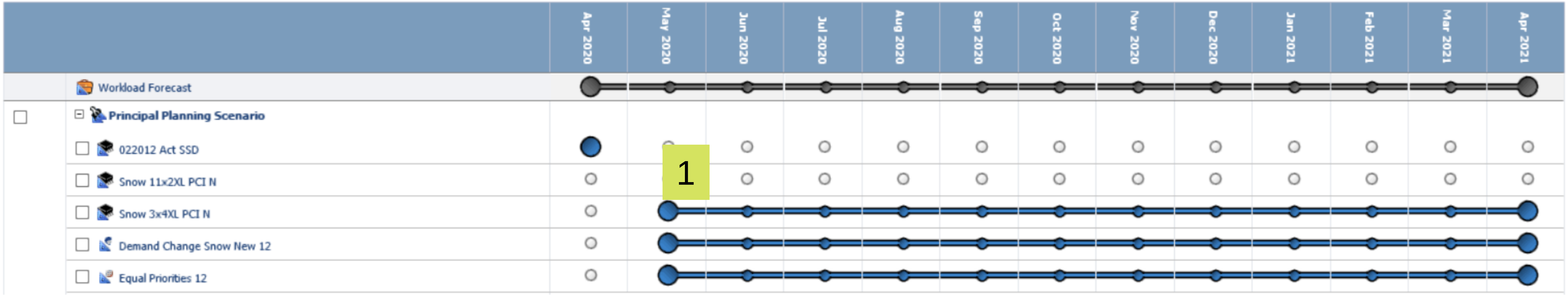
Migration from
On Prem to Redshift

Snowflake 2nd Shift / Month

MIGRATE 4 WORKLOADS INTO ONE SNOWFLAKE VIRTUAL WAREHOUSE:

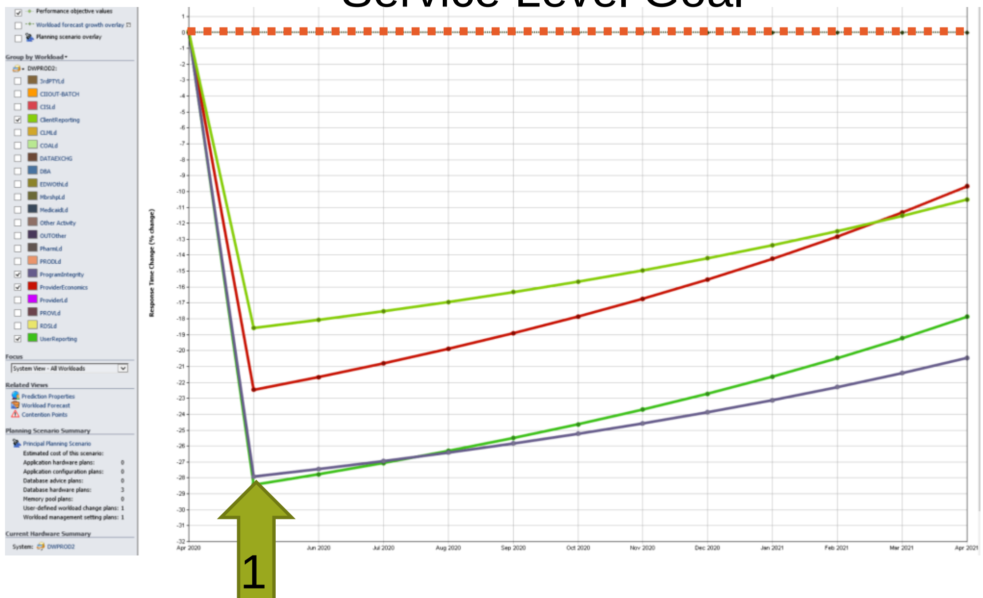
CLIENT REPORTING, USER REPORTING, PROGRAM INTEGRITY AND PROVIDER ECONOMICS

As a result of changing configuration according recommendation, Snowflake Response Time will meet SLGs during next 12 months



		Apr 2020	May 2020	Jun 2020	Jul 2020	Aug 2020	Sep 2020	Oct 2020	Nov 2020	Dec 2020	Jan 2021	Feb 2021	Mar 2021	Apr 2021
	Workload Forecast	●	●	●	●	●	●	●	●	●	●	●	●	●
□	⊟ Principal Planning Scenario													
	□ 022012 Act SSD	●	○	○	○	○	○	○	○	○	○	○	○	○
	□ Snow 11x2XL PCI N	○	○	○	○	○	○	○	○	○	○	○	○	○
	□ Snow 3x4XL PCI N	○	●	●	●	●	●	●	●	●	●	●	●	●
	□ Demand Change Snow New 12	○	●	●	●	●	●	●	●	●	●	●	●	●
	□ Equal Priorities 12	○	●	●	●	●	●	●	●	●	●	●	●	●

Predicted Relative Response Time Change



☑ Performance objective values
□ Workload forecast growth overlay
□ Planning scenario overlay

Group by Workload ▾

DWPROD2:

- □ 3rdPTYLd
- □ CIIOUT-BATCH
- □ CISLd
- ☑ ClientReporting
- □ CLMLd
- □ COALd
- □ DATAEXCHG
- □ DBA
- □ EDWOthLd
- □ MbrshpLd
- □ MedicaidLd
- □ Other Activity
- □ OUTOther
- □ PharmLd
- □ PRODLd
- ☑ ProgramIntegrity
- ☑ ProviderEconomics
- □ ProviderLd
- □ PROVLd
- □ RDSLd
- ☑ UserReporting

Focus

System View - All Workloads

Related Views

- Prediction Properties
- Workload Forecast
- Contention Points

Planning Scenario Summary

Principal Planning Scenario

Estimated cost of this scenario:	
Application hardware plans:	0
Application configuration plans:	0
Database advice plans:	0
Database hardware plans:	3
Memory pool plans:	0
User-defined workload change plans:	1
Workload management setting plans:	1

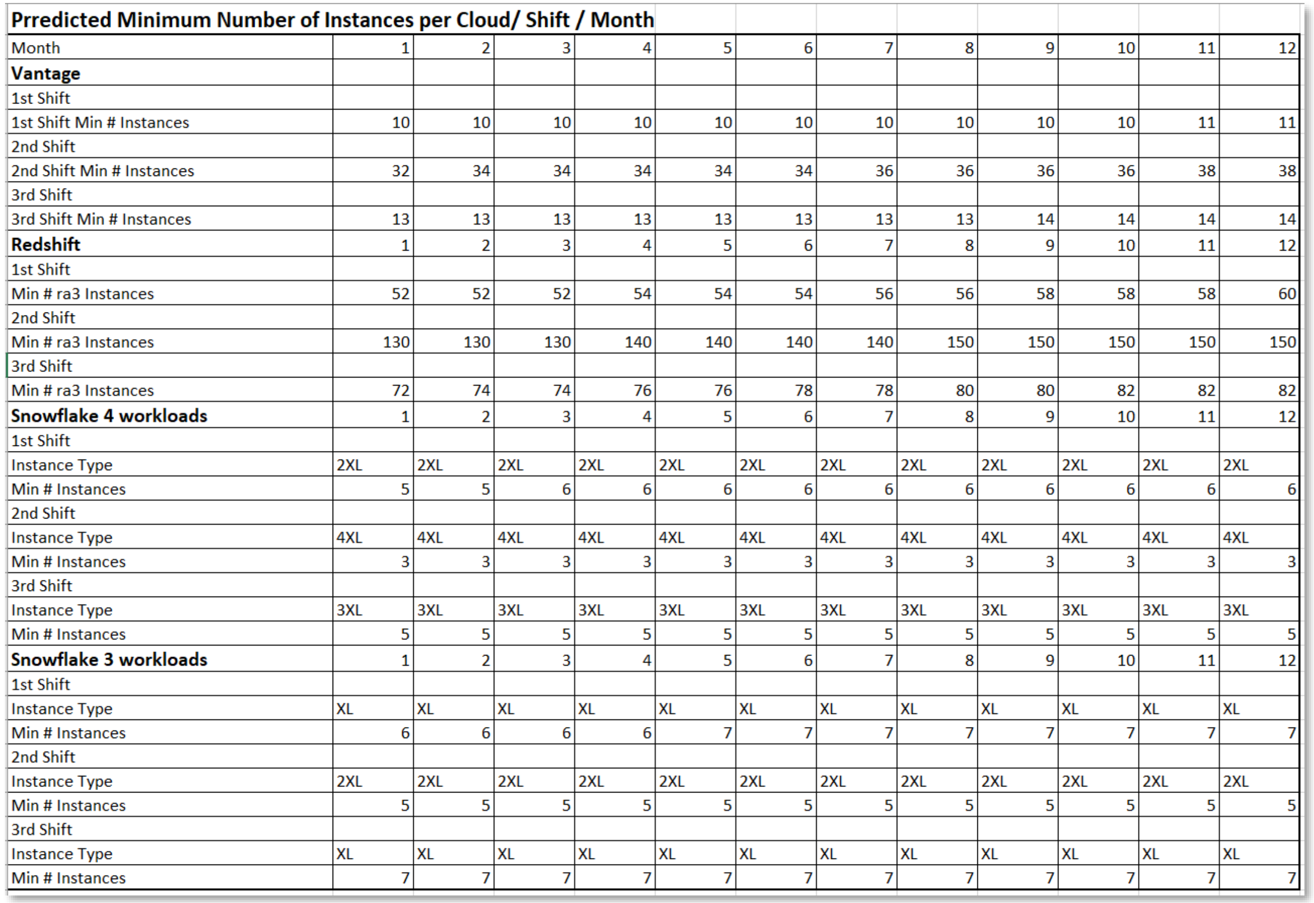
Current Hardware Summary

System: DWPROD2

Service Level Goal

Migration from
On Prem to Snowflake

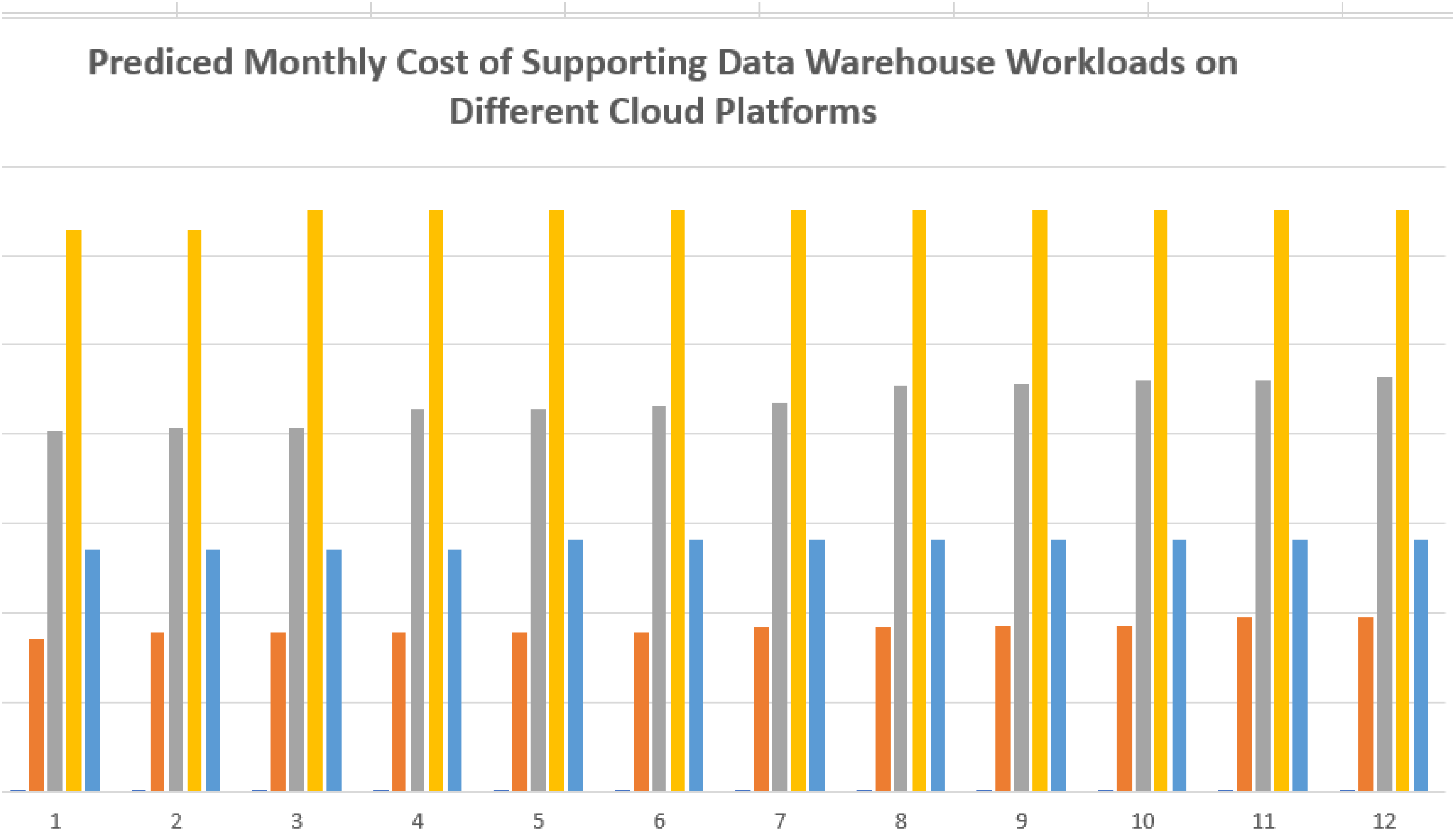
Predicted # of Instances and Instance Type required to meet SLGs for each Cloud / Shift / Month



