USER GUIDE

Uila AA-IPM User Guide

August 2015

Version 1.1

Company Information

Document Revisions

Date	Version Number	Document Changes
4/1/2015	1.0	
8/15/2015	1.5	

Table of Contents

1 Introduction

1.1 Scope and Purpose

The first part of this document describes the system requirements, installation, configuration steps for Uila AA-IPM software.

The second part details how to use AA-IPM console to manage, troubleshoot applications and infrastructure related issues in the data center.

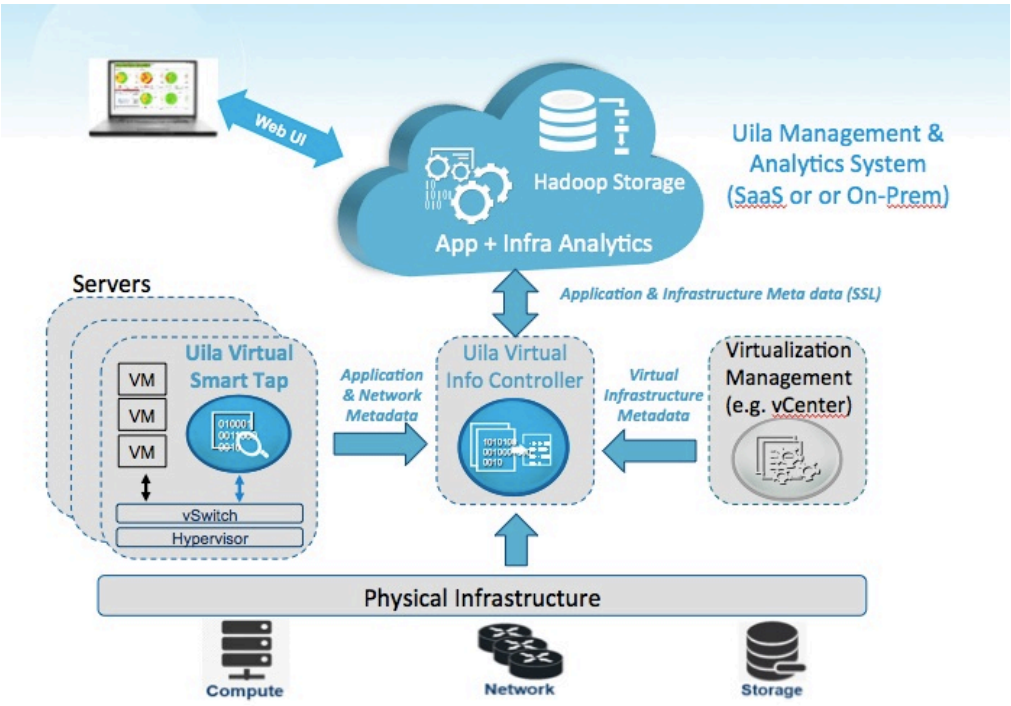
It is assumed reader has already familiar and proficient in VMware installation, configuration and on-going management.

1.2 Architecture Overview

The core of Uila virtual architecture is a Big Data store and analytic engine that are designed from ground up to scale-out to accommodate large data center deployment with thousands of servers, to scale-in to record data in minute resolution, maintain history data (for up to 1 year) while maintaining real time responsiveness. Built-in redundancy offers high availability, removes downtime, and reduce maintenance overhead.

Virtual Information Controller (vIC) is the integration conduit to VMware vCenter. vIC retrieves your infrastructure configuration as a template to build Uila monitoring domain and to streamline deployment. vIC collects network, storage and compute performance metrics that are maintained by vCenter, combines with data from vST, and then transit to Uila Big Data store through encrypted SSL channel.

Virtual Smart Tap (vST) deployed in the Host as small foot print and efficiently designed guest VM with embedded Deep Packet Inspection (DPI) technology to identify unique application and its attributes. vST measures application response time, collects network performance data in meta data form. No packet payload is examined and stored, thus removing the risk of exposing sensitive data.



Unified Central Management Console

Modern virtual technology has improved data center operating efficiency. However, the management tools IT used may not be able to cope with the increase in complexity to monitor application performance effectively. Uila management console Dashboard offers simple and yet powerful view to visualize and reveal the heath of Application and the underlying physical / virtual infrastructure in network, compute and storage to pin point which application performance is degraded, where the bottle neck is.

Automation and Provision

To support the agile environment of the virtual data center, Uila integrates tightly with VMware vCenter to allow DC operators taking advantage of his/her set up to select which tenants and the type of vApp to monitor. Uila takes over the configuration, deployment and provision of Uila guest VM automatically, freeing extra burden of maintenance and support.

1.3 Feature Highlights

1.3.1 Virtual Architecture – Built for Virtual Data Center

Uila architecture is the next generation platform that harnesses the latest Big Data technology to offer unprecedented scalability and flexibility to monitor the mission business critical applications, while maintaining real time responsiveness:

- Scale from small to massive data center with thousands of servers with history records of up to one year with Built-in redundancy for high availability.
- Small foot print virtual Smart Tap (vST) deployed as a guest VM with minimal overheads.

- Collect application response time and more than fifty critical infrastructure performance metrics in minute interval.
- Embedded Deep Packet Inspection (DPI) technology to identify unique application and its attributes.
- Virtual Information Controller (vIC) seamlessly integrates with VMware vCenter. vIC leverages network, storage and compute performance metrics already maintained by vCenter.
- Uila collects data in meta data format. No packet payload is examined and stored. Transmission is through encrypted SSL channel, removing the risk of exposing sensitive data.
- Standard RESTful API to Uila analytic data base

1.3.2 Unified View – Simplify Data Center Operation

The complexity of virtual infrastructure hierarchy that comes with today's virtual data center requires an easy to use but power tool set, such as Uila that helps DC operator to visualize and pin point areas of performance degradation, and identify the root cause immediately:

- Customizable Application and Infrastructure health dashboard that mirrors the logical constructs of the data center hierarchy to aggregate volume of data into meaningful Key Performance Indicators for early warning of poor performance.
- Powerful analytic tool sets of Application Topology, Flow Analyzer, CPU Usage, Memory Usage, and Storage Usage provide unique diagrams that reveals how underlying infrastructure physical or virtual capacity are impacting application performance.
- Innovated browser based UI design simplified navigation and speeds up problem resolution.
- New adaptive baseline technique to create monitoring thresholds that align with actual average performance characteristics for the underline infrastructure dynamically to reduce false positives.
- Integrated alerting, troubleshooting scenario to Help Desk, or Network Operation Center.
- Built-in and customizable C level reporting for service level agreement compliance
- Exportable historical trending data as template for future planning.

1.3.3 SaaS Cloud - Automation and Provisioning

Wide adoption of virtualization and cloud have made SaaS one of consideration for IT, as enterprise and service provider continue to seek better service, lower and predicable cost to service their customers. Uila Cloud helps to reduce IT OpEx and CapEx:

- Integration tightly with VMware vCenter to allow DC operator taking advantage of his/her infrastructure configuration to setup vApp monitoring profile.
- Automated deployment and provision of Uila guest VM to free IT extra burden of maintenance and support.
- SaaS deployment model eliminates the requirement to procure, deploy and maintain appliance and/or hardware probes.
- Multi-tenancy offers easy and common access for IT team.

2 Terminology Used

This section lists common terminology and legend used through out the product and the documentation. It is Uila's goal to use the same terminology as defined by VMware or commonly used by virtualization industry.

Terminology or Legend	Definition	Usage
Application Response Time	Time measured on the server from the arrival of a client request to the transmission of a server response.	
Application Service	Refer to Classifier	
Classifier	It defines the application name as the result of Deep Packet Inspection by vST software agent; for examples, MySQL, iMap. It is interchangeable used as Application Service.	
Cluster	A cluster is a collection of hosts and associated virtual machines. Physical resources from all the hosts in a cluster are jointly owned by the cluster and centrally managed, for example, by the vCenter Server in VMware implementation.	
DPI	Deep Packet Inspection works at the application layer of the OSI reference model of network packets. Ulla DPI uses advanced method of pattern matching and session heuristics to identify applications and their associated attributes, thus helping IT to track many mission critical applications and transaction performance issues and root cause.	
DvPortGroup	DvPort groups provide the same basic functionality as do standard port groups, but offer additional features. For example, administrators can define not just outbound traffic shaping, but inbound traffic shaping as well, when working with DvPort groups.	
DvSwitch	DvSwitches (a VMware concept) is short for Distributed virtual Switches. DvSwitches provide the same features and function as do vSwitches. You can use create and apply a single DvSwitch to all the applicable ESX or ESXi to simply the management of ESX hosts in a Cluster, while vSwitch can only apply to one host at a time. DvSwitches also support port groups, called distributed port groups, or dvport groups.	
Host	A physical server that supports a version of hypervisor; for example, VMware ESXi, or Microsoft Virtual Server.	
pCPU	A PCPU refers to a physical hardware execution context. This can be a physical CPU core if hyperthreading is	

