PERFORMANCE ASSURANCE FOR BIG DATA APPLICATIONS

Boris Zibitsker, PhD

www.beznext.com

bzibitsker@beznext.com

SPEC Presentation

BEZNext

Optimizing Business and IT

INTRODUCTION

Scope

- Development of performance assurance technology, including performance engineering and capacity management and providing performance assurance services
- Incorporation of advanced analytics including descriptive, diagnostic, predictive and prescriptive analytics during data collection, workload characterization, performance evaluation, performance management, workload management and capacity planning during application, system and data life cycle
- Development Recommender, taking into consideration responsiveness, availability and cost requirements
- Incorporation of Big Data capabilities for development enterprise performance assurance platform

PROBLEM

Problem from the Business Prospective

- Business needs
  - To make effective business decisions fast
  - To increase profitability and reduce IT cost
- Business requirements to IT
  - Activity of group of business users, customers and vendors using applications of the line of business is a workload
  - Applications should:
    - Provide Information necessary to support line of business decisions
    - Answer specific What If business questions
    - Generate prescriptions for how to make effective business decisions
  - Each Workload has Service Level Goal (SLG)
    - Responsiveness
    - Demand for resources
    - Data
  - Each Line of Business has:
    - Different budget limitations
    - Different plan of development and implementation of new applications and modification of the existing applications
    - Different plans of growth and increase in volume of data and number of users

Problem from IT Prospective

- How to design and develop Big Data applications satisfying functional and performance requirements (SLGs) of each line of business
- How to plan and cost-effectively manage Big Data infrastructure to meet SLGs of each line of business
- How to set realistic expectations
- How to continuously and cost-effectively meet Service Level Goals for each line of business

Why Rate of Deployment of Big Data Applications is Slower than Expected

Interest in Big Data is high,

so why is the rate of Big Data applications deployment slower than expected?

- Complex Technology
  - Difficult to manage
  - Security and privacy
- Applications
  - Use of advanced analytics
  - Workload growth and new applications like IoT increase demand and contention for resources
- People
  - Difficult to hire experts
- Uncertainty and Risk of Surprises
  - New applications
  - New releases of software
  - Workload management, performance management and capacity planning

SOLUTIONS

BEZNext Solutions

Performance Assurance for Big Data World = Performance Engineering + Capacity Management

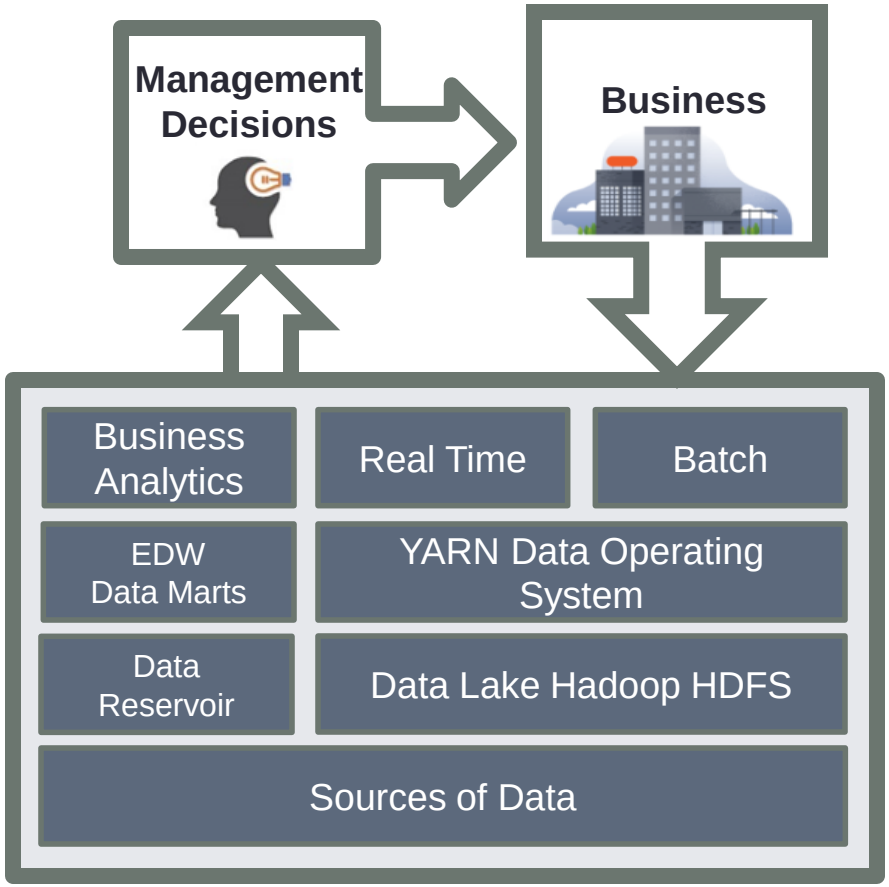
Use of Advanced Analytics

- Descriptive Analytics
- Diagnostic Analytics
- Predictive Analytics
- Prescriptive Analytics

Process

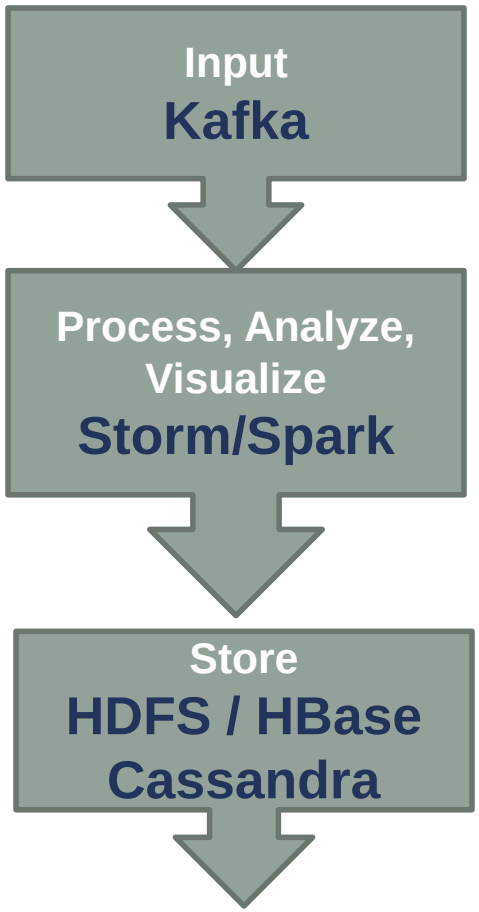
- Data Collection
- Workload Characterization
- Workload Forecasting
- Workload Management
- Performance Management
- Capacity Planning
- Verification
- Control

Hadoop/YARN Data Operating System for Big Data Workloads



- Hadoop 2.x supports concurrent real time, interactive and batch workloads
- Complex multi-tier, distributed, virtualized, parallel processing, interdependent architecture
- YARN rules control cluster resource allocation, and mix workload management

Big Data Real Time Architecture



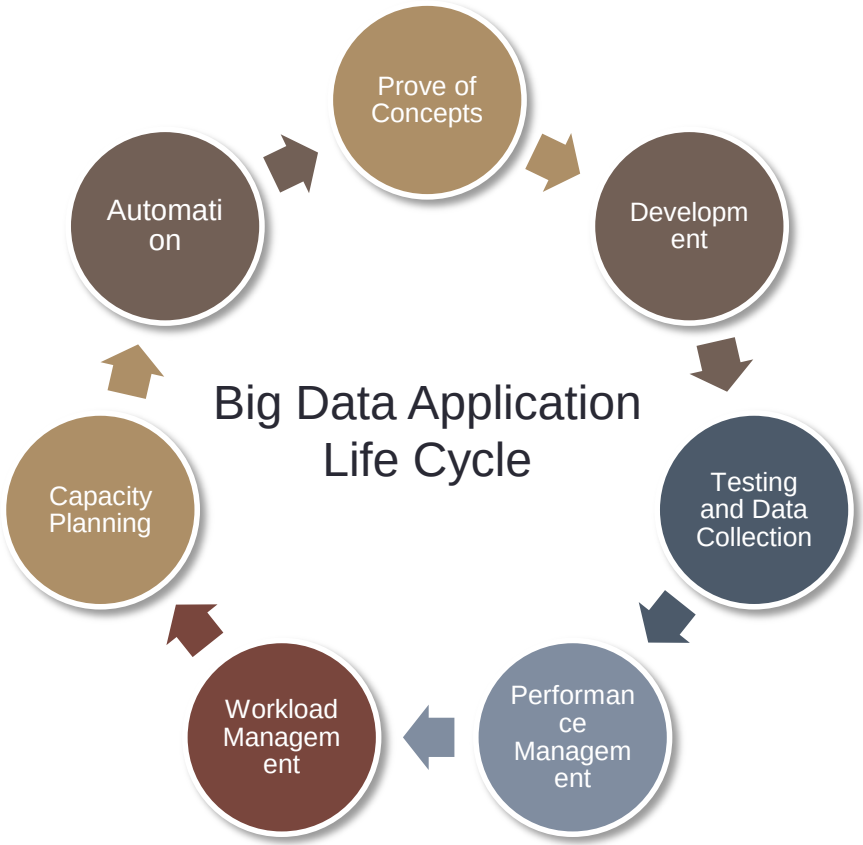
Distributed scalable publish /subscribe system for Big Data

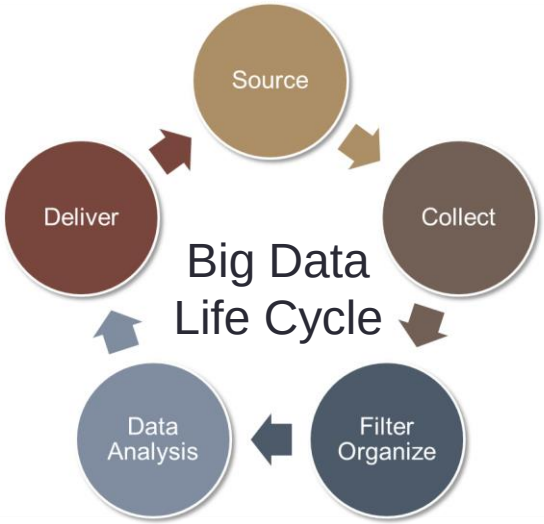
A distributed platform for doing analysis on stream of measurement data in real time

Data Lake – HDFS / HBASE
Cassandra - Open Source distributed DBMS

- Multi-tier
- Distributed
- Virtualized
- Parallel processing
- Mix workloads
- Cloud

Performance Assurance During Application, Data and System Life Cycle Affect



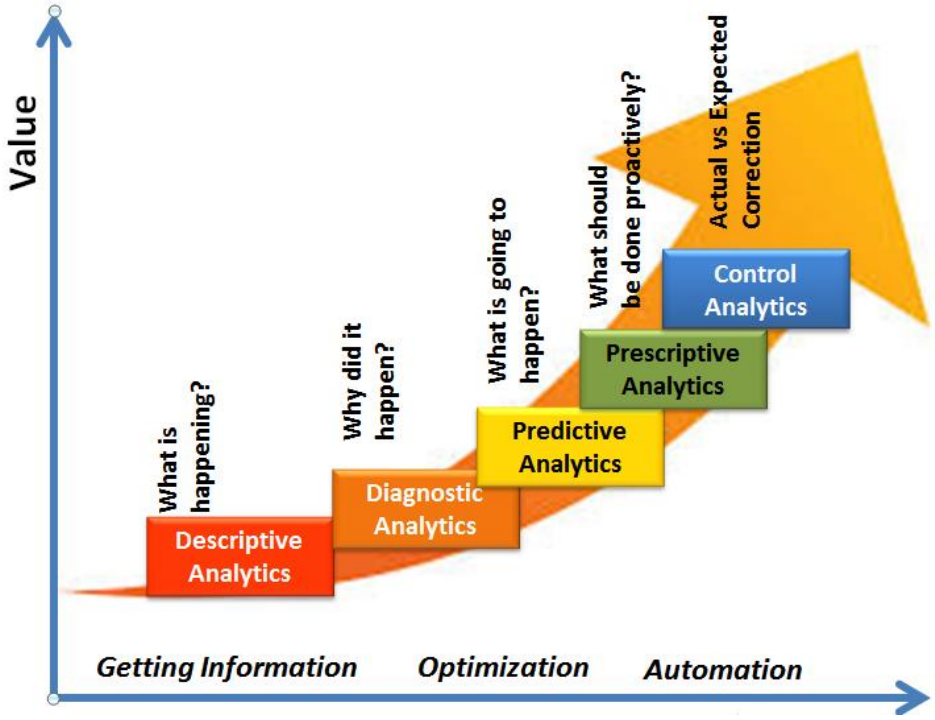


Value of Application Performance Assurance

- Optimization of design and development during application, data and systems life cycle
- Optimization of performance management and workload management
- Optimization of Big Data infrastructure
- Set realistic expectations
- Enables verification
- Business process optimization
- Predictive and prescriptive analytics enables automatic proactive performance assurance process focusing on continuously meeting SLGs
- Reduce uncertainty and risk of performance surprises
- Enables collaborative capacity management process providing better aliment between business and IT

Advanced Analytics

Decision Optimization During Application and Data Lifecycle



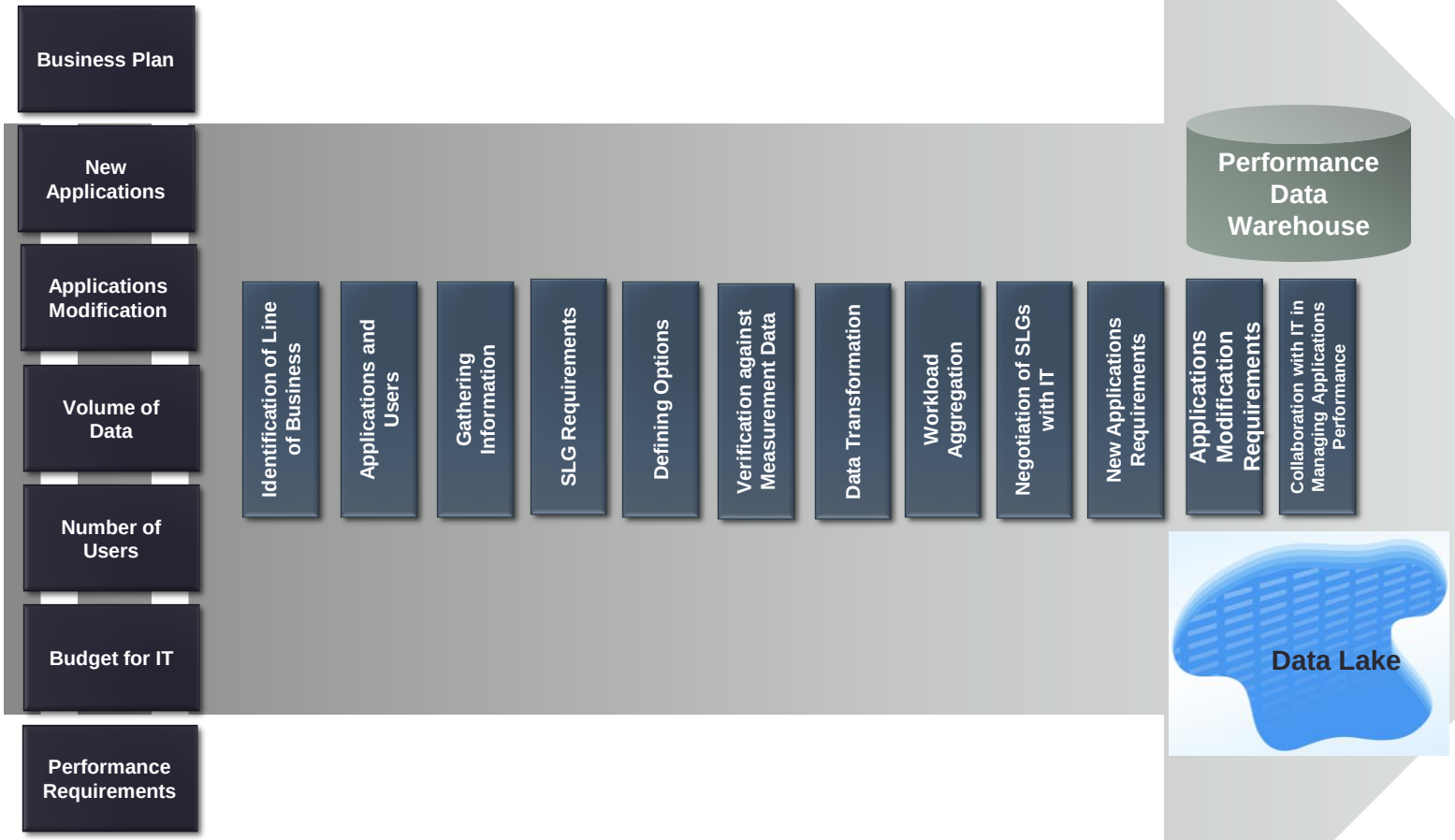
- **Descriptive analytics** to identify significant changes in applications performance, resource utilization and data usage profiles
- **Diagnostic analytics** to identify current problems and the root causes of those problems
- **Predictive analytics** to answer What If questions and to predict the outcome of anticipated changes and identify potential problems
- **Prescriptive analytics** to evaluate different options, provide proactive recommendations and generate automated advice in order to set realistic expectations
- **Control analytics** compares the actual results with expected in order to develop corrective actions and feed results into a continuous management process