Prredicted Minimum Number of Instances per Cloud/ Shift / Month

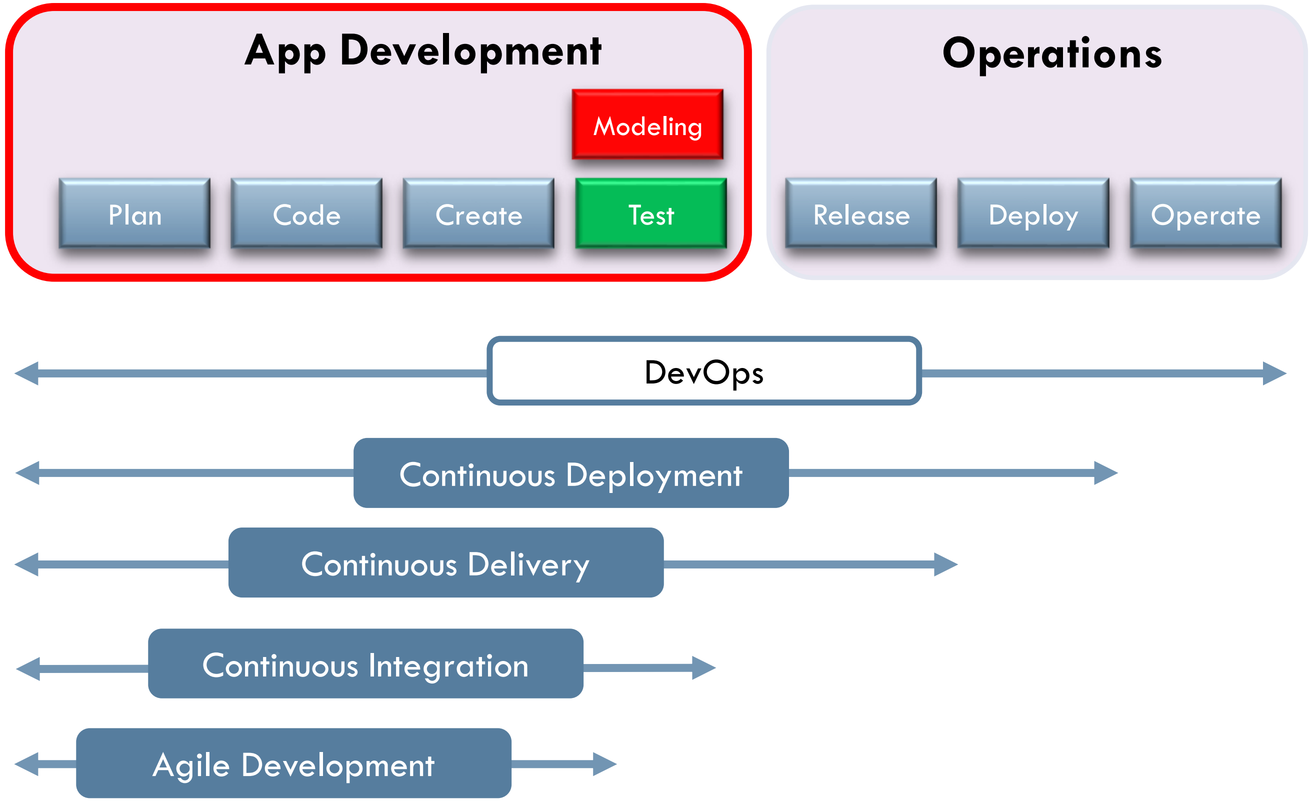
Month	1	2	3	4	5	6	7	8	9	10	11	12
Vantage												
1st Shift												
1st Shift Min # Instances	10	10	10	10	10	10	10	10	10	10	11	11
2nd Shift												
2nd Shift Min # Instances	32	34	34	34	34	34	36	36	36	36	38	38
3rd Shift												
3rd Shift Min # Instances	13	13	13	13	13	13	13	13	14	14	14	14
Redshift	1	2	3	4	5	6	7	8	9	10	11	12
1st Shift												
Min # ra3 Instances	52	52	52	54	54	54	56	56	58	58	58	60
2nd Shift												
Min # ra3 Instances	130	130	130	140	140	140	140	150	150	150	150	150
3rd Shift												
Min # ra3 Instances	72	74	74	76	76	78	78	80	80	82	82	82
Snowflake 4 workloads	1	2	3	4	5	6	7	8	9	10	11	12
1st Shift												
Instance Type	2XL	2XL	2XL	2XL	2XL	2XL	2XL	2XL	2XL	2XL	2XL	2XL
Min # Instances	5	5	6	6	6	6	6	6	6	6	6	6
2nd Shift												
Instance Type	4XL	4XL	4XL	4XL	4XL	4XL	4XL	4XL	4XL	4XL	4XL	4XL
Min # Instances	3	3	3	3	3	3	3	3	3	3	3	3
3rd Shift												
Instance Type	3XL	3XL	3XL	3XL	3XL	3XL	3XL	3XL	3XL	3XL	3XL	3XL
Min # Instances	5	5	5	5	5	5	5	5	5	5	5	5
Snowflake 3 workloads	1	2	3	4	5	6	7	8	9	10	11	12
1st Shift												
Instance Type	XL	XL	XL	XL	XL	XL	XL	XL	XL	XL	XL	XL
Min # Instances	6	6	6	6	7	7	7	7	7	7	7	7
2nd Shift												
Instance Type	2XL	2XL	2XL	2XL	2XL	2XL	2XL	2XL	2XL	2XL	2XL	2XL
Min # Instances	5	5	5	5	5	5	5	5	5	5	5	5
3rd Shift												
Instance Type	XL	XL	XL	XL	XL	XL	XL	XL	XL	XL	XL	XL
Min # Instances	7	7	7	7	7	7	7	7	7	7	7	7

Predicted Monthly Cost to Maintain SLGs on Different Cloud Platforms



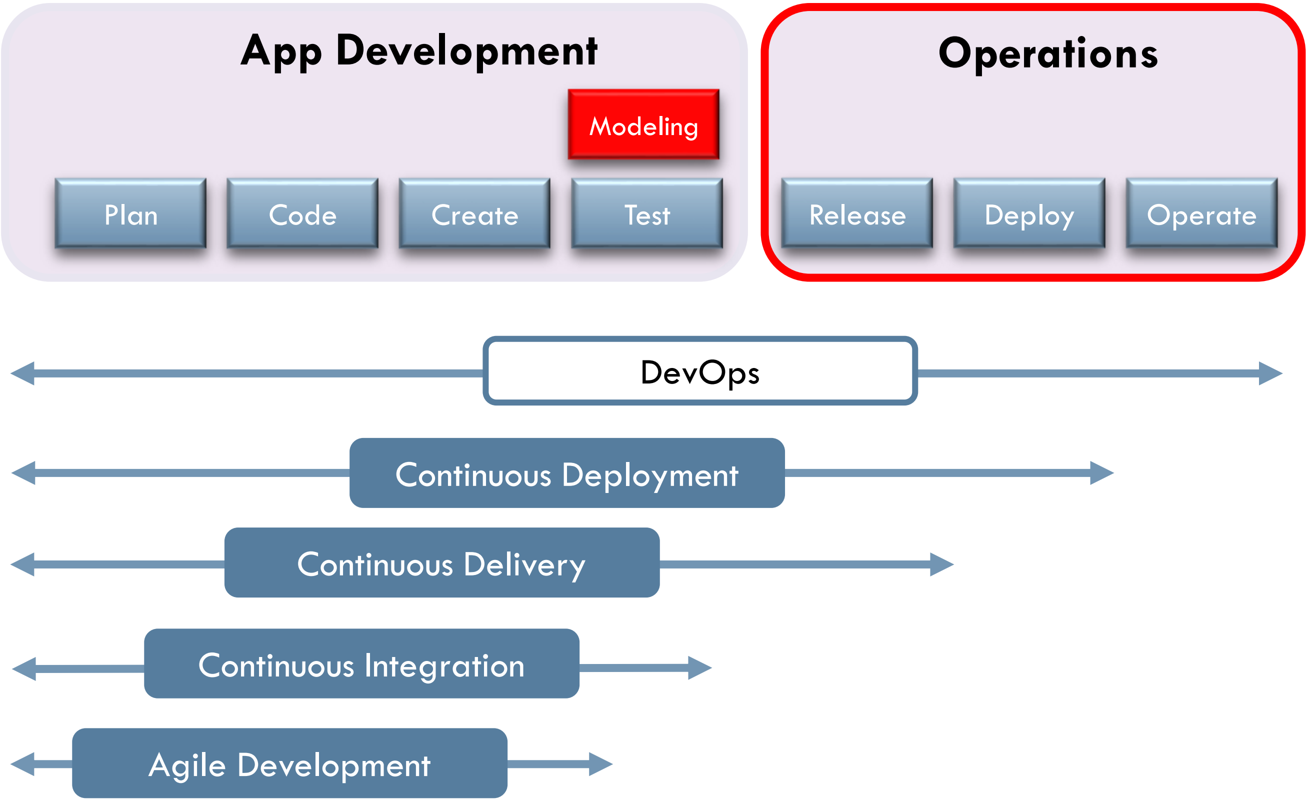
Performance Engineering focusing on Devops

ROLE OF MODELING DURING APPLICATION DEVELOPMENT



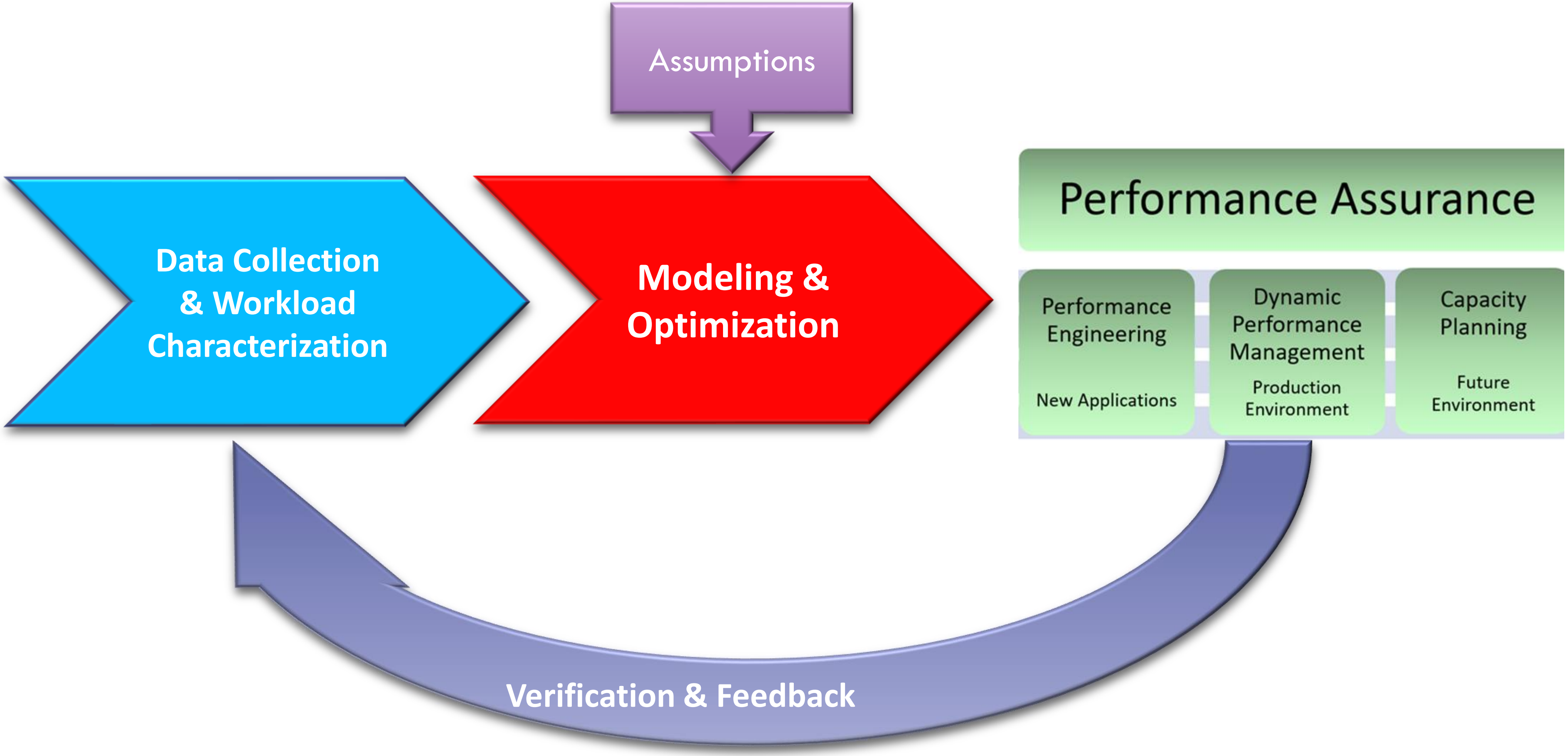
- **Predict new applications implementation impact**
  - Predict how new application will perform in production environment
  - Identify Anomalies and their Root Causes during testing of new applications
  - Develop recommendations to Application Developers
- **Predict how new application will affect existing production applications**
  - Predict how implementation of new applications will affect Response Time and Throughput of existing applications
  - Develop capacity planning recommendations
  - Set up realistic expectations