	unavailable or disabled, or a logical CPU (LCPU or SMT thread) if hyperthreading is enabled. For example, a server equipped with a CPU with 4 cores without hyperthreading will have 4 pCPU. If hyperthreading has been enabled then a pCPU would consitute a logical CPU. This is because hyperthreading enables a single processor core to act like two processors i.e. logical processors. So for example, if an ESX 8-core server has hyper-threading enabled it would have 16 threads that appear as 16 logical processors and that would constitute 16 pCPUs.
Port Group	Port group is a template for creating virtual ports of a vSwitch with particular sets of specifications, which gives virtual machines (VM) common attributes for connectivity on every host on which it might run. Port group is an important concept in VMware Infrastructure virtual networking. However, Port group does not correspond exactly to features commonly found in physical networks.
TCP Fatal Retry	
Tenant	Tenant is defined as a Customer or a User Group that uses the same Port Group. As a service provider, it allows the setup of protect containers to keep their customers' (or Tenants) data and existence from other customers that are sharing the same physical data center. In the enterprise private data center, Sales, HR, Manufacturing, or Financial is some common example of a Tenant.
ToR Switch	A Top of the Rack or (ToR) switch is a high port count switch, typically 48 1G or 10G ports plus 4 additional up link ports that sits on the top of server rack in Data Centers or Co-location facilities. ToR switches are then connected to the next level aggregation switch or core router to allow communication between servers in different rack or to internet.
vApp	vApp is a collection of pre-configured virtual machines (VMs) that combine applications with the operating systems that they require. VApps allow disparate VMs to work together in a stack as an application, and support cloud computing architectures. vApp is VMware defined term, and may be used in other similar products.
vCPU	A vCPU stands for Virtual Central Processing Unit. One or more vCPUs are assigned to every Virtual Machine (VM) within a cloud environment. Each vCPU is seen as a single physical CPU core by the VM's operating system. If the host machine has multiple CPU cores at its disposal, then the vCPU is actually made up of a number of time slots across all of the available cores, thereby allowing multiple VMs to be hosted on a smaller number of physical cores.
VM	A virtual machine (VM) is a software emulating of a complete system platform (for example, a server) that

	supports the execution of a complete operating system (OS).
vIC	Virtual Information Manager is a Uila software agent implement as a guest (VM) which (1) interfaces to vCenter to retrieve compute and storage performance data, (2) acts as a proxy for vST to transfer vST meta data to Uila Cloud, (3) receives Uila management commands to install and configure vST. There is only one instance of vIC per vCenter.
vST	Virtual Smart Tap is a Uila software agent implement as a guest (VM) resides in the same Host as other application VM. It captures and analyzes all traffics between VM's within the same host, and other hosts .
vSwitch	vSwitch is short for Virtual Switch and represents networking entities connecting Virtual Machines in a virtual network at layer 2. The Virtual Switch is fully virtualized and connected to a NIC (Network Interface Card) inside a server. The vSwitch merges physical switches into a single logical switch. This helps to increase bandwidth and create an active mesh between server and switches. The VMware Virtual Switch is a switching fabric built into the VMware infrastructure (ESX) that allows you to network your Virtual Machines (VMs).

Figure X-X: Uila Legend and Terminology Definitions

3 Icon Definitions

This section lists used through out the product and the documentation.

ICon	Definition	Usage
	Maximize display viewing area by hiding browser menu and other title bars. Toggle to restore original display view.	
	Logout your Uila session.	
	Launch help.	
	Select color for the title bar.	
	Collapse or minimize the individual sub-view within the Dashboard.	
	Restore the minimized the sub-view within the Dashboard.	
	Toggle between full screen and normal mode.	
	Re-layout the Application Topology view.	
	Select infrastructure component to display in the Flow Analysis view.	

4 Getting Started

This chapter describes the minimum system requirement to install and operate Uila IPM, initial registration steps, and how to install and configure Uila software in vCenter and vSphere environments.

For the following sections, please refer to *Uila vIC and vST Installation Guide* for System Requirements, Registration Instructions, and Instructions to install Uila software.

4.1 System Requirements

4.2 Registration Instruction to Download Software

4.3 Install Software and Prepare VMware vCenter

5 Baseline

A baseline is a process for monitoring the data center infrastructure of its network, compute and storage resources at regular intervals to ensure that the infrastructure which supports business applications are working as designed. It is a process of continuing monitoring the key performance indicators to report the health of all applications and its associated data center at a certain point in time. Properly constructing the baseline for your data center, you can obtain the following information:

- Monitor application response time and availability
- Reveal the health state of the infrastructure resources both virtual and physical
- Obtain the current utilization of system resources
- Determine and set alarm thresholds that are unique to your data center operation characteristics
- Alert and identify current system problems that impacts Application performance
- Plan for future upgrades and expansions

Another way of looking at the baseline is illustrated in the following diagram.

[add Diagram] the Application performance break point, is the point at which the one or a combination of resources load will break, which is determined through the knowledge of how the physical and virtual infrastructure entities interact by responding to Application resource demands. The green line, the infrastructure load, is progression trend of the resource load on the infrastructure as new applications are added, and other systems constraints imposed.

"By performing a baseline on a regular basis, you can find out the current state *and* extrapolate when potential failures might occur and prepare for them in advance. This also helps you to make more informed decisions about when, where, and how to spend budget money on infrastructure upgrades."

5.1 Uila Baseline

Uila IPM implements baseline methodology extensively, as it is the foundation from which *Performance Grades (Infrastructure health performance index) are calculated* and *Alarms* are generated in real time.

Uila maintains a group of *Performance Metrics (See Chapter x.x)*; for example, Application Response Time, Network Response Time, TCP/IP fatal retry, CPU usage, Memory usage, Disk latency, and many more in its Hadoop data base. Virtual Smart Taps and Virtual Information Manager deployed in user's data center analyze, collect, and transmit these Performance Metrics every minute to Uila Cloud.

Every Metric in per minute interval is compared to a Baseline value for that Metric in real time and a Health Score is calculated based on the formula listed in Figure (X.X).

Delta From Baseline	Alarm Severity	Health Score	Color
Less or equal to 5%	**Normal**	75-100	**Green**
Between 5% and 10%, including 10%	**Minor (1)**	50-74	**Yellow**
Between 10% and 20%, including 20%	**Major (2)**	25-49	**Orange**
Above 20%	**Critical (3)**	0-24	**Red**

Uila maintains two kinds of Baseline record for each of Performance Metric monitored;

- Fixed value: it is a constant value; based on VMware best practices, for example, CPU usage for VM is pre-defined as 80%.
- Variable value: it is an average of measured metric (per minute) within an hour, i.e. 60 data points. Example of variable metrics are Application Response time, and Network Round Trip time.

Uila Baseline is a record of full 7 days (one week) of hourly data, beginning from Sunday 12:00am and ending at Saturday 11:59pm. Each of Metrics is compared to its corresponding Baseline record for the same hour of the same week day. e.g. Monday 12:37 pm's metric data is compared to Monday 12PM's Baseline value.

During the first day of starting up, current Metrics will be compared to previous hour's value as the default baseline value.

Method of Building Baseline record

Here are the choices you can change how Uila baseline values are defined.

Baseline Metrics	Remarks
Last Hour's value	This is the system default.
Yesterday's value	Select Yesterday's value as the Baseline.
Last Week's value	Select the entire week for the weekly baseline setting. One week after deployment, this baseline option is used automatically by default.
User Configuration option	User selects and locks to a specific week's performance metrics as baseline.

5.2 Health Score and Alarm Definition

Performance Grades are for visual display only and typically color-coded to show the health scores where low score (red) is poor health, and high score (green) is good health. (see Figure x.x), and are updated every minute.

Here is an example of the Data Center Application Performance summary in color:

Alarm is generated when a performance metric cross the Critical threshold. Alarm is generated every 15 minutes by default, and can be configured as one shot, or re-generated in more or less frequently.

Threshold is defined as the % value that crosses the baseline.

Severity is a user definable indicator to help identify the criticality of the performance metrics monitored to alert user if a entity or entities in his/her data center infrastructure is (are) about to impact the Application's performance.

Delta From Baseline	Alarm Severity	Health Score	Color
Less or equal to 5%	**Normal**	75-100	Green
Between 5% and 10%, including 10%	**Minor (1)**	50-74	Yellow
Between 10% and 20%, including 20%	**Major (2)**	25-49	Orange

Above 20%	**Critical (3)**	0-24	**Red**

Note: These standard color definitions are applied through out Uila User Interfaces for consistence and ease of recognition.