DATA COLLECTION

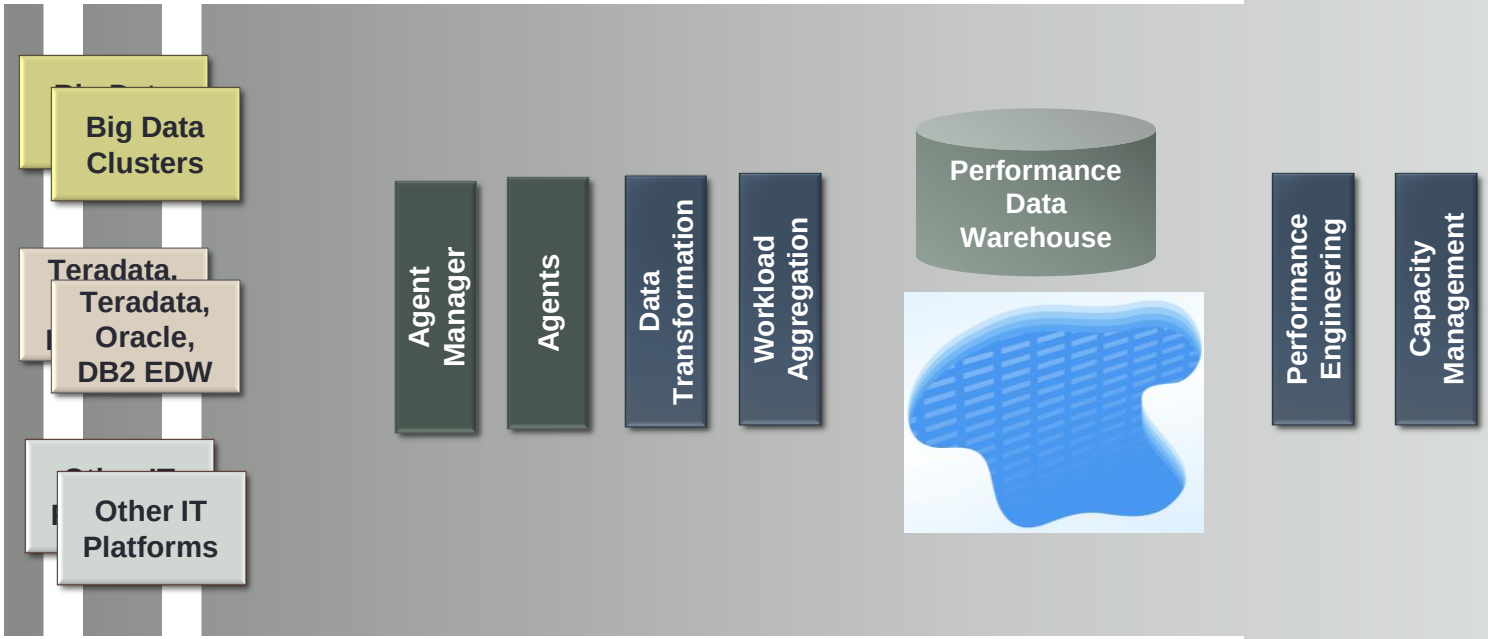
Goals

- Organize continuous data collection from different systems
- Transform and aggregate data into workloads representing line of business with ability to drill down to users, applications, and so on
- For each workload, build performance, resource utilization and data usage profiles, and calibrate profiles to make data collected from different sources correspond to each other

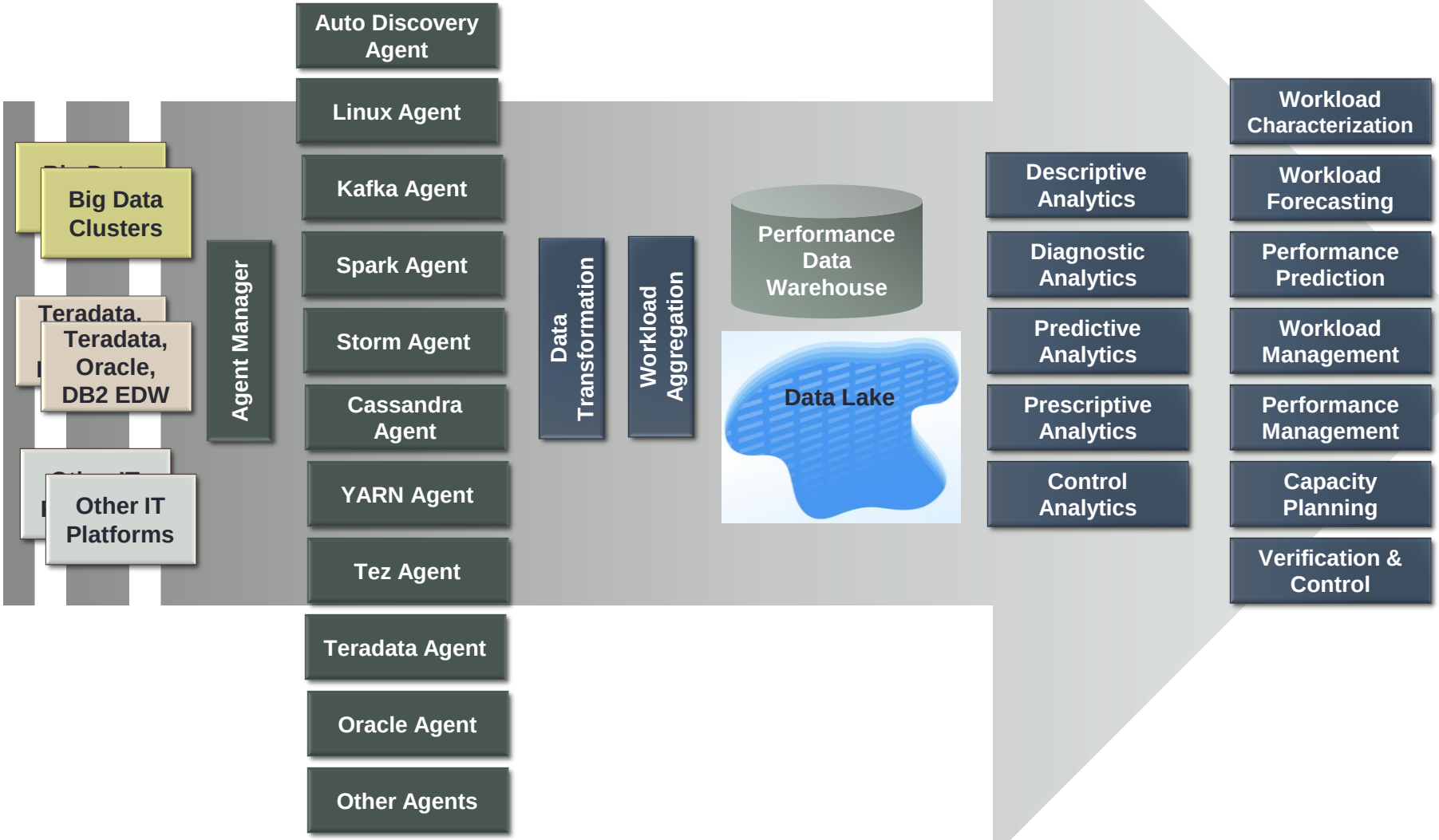
Business Data Collection Stages



IT Data Collection Stages



BEZNext Data Collection Components



Sources of Data

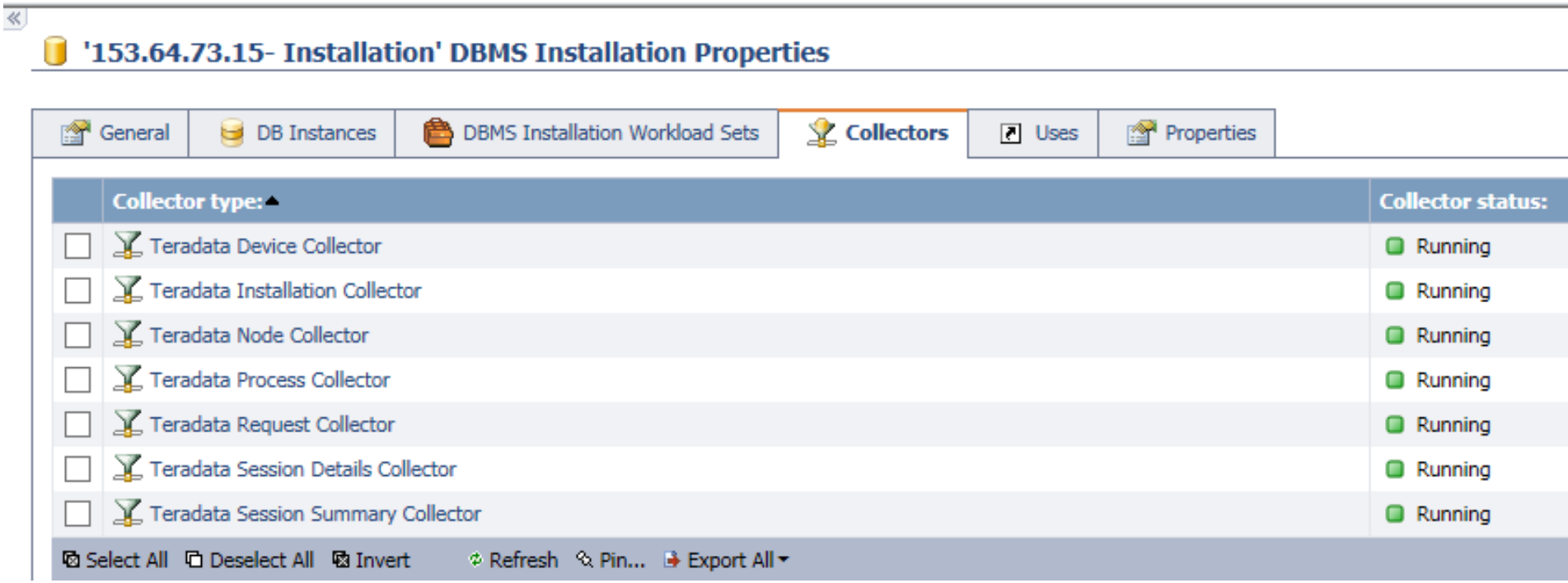
- Configuration
  - Customer
  - Ganglia / Ambari / Hadoop Hadoop and subsystems / YARN / Zookeeper
  - Linux
- Resource Consumption
  - Operating Systems
    - Linux /proc directory (CPU, memory, IO and network traffic for each host as a whole and individual processes)
    - Windows, and so on
  - HDFS NameNode (disk space)
- Performance
  - Subsystems like YARN, Kafka, Spark, Storm, Cassandra, Hbase through API and JMX
  - Teradata, Oracle, DB2, SQL Server
  - Operating system

BEZNext Agents

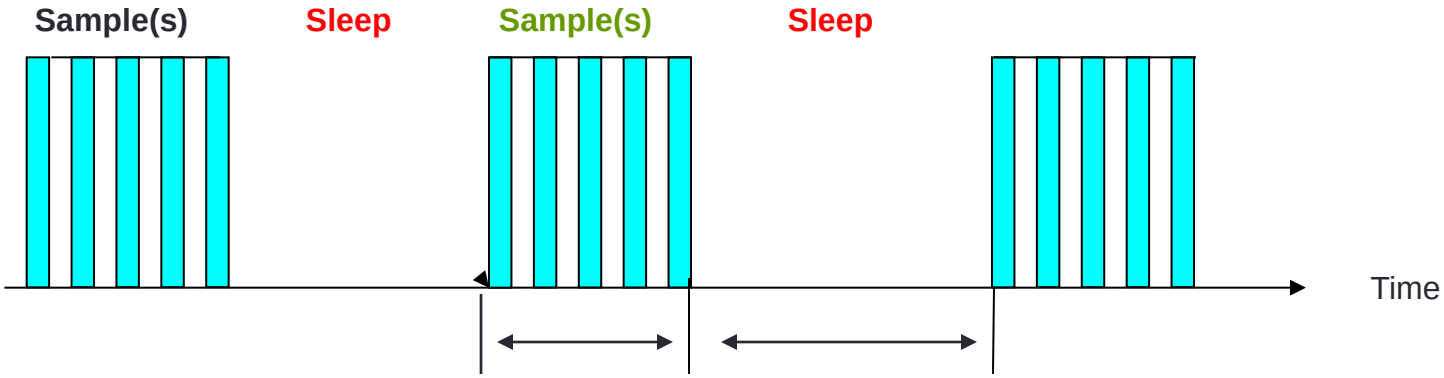
- Operating System
  - Remote connection to the monitored server
  - Data retrieval from existing 3rd party source
- Oracle
  - JDBC connection to one of the instances associated with the monitored database. Retrieval of data from GV$ tables
- Teradata
  - JDBC connection to the monitored Teradata system. Retrieval of data from Resusage, DBQL and TDWM
- Big Data
  - Retrieval of metric sets from API or JMX interfaces to each specific technology installed on the cluster (that is, YARN, Cassandra, Spark, Kafka, and so on)

BEZNext Collectors / Agents

Configuration -> Overview Level -> Detailed Activity



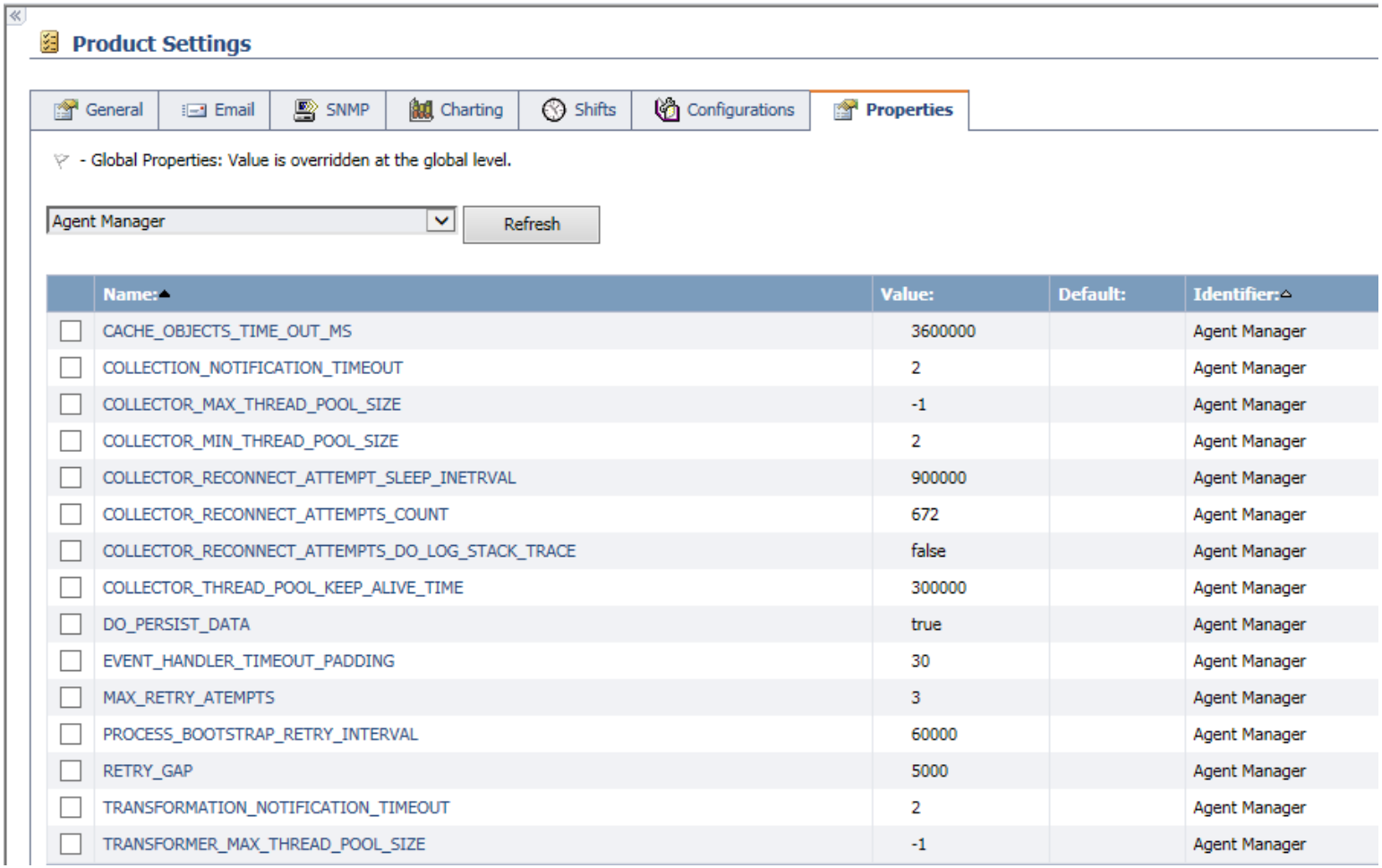
Sampling Example



Default Properties for Oracle Collection

Collector	Sleep Interval
Process	1 min
Node	5 min
Device (I/O)	5 min
Instance	15 min
Session	2 min
Request	10 sec
Response Time	15 min