ROLE OF MODELING FOR OPERATIONS

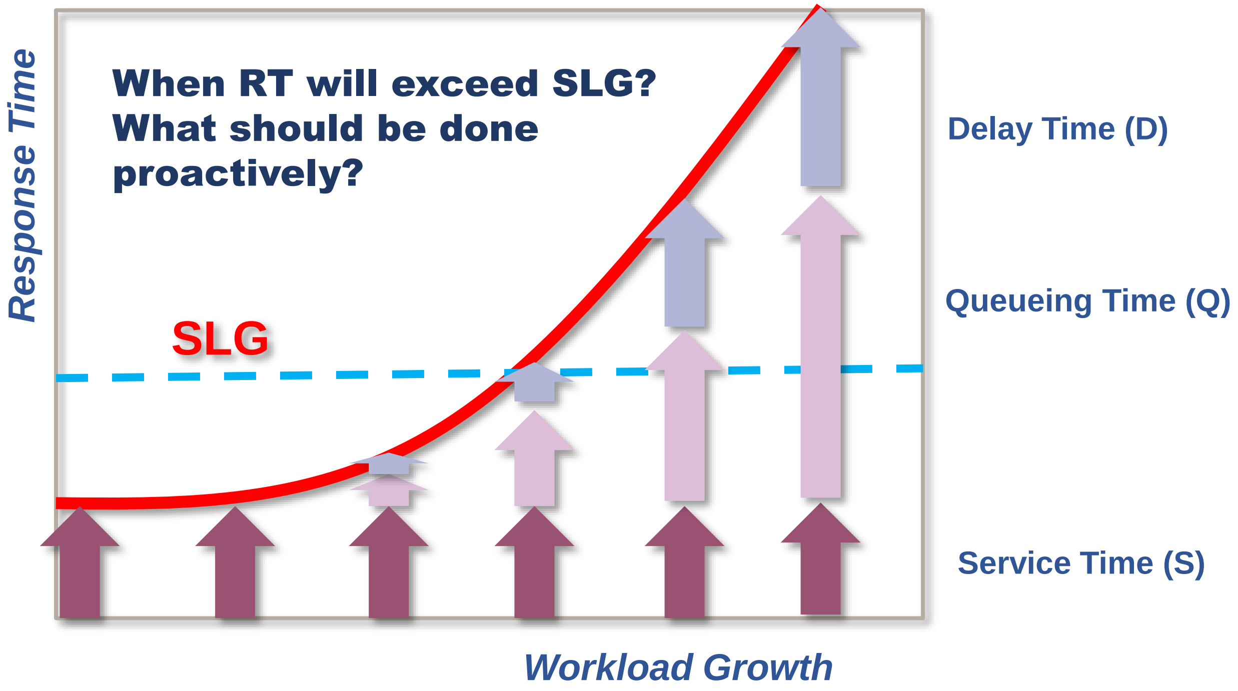


- **Develop Proactive Performance Management and Workload Management Recommendations**
  - Compare performance measurement results after implementation of the new application with expected
  - Develop proactive performance tuning recommendations
  - Develop proactive workload management recommendations
  - Reevaluate Capacity Planning recommendations

MODELING IS A BASE FOR PERFORMANCE ASSURANCE FOR DEVOPS



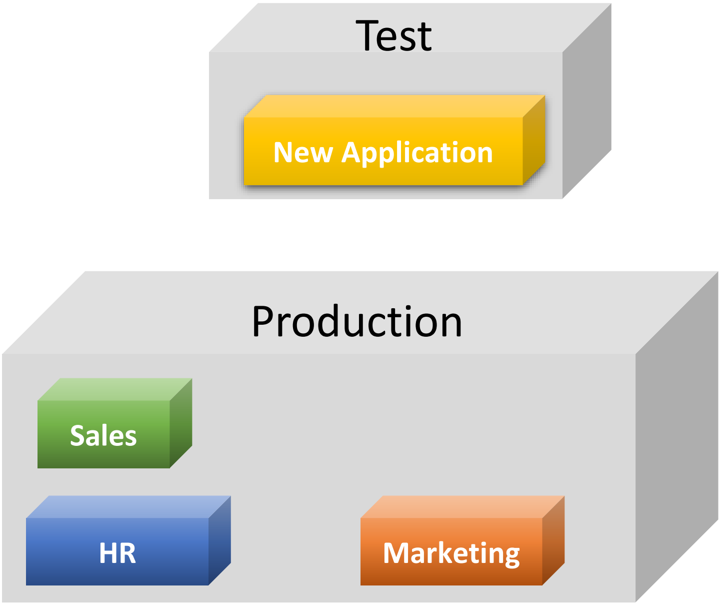
TEN STEPS OF APPLYING MODELING TO OPTIMIZE APPLICATION DEVELOPMENT AND OPERATIONAL DEVOPS DECISIONS



FIRST STEP

DATA COLLECTION DURING PERFORMANCE TESTING AND FOR PRODUCTION WORKLOADS

Data Collection during Performance Testing of New Application on Test System and for all workloads in Production Environments



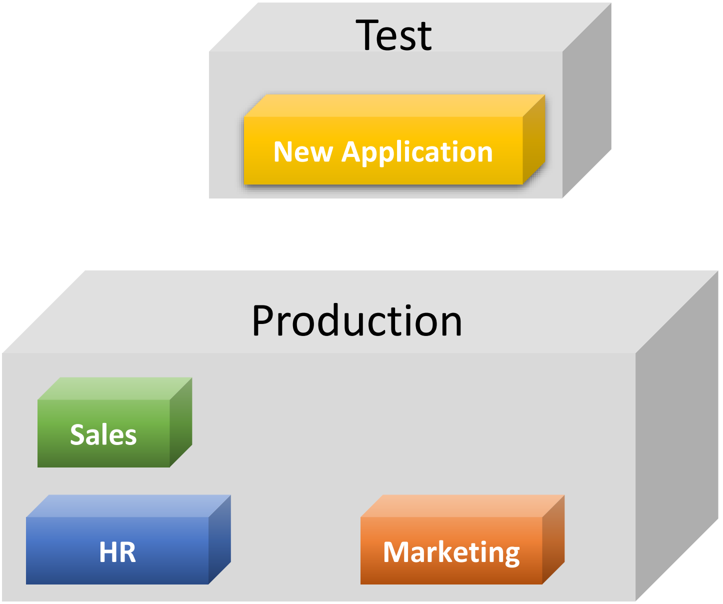
Measurement Data Types

- Hardware and Software Configuration
- Response Time
- Throughput
- CPU Utilization and CPU Service Time per request
- Disk Utilization, I/O rate , #I/O operations per request and KB/Request, Channel Utilization
- Memory utilization
- Network utilization
- Level of concurrency

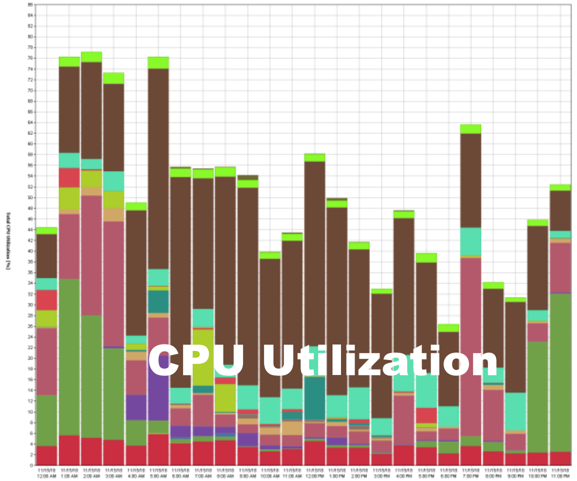
SECOND STEP

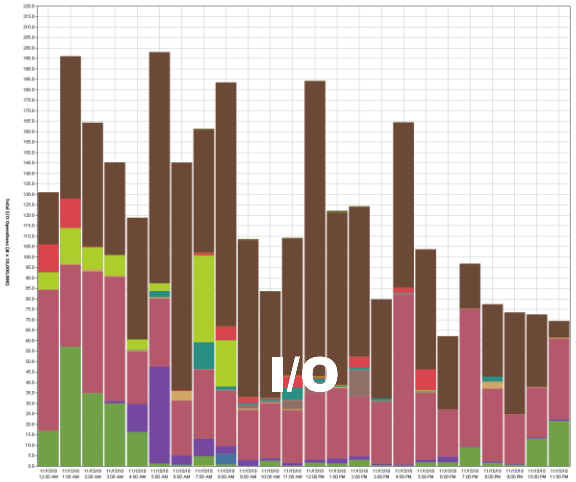
WORKLOAD CHARACTERIZATION

Test and Production Environments

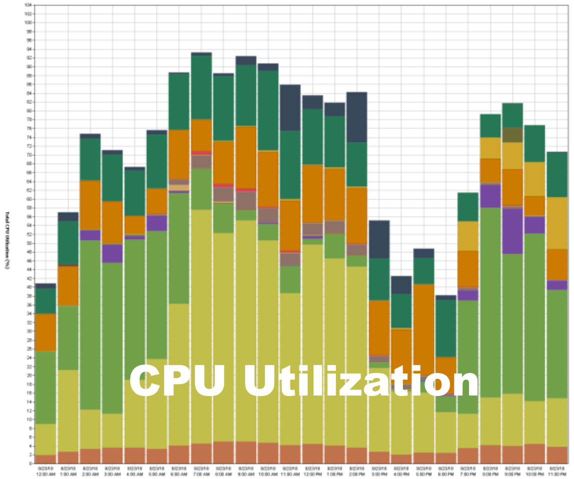


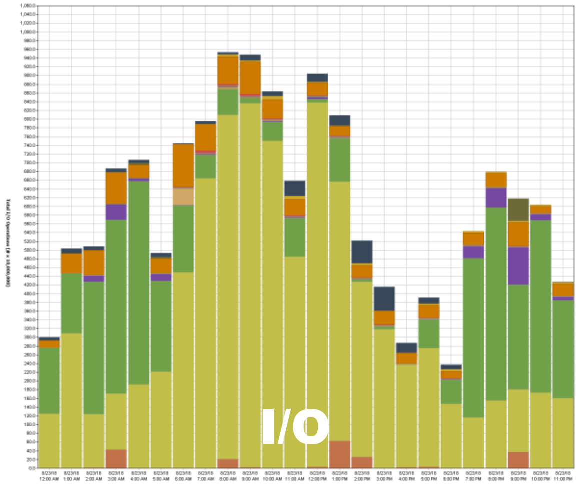
Test





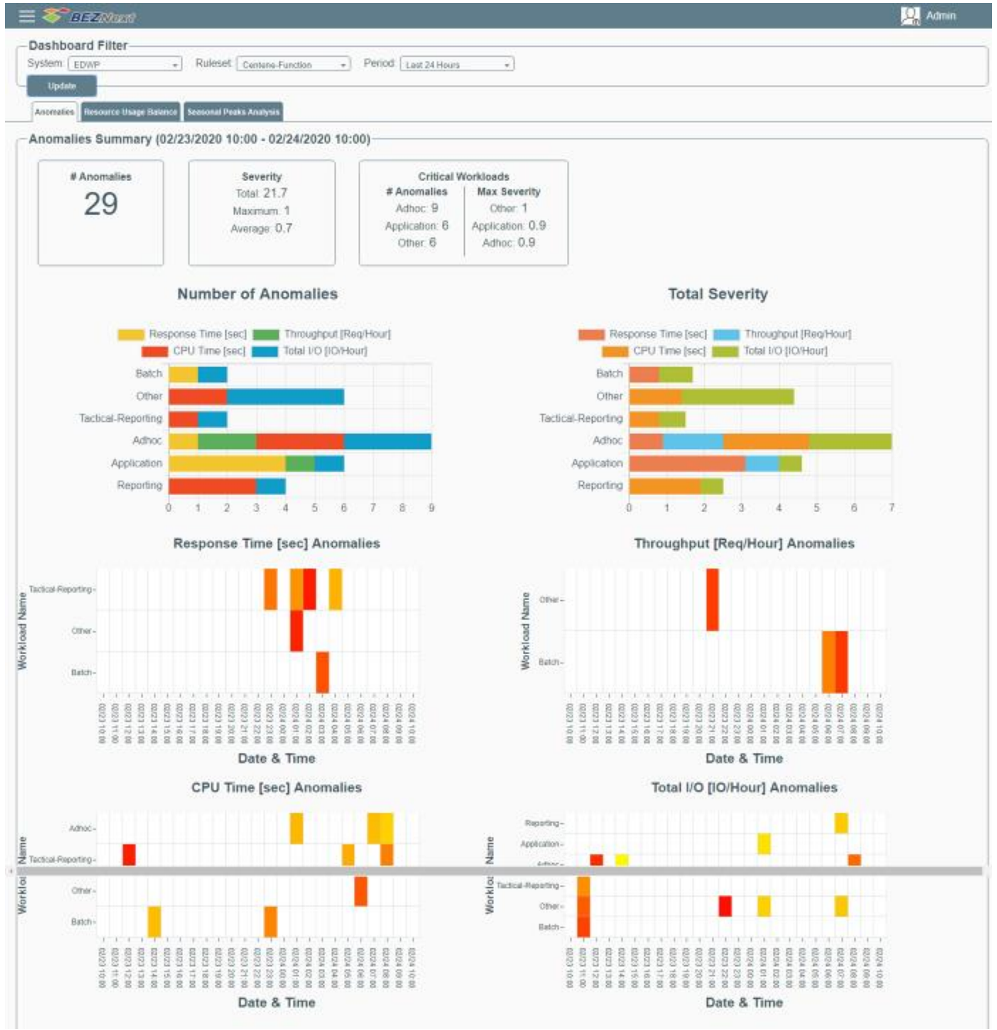
Production





THIRD STEP

ANOMALY AND ROOT CAUSE DETECTION



Dashboard Filter

System: EDWP

Ruleset: Centene-Function

Period: Last 24 Hours

Anomalies Summary (02/23/2020 10:00 - 02/24/2020 10:00)