6 Managing Your Work from the Console Home Page

Uila console home page is your default infrastructure monitor where you perform your daily tasks, such as:

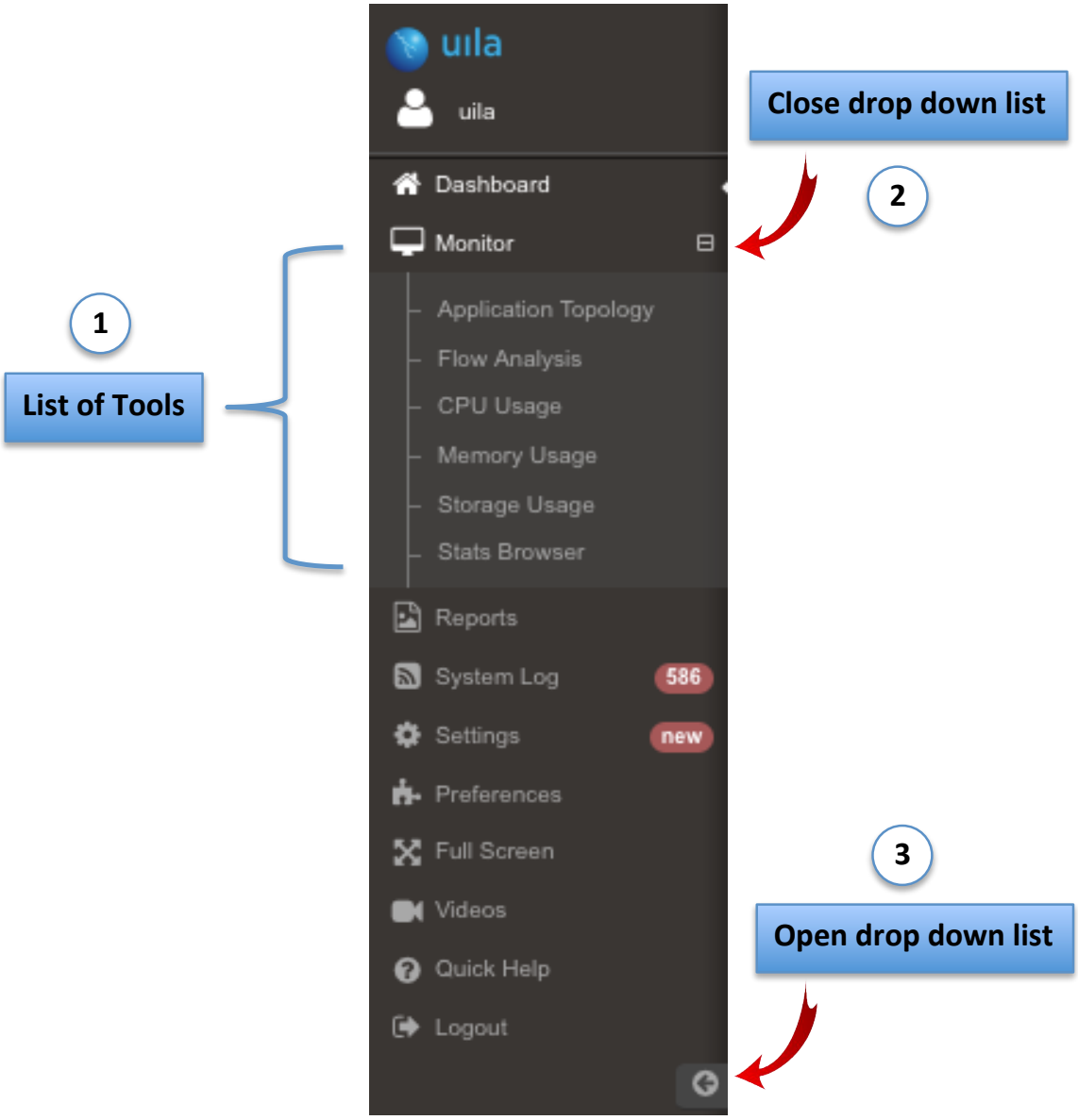
- View Application and Infrastructure health dashboard, investigate performance degradation, troubleshooting, and identifying root cause in real time
- Launch additional monitor applications
- Generate reports
- View Syslog
- Change Settings
- Set Preferences
- Go to Full Screen
- See On-line Videos
- Quick Helps
- Logout

Here is the default Console home page is layout:

6.1 Tools Pane

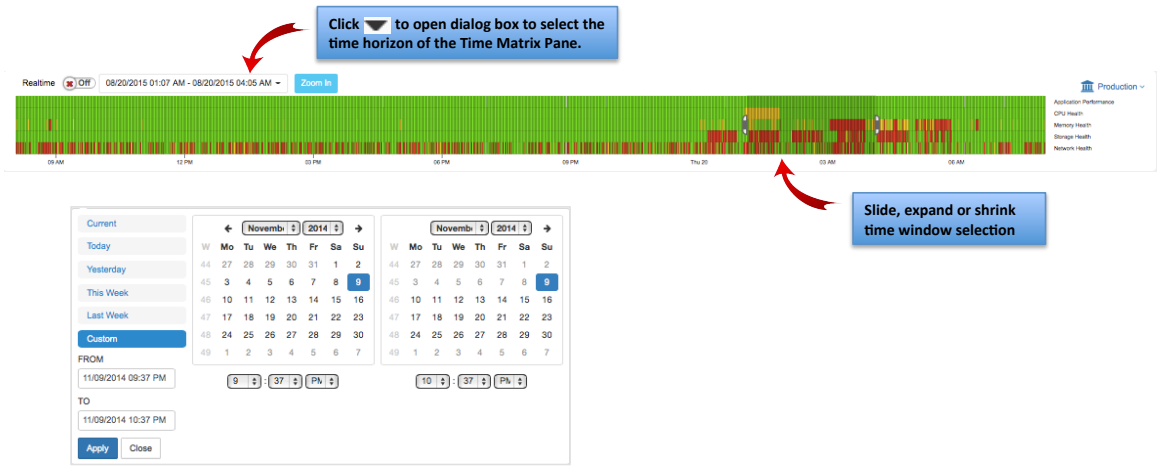
The Tool Pane consists of menu to set up the User profile, and a list of Uila tools for monitoring, report and configuration.

- Indicated by (1): List a menu of Uila Tools for user to launch directly.
- Indicated by (2): Click the - box Symbol to close the menu drop down list.
- Indicated by (3): Click the <- box to close the Tool Pane, and make more space for the Monitor pane.



6.2 Time Matrix Pane

The Time Matrix tool bar allow you to set up a Time Bracket within your timeline horizon where your entire infrastructure performance data are calculated, summarized, compared to prior baseline and displayed in the Monitor pane. You can customize your time window in minutes, hours, or days depending on how you wish to perform real time monitoring, or root cause analysis.



The Time Matrix pane consists of:

- Calendar box to set up time window which you can select between 'Real Time' and 'Time Travel' mode. Select *Current* for Real Time mode.
- Time line window with slide brackets; which can be dragged along the time line to widen or narrow the monitoring window (time range between the brackets)
- Up to five (5) user definable key performance index (KPI) to be monitored. The default KPI are *App Performance*, *CPU Health, Memory Health*, *Storage Health* and *TCP Fatal retry.* The Definition of the first four (4) KPIs are described in the Dashboard chapter in details.

Real Time Mode

In real time mode, all the performance counters are calculated and updated every minute. Typically, you use real time mode to identify root causes of critical applications that exhibit performance degradation in short term, typically in past hours or minutes. System defaults to Real Time mode.

Time Travel Mode

In Time Travel mode, performance data and health measurement metrics are aggregated and calculated based on the Time Bracket you selected. Screen update is stopped. However, data collection continues in real time in the background. Time Travel mode is commonly used for

- Setting infrastructure Baseline to monitor for exceptional events that impact Application performance health. We recommend that you set the larger window bracket what is large enough to obtain a Baseline to represent your infrastructure health that is under normal operation. Common best practice is use a full week that average over several weeks to smooth out exceptional conditions.
- Real time troubleshooting where you may need to travel back in time to look for similar alerting event patterns that impacted performance currently.

6.3 Monitor Pane

The Monitor pane is the working space where Uila tools; such as, Dashboard, Flow Analysis, Application Topology, reports, and other Uila Tool displays its contents as a result of your drill down action. By default, a Dashboard that highlights your infrastructure performance health is displayed after you log in to the system.

6.4 Settings

The Settings maintains Uila systems configurations for; (1) vST and vIC software initial installation, and new software updates and upgrades, (2) Interface to physical devices, (3) External systems to receive Alarms.

Here is a list of Configuration Settings Menu:

Menu	Definition
VST Configuration	Use to select which vSwitch(s) in a host to install vST guest VM.
Alarm Configuration	(1) Select Baseline from - Last Hour - Yesterday - Last Week - Any Week since Uila keeps trending records (2) Define Alarm Action. Support delivery alarm by e-mail.
Software Update	List your Uila software version installed, and if new update is available.
SNMP Configuration	Define IP Address of the top of rack switch to be monitored. User has the option to select SNMP protocol version 2 or 3 as the method for Uila to access the switch.
Global Configuration	Define SNMP server IP address, port number, user name, and password to receive the Uila Alarms.
VIC Installation	Step by step instructions to install VIC either the first time, or user wish to deploy VIC in more data centers.

7 Dashboard

Dashboard is the first screen displayed after you login. It allows you to have a unified high-level view of the overall Health of your virtual infrastructure key components in real time and critical alerts that impacted your Application performance, and decides on the next area for focus to investigate which Application is slowed, which area is the main culprit, and what issues are impacting the Applications performance.

The center of the screen shows you the overall health scores in five (5) key areas; ***Application***, ***Network***, ***Storage***, ***CPU*** and ***Memory*** in your virtual infrastructure components, and organized by hierarchical structure relevant to each component in sun burst (color wheel) format.

Figure X-X: Dashboard View

7.1 Summary of Key Performance Index

The Application and the related infrastructure Health score are monitored according the metric listed in Figure x.x.