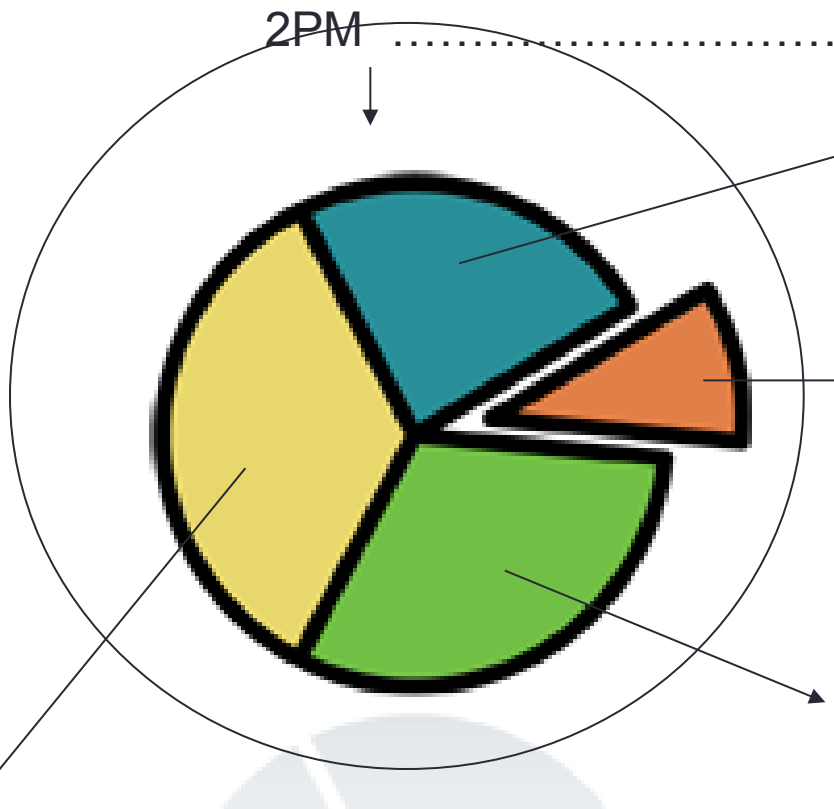
BEZVision Parameters



	Name:	Value:	Default:	Identifier:
☐	CACHE_OBJECTS_TIME_OUT_MS	3600000		Agent Manager
☐	COLLECTION_NOTIFICATION_TIMEOUT	2		Agent Manager
☐	COLLECTOR_MAX_THREAD_POOL_SIZE	-1		Agent Manager
☐	COLLECTOR_MIN_THREAD_POOL_SIZE	2		Agent Manager
☐	COLLECTOR_RECONNECT_ATTEMPT_SLEEP_INETRVAL	900000		Agent Manager
☐	COLLECTOR_RECONNECT_ATTEMPTS_COUNT	672		Agent Manager
☐	COLLECTOR_RECONNECT_ATTEMPTS_DO_LOG_STACK_TRACE	false		Agent Manager
☐	COLLECTOR_THREAD_POOL_KEEP_ALIVE_TIME	300000		Agent Manager
☐	DO_PERSIST_DATA	true		Agent Manager
☐	EVENT_HANDLER_TIMEOUT_PADDING	30		Agent Manager
☐	MAX_RETRY_ATEMPTS	3		Agent Manager
☐	PROCESS_BOOTSTRAP_RETRY_INTERVAL	60000		Agent Manager
☐	RETRY_GAP	5000		Agent Manager
☐	TRANSFORMATION_NOTIFICATION_TIMEOUT	2		Agent Manager
☐	TRANSFORMER_MAX_THREAD_POOL_SIZE	-1		Agent Manager

Transformation / Profile Creation

Hourly Profiles Building Steps



Data retrieval for the 1PM to 2PM activity takes place.

Data is aggregated in hourly summary views.

Workload assignment rules are applied using the default rule set and individual workload profiles are created.

Workload profiles are calibrated,
the “automatic” profile is created and the
1PM – 2PM data is made available in the product.

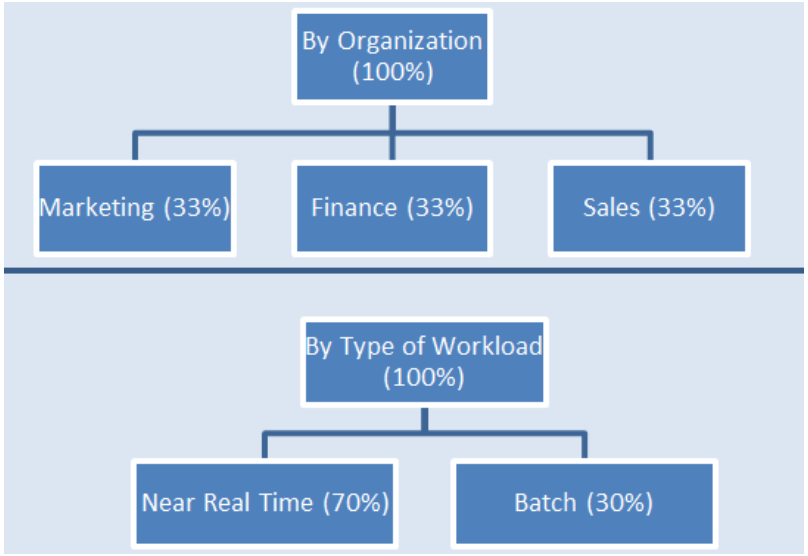
Performance, Resource and Data Usage Profiles for Each Workload

Workload Aggregation Rules
Username='salesops' and Program='finapp.exe'

Performance
Average Response Time
Throughput
Etc ..

Resource Utilization
CPU Usage
I/O Rate
Memory Usage
Network

Data Usage
Read/Write
Frequency of Data Accesses
Parallelism
Join / Sorts Operations
Etc ...



- *Workload Aggregation Rules are used to aggregate measurement data into workload*
- *Each workload has performance, resource utilization and data usage hourly profiles*
  - *Line of Business (Marketing, Finance, and so on)*
  - *Type of Activity Near Real Time, Batch, and so on*

Level of Detail

- Depends on the problem you need to solve and available sources of data
- Options:
  - By System or Node
  - By Subsystem
  - By Workload / Application / User

Frequency of Data Collection

- Frequency of sampling depends on data variability and overhead of data collection
- Options
  - Continuous data collection of basic performance and resource consumption data
  - Variable rate of collection
  - Collection of detail data only when anomaly is detected or predicted

Data Transformation for Big Data

- Group individual data samples (like every minute) into modeling intervals (like every hour)
- Summarize resources consumed by child processes up to the parent process
- Map Linux processes to users, applications and Hadoop subsystems
- Match Linux processes with subsystems' applications to create both performance and resource usage profiles
- Fill in information ("workload elements") allowing grouping individual units of work into business workloads
- Prepare configuration and workload information to import into BEZVision

What Data is Stored in Performance Data Warehouse?

- Aggregated data representing hourly workloads' performance, resource utilization and data usage profiles
- Results of auto-discovery characterizing configuration
- By streaming measurement data using Kafka, and by doing in-memory data aggregation and calculation of hourly average, STD, 95 percentile and implementing diagnostic analysis with Storm or Spark, you can reduce overhead, implement near real time capacity management and reduce the volume of data stored

Service Level Goal (SLGs)

- Performance
  - Response time, throughput
- Resource Utilization
  - CPU, memory, SSD, HDD, network
- Data usage profile
  - Read/write, parallelism, and so on
- Disk Space Usage
  - Total, allocated, used
- Availability
  - % of time when devices are available
- Reliability
  - Frequency of errors and outages, including CPU, memory, SSD, HDD, network, software and applications
- Power usage
  - Correlation between power consumption and utilization of hardware

Implementation for Big Data

- Shell scripts / Python scripts / C executables to collect Linux data on each host
- Python scripts on the remote control node to collect the whole cluster and subsystem level data and to organize continuous data collection process on changeable cluster configuration
- Java applications in scalable Kafka and Spark environment to pick data from the cluster hosts and transform
- Additional module in BEZVision to import transformed data and create performance and storage profiles

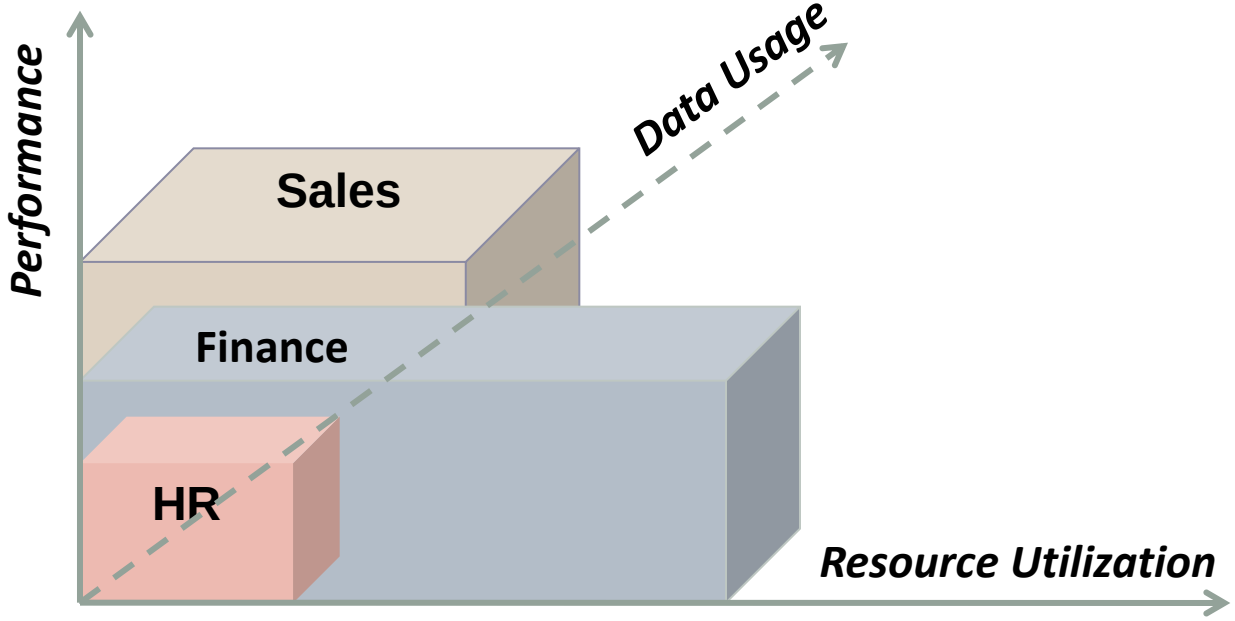
Data Collection Summary

- BEZNext Agents incorporate Big Data capability to achieve scalable solutions supporting continuous 24 X 7 data collection from distributed, multi-tier, parallelized systems, data transformation and processing
- BEZNext Agents incorporate advanced analytics to clean data and reconstruct missing data

Data Collection URL Links

- Linux
  - /proc directory: http://www.tldp.org/LDP/Linux-Filesystem-Hierarchy/html/proc.html
- YARN
  - REST API: https://hadoop.apache.org/docs/stable/hadoop-yarn/hadoop-yarn-site/WebServicesIntro.html
- Kafka
  - JMX: http://kafka.apache.org/documentation.html#monitoring
- Spark
  - http://spark.apache.org/docs/latest/monitoring.html
- Storm
  - REST API: https://github.com/Parth-Brahmbhatt/incubator-storm/blob/master/STORM-UI-REST-API.md, https://github.com/apache/storm/blob/master/STORM-UI-REST-API.md
- Cassandra
  - https://docs.datastax.com/en/cassandra/2.0/cassandra/operations/ops_monitoring_c.html
- Hbase
  - JMX: https://hbase.apache.org/metrics.html
- HDFS
  - REST API: https://hadoop.apache.org/docs/stable/hadoop-project-dist/hadoop-hdfs/WebHDFS.html
  - JMX: http://docs.hortonworks.com/HDPDocuments/HDP2/HDP-2.3.0/bk_hdfs_admin_tools/content/ch07.html

WORKLOAD CHARACTERIZATION



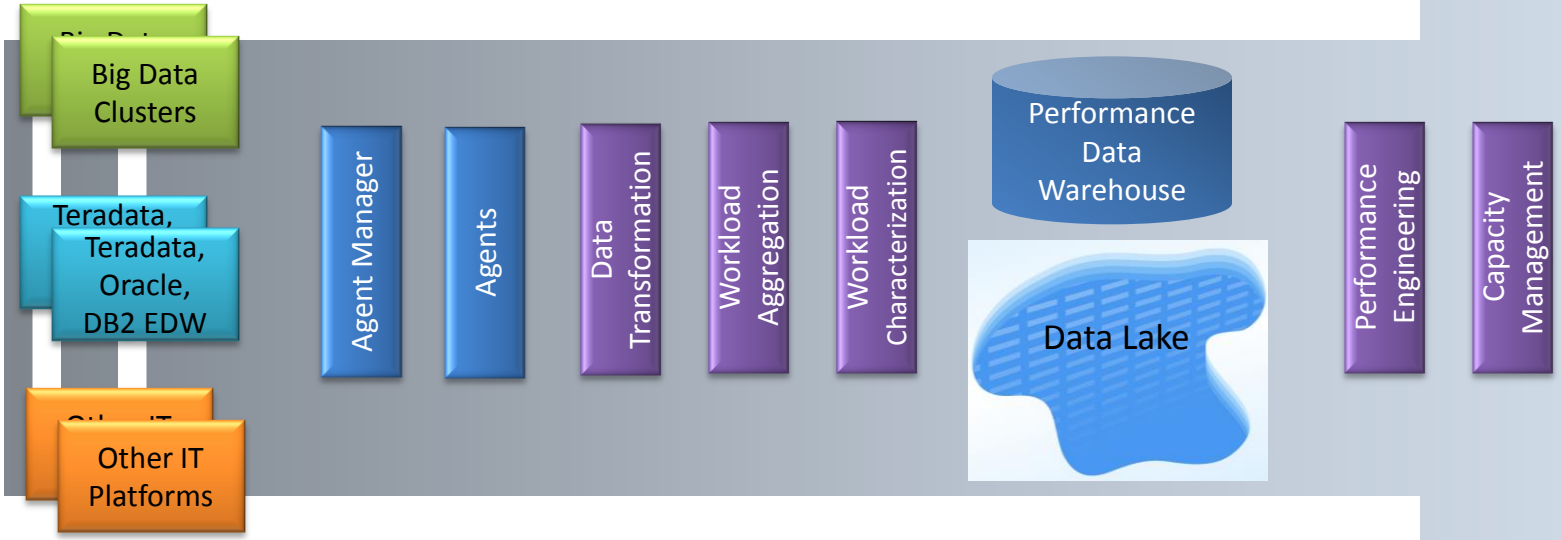
Workload Aggregation and Characterization Process

- Create Workload Aggregation rules
- Build Workload profiles
  - Performance, resource utilization and data usage profiles for each workload
- Results of Workload Characterization are used for
  - Diagnostic and root cause analysis
  - Determining seasonal peaks and workload forecasting
  - Workload management
  - Performance Management
  - Capacity planning
  - Generating prescriptions

What is a Workload?

- A workload represents aggregated activity of a group of users or applications supporting a specific line of business, business function or department
- Workloads characterization provides an integrated view of the business demand for IT Resources and Data on one hand and level of service or performance provided by IT in servicing the workload
- Each Workload has unique performance, resource utilization and data usage profiles
  - Performance profile – the average response time and throughput
  - Resource utilization profile – average CPU utilization, I/O rate, Memory and disk utilization, level of concurrency, level of parallelism and network utilization
  - The data usage profile includes the frequency and type of data access
- Increase in number of users, volume of data, implementation of new applications and modification of existing applications changes workloads' profiles

Workload Characterization Process



Aggregate Data into Workloads

Each workload represents a line of business, business process, department or group of users

Metrix

Total CPU Seconds Consumed

Total I/O Operations

Total number of Requests - Throughput (requests/second)

Parallel Sessions (concurrent connections)

Delay Time (seconds)