Anomalies
29

Severity
Total: 21.7
Maximum: 1
Average: 0.7

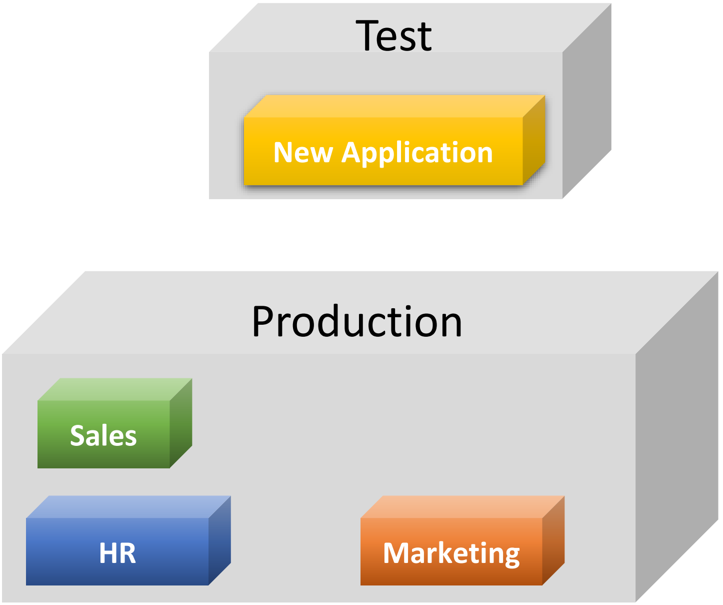
Critical Workloads

# Anomalies	Max Severity
Adhoc: 9	Other: 1
Application: 6	Application: 0.9
Other: 6	Adhoc: 0.9

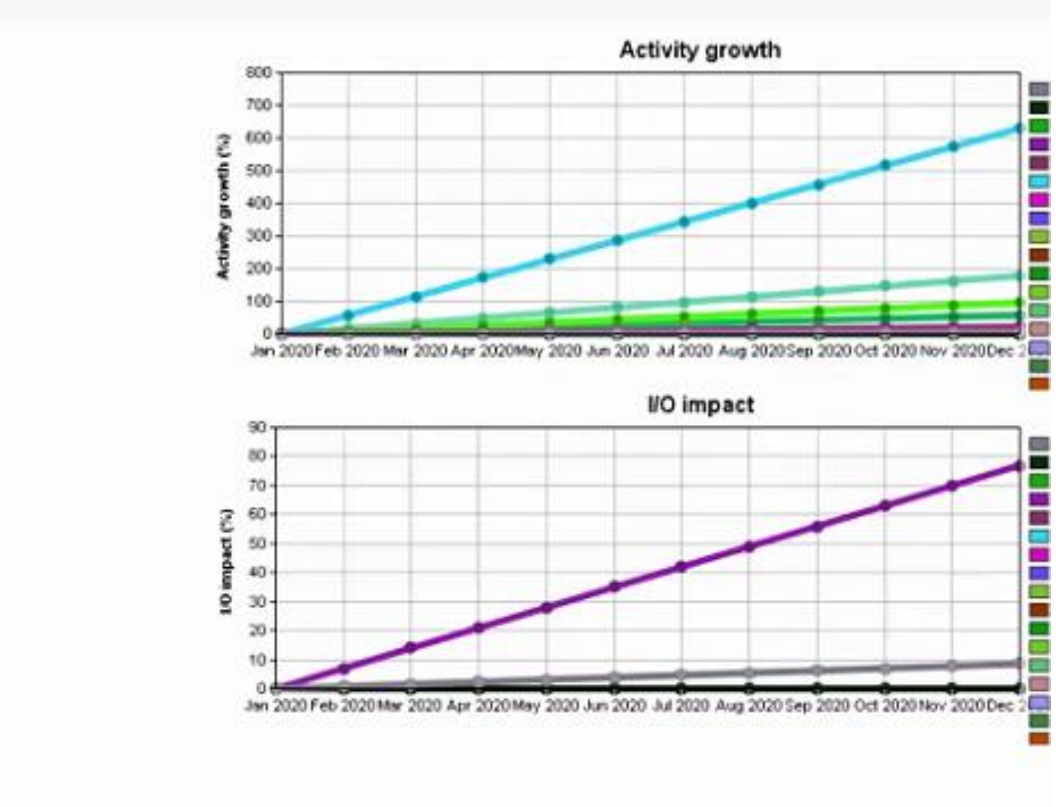
FOURTH STEP

WORKLOAD FORECASTING FOR NEW AND PRODUCTION WORKLOADS

Test and Production Environment

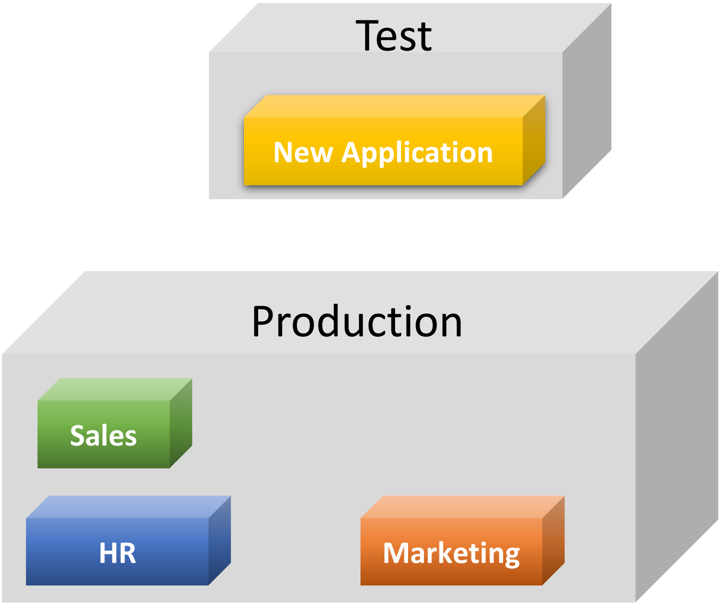


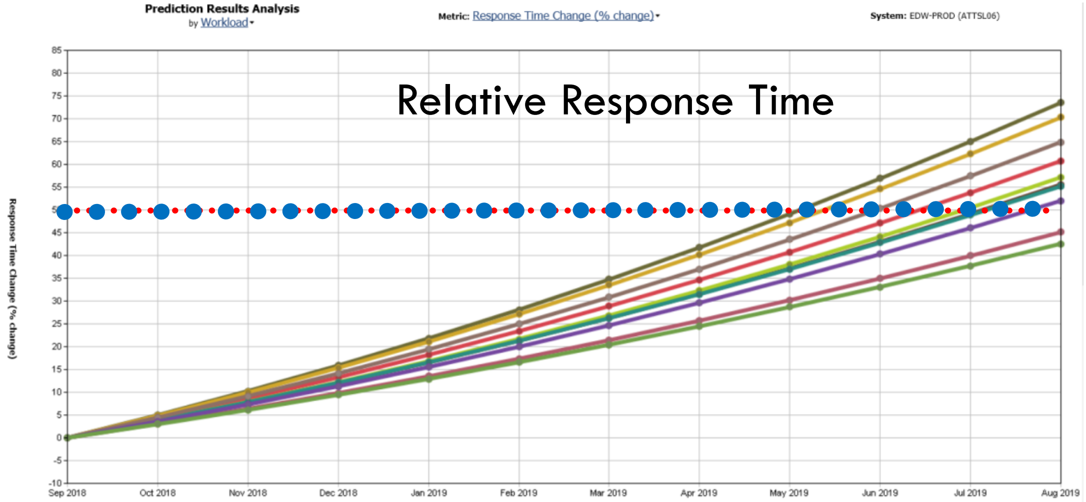
Expected Workload and Volume of Data Growth