KPI	Metric Monitored	Measurement Method	Unit of Measure	Baseline Value
Application Performance	Application Response Time	Time measured on the server from the arrival of a client request to the transmission of a server response		
Network Health	Network Round Trip Time	Packet round trip time spent in the network		
	TCP Fatal Retry	TCP re-transmit the same packet for the forth time or greater		
Storage Health	Disk Read Latency	Average mount of time taken process a read command issued from the Guest OS to the virtual machine. The sum of kernelReadLatency and deviceReadLatency in VCDB		
	Disk Write Latency	Average amount of time taken processing a Write command issued from the Guest OS to the virtual machine. The sum of kernelWriteLatency and deviceWriteLatency in VCDB		
CPU Health	CPU Ready	Percentage of time that the VM was ready, but could not get scheduled to run on the physical CPU due to physical CPU resource congestion		
	CPU Usage	Average CPU utilization over all available virtual CPUs in the VM		
Memory Health	Swap Wait Time	Time the virtual machine is waiting for memory pages to be swapped in		
	VM Memory Usage	Memory usage as percentage of total configured or available memory		

Figure x.x : Infrastructure Health Measurement Metrics and Definitions

7.2 Application Performance Metric

The Application Performance color wheel displays the health of Applications currently running in your data center. The rings present the hierarchical constructs of a virtual Data Center, where you may configure your data center in multiple logical Port Groups. Each Port Group consists of a series of Applications (vApp); such as MySQL, business logics, and web service to perform a specific application function for the end user. These applications depending on the business requirement may run on one or more than VMs.

Application Performance Health Metric

Measurement Metric	Measurement Method	Definition	Default Grade
Application Response Time (in millisecond)	Monitored at packet transaction level	Time measured on the server from the arrival of a client request to the transmission of a server response	

Ring Structure and Size Definition

Ring Structure		Color	Size
Ring Center	Data Center	Color represents the averaged Application Performance for the group over the time range selection in the Time Matrix bar. See color and baseline definition in Time Matrix Bar (Figure	Application Transaction Volume
Ring 1 (inner ring)	Cluster		
Ring 2	Host		

Ring 3 (outer ring)	VM	X.X)

7.3 Network Performance Metric

The Network Health color wheel displays the health of network with respect to the infrastructure currently running in your data center. The rings present the hierarchical constructs of a virtual Data Center, where it typically structures from TOR Switches, Host, to VM's. Each TOR Switch is connected to a number of Hosts, where one or more VM's resides.

Network Health Metric

Measurement Metric	Measurement Method	Definition	Default Grade
Network Round Trip time (in millisecond)	Monitored at packet level	Packet Round trip time spent in the network	
TCP Fatal Retry (in count)	Monitored at packet level	TCP Fatal retry is the TCP packet retransmission for the same packet for the forth time, which triggers TCP back off algorithm and significant application delay in response.	

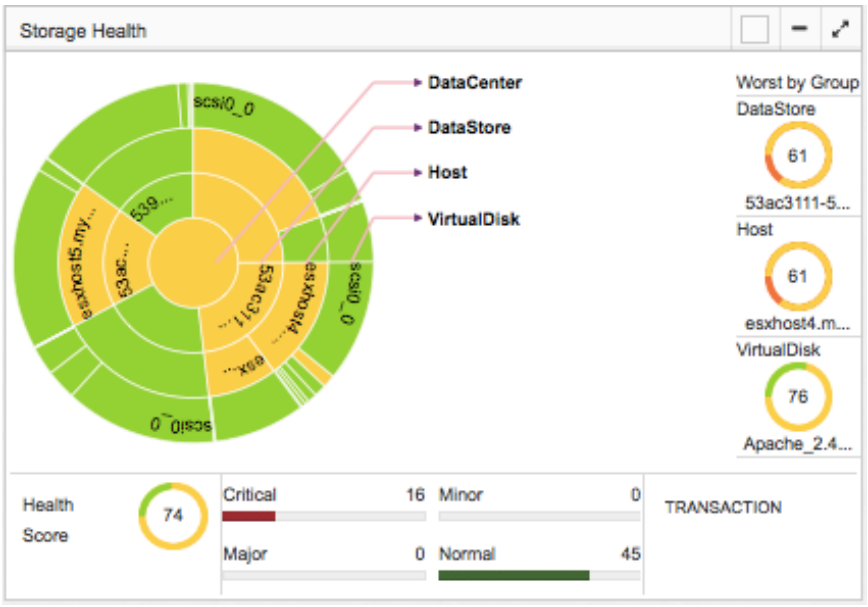
Note: Receiver Packet Drop. See ……

Ring Structure and Size Definition

Ring Structure		Color	Size
Ring Center	Data Center	Color represents the average weighted Network Health score for each respective group over the time range selection in the Time Matrix bar. See color and baseline definition in Time Matrix Bar (Figure X.X)	Network Traffic Volume
Ring 1 (inner ring)	Cluster		
Ring 2	Host		
Ring 3 (outer ring)	VM		

7.4 Storage Performance Metric

The Storage Health color wheel displays the health of storage systems currently running in your data center. The rings present the hierarchical constructs of a storage system within your Data Center, where it typically owns multiple Data Stores. Each Data Store groups together a number of Hosts. Virtual Disk……..



Storage Health Metric

Measurement Metric	Measurement Method	Definition	Default Grade
Disk Read Latency (in millisecond)	Sourced from vCenter (VCDB)	Time taken complete a Read command issued from the Guest OS. This Disk Read Latency includes VM kernel Read Latency and Device Read Latency.	
Disk Write Latency (in millisecond)	Sourced from vCenter (VCDB)	Same as the above for Write command.	

Ring Structure and Size Definition

Ring Structure		Color	Size
Ring Center	Data Center	Color represents the average weighted Storage Health score for each respective group over the time range selection in the Time Matrix bar. See	Number of Storage I/O Operations
Ring 1 (inner ring)	Data Store		
Ring 2	Host		

Ring 3 (outer ring)	Virtual Disk	color and baseline definition in Time Matrix Bar (Figure X.X)

7.5 CPU Performance Metric

The CPU Health color wheel displays the performance of all CPU in your Hosts with respect to the infrastructure currently running in your data center. The rings present the hierarchical constructs of a virtual Data Center, where it typically structures to form a number of Cluster of Hosts, under which one or more VM's resides.

CPU Health Metric

Measurement Metric	Measurement Method	Definition	Default Grade
CPU-Ready (%)	Sourced from vCenter (VCDB)	Percentage of time that the VM was ready to run, but could not get scheduled to run on the physical CPU due to physical CPU resource congestion.	
CPU Usage (%)	Sourced from vCenter (VCDB)	Average CPU utilization over allocated CPU capacity for the VM. For Host and Cluster, it is over physical CPU capacity.	

Host CPU Metric Calculation

Measurement	Normal	Minor	Major	Critical

Metric	(Green)	(Yellow)	(Orange)	(Red)
CPU-Ready (%) (X = CPU.Ready/ # of pCPU)	X < 6,000 ms (10% per 1 min)	6,000 ms <= X < 9,000ms (10% ~ 15%)	9,000 ms <= X < 15,000ms (15% ~ 25%)	X >= 15,000 ms (>= 25%)
Y=CPU Usage (%)	Y <= 80%	80% < Y <= 85%	85% < Y <= 90%	Y > 90%

Note:

Host CPU Ready Time = Sum of all pCPU's Ready Time.

VM CPU Metric Calculation

Measurement Metric	Normal (Green)	Minor (Yellow)	Major (Orange)	Critical (Red)
CPU-Ready (%) (X = CPU.Ready/ # of vCPU)	X < 3,000 ms (5% per 1 min)	3,000 ms <= X < 6,000ms (5% ~ 10%)	6,000 ms <= X < 12,000ms (10% ~ 20%)	X >= 12,000 ms (>= 20%)
Y=CPU Usage (%)	Y <= 80%	80% < Y <= 85%	85% < Y <= 90%	Y > 90%

Ring Structure and Size Definition

Ring Structure		Color	Size
Ring Center	Data Center	Color represents the average weighted CPU Health score for each respective group over the time range selection in the Time Matrix bar. See color and baseline definition in Time Matrix Bar (Figure X.X)	?
Ring 1 (inner ring)	Cluster		Physical CPU capacity (MHz)
Ring 2	Host		Physical CPU capacity (MHz)
Ring 3 (outer ring)	VM		?

7.6 Memory Performance Metric

The Memory Health color wheel displays the performance of all memory arrays in your Hosts with respect to the infrastructure currently running in your data center. The rings present the hierarchical constructs of a virtual Data Center, where it typically structures to for a number of Cluster of Hosts, under which one or more VM's resides.

Memory Health Metric

Measurement Metric	Measurement Method	Definition	Default Grade
Swap Wait time (milliseconds)	Sourced from vCenter (VCDB)	Time the virtual machine is waiting for memory pages to be swapped in.	
Memory Usage (%)	Sourced from vCenter (VCDB)	VM Memory usage is the percentage of active memory to total configured memory. Host and Cluster Memory Usage is the percentage of consumed memory (including VMkernel and Guest VMs) to physical memory capacity.	
Swap-in Rate (kbps)	Sourced from vCenter (VCDB)	Average amount of memory (kbps) swapped in from disk into memory for VM to run.	