Performance Profile

Response time
Throughput

Resource Utilization Profile

CPU
I/O
Memory
Internode communication

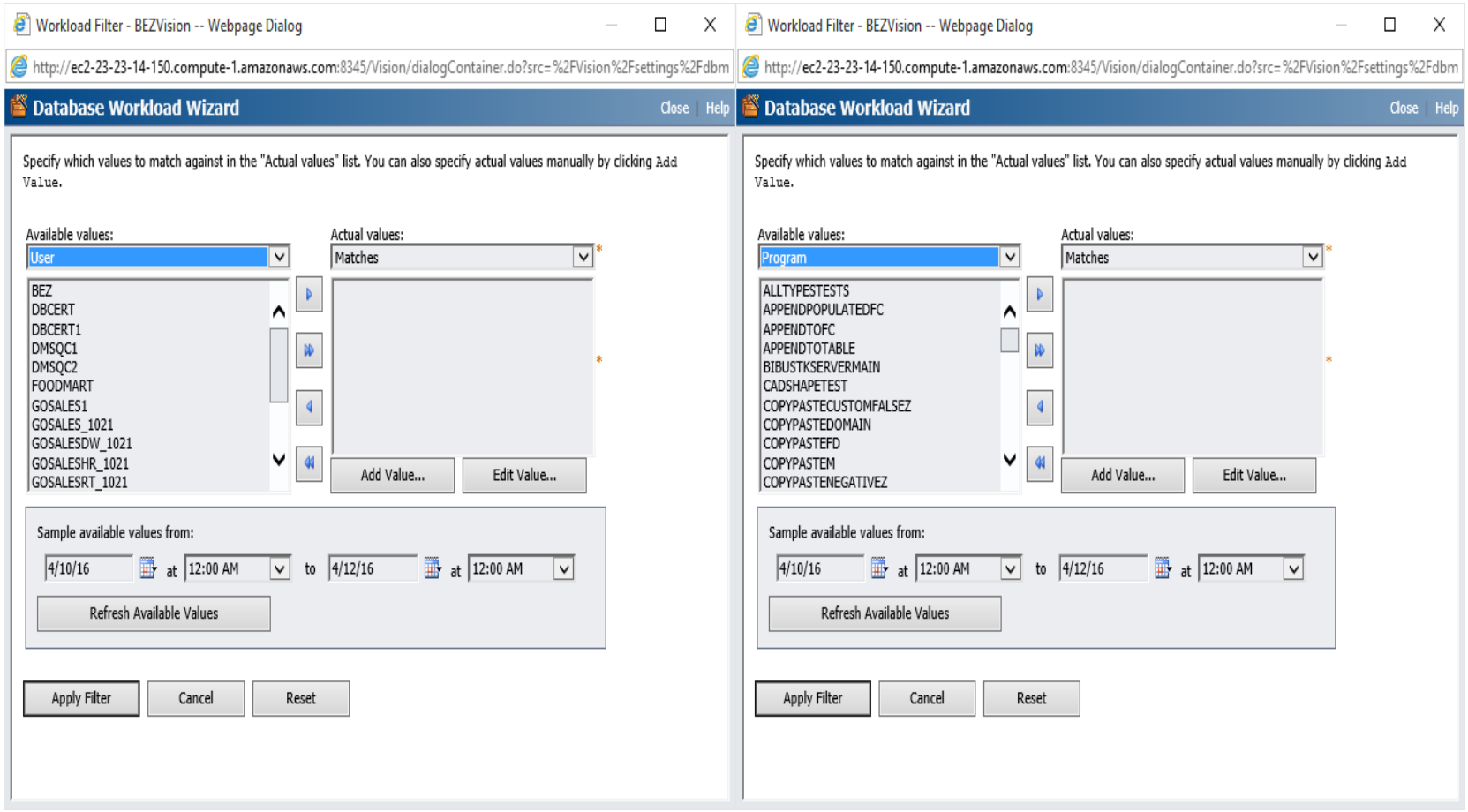
Data Access Profile

Read/Write
Level of parallelism

Workload Aggregation Rules Hourly Profiles

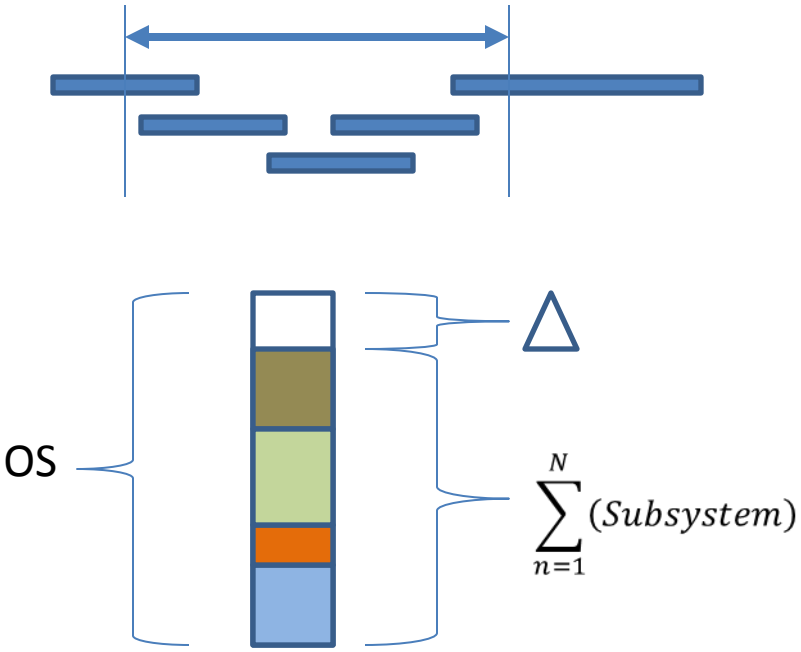
- Input for workload aggregation: transformed measurement data and Workload Aggregation Rules (WAG)
- WAG use:
  - Users' names, application/program names or other common parameters
  - Cluster analysis results based on performance and usage of resources
- WAG aggregate detail measurement data into business workloads

Workload Aggregation (WAG)



Workload Aggregation Challenges

- End effect
- Distribute delta between OS and subsystem measurement data between workloads or create a separate workload for OS own activity, or unrecognized activity (misc workloads in BV)
- Coordination of workloads between tiers, clusters



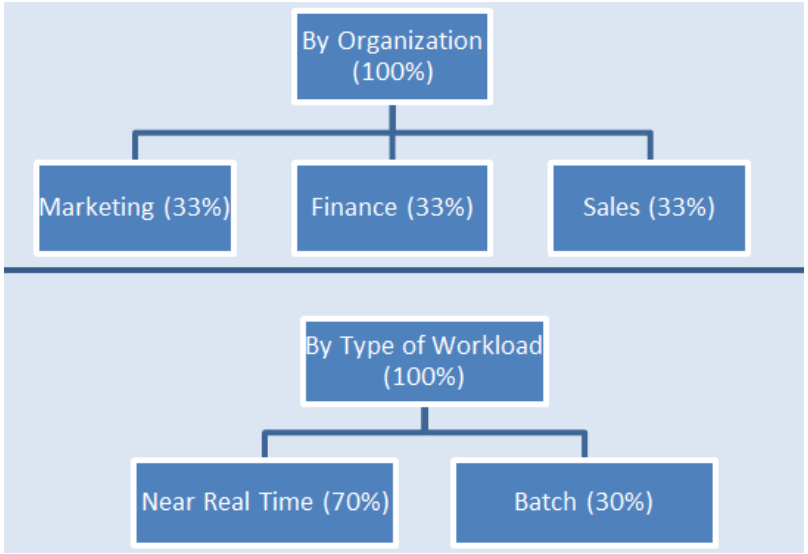
Performance, Resource and Data Usage Profiles for Each Workload

Workload Aggregation Rules
Username='salesops' and Program='finapp.exe'

Performance
Average Response Time
Throughput

Resource Utilization
CPU Usage
I/O Rate
Memory Usage
Network

Data Usage
Read/Write
Frequency of Data Accesses
Parallelism
Join / Sorts Operations
...



- Workload aggregation rules are used to aggregate measurement data into Workload
- Each workload has performance, resource utilization and data usage hourly profiles
  - Line of business (Marketing, Finance, etc.)
  - Type of activity (near real time, batch, etc.)

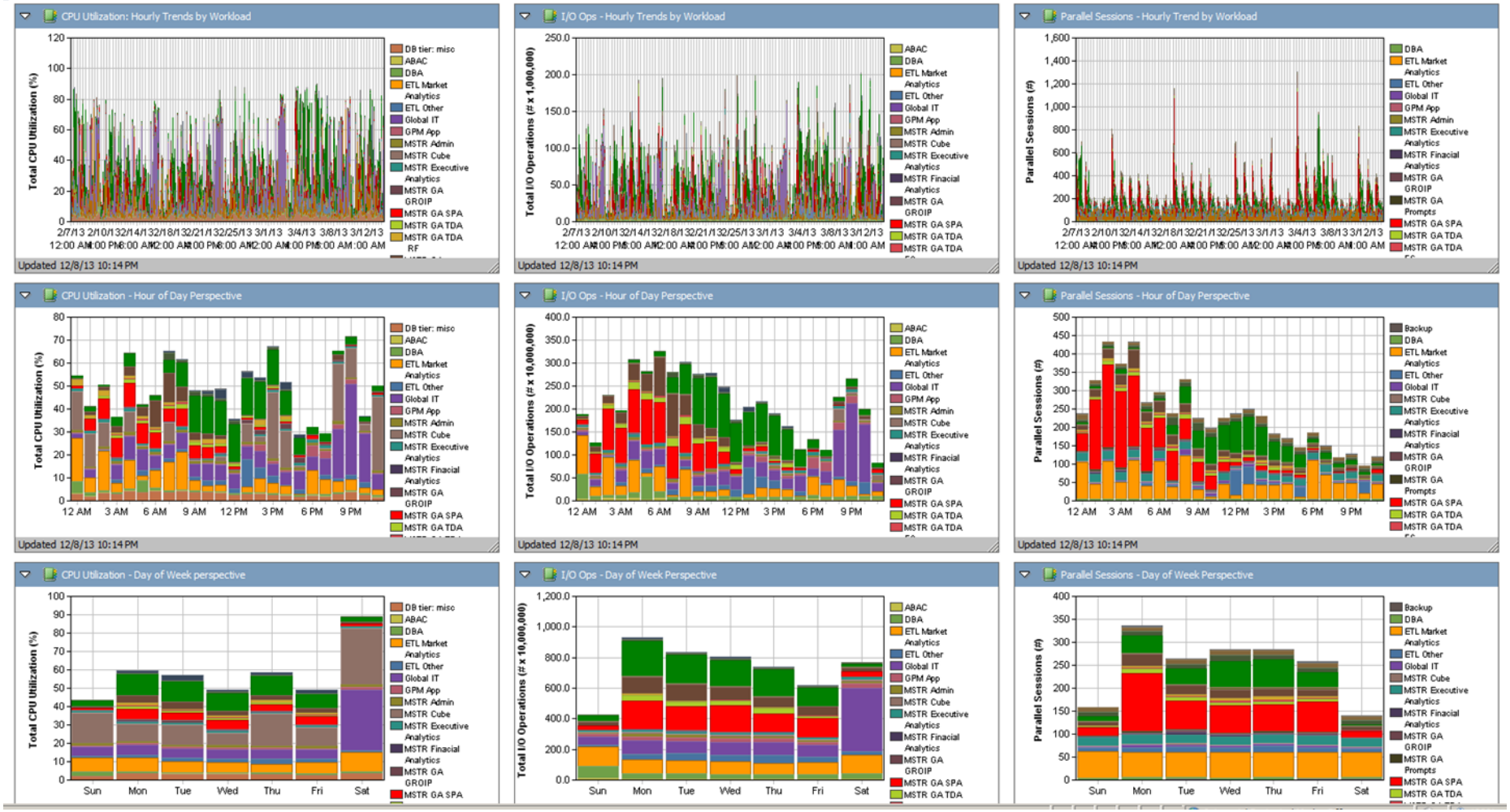
Output of the Workload Characterization

- Workload characterization is performed continuously 24 X 7
- Performance, resource utilization and data usage profiles are created hourly for each workload.
- Performance profile includes workload average response time, user think time and throughput during different representative time intervals; for example, prime shift during holiday season, prime shift end of month processing and typical week day
- Workload's resource usage profile of each workload includes average number of active users, average priority of requests within the workload, average CPU utilization by application, inter-node utilization, I/O rate to disks, read/write ratio, average level of parallelism
- Workload's sata usage profile includes the list of files, databases and tables accessed by applications
- Disk space usage is determined periodically
- Advanced analytics identify the trends and significant changes in performance, usage of resources and data, enable root cause analysis and focus performance tuning on the most critical problems affecting performance of the most critical workloads

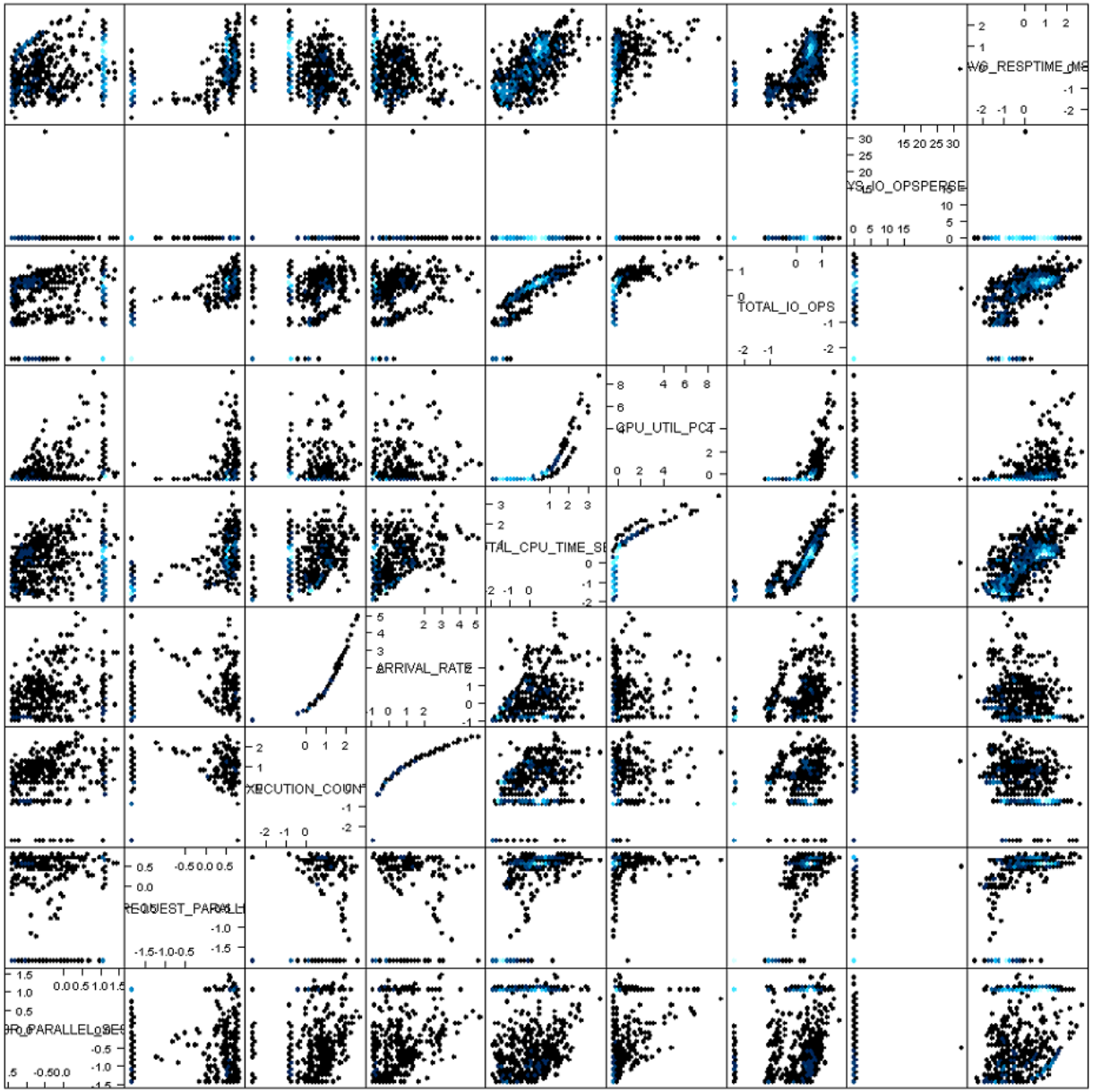
Demo

- **Examples of workload characterization results**
- **Examples of rules for data aggregation**