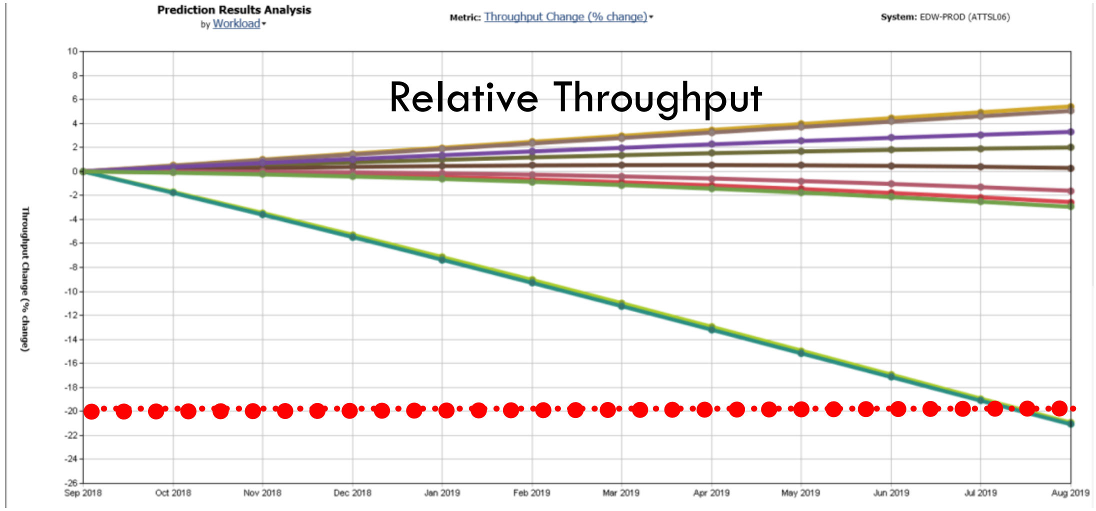


FIFTH STEP

PREDICTING IMPACT OF EXPECTED WORKLOAD AND VOLUME OF DATA GROWTH IN PRODUCTION ENVIRONMENT

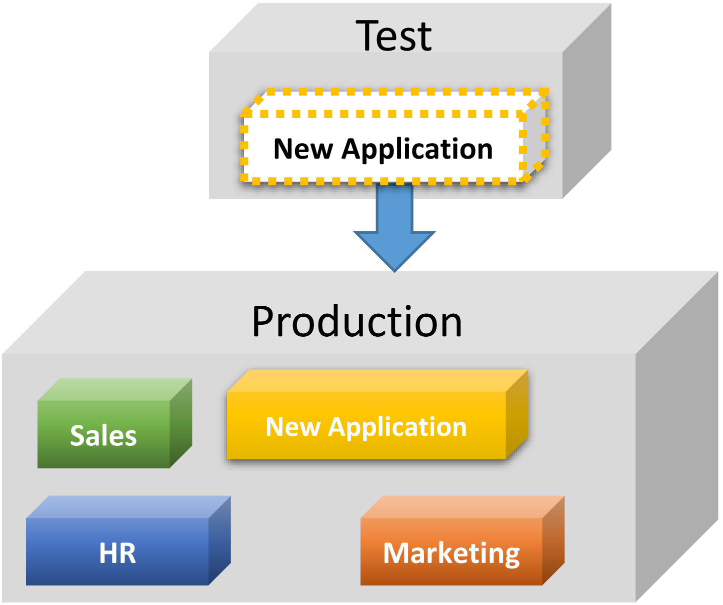


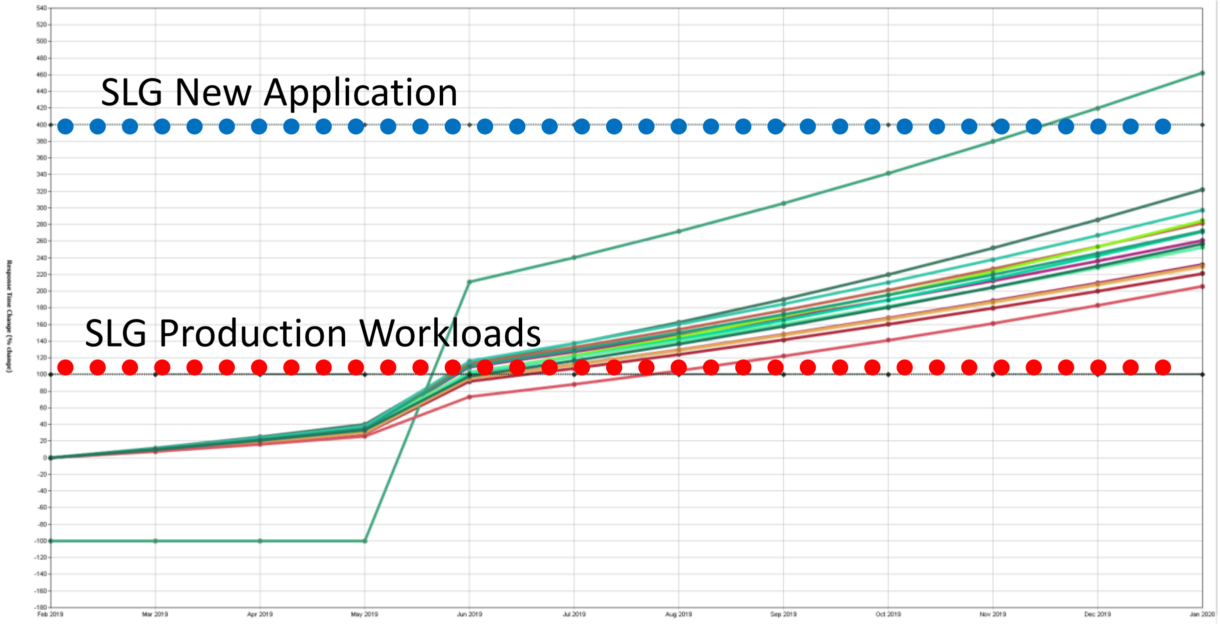




SIX STEP

PREDICTING IMPACT OF NEW APPLICATION IMPLEMENTATION

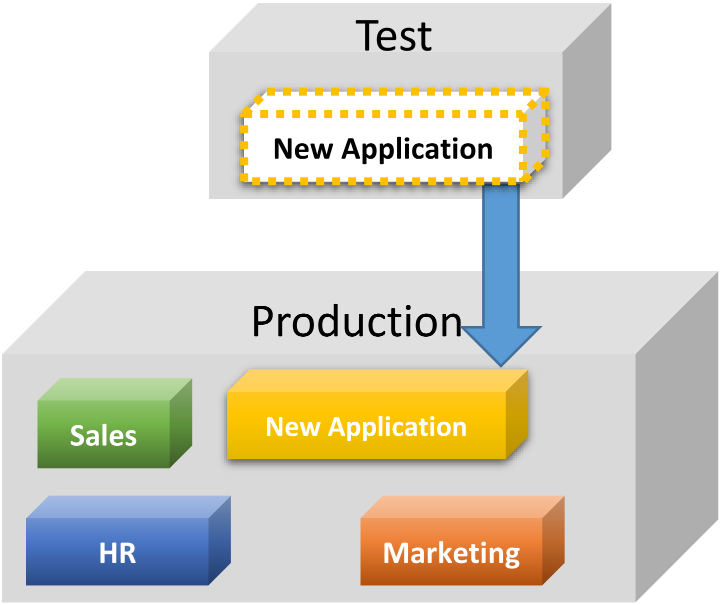


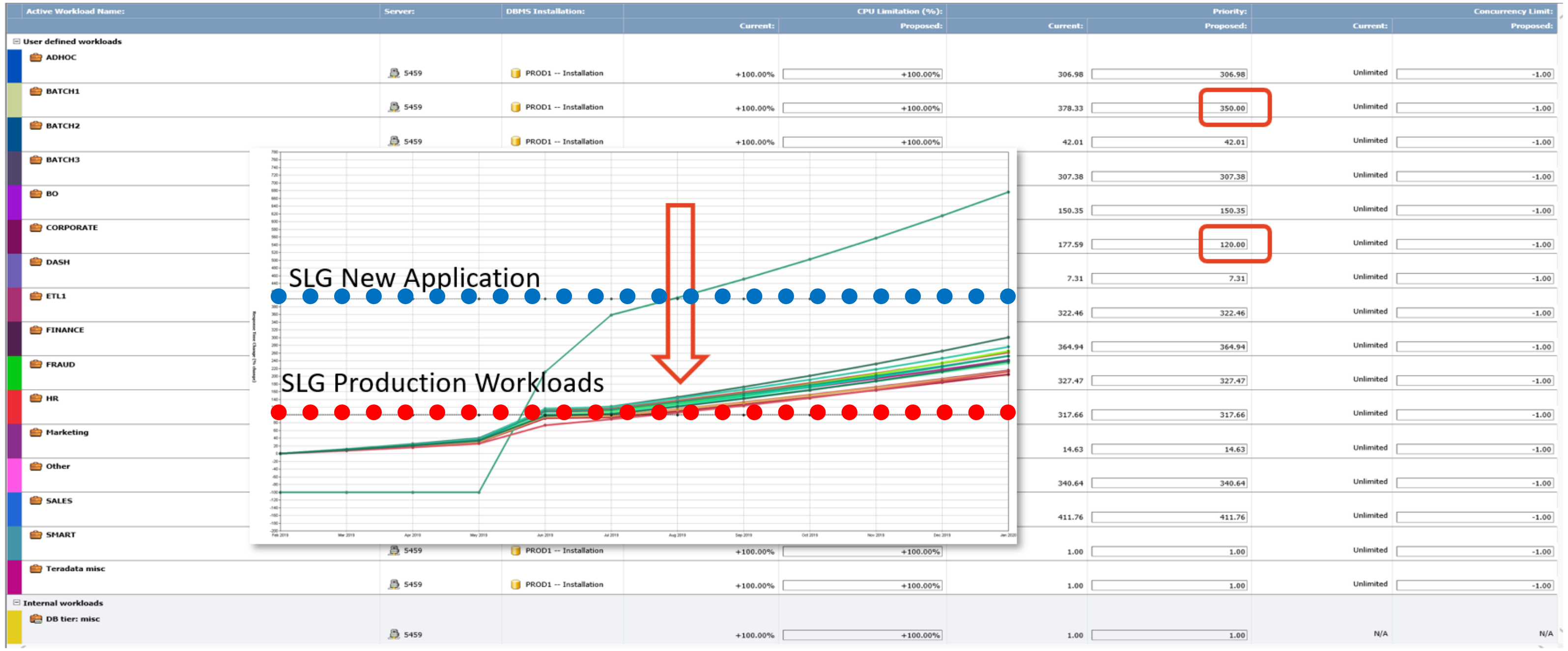


SEVENTH STEP

PREDICTING IMPACT OF THE WORKLOAD MANAGEMENT OPTIMIZATION

WORKLOAD MANAGEMENT OPTIMIZATION WILL NOT BE SUFFICIENT TO MEET SLG



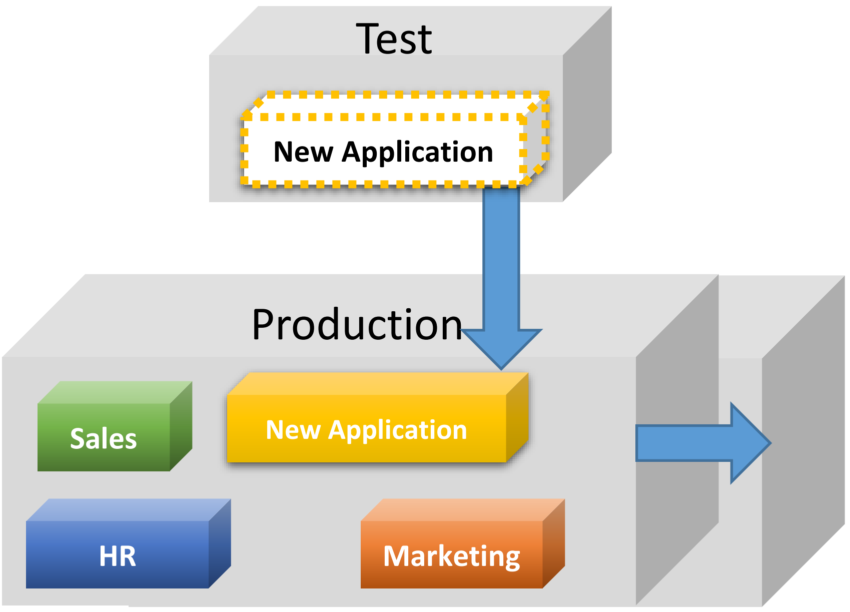


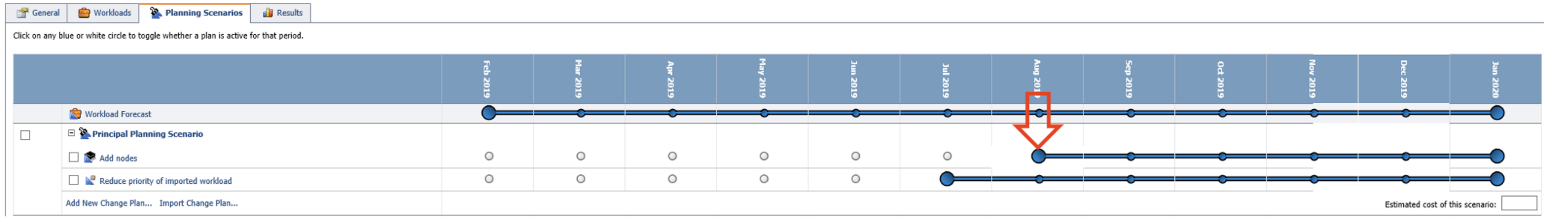
Active Workload Name:	Server:	DBMS Installation:	CPU Limitation (%):		Priority:		Concurrency Limit:	
			Current:	Proposed:	Current:	Proposed:	Current:	Proposed:
⊟ User defined workloads								
ADHOC	5459	PROD1 -- Installation	+100.00%	+100.00%	306.98	306.98	Unlimited	-1.00
BATCH1	5459	PROD1 -- Installation	+100.00%	+100.00%	378.33	350.00	Unlimited	-1.00
BATCH2	5459	PROD1 -- Installation	+100.00%	+100.00%	42.01	42.01	Unlimited	-1.00
BATCH3					307.38	307.38	Unlimited	-1.00
BO					150.35	150.35	Unlimited	-1.00
CORPORATE					177.59	120.00	Unlimited	-1.00
DASH					7.31	7.31	Unlimited	-1.00
ETL1					322.46	322.46	Unlimited	-1.00
FINANCE					364.94	364.94	Unlimited	-1.00
FRAUD					327.47	327.47	Unlimited	-1.00
HR					317.66	317.66	Unlimited	-1.00
Marketing					14.63	14.63	Unlimited	-1.00
Other					340.64	340.64	Unlimited	-1.00
SALES					411.76	411.76	Unlimited	-1.00
SMART	5459	PROD1 -- Installation	+100.00%	+100.00%	1.00	1.00	Unlimited	-1.00
Teradata misc	5459	PROD1 -- Installation	+100.00%	+100.00%	1.00	1.00	Unlimited	-1.00
⊟ Internal workloads								
DB tier: misc	5459		+100.00%	+100.00%	1.00	1.00	N/A	N/A

EIGHTH STEP

PREDICTING MINIMUM ON PREM UPGRADE REQUIRED TO MEET SLG AFTER NEW APPLICATION IMPLEMENTATION

ADDITIONAL 14 NODES WILL BE REQUIRED TO MEET SLG



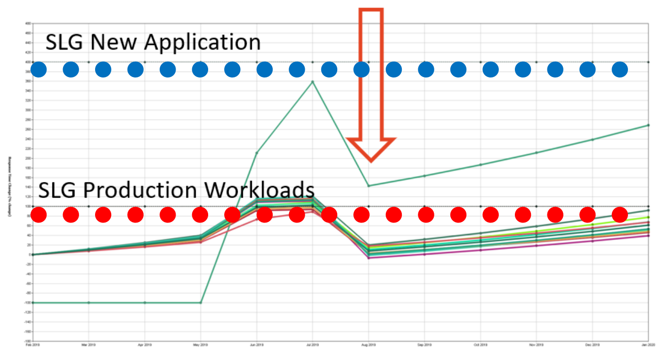


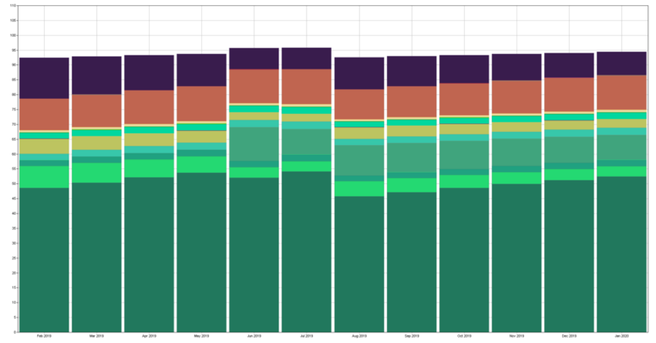
Click on any blue or white circle to toggle whether a plan is active for that period.

	Feb 2019	Mar 2019	Apr 2019	May 2019	Jun 2019	Jul 2019	Aug 2019	Sep 2019	Oct 2019	Nov 2019	Dec 2019	Jan 2020
Workload Forecast	●	●	●	●	●	●	●	●	●	●	●	●
□ ⊟ **Principal Planning Scenario**												
□ Add nodes	○	○	○	○	○	○	●	●	●	●	●	●
□ Reduce priority of imported workload	○	○	○	○	○	●	●	●	●	●	●	●

Add New Change Plan... Import Change Plan...