Host Memory Metric Calculation

Measurement	Normal	Minor	Major	Critical

Metric	(Green)	(Yellow)	(Orange)	(Red)
Swap-Wait (%) (X = Swap-Wait/ # of pCPU)	X < 6,000 ms (10% per 1 min)	6,000 ms <= X < 9,000ms (10% ~ 15%)	9,000 ms <= X < 15,000ms (15% ~ 25%)	X >= 15,000 ms (>= 25%)

Where:

X=CPU.SwapWait /# pCPU (ref %SWPWT in ESXTOP)

VM Memory Metric Calculation

Measurement Metric	Normal (Green)	Minor (Yellow)	Major (Orange)	Critical (Red)
Swap-Wait (%) (X = Swap-Wait/ # of vCPU)	X < 3,000 ms (5% per 1 min)	3,000 ms <= X < 6,000ms (5% ~ 10%)	6,000 ms <= X < 12,000ms (10% ~ 20%)	X >= 12,000 ms (>= 20%)
Y= Mem Usage (%)	Y <= 70%	70% < Y <= 75%	75% < Y <= 85%	Y > 85%

Note:

VM Memory Swap Wait Time = Sum of all vCPU's Swap Wait Time.

VM Mem Usage = Active / Virtual machine configured size.

Ring Structure and Size Definition

Ring Structure		Color	Size
Ring Center	Data Center	Color represents the average weighted CPU Health score for each respective group over the time range selection in the Time Matrix bar. See color and baseline definition in Time Matrix Bar (Figure X.X)	?
Ring 1 (inner ring)	Cluster		Physical CPU capacity (MHz)
Ring 2	Host		Physical CPU capacity (MHz)
Ring 3 (outer ring)	VM		?

The consolidation ratio is a measure of the number of VMs placed on a physical machine. ESX Server's over commitment technology is an enabling technology allowing users to achieve a higher consolidation ratio, thus reducing the total cost of operation. Over commitment is the ability to allocate more virtual resources than available physical resources. ESX Server offers users the ability to overcommit memory and CPU resources on a physical machine.

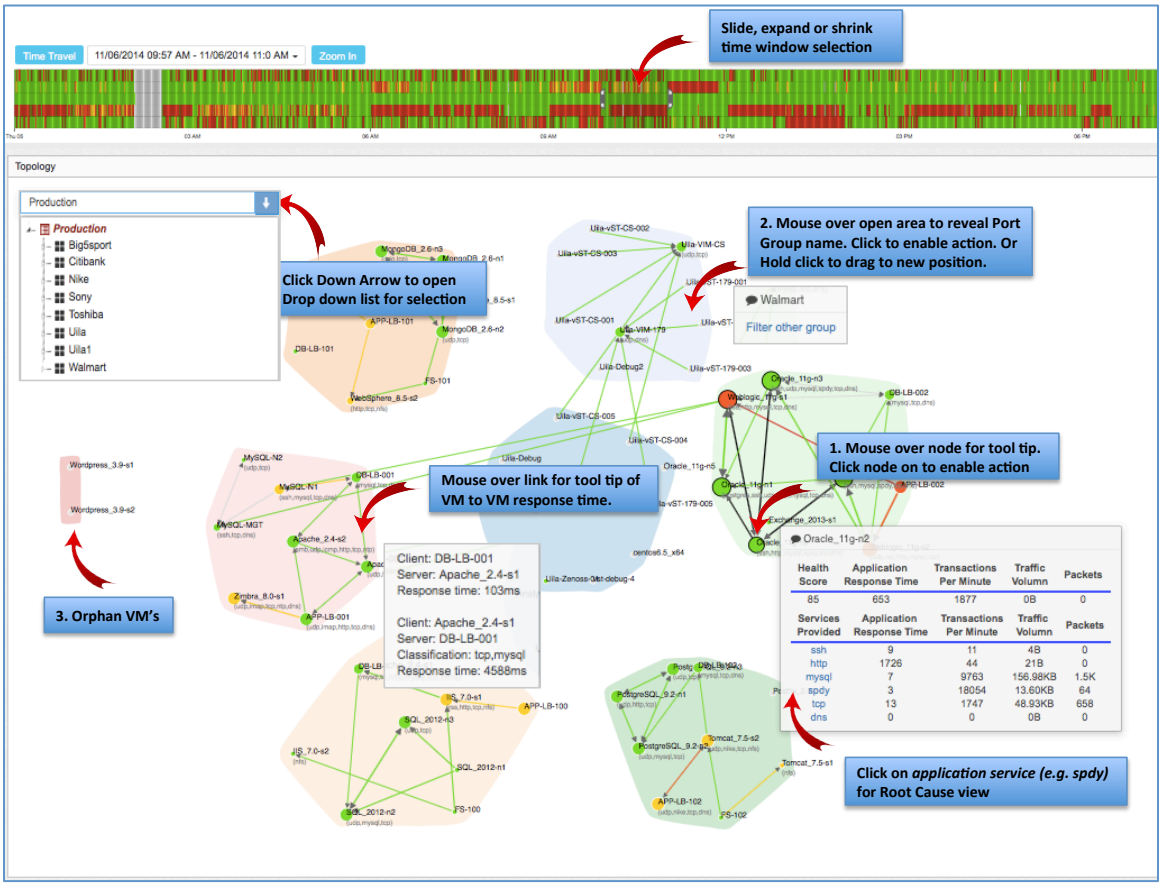
8 Application Topology

Application Topology provides you a visual view of ALL virtual Application (vAPP) service chains within your data center in real time. Applications within a defined Port Group are grouped together to help you quickly identify how each Application and its associated VM is communication with each other, and its performance health.

Application Topology view is started from Dashboard's Application Performance color wheel, or directly launched from the Tool Pane menu.

You can use Application Topology to

1. Select which vAPP (VM) performance is impacted by identifying the VM with the red color for further drill down for root causes,
2. Reveal how and if Port Groups are interconnected,
3. Identify any orphan VM's (VM's are standalone without communication with any other VM), which are the result of misconfiguration.



[Application Topology View]

Symbol	Definition	Mouse Over Information	Click Action
	Application VM name with list of protocol identified.	Highlight connections between this Application VM and neighbor VM's Show a list of active Application protocols and associated response time	Select one of the protocols to identify the root cause of slow response time
	Traffic flow between Applications	Displays average transaction response time between two VM's for each of the application service running.	None

Note: See standard color-codes in Table x.x

8.1 Navigation Tips

The scenario below describes the steps of finding the root cause of application performance degradations.

9 Flow Analysis

Flow Analysis diagram (also called Sankey diagram) is a powerful visualization tool to show you how your vAPP network traffic are traversing through physical devices (ToR switches, hosts, etc.), virtual entities (vSwitch, Port Group, vAPP, VM), and finally, to Application Services (or Classifier). Flow Analysis view can be launched from Dashboard's Network Health color wheel, or directly from the Tool Pane menu.

It can quickly

- Identify which infrastructure entities are impacting the Network Health in the Time Frame that is being monitored (one with the Red or Orange color)
- Review network round trip time, application response time and traffic volume of each application service (Classifier) of the respective entity.
- Facilitate further drill down to correlate Application performance impacts.

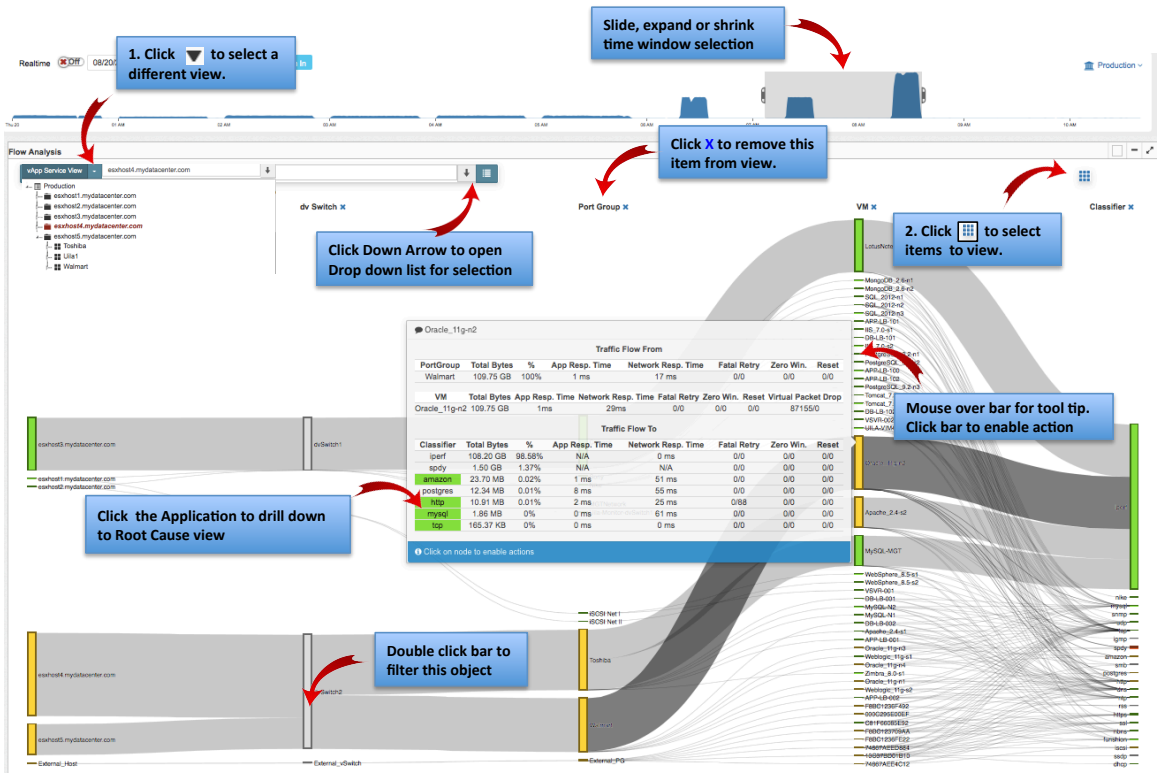
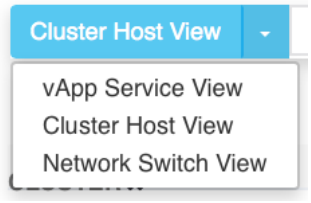


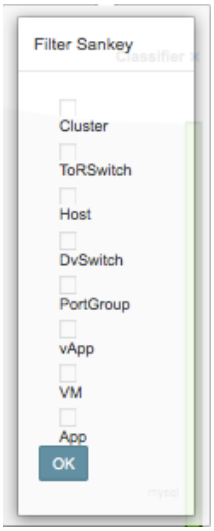
Figure X-X: Flow Analysis View

Additional Drop Down list and Button Definitions (on Figure X.X):

1. Click to display a drop down list to select a specific view of :



2. Click  to display a selection box to select which infrastructure components to display



-
- Select the entities that you wish to display in the Flow Analysis diagram.