Workloads Profiles

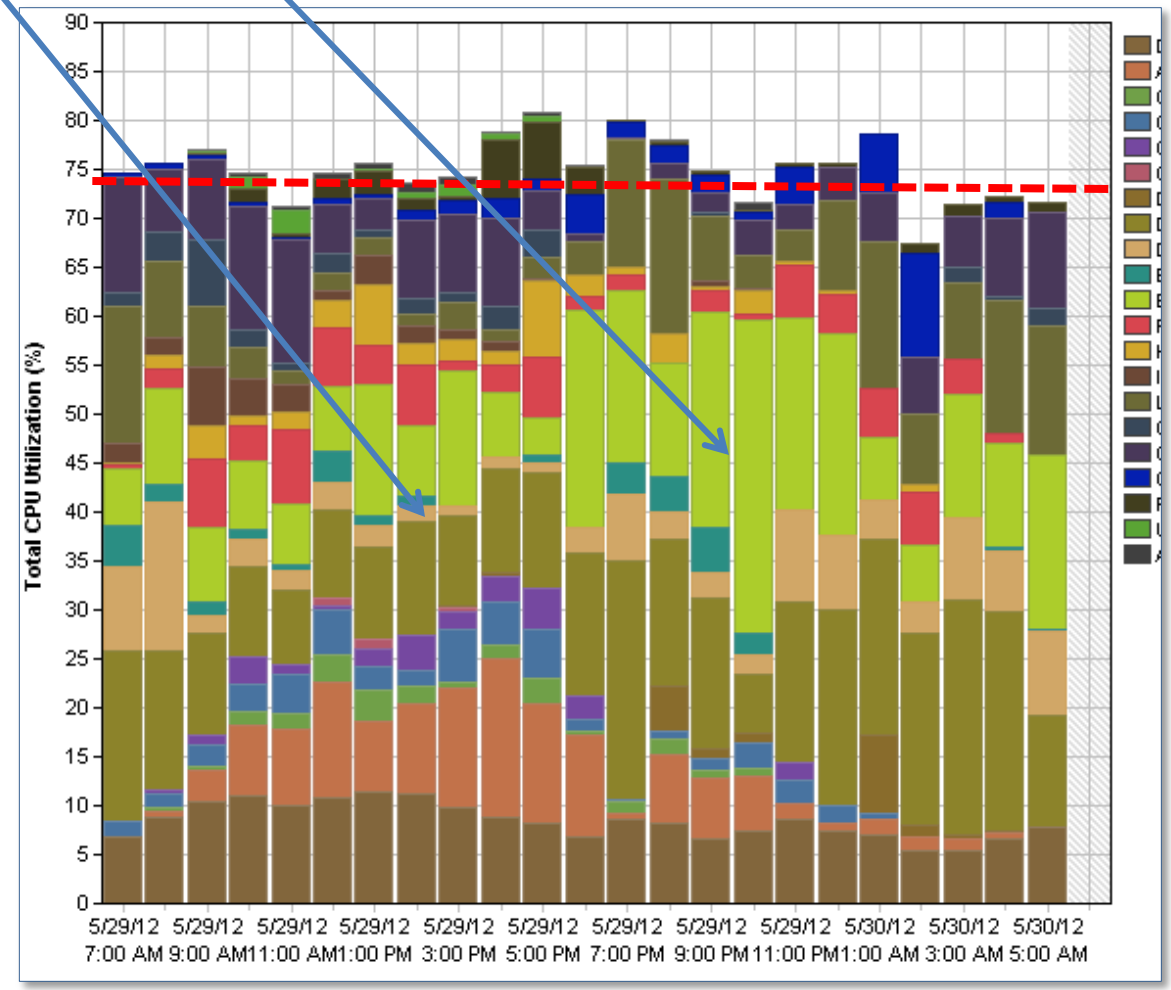


Scatter Plot Matrix Grouped / binned into hexagonal tiles

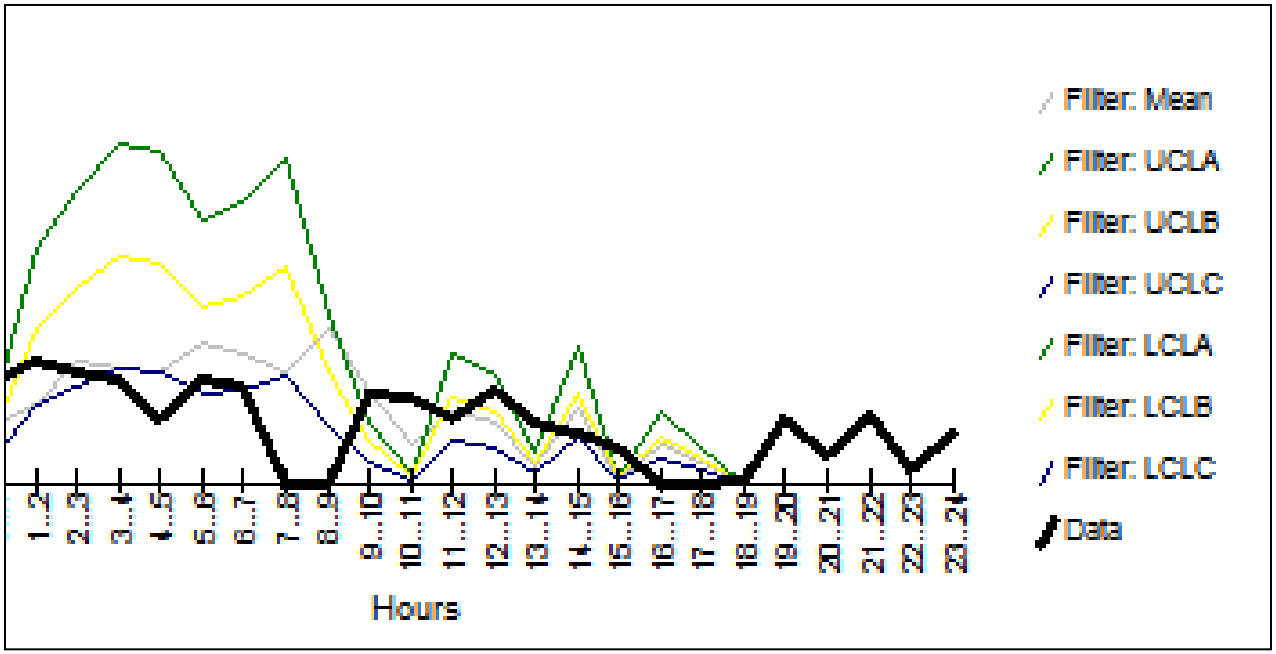


CPU Utilization by Business Workloads

ETL Sales and ETL Marketing Use Almost 40% of Resources



Determining Anomalies -
Statistical Process Control



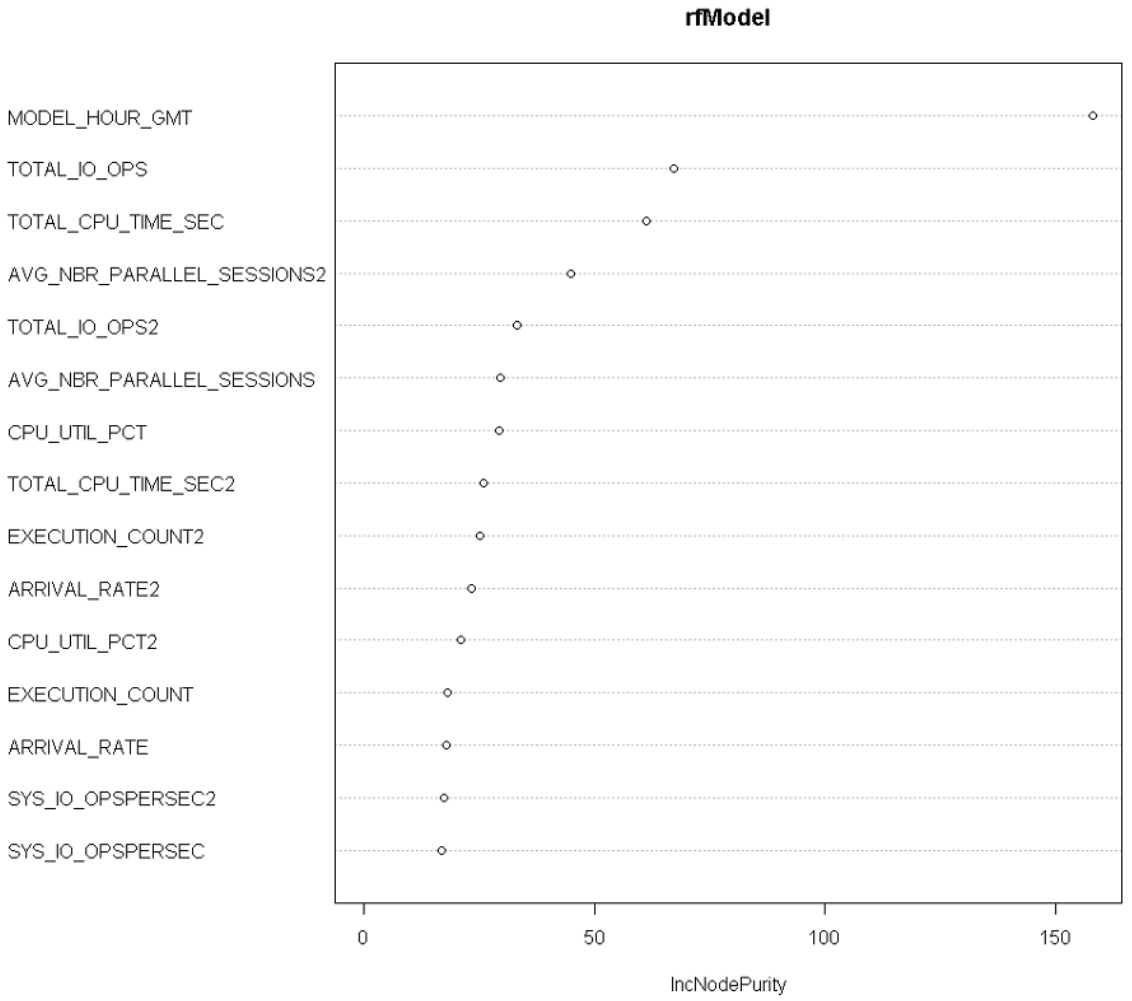
Root Cause Analysis

Random Forest

In this example, using a Random Forest model yielded a model with similar fits, but different insights into the data.

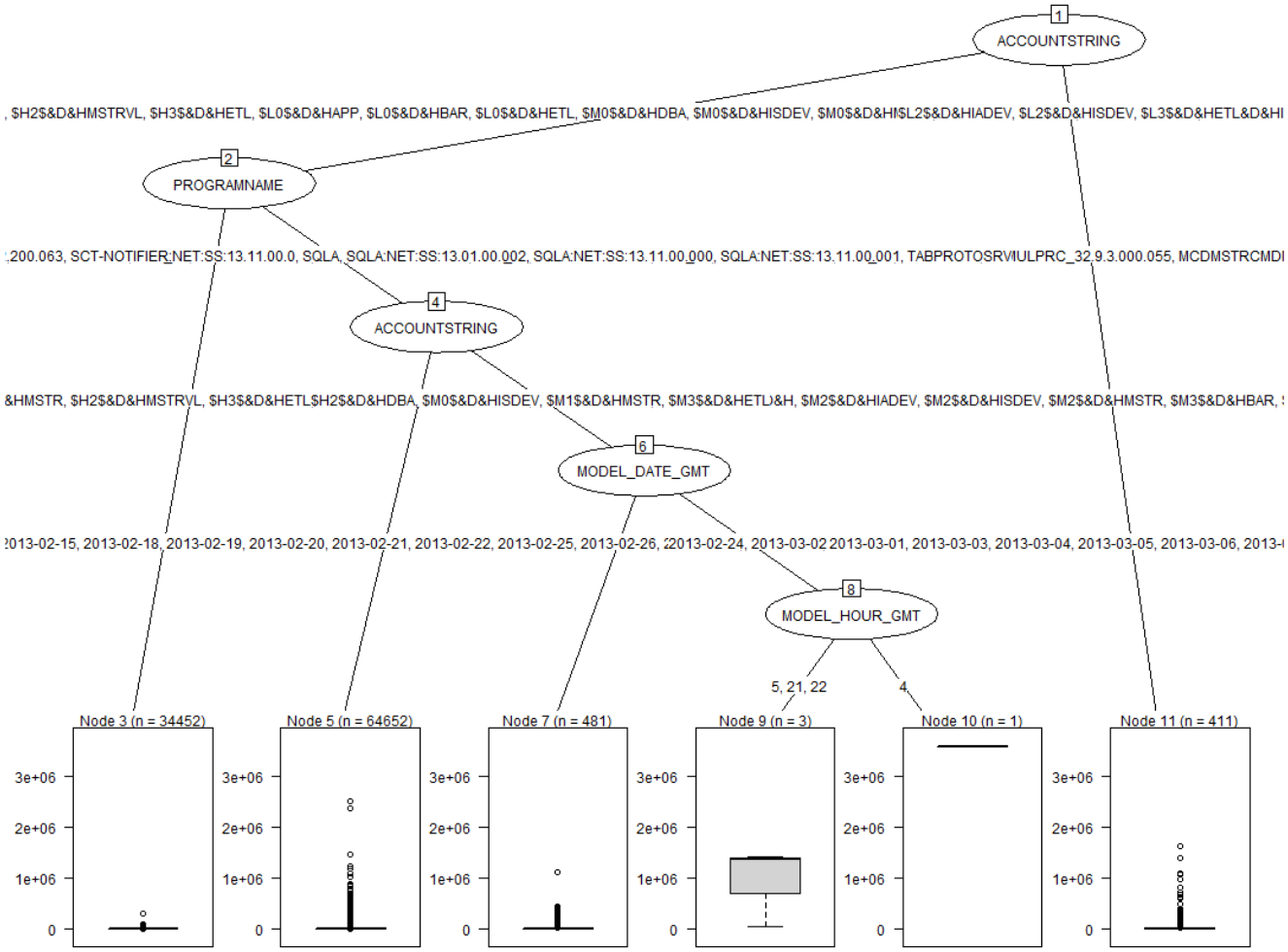
We can see which variables the model found to be important.

Random Forest models use an ensemble of trees to make predictions.



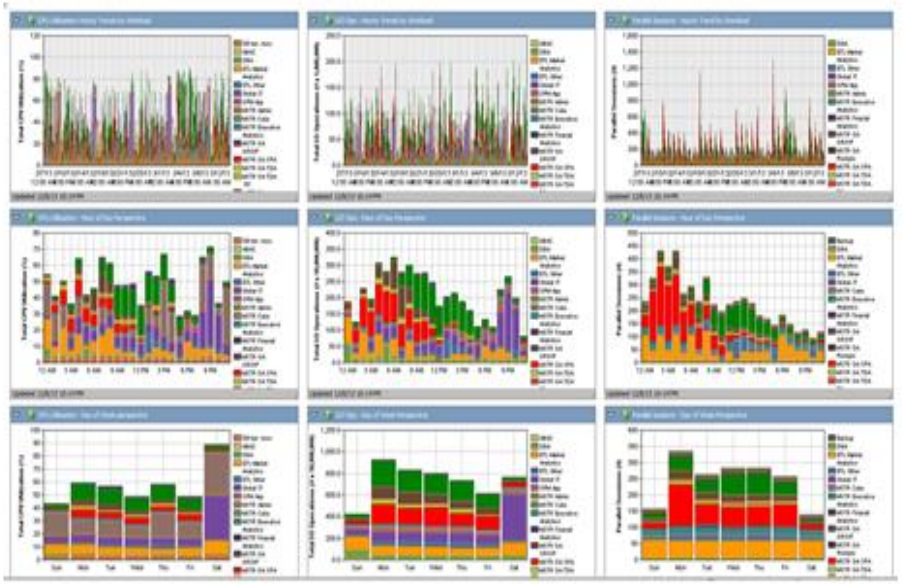
Root Cause Analysis – Decision Tree

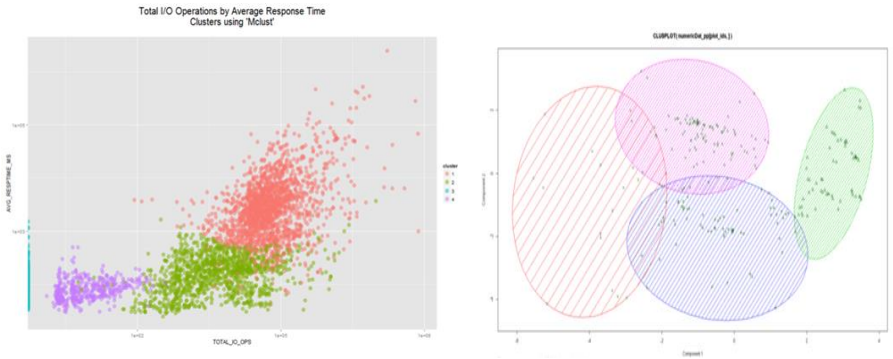
Leaf page and branches identify the cause



Workload Characterization

- Data Aggregation
- Building workloads' profiles
  - Performance
  - Resource utilization
  - Data usage
- Results are used as input for:
  - Workload forecasting
  - Performance management
  - Workload management
  - Capacity planning

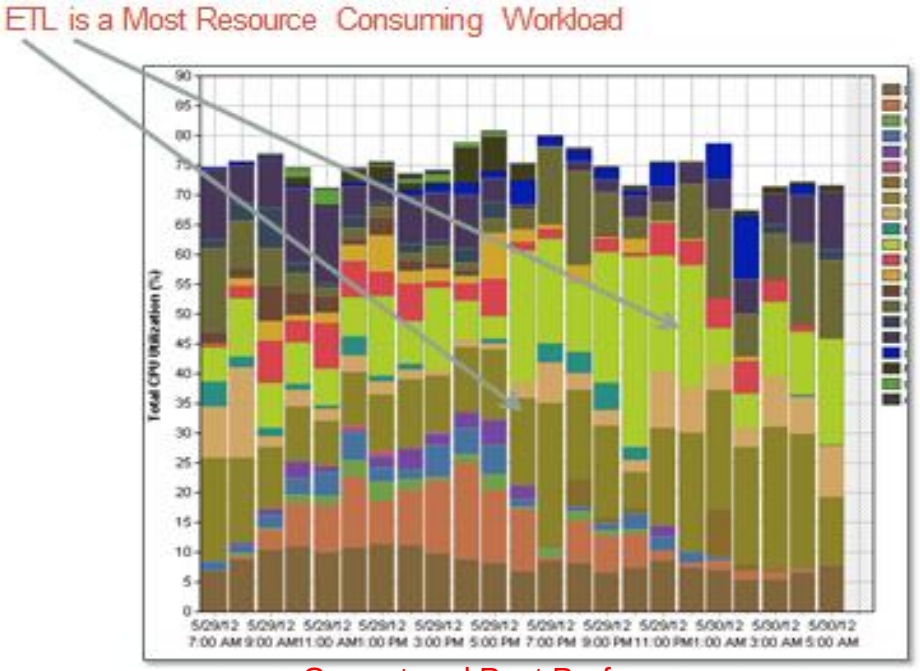




PERFORMANCE MANAGEMENT

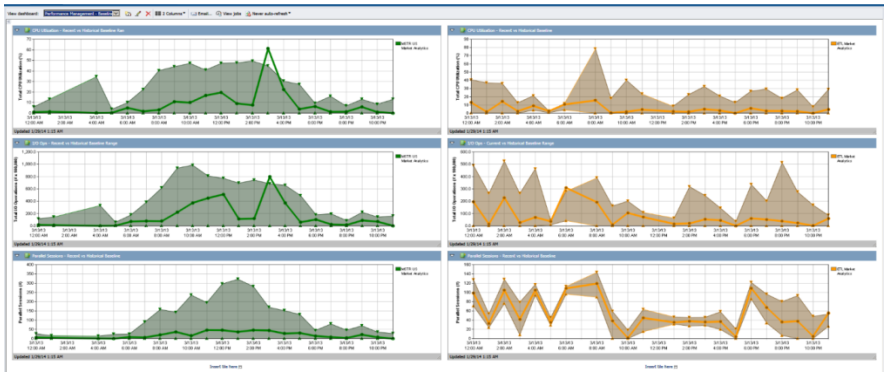
Performance Management

- Descriptive Analytics
  - Current and past performance
- Diagnostic Analytics
  - Anomalies detection
  - Root cause Analysis
- Predictive Analytics
  - Discover future bottlenecks