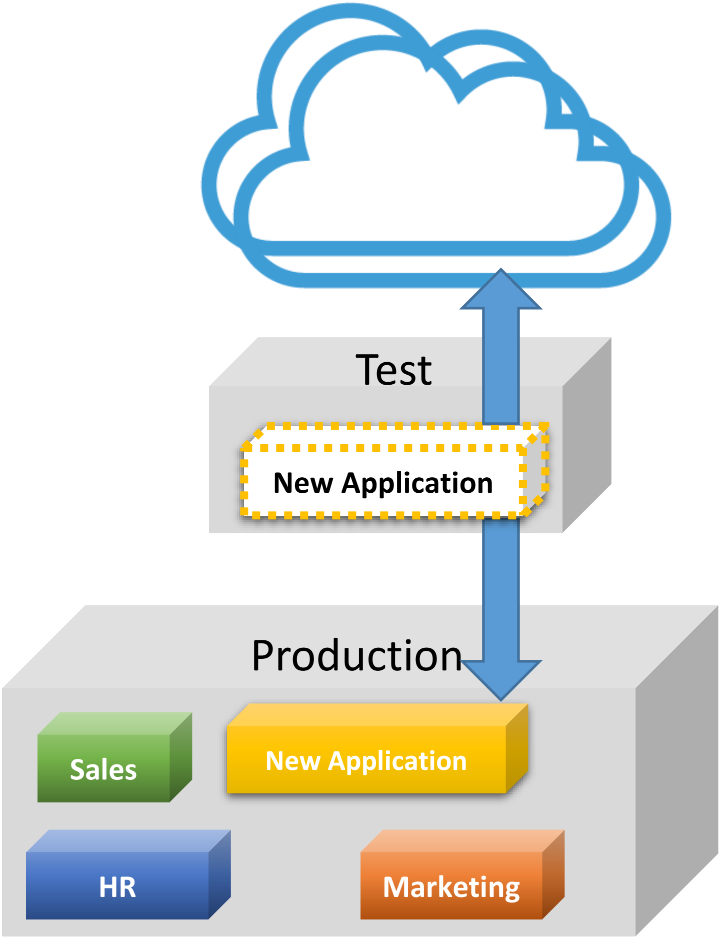
Estimated cost of this scenario: □





NINTH STEP

DETERMINING APPROPRIATE CLOUD PLATFORM FOR NEW APPLICATION

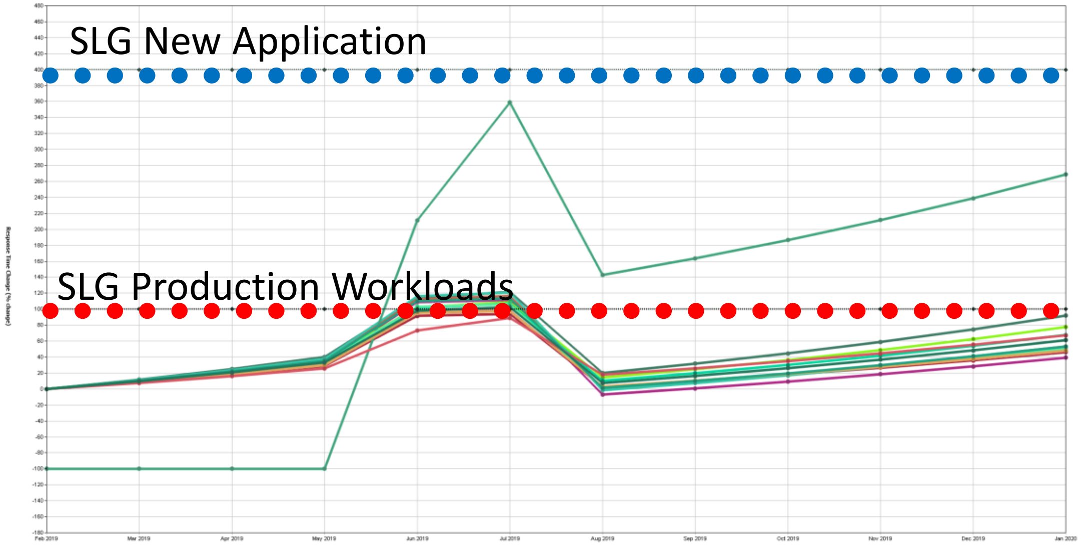


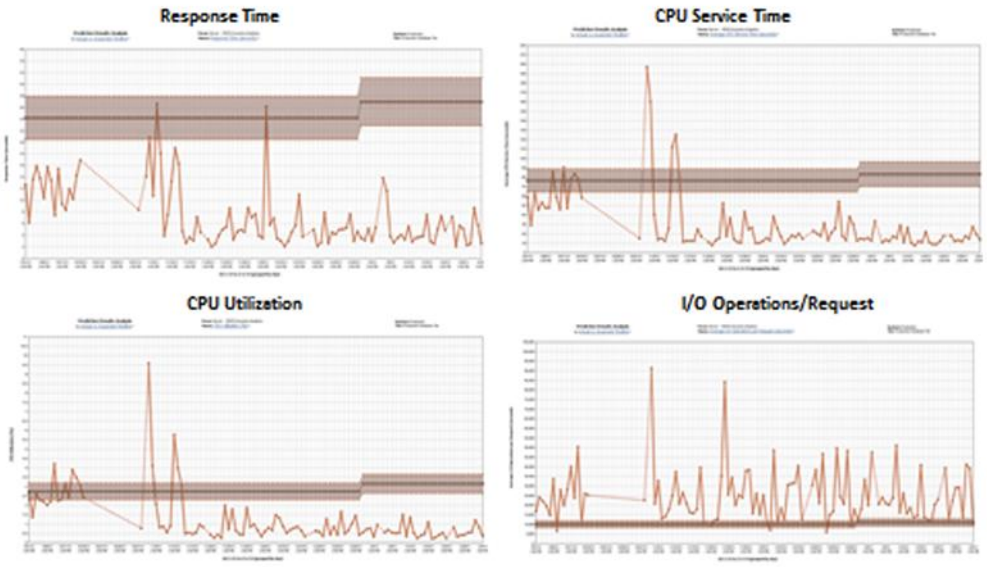
BEZNext Approach to Selection of the Appropriate Cloud

- ❑ Predict the minimum configuration required to meet SLGs
  - Instance type and # of instances which will be required Hour by Hour, Shift by Shift, Month by Month to meet SLGs for each of On Prem Production workload on each of the optional Cloud Platform
- ❑ Predict cost of running On Prem Data Warehouse Workloads on each of the optional Cloud Platforms
- ❑ Select Cloud platform capable to meet SLGs for all of the growing workloads with the lowest cost

TENTH STEP

AUTOMATIC RESULT VERIFICATION AND CREATION OF CONTINUOUS PERFORMANCE ASSURANCE PROCESS





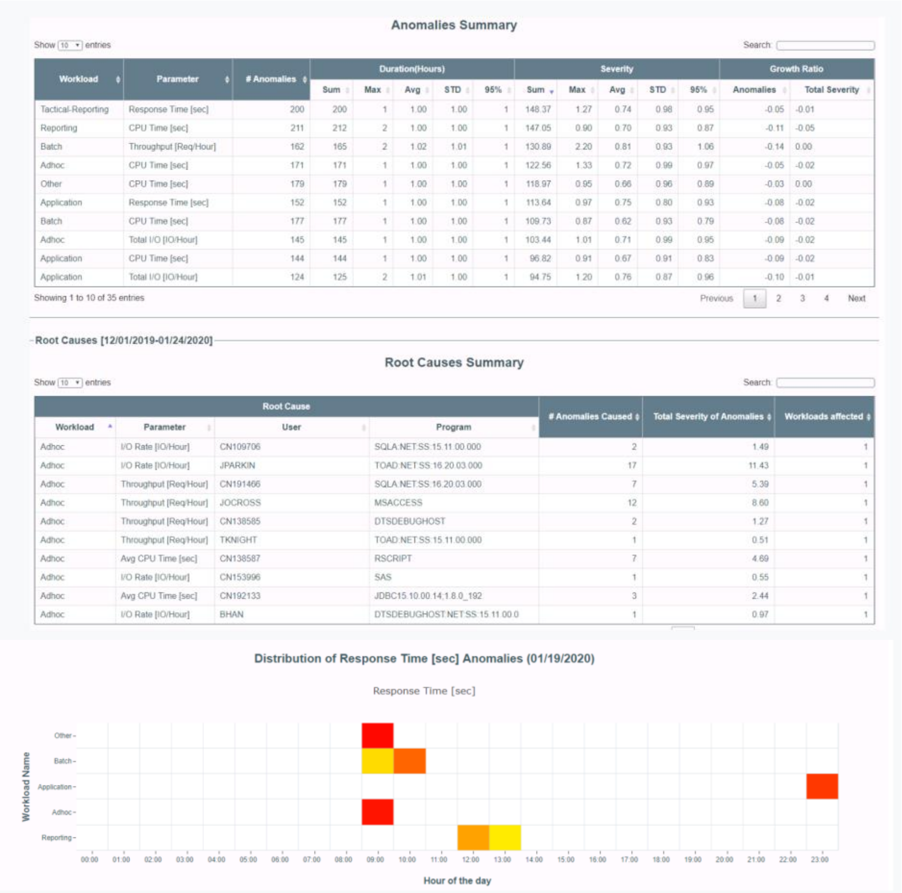
HOW TO OPTIMIZING DATA WAREHOUSE AND BIG DATA APPLICATIONS PERFORMANCE ON PREM AND IN THE CLOUD

DYNAMIC PERFORMANCE MANAGEMENT FOR DATA WAREHOUSES, AND BIG DATA APPLICATIONS ON PREM AND IN THE CLOUD

- ❑ Set realistic Service Level Goals
  - ❑ Formal SLG are based on business requirements
  - ❑ Informal SLGs are based on analysis of historical data
  - ❑ Without SLG impossible to manage and plan effectively
- ❑ Data collection and Workload Aggregation
  - ❑ Automatic data collection across all platform and transforming to universal format reduce time required to coordinate and interpret data
  - ❑ WAG by line of business allows to present results of analysis and recommendation clear to business people and IT management
- ❑ Workload Characterization
  - ❑ Automatic generation of Performance, Resource utilization and Data usage by Line of Business/Workloads enables automation of identification of problems and their root causes and use of modeling and optimization to generate proactive recommendations, including determining:
    - The most frequent performance anomalies/problems and their root causes
    - Pattern and balance of performance and resource utilization
    - Application availability
    - Seasonality for each workload
- ❑ Evaluate solutions for fixing the problems
- ❑ Verify results

DETERMINE MOST FREQUENT ANOMALIES AND ROOT CAUSES TO NARROW DOWN THE SCOPE OF PERFORMANCE TUNING

Data Warehouse



Anomalies Summary

Show 10 ▾ entries   Search: ▭

Workload	Parameter	# Anomalies	Duration(Hours)					Severity					Growth Ratio	
			Sum	Max	Avg	STD	95%	Sum	Max	Avg	STD	95%	Anomalies	Total Severity
Tactical-Reporting	Response Time [sec]	200	200	1	1.00	1.00	1	148.37	1.27	0.74	0.98	0.95	-0.05	-0.01
Reporting	CPU Time [sec]	211	212	2	1.00	1.00	1	147.05	0.90	0.70	0.93	0.87	-0.11	-0.05
Batch	Throughput [Req/Hour]	162	165	2	1.02	1.01	1	130.89	2.20	0.81	0.93	1.06	-0.14	0.00
Adhoc	CPU Time [sec]	171	171	1	1.00	1.00	1	122.56	1.33	0.72	0.99	0.97	-0.05	-0.02
Other	CPU Time [sec]	179	179	1	1.00	1.00	1	118.97	0.95	0.66	0.96	0.89	-0.03	0.00
Application	Response Time [sec]	152	152	1	1.00	1.00	1	113.64	0.97	0.75	0.80	0.93	-0.08	-0.02
Batch	CPU Time [sec]	177	177	1	1.00	1.00	1	109.73	0.87	0.62	0.93	0.79	-0.08	-0.02
Adhoc	Total I/O [IO/Hour]	145	145	1	1.00	1.00	1	103.44	1.01	0.71	0.99	0.95	-0.09	-0.02
Application	CPU Time [sec]	144	144	1	1.00	1.00	1	96.82	0.91	0.67	0.91	0.83	-0.09	-0.02
Application	Total I/O [IO/Hour]	124	125	2	1.01	1.00	1	94.75	1.20	0.76	0.87	0.96	-0.10	-0.01

Showing 1 to 10 of 35 entries   Previous 1 2 3 4 Next

Root Causes [12/01/2019-01/24/2020]