Graphic	Definition	Mouse Over Information	Click Action
	Name of physical or virtual entity. Color reflects the network round trip time grading at this entity.	Review network round trip time, application response time and traffic volume of each application service (Classifier) of the respective entity.	Enable *Analyze Application Performance*. Launch Application Topology with filtered view.

Note: See standard color-codes in Table x.x

10 CPU Usage

CPU Usage diagram is a visualization tool to show you CPU usage and Health Score within your data center physical or virtual entities. CPU Usage view can be launched from Dashboard's CPU Health color wheel, or directly from the Tool Pane menu. User can select from two views:

1. Physical View: to see Cluster, Host and VM
2. Virtual View: to see Port Group, vAPP, VM

Below diagram shows the navigation method and tool tips:

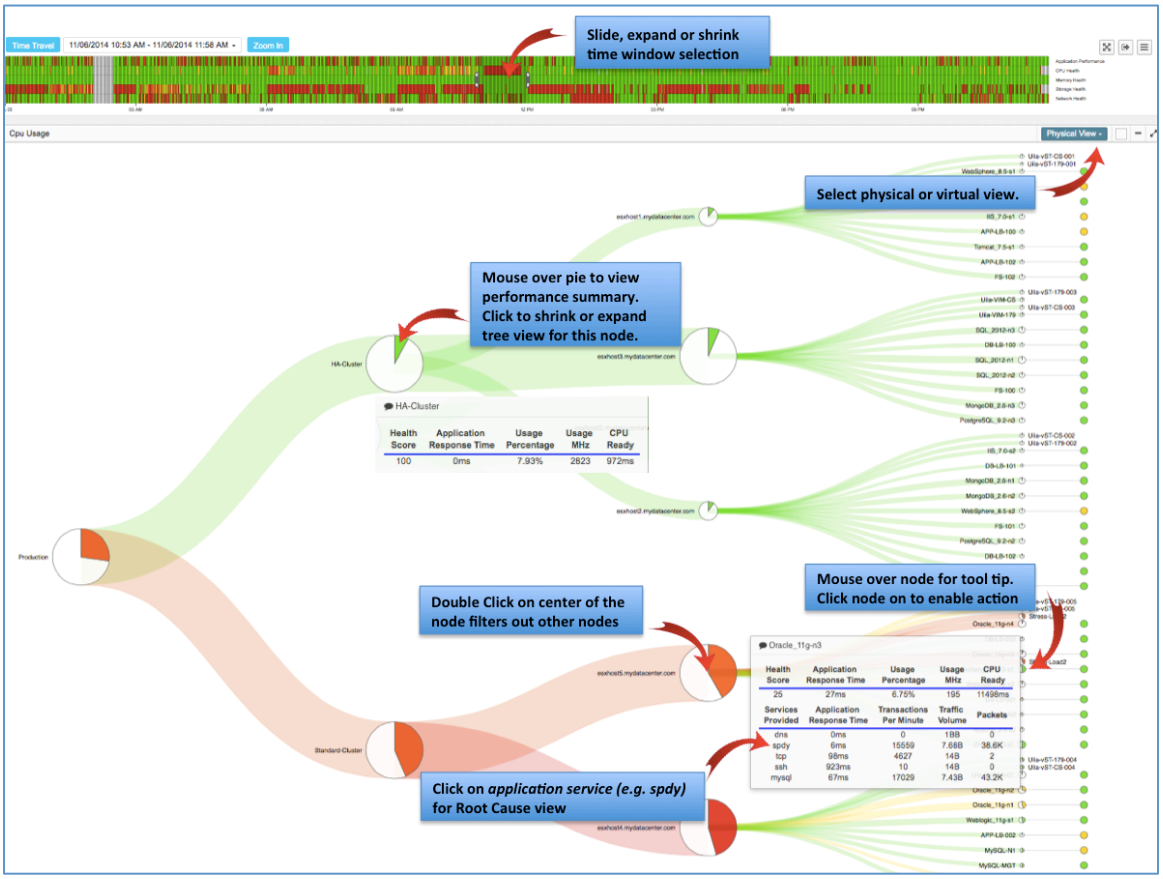


Figure X-X: CPU Usage View

Refer to Section x.x CPU Performance Metric for CPU metric definition, and how metrics are calculated to determine health score and the associated base line values.

11 Storage Usage

Storage Usage diagram is a visualization tool to show you Storage usage and Health Score within your data center physical or virtual entities. Storage Usage view can be launched from Dashboard's Storage Health color wheel, or directly from the Tool Pane menu.

The figure below shows the navigation method and tool tips in the Storage Usage view.