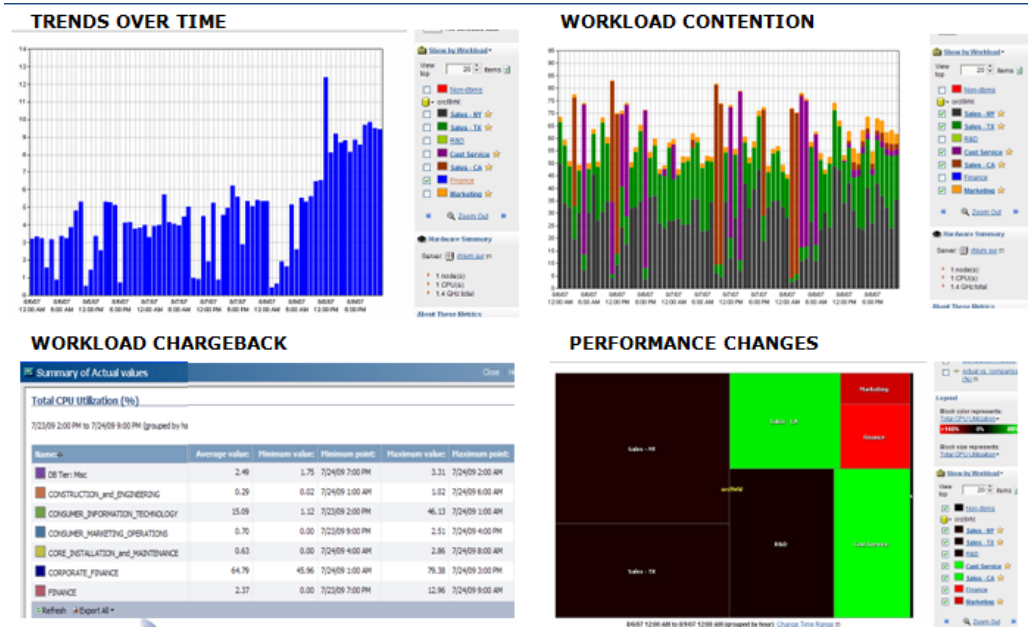


Current and Past Performance

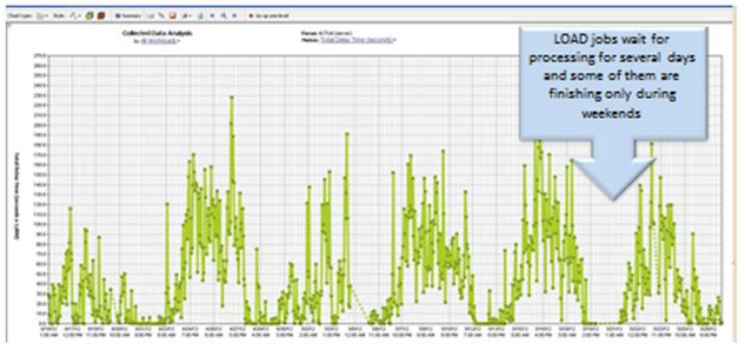
Diagnostic analytics identifies significant changes in performance and resource utilization of the individual workloads



Performance Analysis



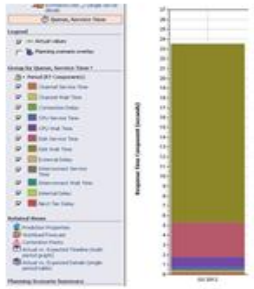
Long Time to Load Data and Inconsistent Response Time



Tactical Performance Management Options

- Software configuration
- Database tuning
- Application tuning
- Balancing workloads between systems

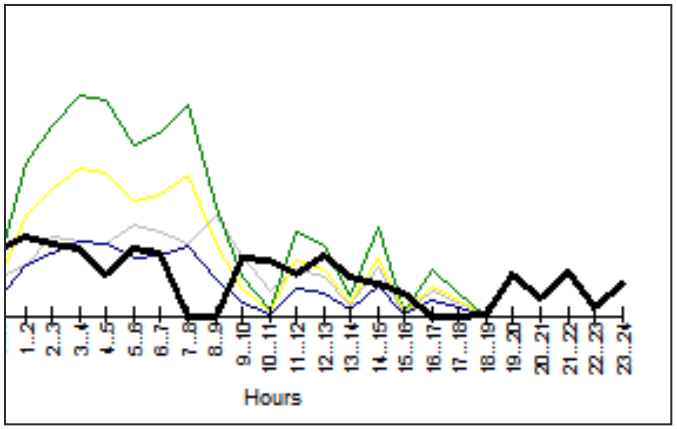
Example of Root Cause Analysis

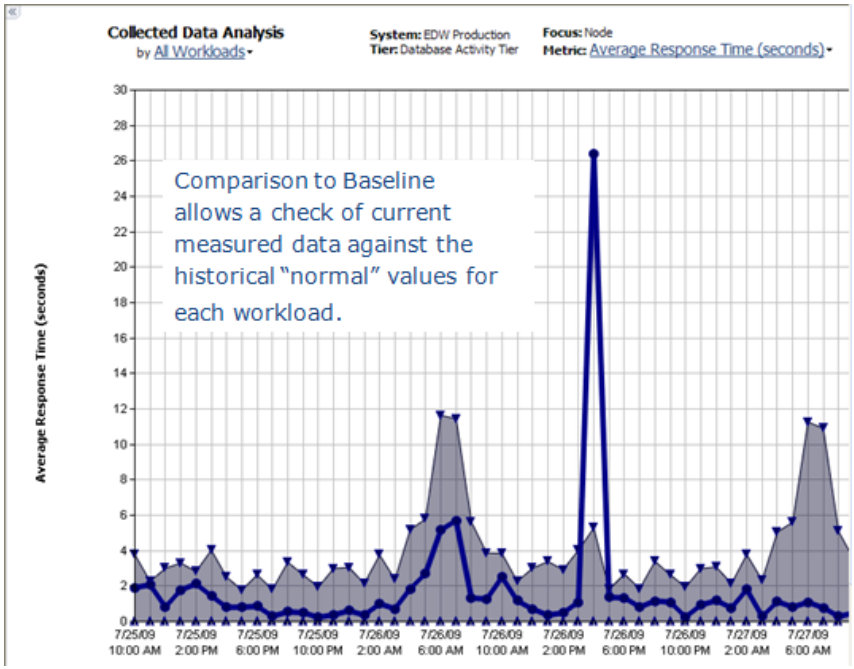


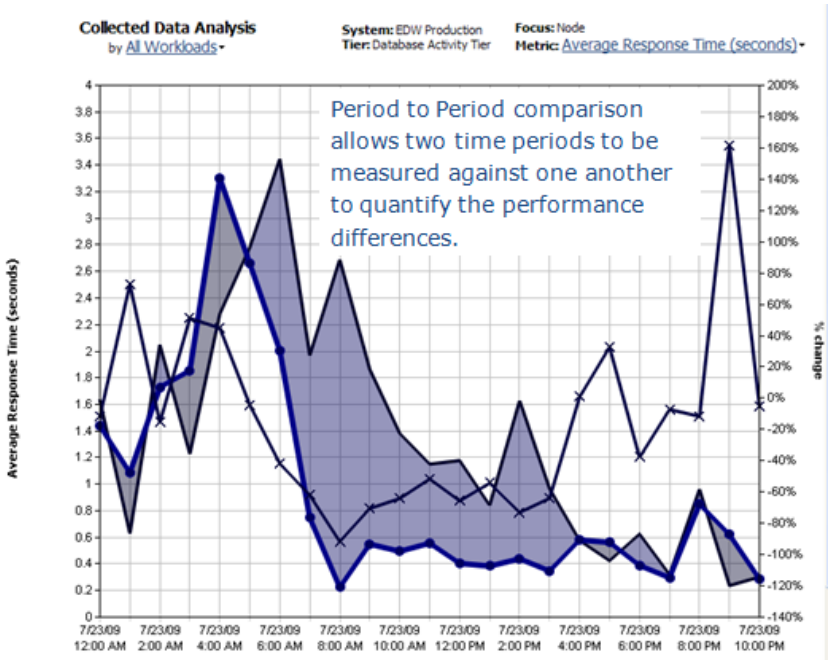
The largest component of the Load Response Time is Disk Wait Time

Anomaly Detection

- Diagnostic analytics identifies anomalies
  - Determining significant Changes with RT, throughput and resource utilization diagnostic analytics

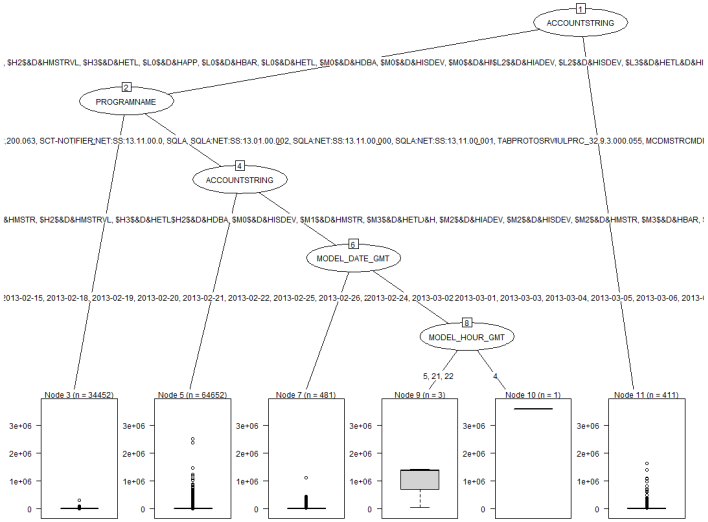






Root Cause Analysis

- Determine causes of performance degradation
  - Decision trees and
  - Logistic regression analysis
- Predict future bottlenecks
  - Predictive Analytics



Decision Tree - Leaf page and branches identify the root cause

WORKLOAD MANAGEMENT

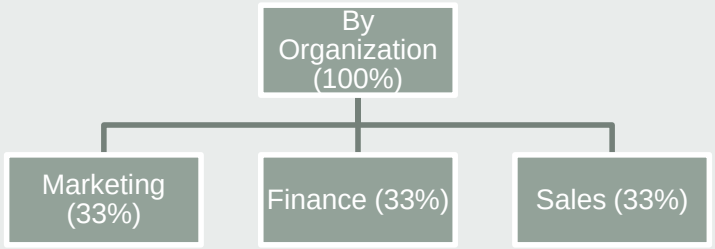
Priority

Resource Allocation

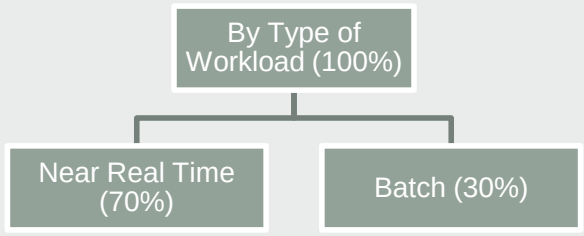
Concurrency

Example of Workload Management by Queues

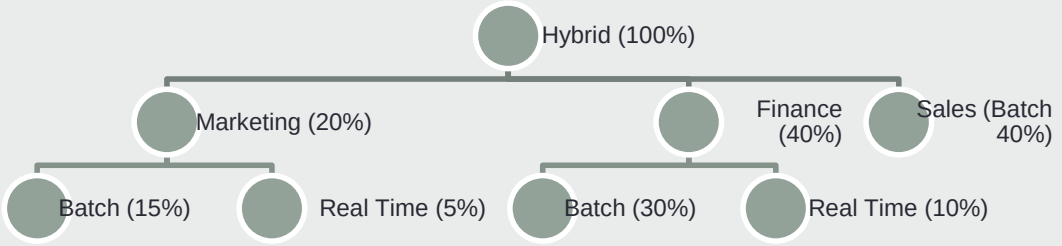
- By Organization 100%
  - Marketing 33%
  - Finance 33%
  - Sales 33%



- By Type of Workload 100%
  - Near Real Time 70%
  - Batch 30%

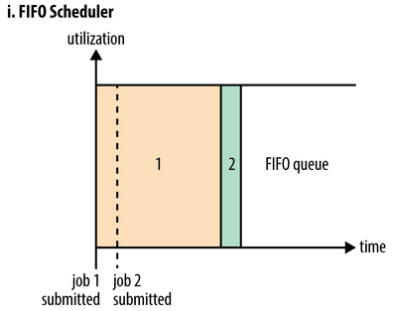


- Hybrid 100%
  - Marketing 20%
    - Batch 15%
    - Real-Time 5%
  - Finance – 40%
    - Real-Time 10%
    - Batch 30%
  - Sales – Batch 40%

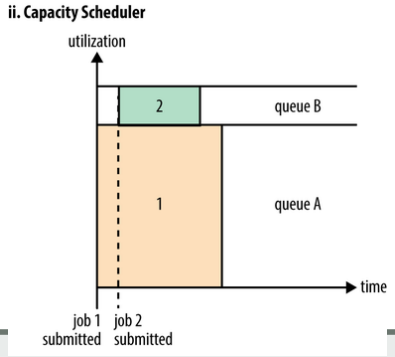


Resource Manager Schedulers

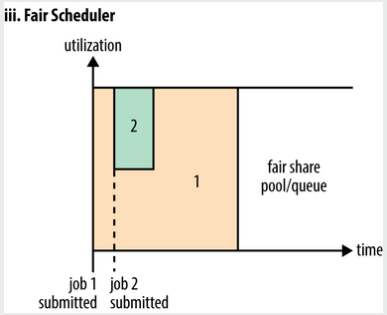
- FIFO Scheduler
  - Processing jobs in order

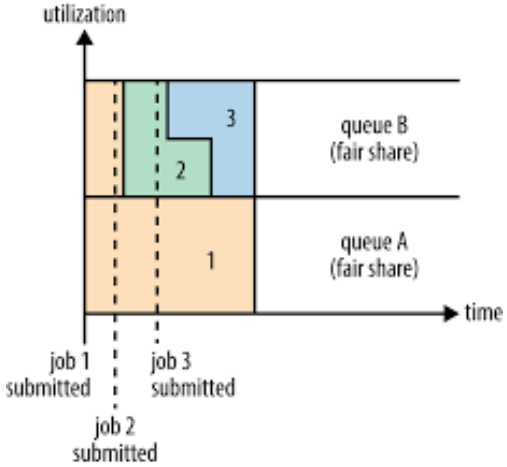


- Capacity Scheduler (Default)
  - Queue shares as percentage of clusters
  - FIFO scheduling within each queue
  - Supporting preemption

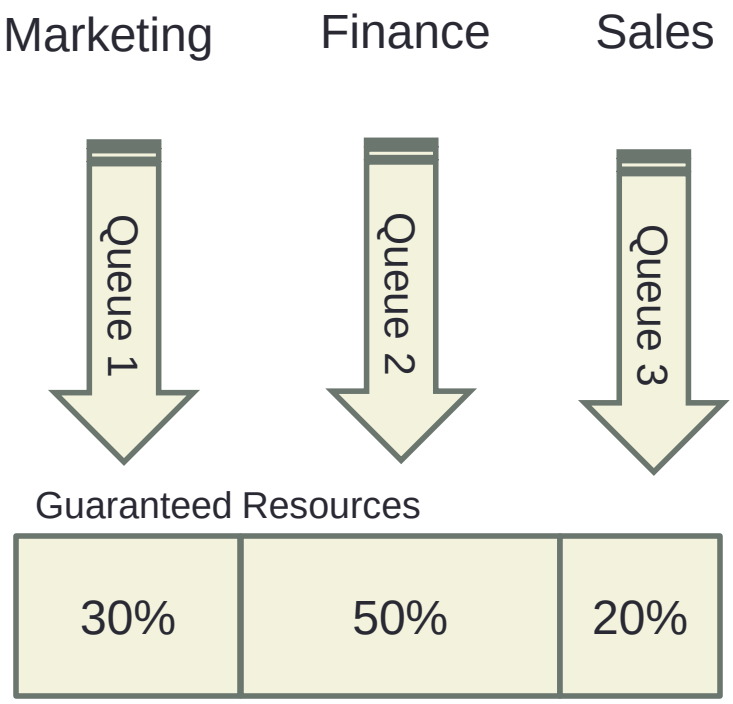


- Fair Scheduler
  - Fair to all users





YARN Capacity Scheduler



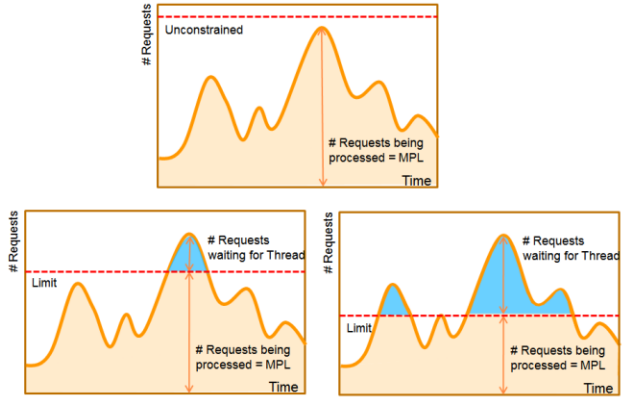
Set limits on capacity:

- Minimum capacity for the queue
- Maximum capacity (% of cluster resources) for a queue
- Resource elasticity when not being used by other queues
- Minimum user limits – user sharing for a given queue
- User limit factor – maximum queue capacity that one user can take up
- Application limit – maximum # of applications submitted to one queue

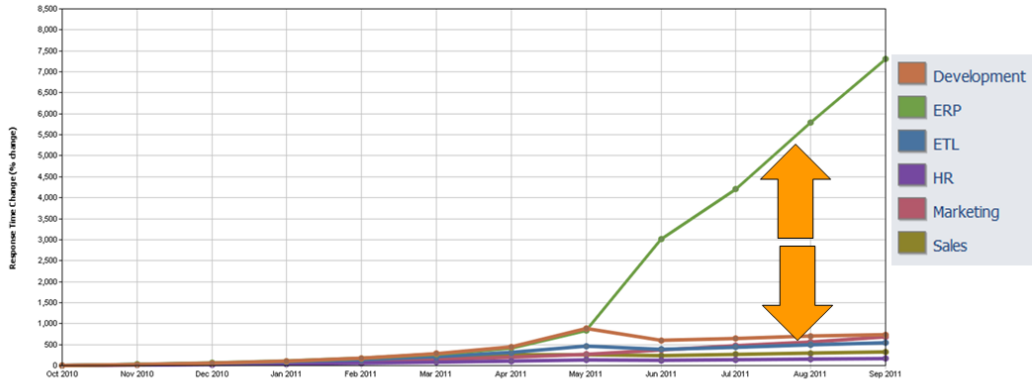
Example of Predicting Workload Concurrency Change Impact

Concurrency Tradeoffs:

Reduction in Contention for Resources vs Increase of Delay Time

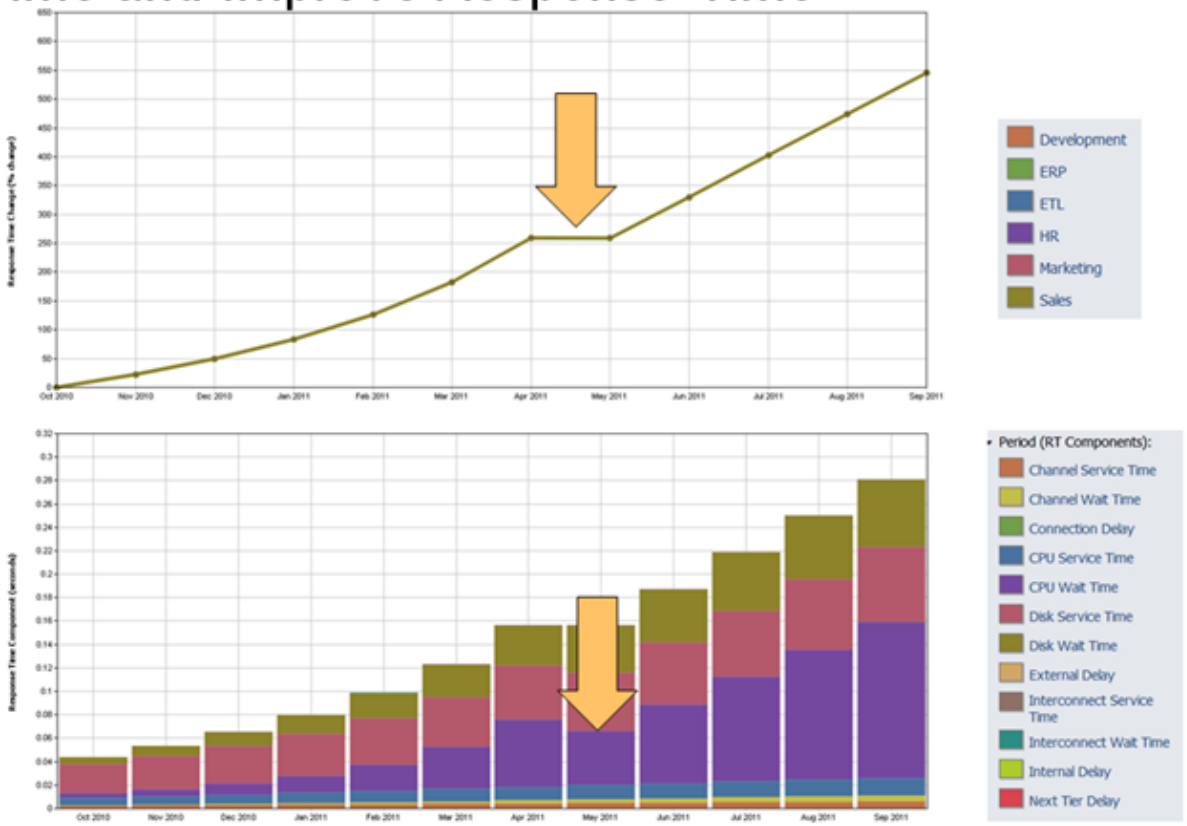


ERP and Marketing Throttling Will Elongate Their Response Time but Improve It for All Other Workloads



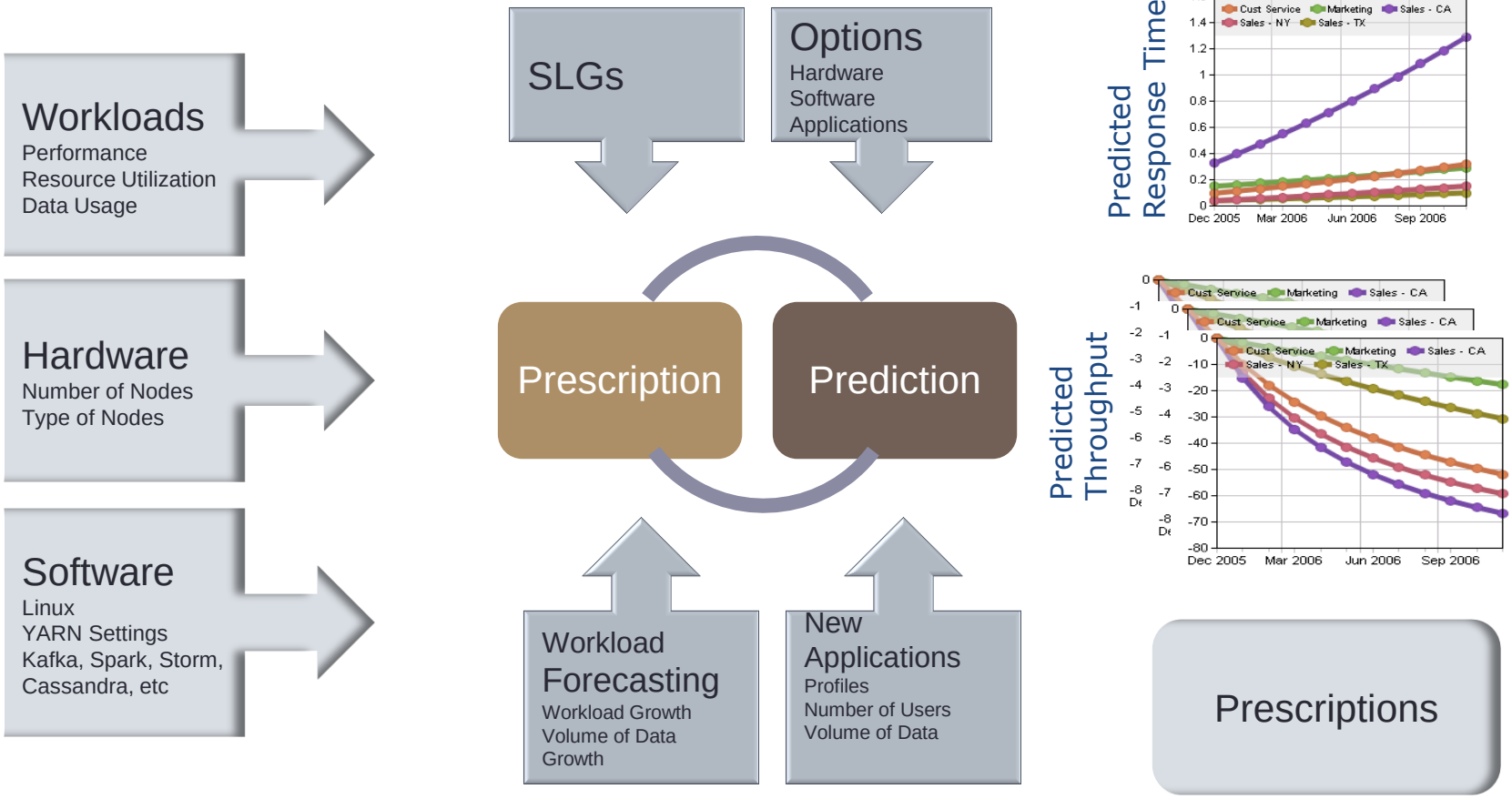
Example of Predicting Workload Priority Change Impact

Increasing Priority for Sales Will Reduce CPU Wait Time and Improve Response Time

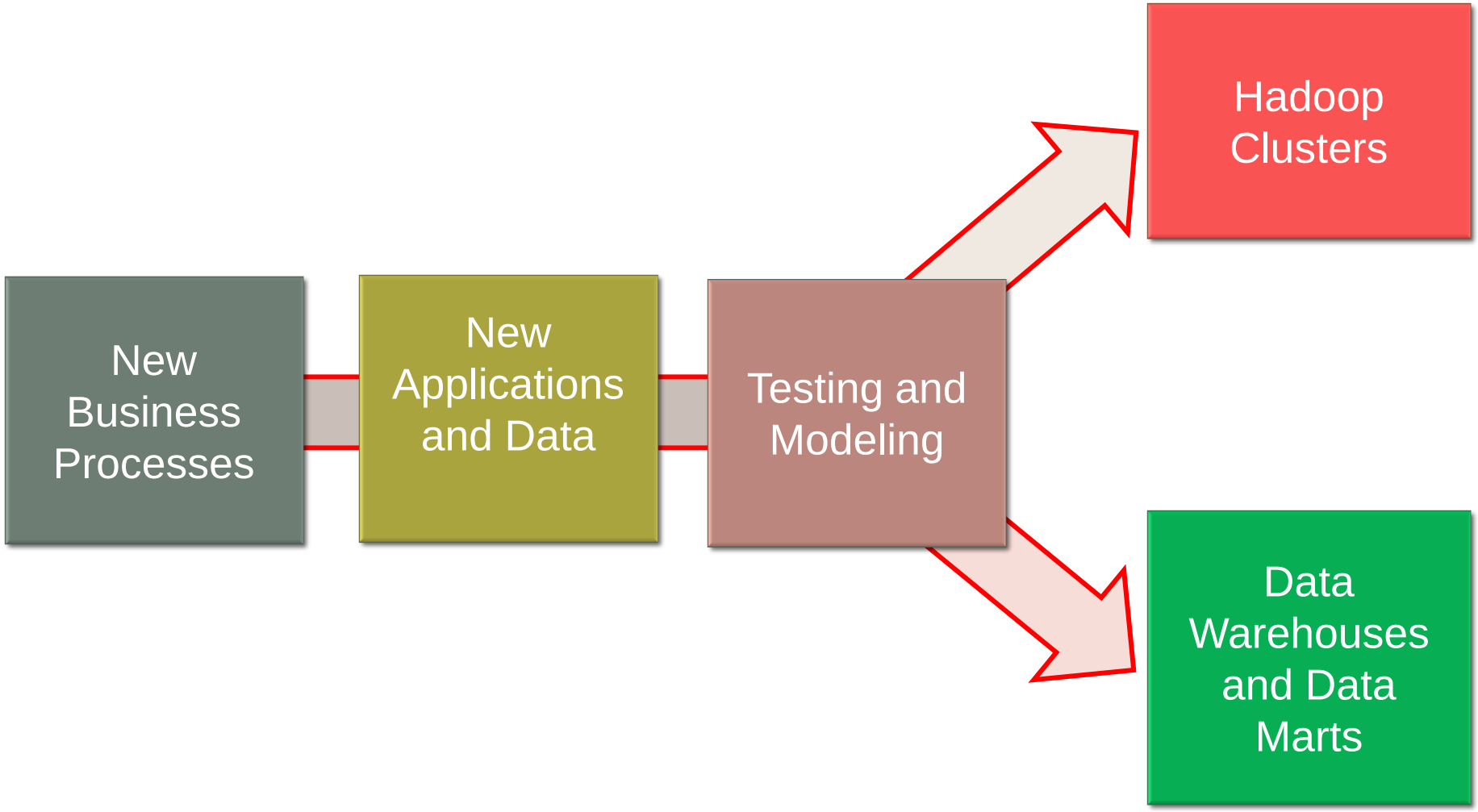


CAPACITY PLANNING

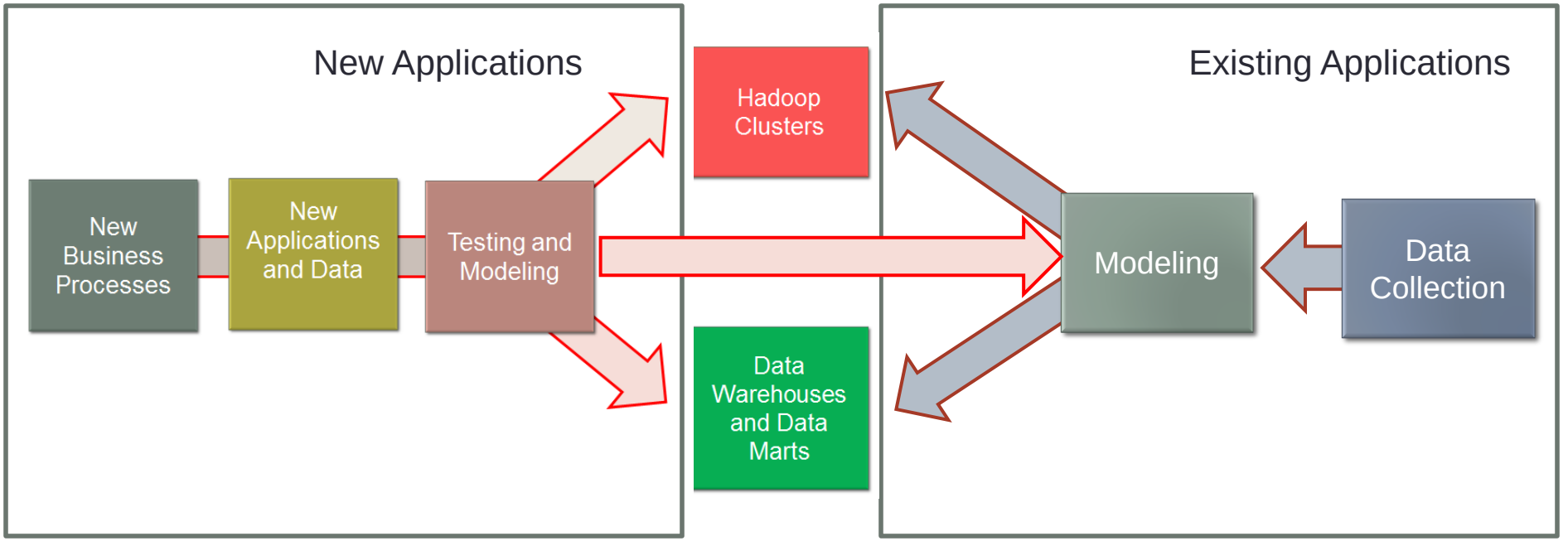
Performance Prediction and Prescription



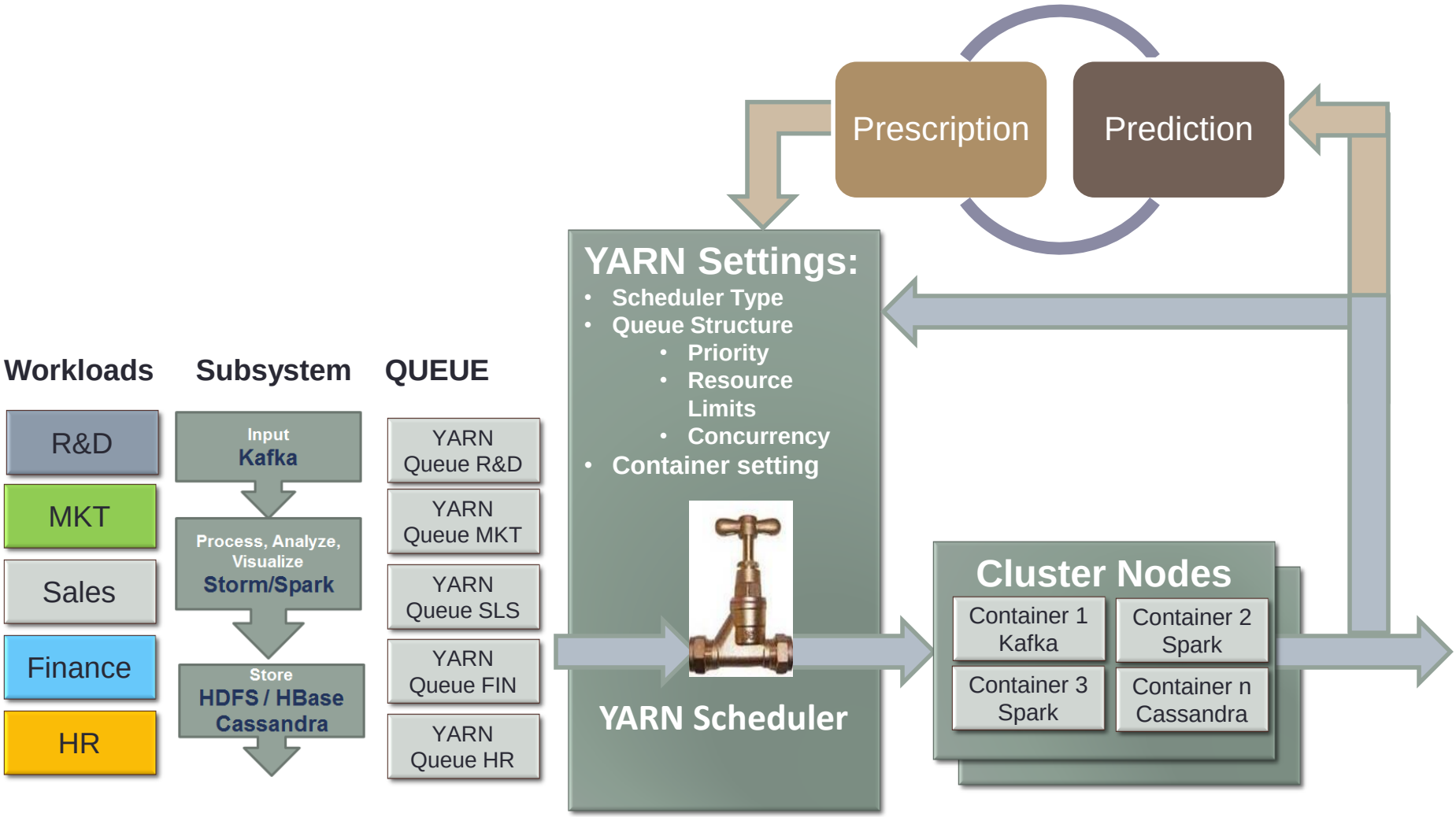
Collecting Data and Modeling in Test Environment