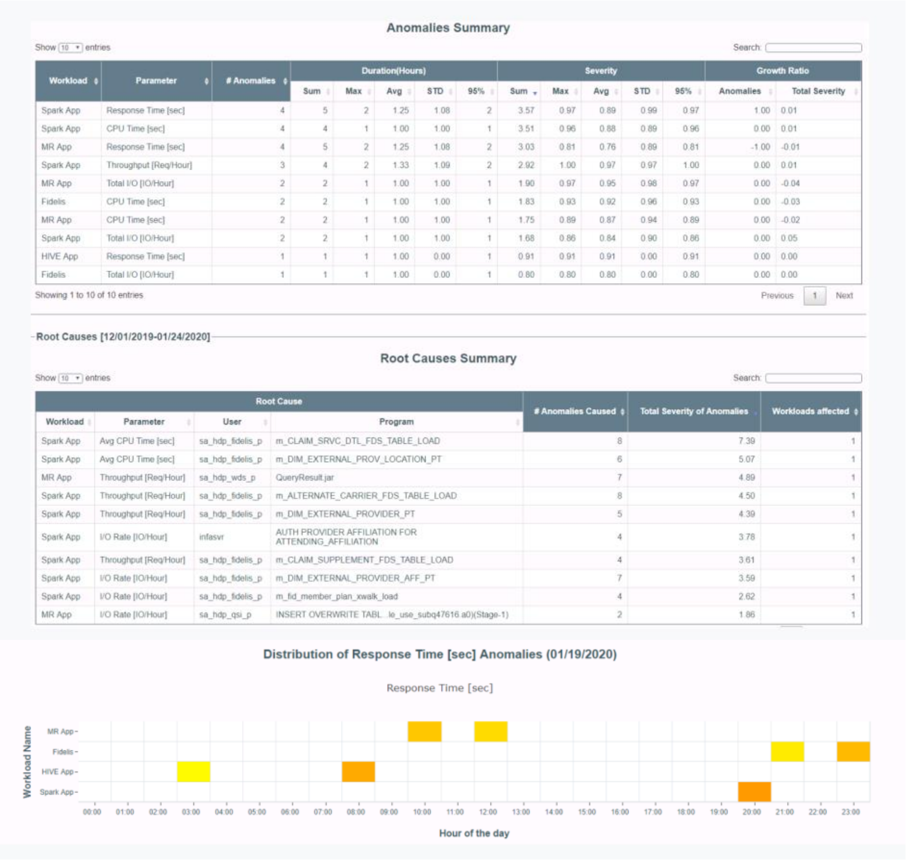
Root Causes Summary

Show 10 ▾ entries   Search: ▭

Root Cause				# Anomalies Caused	Total Severity of Anomalies	Workloads affected
Workload	Parameter	User	Program			
Adhoc	I/O Rate [IO/Hour]	CN109706	SQLA:NET:SS:15.11.00.000	2	1.49	1
Adhoc	I/O Rate [IO/Hour]	JPARKIN	TOAD:NET:SS:16.20.03.000	17	11.43	1
Adhoc	Throughput [Req/Hour]	CN191466	SQLA:NET:SS:16.20.03.000	7	5.39	1
Adhoc	Throughput [Req/Hour]	JOCROSS	MSACCESS	12	8.60	1
Adhoc	Throughput [Req/Hour]	CN138585	DTSDEBUGHOST	2	1.27	1
Adhoc	Throughput [Req/Hour]	TKNIGHT	TOAD:NET:SS:15.11.00.000	1	0.51	1
Adhoc	Avg CPU Time [sec]	CN138587	RSCRIPT	7	4.69	1
Adhoc	I/O Rate [IO/Hour]	CN153996	SAS	1	0.55	1
Adhoc	Avg CPU Time [sec]	CN192133	JDBC15.10.00.14;1.8.0_192	3	2.44	1
Adhoc	I/O Rate [IO/Hour]	BHAN	DTSDEBUGHOST:NET:SS:15.11.00.0	1	0.97	1

Distribution of Response Time [sec] Anomalies (01/19/2020)

Big Data



Anomalies Summary

Show 10 ▾ entries   Search: ▭

Workload	Parameter	# Anomalies	Duration(Hours)					Severity					Growth Ratio	
			Sum	Max	Avg	STD	95%	Sum	Max	Avg	STD	95%	Anomalies	Total Severity
Spark App	Response Time [sec]	4	5	2	1.25	1.08	2	3.57	0.97	0.89	0.99	0.97	1.00	0.01
Spark App	CPU Time [sec]	4	4	1	1.00	1.00	1	3.51	0.96	0.88	0.89	0.96	0.00	0.01
MR App	Response Time [sec]	4	5	2	1.25	1.08	2	3.03	0.81	0.76	0.89	0.81	-1.00	-0.01
Spark App	Throughput [Req/Hour]	3	4	2	1.33	1.09	2	2.92	1.00	0.97	0.97	1.00	0.00	0.01
MR App	Total I/O [IO/Hour]	2	2	1	1.00	1.00	1	1.90	0.97	0.95	0.98	0.97	0.00	-0.04
Fidelis	CPU Time [sec]	2	2	1	1.00	1.00	1	1.83	0.93	0.92	0.96	0.93	0.00	-0.03
MR App	CPU Time [sec]	2	2	1	1.00	1.00	1	1.75	0.89	0.87	0.94	0.89	0.00	-0.02
Spark App	Total I/O [IO/Hour]	2	2	1	1.00	1.00	1	1.68	0.86	0.84	0.90	0.86	0.00	0.05
HIVE App	Response Time [sec]	1	1	1	1.00	0.00	1	0.91	0.91	0.91	0.00	0.91	0.00	0.00
Fidelis	Total I/O [IO/Hour]	1	1	1	1.00	0.00	1	0.80	0.80	0.80	0.00	0.80	0.00	0.00

Showing 1 to 10 of 10 entries   Previous 1 Next

Root Causes [12/01/2019-01/24/2020]

Root Causes Summary

Show 10 ▾ entries   Search: ▭

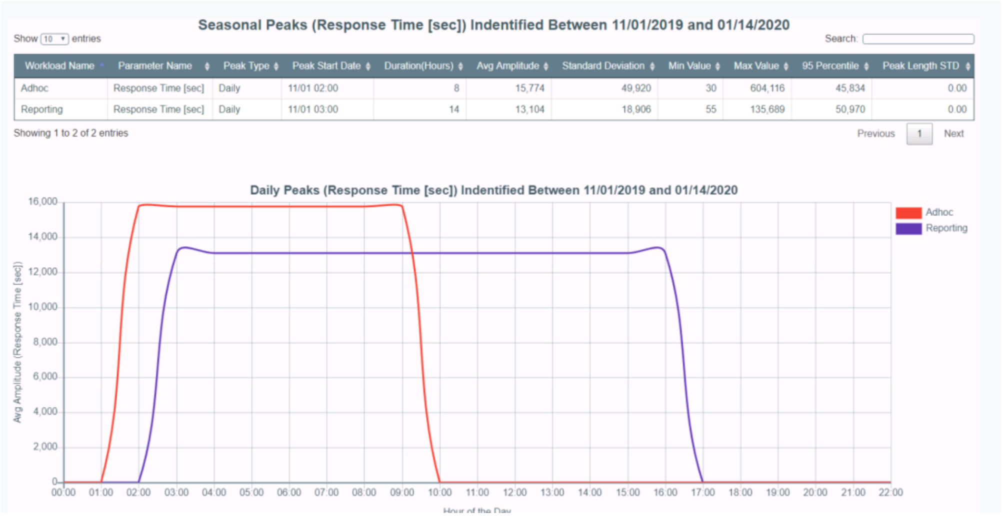
Root Cause				# Anomalies Caused	Total Severity of Anomalies	Workloads affected
Workload	Parameter	User	Program			
Spark App	Avg CPU Time [sec]	sa_hdp_fidelis_p	m_CLAIM_SRVC_DTL_FDS_TABLE_LOAD	8	7.39	1
Spark App	Avg CPU Time [sec]	sa_hdp_fidelis_p	m_DIM_EXTERNAL_PROV_LOCATION_PT	6	5.07	1
MR App	Throughput [Req/Hour]	sa_hdp_wds_p	QueryResult.jar	7	4.89	1
Spark App	Throughput [Req/Hour]	sa_hdp_fidelis_p	m_ALTERNATE_CARRIER_FDS_TABLE_LOAD	8	4.50	1
Spark App	Throughput [Req/Hour]	sa_hdp_fidelis_p	m_DIM_EXTERNAL_PROVIDER_PT	5	4.39	1
Spark App	I/O Rate [IO/Hour]	infasvr	AUTH PROVIDER AFFILIATION FOR ATTENDING_AFFILIATION	4	3.78	1
Spark App	Throughput [Req/Hour]	sa_hdp_fidelis_p	m_CLAIM_SUPPLEMENT_FDS_TABLE_LOAD	4	3.61	1
Spark App	I/O Rate [IO/Hour]	sa_hdp_fidelis_p	m_DIM_EXTERNAL_PROVIDER_AFF_PT	7	3.59	1
Spark App	I/O Rate [IO/Hour]	sa_hdp_fidelis_p	m_fid_member_plan_xwalk_load	4	2.62	1
MR App	I/O Rate [IO/Hour]	sa_hdp_qsi_p	INSERT OVERWRITE TABL...le_use_subq47616.a0)(Stage-1)	2	1.86	1

Distribution of Response Time [sec] Anomalies (01/19/2020)

DETERMINE SEASONAL RESOURCE UTILIZATION PEAKS

TO OPTIMIZE WORKLOAD MANAGEMENT AND RESOURCE ALLOCATION RULES ON PREM AND IN THE CLOUD (TASM, YARN, ALLOCATION AND DEALLOCATION RESOURCES IN THE CLOUD)

Data Warehouse



Seasonal Peaks (Response Time [sec]) Indentified Between 11/01/2019 and 01/14/2020

Show 10 entries

Search:

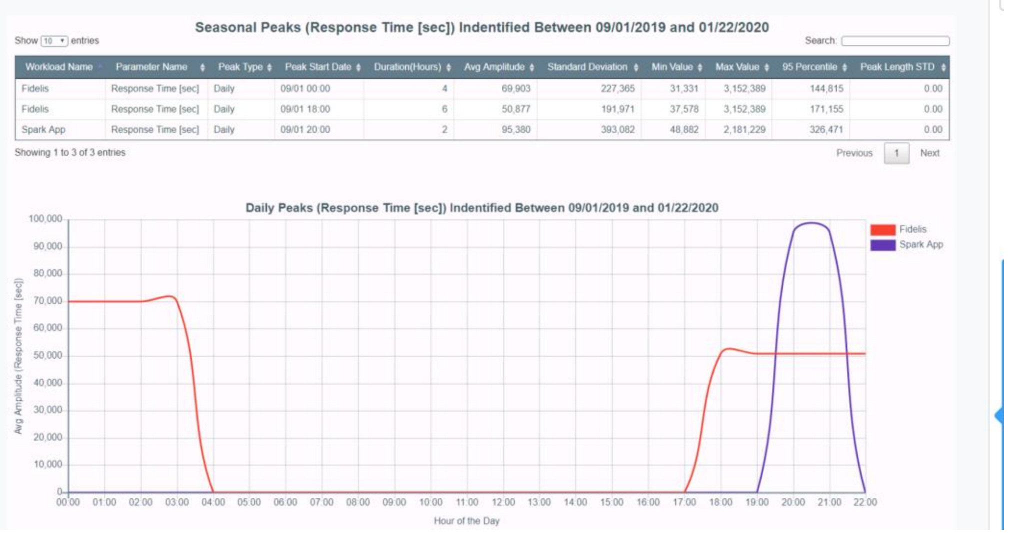
Workload Name	Parameter Name	Peak Type	Peak Start Date	Duration(Hours)	Avg Amplitude	Standard Deviation	Min Value	Max Value	95 Percentile	Peak Length STD
Adhoc	Response Time [sec]	Daily	11/01 02:00	8	15,774	49,920	30	604,116	45,834	0.00
Reporting	Response Time [sec]	Daily	11/01 03:00	14	13,104	18,906	55	135,689	50,970	0.00

Showing 1 to 2 of 2 entries

Previous 1 Next

Daily Peaks (Response Time [sec]) Indentified Between 11/01/2019 and 01/14/2020

Big Data



Seasonal Peaks (Response Time [sec]) Indentified Between 09/01/2019 and 01/22/2020

Show 10 entries

Search:

Workload Name	Parameter Name	Peak Type	Peak Start Date	Duration(Hours)	Avg Amplitude	Standard Deviation	Min Value	Max Value	95 Percentile	Peak Length STD
Fidelis	Response Time [sec]	Daily	09/01 00:00	4	69,903	227,365	31,331	3,152,389	144,815	0.00
Fidelis	Response Time [sec]	Daily	09/01 18:00	6	50,877	191,971	37,578	3,152,389	171,155	0.00
Spark App	Response Time [sec]	Daily	09/01 20:00	2	95,380	393,082	48,882	2,181,229	326,471	0.00