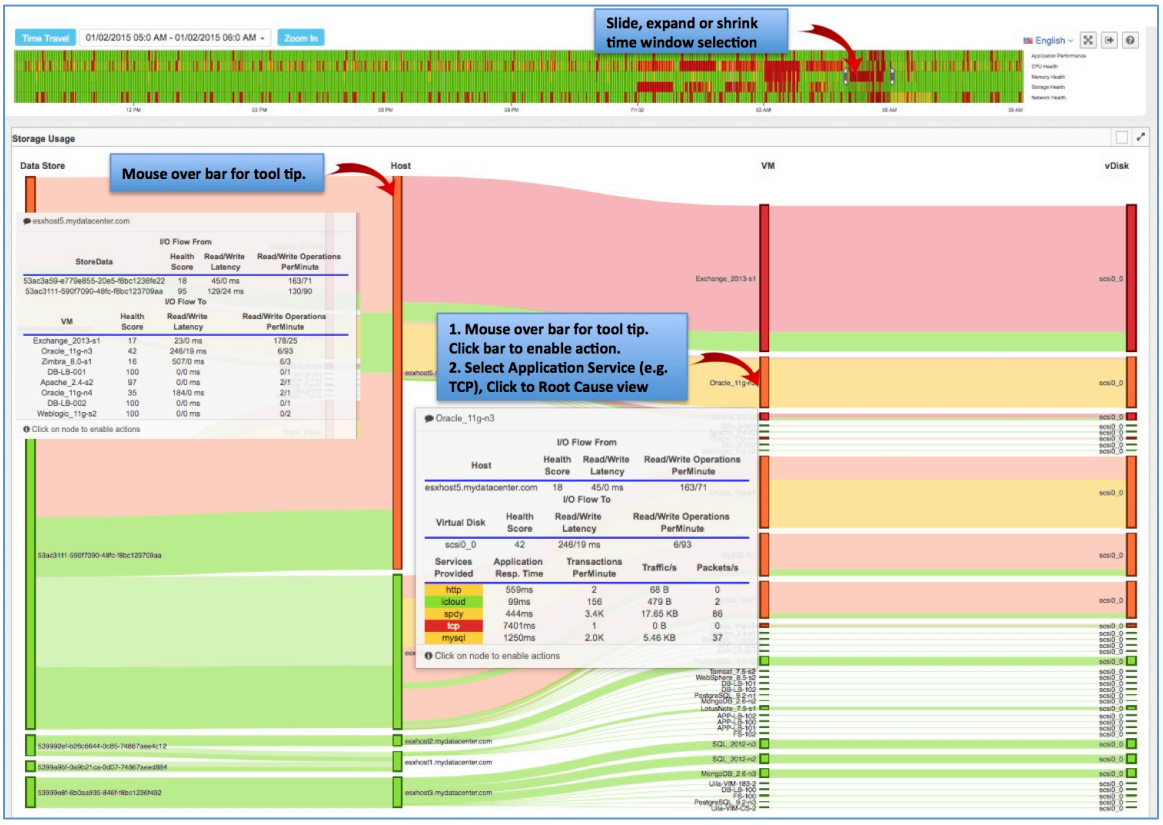


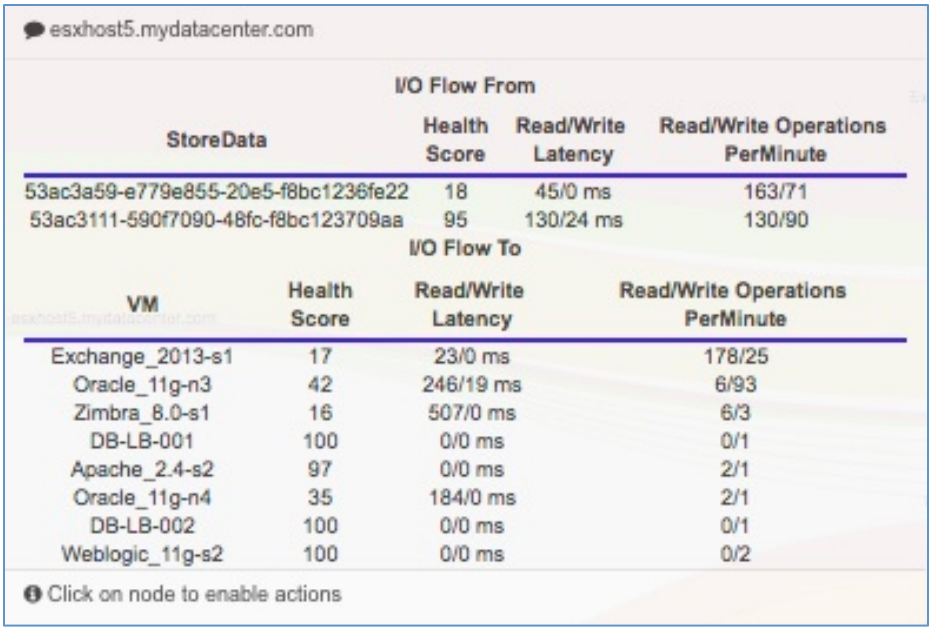
Figure X-X: Storage Usage View

Refer to Section x.x, Storage Performance Metric for Storage metric definition, and how metrics are calculated to determine health score and the associated base line values.

To help investigate performance issues, you can place the mouse over the vertical bar of each storage infrastructure component to reveal the health and performance summary of its upstream and downstream neighbors in a Tool Tip.

There are two types of Tool Tips:

1. One with Summary Information only:



esxhost5.mydatacenter.com

I/O Flow From

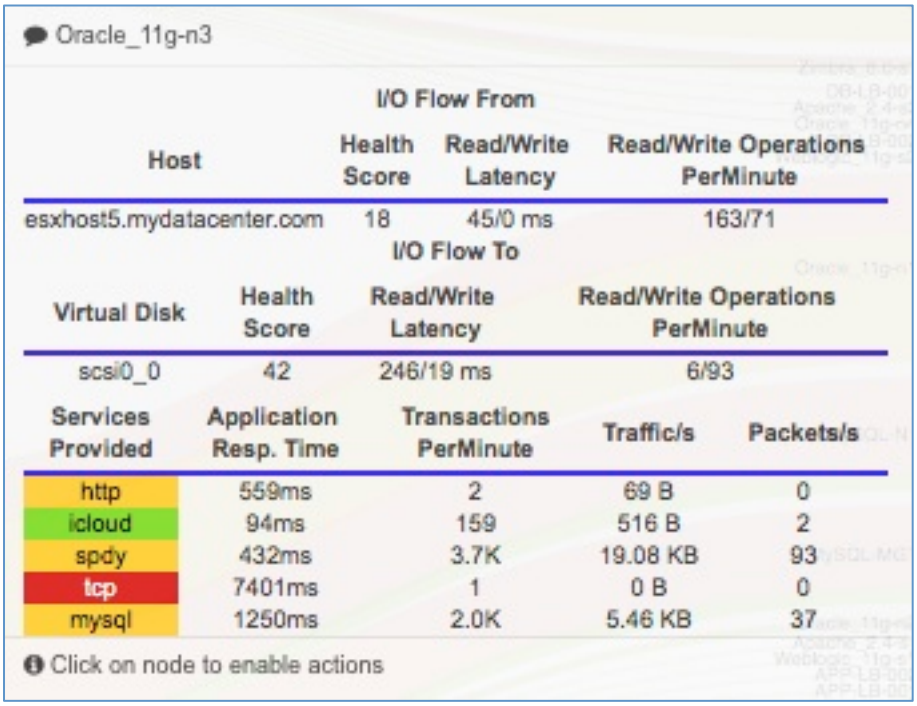
StoreData	Health Score	Read/Write Latency	Read/Write Operations PerMinute
53ac3a59-e779e855-20e5-f8bc1236fe22	18	45/0 ms	163/71
53ac3111-590f7090-48fc-f8bc123709aa	95	130/24 ms	130/90

I/O Flow To

VM	Health Score	Read/Write Latency	Read/Write Operations PerMinute
Exchange_2013-s1	17	23/0 ms	178/25
Oracle_11g-n3	42	246/19 ms	6/93
Zimbra_8.0-s1	16	507/0 ms	6/3
DB-LB-001	100	0/0 ms	0/1
Apache_2.4-s2	97	0/0 ms	2/1
Oracle_11g-n4	35	184/0 ms	2/1
DB-LB-002	100	0/0 ms	0/1
Weblogic_11g-s2	100	0/0 ms	0/2

Click on node to enable actions

2. One with Summary Information plus actionable list:



Oracle_11g-n3

I/O Flow From

Host	Health Score	Read/Write Latency	Read/Write Operations PerMinute
esxhost5.mydatacenter.com	18	45/0 ms	163/71

I/O Flow To

Virtual Disk	Health Score	Read/Write Latency	Read/Write Operations PerMinute
scsi0_0	42	246/19 ms	6/93

Services Provided	Application Resp. Time	Transactions PerMinute	Traffic/s	Packets/s
http	559ms	2	69 B	0
icloud	94ms	159	516 B	2
spdy	432ms	3.7K	19.08 KB	93
tcp	7401ms	1	0 B	0
mysql	1250ms	2.0K	5.46 KB	37

Click on node to enable actions

12 Stats Browser

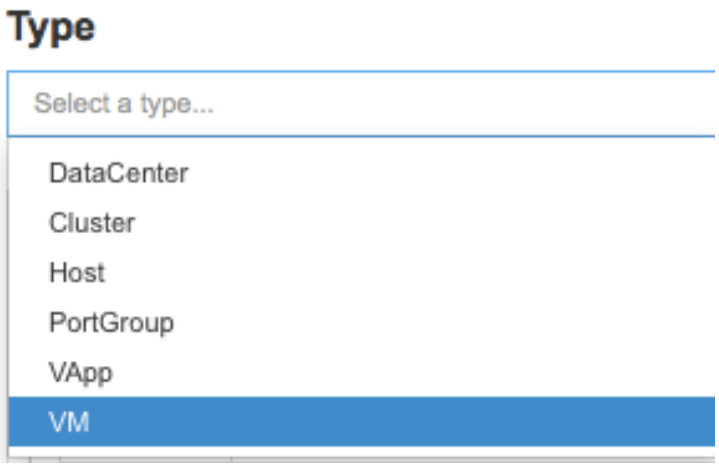
Stats Browser is another powerful visualization tool that places all the metrics collected for any of the infrastructure components; Cluster, Host, and VM in one single unified screen view. It is a particular useful when

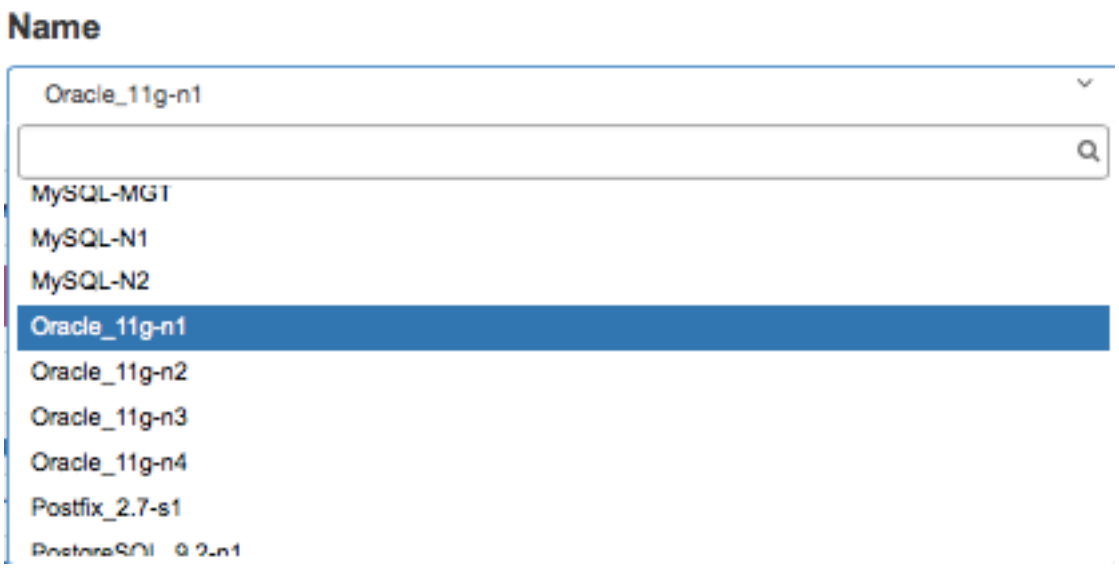
(1). You have identified the Root Cause of an application performance issue, and wish to further validate across all the infrastructure metrics,

(2).