Predicting How New Applications will Perform in Production Environment

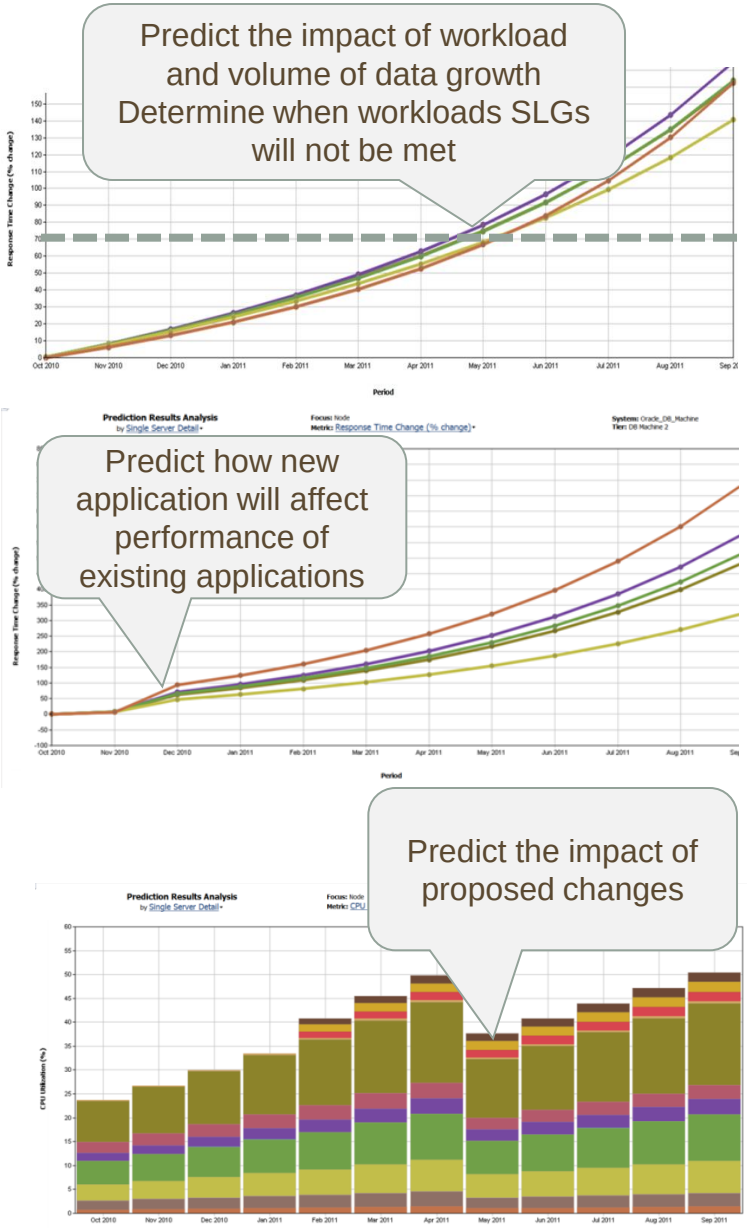


Dynamic Capacity Management

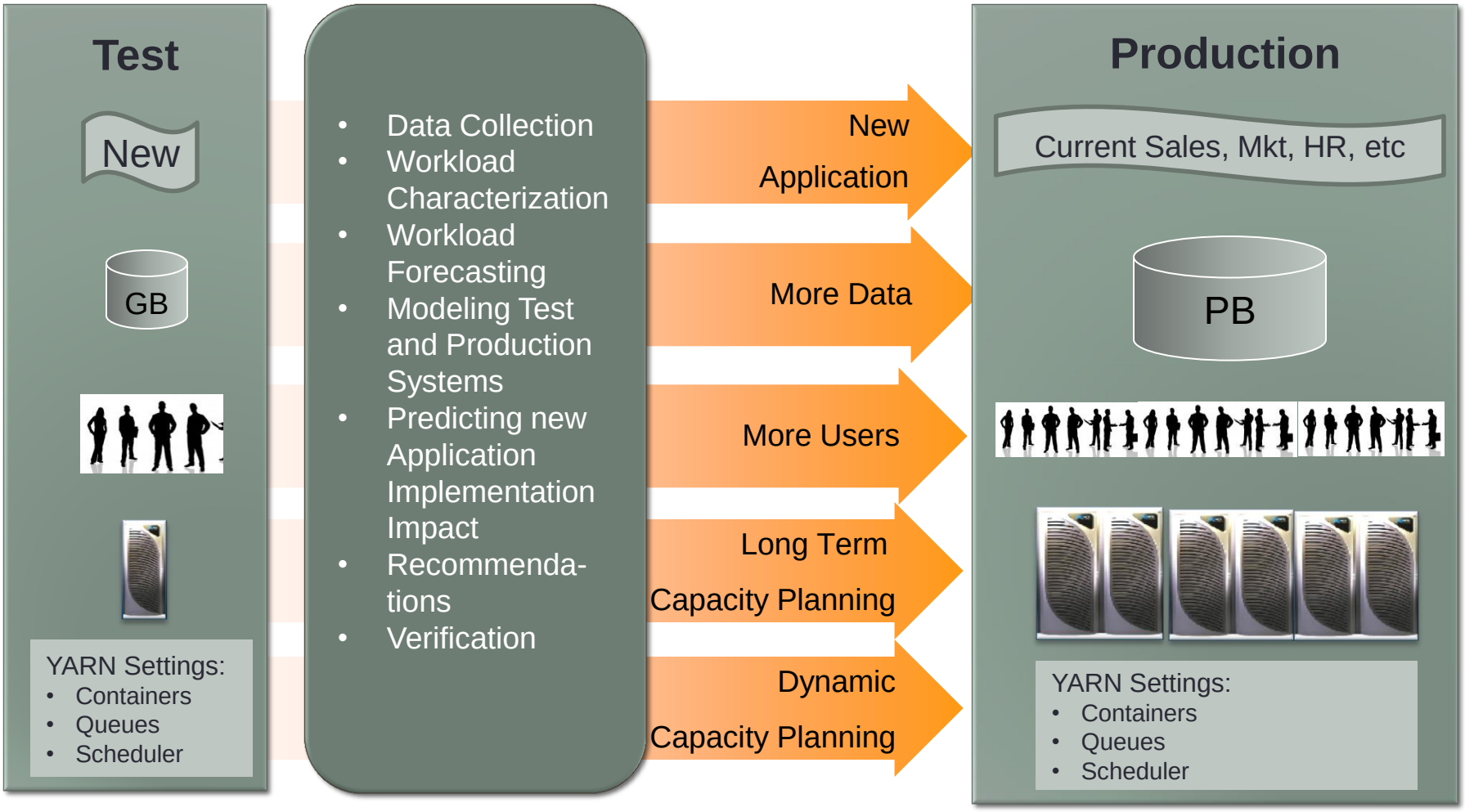


Capacity Planning

- Long Term Planning
  - Apply Predictive Analytics to determine number of Nodes in Cluster required to support expected workload and volume of data growth
  - Predict how new application will perform on production system
- Dynamic – Real Time Capacity Planning
  - Apply Prescriptive Analytics to evaluate options and determine how to dynamically change YARN Settings, including Containers, Queues and Scheduler to meet individual workloads SLGs
- Set realistic expectations
- Verify results



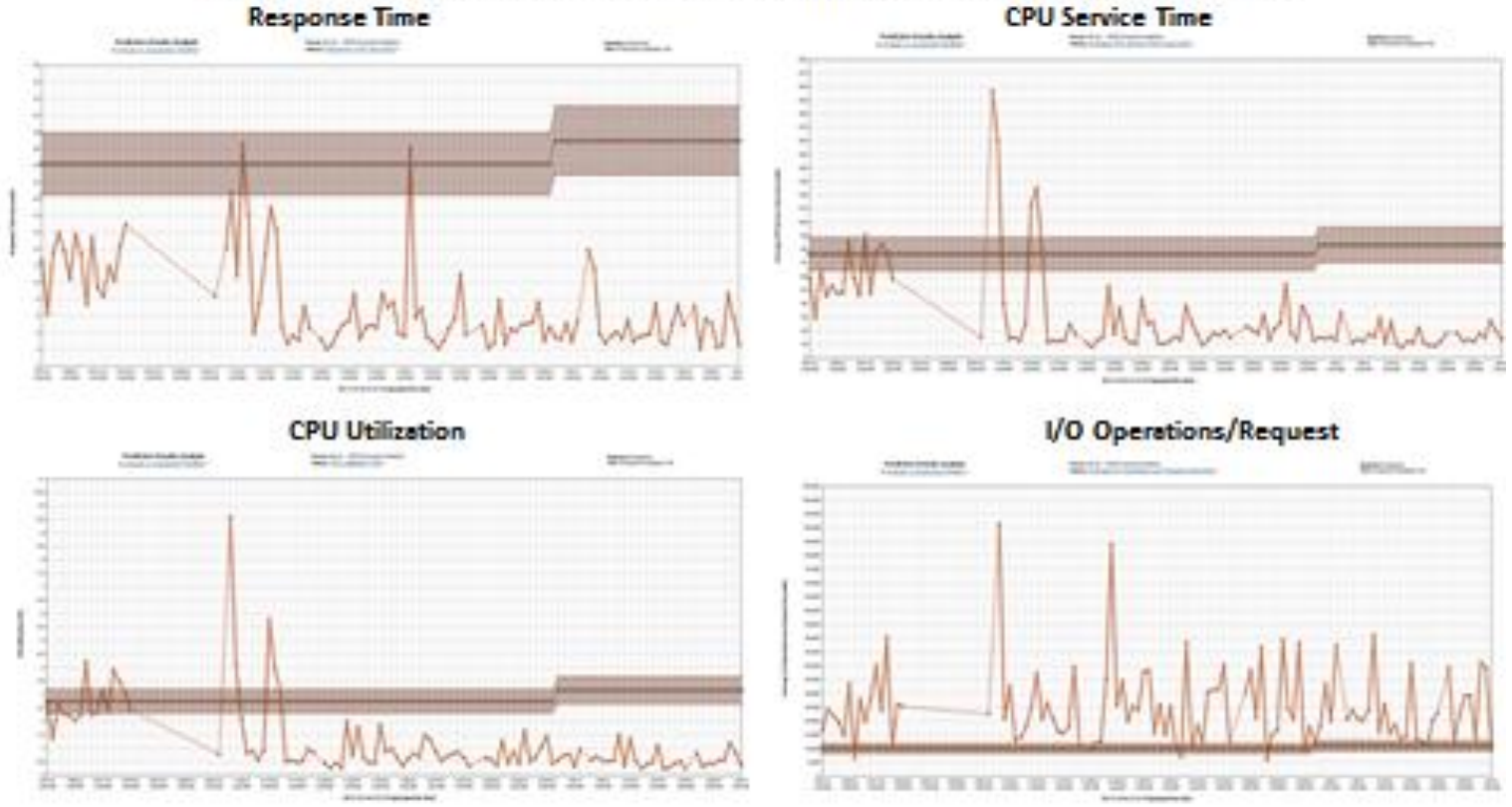
Predicting New Application Implementation Impact



VERIFICATION AND AUTOMATION

Verification – Actual vs Expected (A2E)

The Root Cause analysis shows that increase in complexity of workload caused increase in CPU utilization and difference in RT A2E



SUMMARY

Value of Application Performance Assurance

- Optimization of Performance Management, Workload Management and Capacity Planning Decisions during Application, Data and Systems Life Cycle
- Set Realistic Expectations
- Enables Verification
- Automation
  - Predictive and Prescriptive analytics enables automatic proactive Performance Assurance process focusing on continuous meeting SLGs
- Reduce uncertainty and risk of performance surprises
- Collaboration
  - Better aliment between business and IT

References

- B. Zibitsker, Big Data Advanced Analytics, Minsk 2016, Key Note Presentation on Big Data Analytics
- B. Zibitsker, CMG 2016, Performance Assurance for Real Time Applications
- B. Zibitsker, CMG 2016, Enterprise Performance Assurance Platform
- B Zibitsker, IEEE Conference, Delth Netherlands, March 2016, Big Data Performance Assurance
- B. Zibitsker, Big Data Predictive Analytics Conference, Minsk 2015, Key note presentation “Role of Big Data Predictive Analytics”
- B. Zibitsker, Big Data Predictive Analytics Conference, Minsk 2015, Workshop on “Big Data Predictive Analytics”
- B. Zibitsker, Big Data Conference, Riga 2014, “Application of Predictive Analytics for Better Alignment of Business and IT”
- B. Zibitsker, T. Jung, Teradata Partners, 2012, “Collaborative Capacity Management”
- B. Zibitsker, A. Lupersolsky, OOW 2009, “Modeling and Optimization in Virtualized Multi-tier Distributed Environment”
- B. Zibitsker, Teradata Partners 2008, “Proactive Performance Management of Data Warehouses with Mixed Workloads”
- B. Zibitsker, DAMA 2007, “Enterprise Data Management and Optimization”
- J. Buzen, B. Zibitsker, CMG 2006, “Challenges of Performance Prediction in Multi-tier Parallel Processing Environments”
- B. Zibitsker, CMG 2008, 2009 “Hands on Workshop on Performance Prediction for Virtualized Multi-tier Distributed Environments”

Boris Zibitsker, PhD

- **Founder and CEO of BEZNext**, 2011 - present
- Current focus of research is on applying predictive and prescriptive analytics for optimization of business and IT decisions during applications and data life cycle
- Manage development of the Performance Assurance technology incorporating advanced analytics for optimization of Big Data and Data Warehouse applications in complex multi-tier, distributed, virtualized, parallel processing environment
- Consulted many of Fortune 500 companies
- **CTO of Modeling and Optimization at Compuware** (2010-2014)
- Participated in development of Application Performance Management software incorporating Machine Learning algorithms for performance and availability problems detection, and root cause analysis determination for web applications
- **Founder, President and Chairman of BEZ Systems** (1983 - 2010), acquired by Compuware in 2010
- Managed development of BEZVision Performance Prediction and Capacity Management software for Teradata, Oracle, DB2 and SQL Servers
- **Performance Analyst:**
- Started out as engineer at Computer Systems Research Institute working on modeling and performance evaluation of large computer systems and applying modeling results for optimization of jobs scheduling and storage performance management
- Worked in capacity management departments at FNBC and CNA Insurance company in Chicago
- **Adjunct Associate Professor, DePaul University in Chicago** (1983 – 1990)
- Taught graduate courses on Modeling of Computer Systems, Queueing Theory with Computer Applications, Computer Communication Systems Design and Analysis
- Taught seminars at Northwestern University, University of Chicago and Relational Institute - North and South America, Europe, Asia, and Africa
- Author of papers on applying modeling and optimization for performance evaluation, performance assurance, performance management, workload management and capacity planning for Big Data and Data Warehouse environments
- **Education:** MS and PhD research at BSUIR and NIIEVM

ARE THERE ANY QUESTIONS?

Boris Zibitsker, PhD
bzibitsker@beznext.com