Showing 1 to 3 of 3 entries

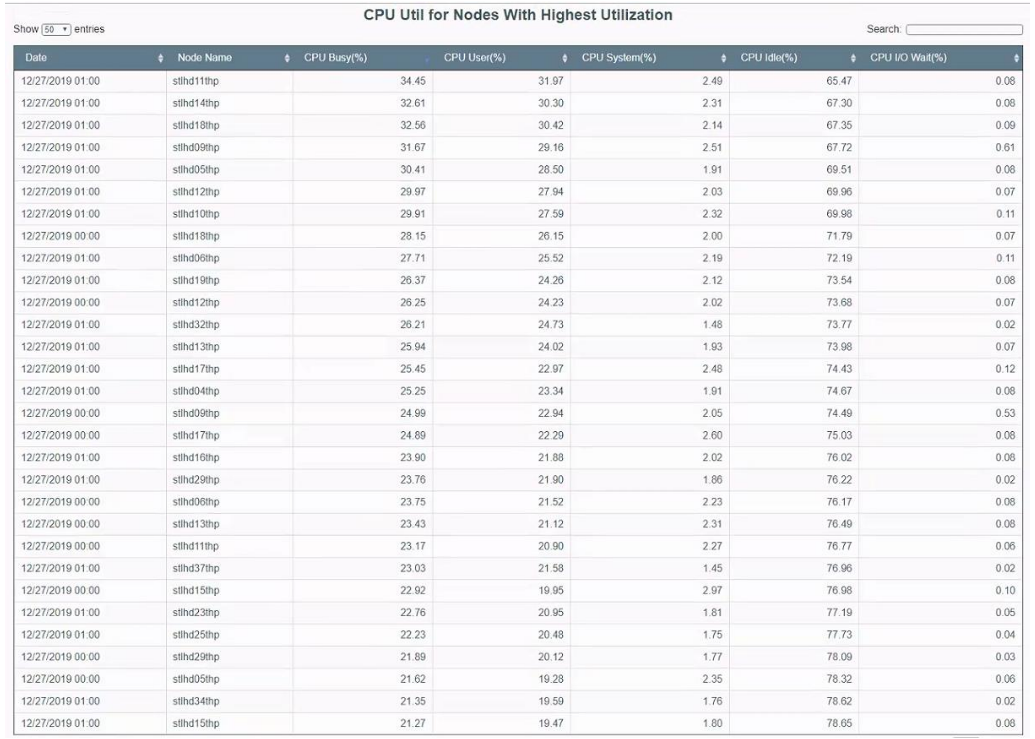
Previous 1 Next

Daily Peaks (Response Time [sec]) Indentified Between 09/01/2019 and 01/22/2020

ANALYZE BIG DATA NODE UTILIZATION VARIABILITY

RECOMMEND CHANGING OF DATA AND APPLICATIONS PLACEMENT TO IMPROVE RESOURCE UTILIZATION BALANCE

Difference in Big Data high and low nodes utilization



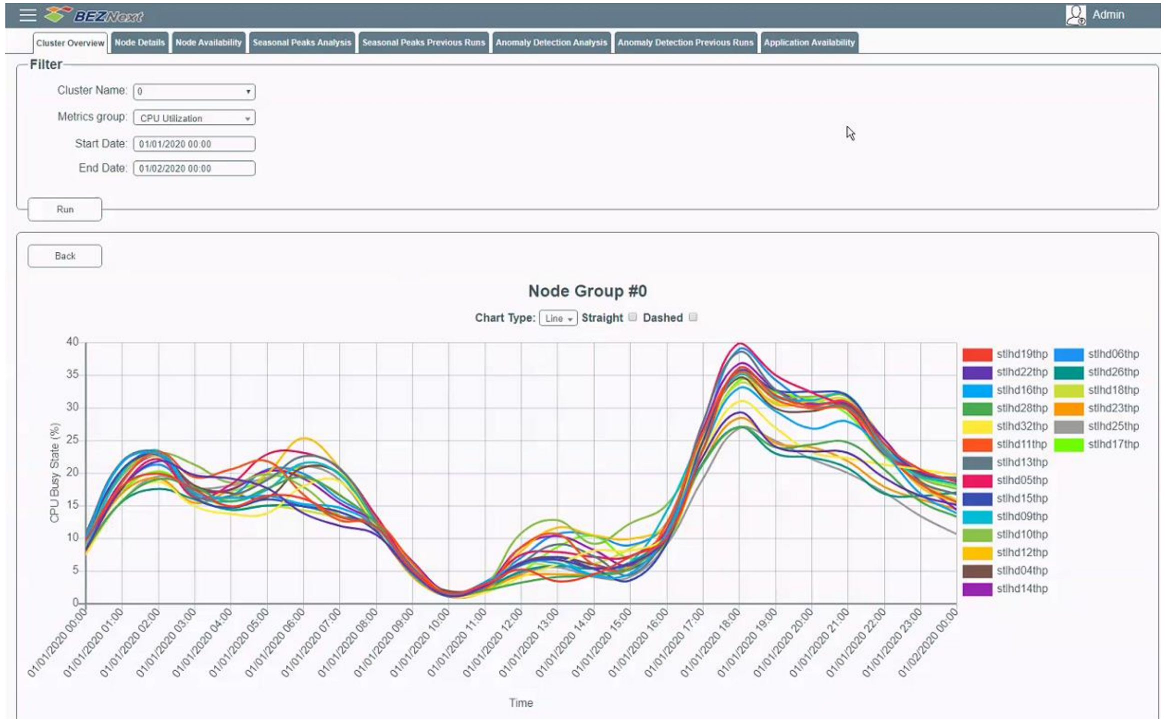
CPU Util for Nodes With Highest Utilization

Show 50 entries

Search:

Date	Node Name	CPU Busy(%)	CPU User(%)	CPU System(%)	CPU Idle(%)	CPU I/O Wait(%)
12/27/2019 01:00	stlhd11thp	34.45	31.97	2.49	65.47	0.08
12/27/2019 01:00	stlhd14thp	32.61	30.30	2.31	67.30	0.08
12/27/2019 01:00	stlhd18thp	32.56	30.42	2.14	67.35	0.09
12/27/2019 01:00	stlhd09thp	31.67	29.16	2.51	67.72	0.61
12/27/2019 01:00	stlhd05thp	30.41	28.50	1.91	69.51	0.08
12/27/2019 01:00	stlhd12thp	29.97	27.94	2.03	69.96	0.07
12/27/2019 01:00	stlhd10thp	29.91	27.59	2.32	69.98	0.11
12/27/2019 00:00	stlhd18thp	28.15	26.15	2.00	71.79	0.07
12/27/2019 01:00	stlhd06thp	27.71	25.52	2.19	72.19	0.11
12/27/2019 01:00	stlhd19thp	26.37	24.26	2.12	73.54	0.08
12/27/2019 00:00	stlhd12thp	26.25	24.23	2.02	73.68	0.07
12/27/2019 01:00	stlhd32thp	26.21	24.73	1.48	73.77	0.02
12/27/2019 01:00	stlhd13thp	25.94	24.02	1.93	73.98	0.07
12/27/2019 01:00	stlhd17thp	25.45	22.97	2.48	74.43	0.12
12/27/2019 01:00	stlhd04thp	25.25	23.34	1.91	74.67	0.08
12/27/2019 00:00	stlhd09thp	24.99	22.94	2.05	74.49	0.53
12/27/2019 00:00	stlhd17thp	24.89	22.29	2.60	75.03	0.08
12/27/2019 01:00	stlhd16thp	23.90	21.88	2.02	76.02	0.08
12/27/2019 01:00	stlhd29thp	23.76	21.90	1.86	76.22	0.02
12/27/2019 00:00	stlhd06thp	23.75	21.52	2.23	76.17	0.08
12/27/2019 00:00	stlhd13thp	23.43	21.12	2.31	76.49	0.08
12/27/2019 00:00	stlhd11thp	23.17	20.90	2.27	76.77	0.06
12/27/2019 01:00	stlhd37thp	23.03	21.58	1.45	76.96	0.02
12/27/2019 00:00	stlhd15thp	22.92	19.95	2.97	76.98	0.10
12/27/2019 01:00	stlhd23thp	22.76	20.95	1.81	77.19	0.05
12/27/2019 01:00	stlhd25thp	22.23	20.48	1.75	77.73	0.04
12/27/2019 00:00	stlhd29thp	21.89	20.12	1.77	78.09	0.03
12/27/2019 00:00	stlhd05thp	21.62	19.28	2.35	78.32	0.06
12/27/2019 01:00	stlhd34thp	21.35	19.59	1.76	78.62	0.02
12/27/2019 01:00	stlhd15thp	21.27	19.47	1.80	78.65	0.08

Big Data Top 20 Nodes Utilization in Time



Cluster Name: 0

Metrics group: CPU Utilization

Start Date: 01/01/2020 00:00

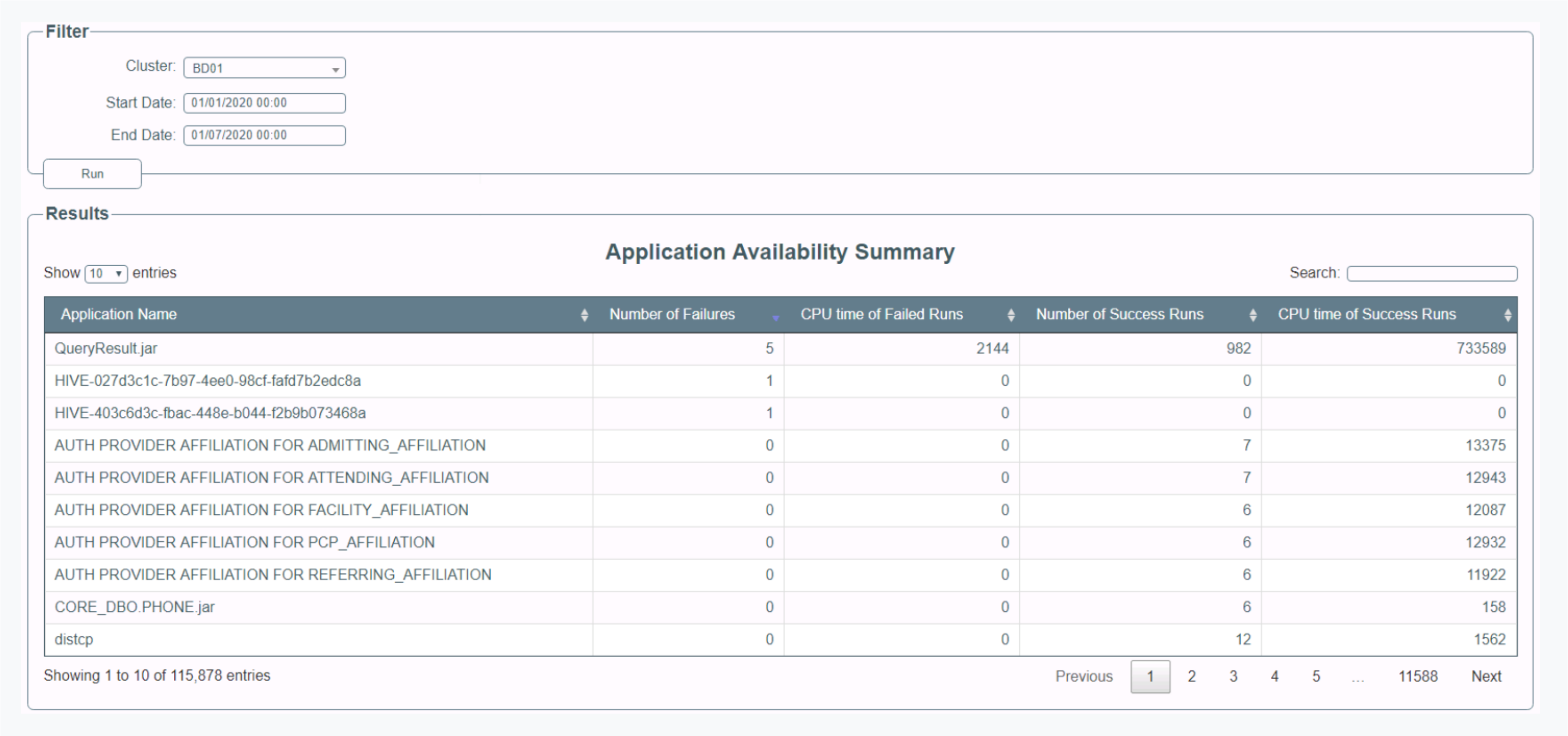
End Date: 01/02/2020 00:00

Node Group #0

Chart Type: Line Straight □ Dashed □

MEASURE APPLICATIONS AVAILABILITY

IDENTIFY APPLICATIONS WITH THE HIGHEST FREQUENCY OF FAILURES, WASTING RESOURCES - CANDIDATES FOR TUNING



Filter

Cluster: BD01

Start Date: 01/01/2020 00:00

End Date: 01/07/2020 00:00

Results

Application Availability Summary

Show 10 entries

Search:

Application Name	Number of Failures	CPU time of Failed Runs	Number of Success Runs	CPU time of Success Runs
QueryResult.jar	5	2144	982	733589
HIVE-027d3c1c-7b97-4ee0-98cf-fafd7b2edc8a	1	0	0	0
HIVE-403c6d3c-fbac-448e-b044-f2b9b073468a	1	0	0	0
AUTH PROVIDER AFFILIATION FOR ADMITTING_AFFILIATION	0	0	7	13375
AUTH PROVIDER AFFILIATION FOR ATTENDING_AFFILIATION	0	0	7	12943
AUTH PROVIDER AFFILIATION FOR FACILITY_AFFILIATION	0	0	6	12087
AUTH PROVIDER AFFILIATION FOR PCP_AFFILIATION	0	0	6	12932
AUTH PROVIDER AFFILIATION FOR REFERRING_AFFILIATION	0	0	6	11922
CORE_DBO.PHONE.jar	0	0	6	158
distcp	0	0	12	1562

Showing 1 to 10 of 115,878 entries

SUMMARY

- We reviewed how modeling and optimization technology in predicting the minimum configurations required for each Cloud to meet SLGs for growing workloads during next 12 months and how to predict the corresponding cost.
- This approach can be used for other Cloud platform
- We also reviewed role of Performance Engineering for new applications and Dynamic Performance management of production workloads in the Cloud

Thank you!

Questions?

bzibitsker@beznext.com