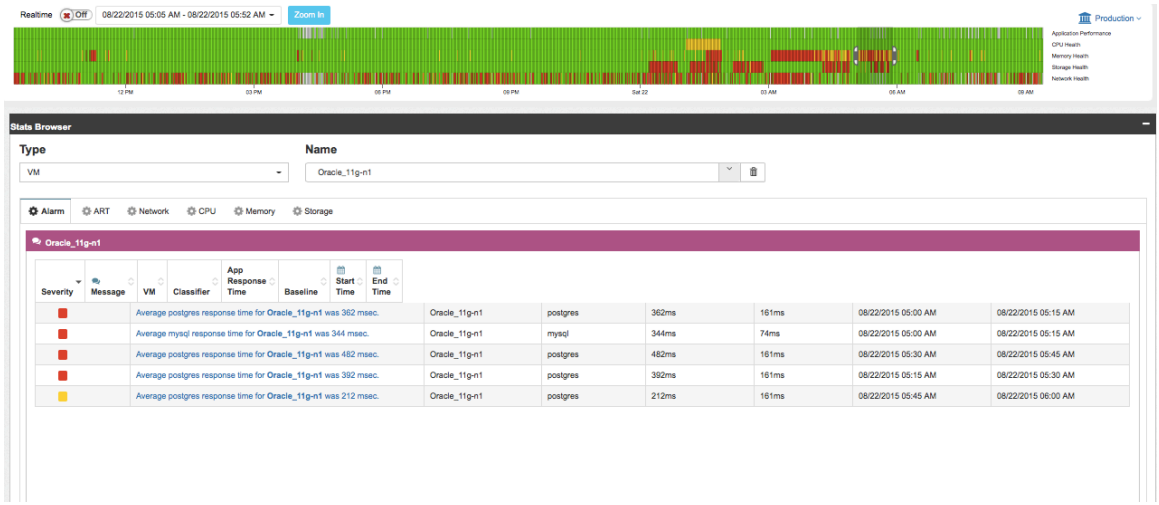
The figure below shows the navigation method and tool tips in the Stats Browser view:

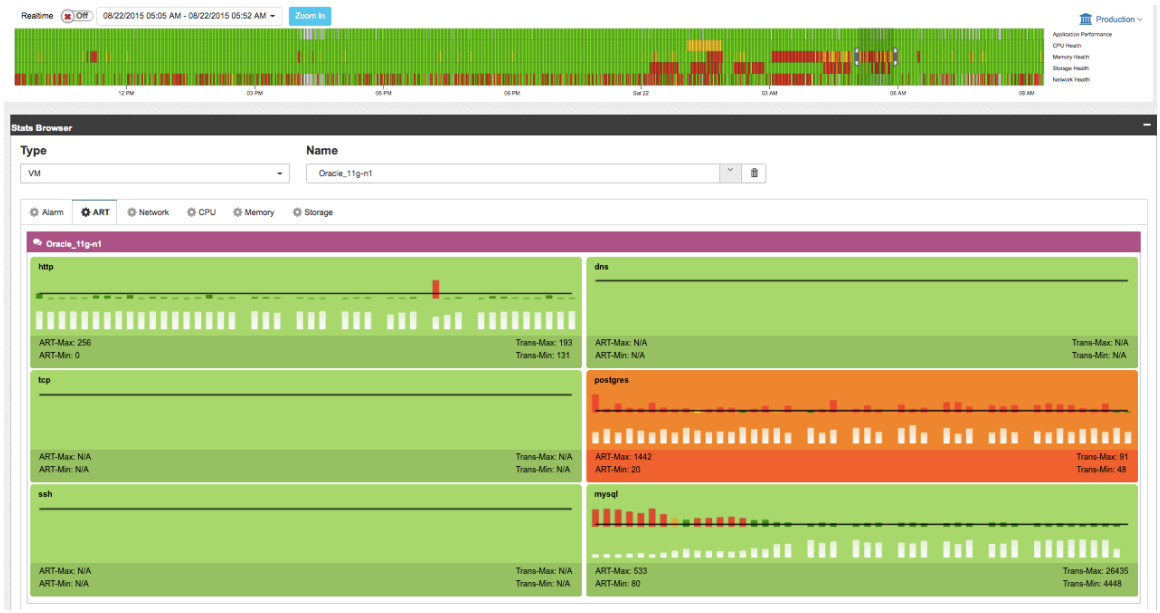
Use the Drop down box below to select Type and Name of the specific infrastructure units to view the summary of metrics over time bracket selected:

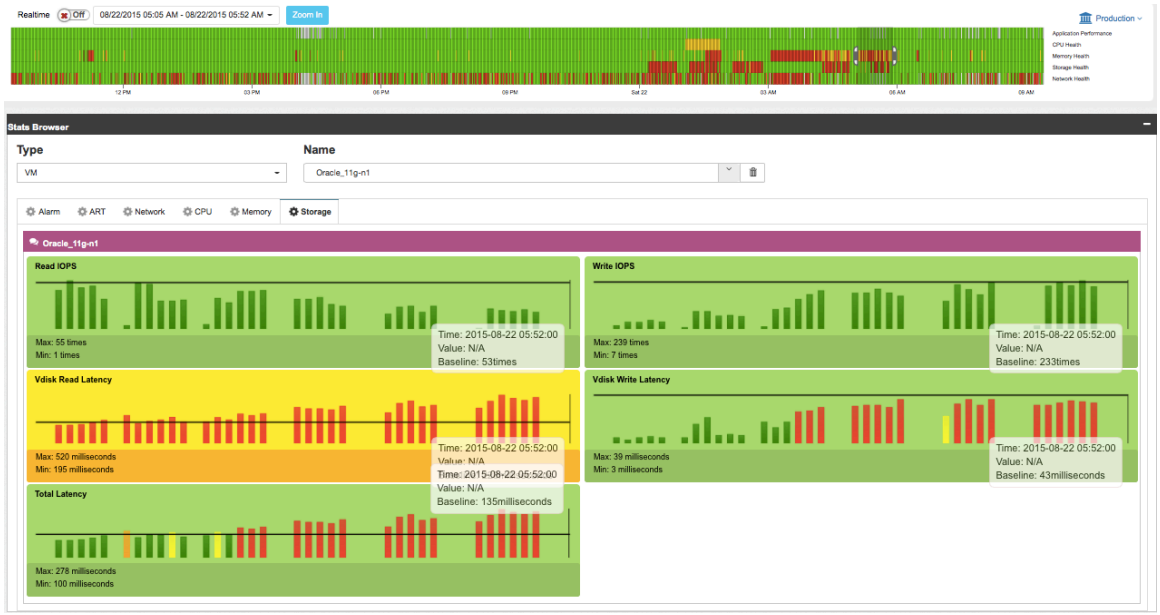




Here is the Example of the Metric summary selected for VM 'Oracle_11g-n1' between 5:05am to 5:52am, when applications *postgres* and *mysql* performance are degraded, and where the root cause is pinpointed.





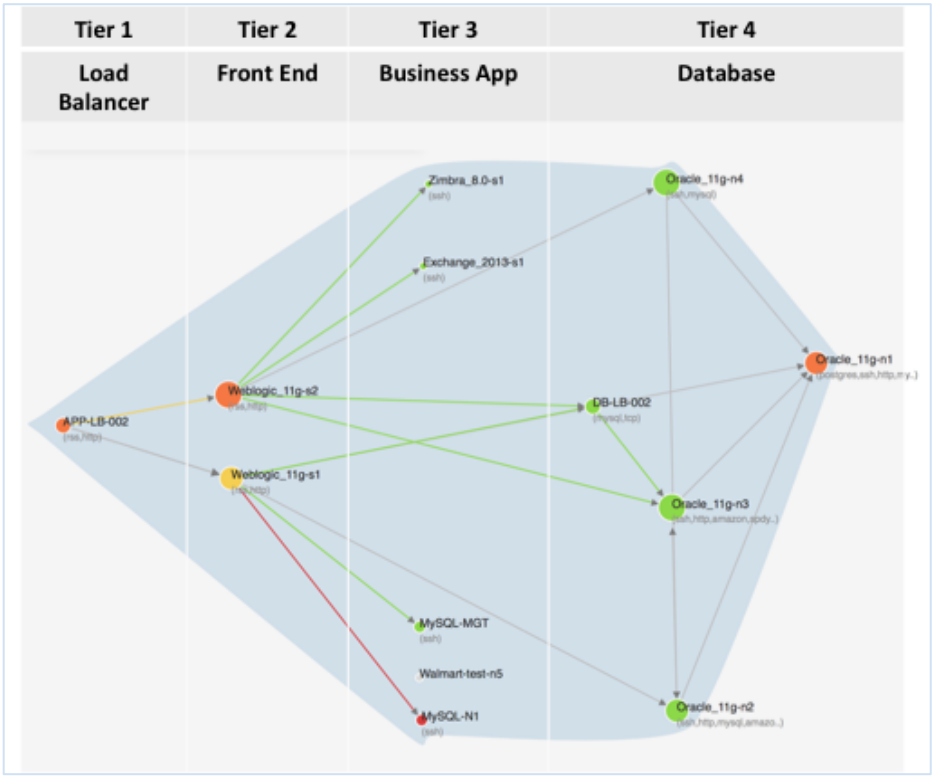


13 How to Conduct Root Cause Analysis

One of the most powerful features in Uila AIPM is the ability to perform application Root Cause analysis with just one or two clicks from various Infrastructure views (e.g. Application Topology) or from the Alarm List.

Before you conduct the Root Cause analysis, you should understand how modern applications are virtualized and deployed as multi-tiered applications cross multiple VM's and resides on different hosts, which will help you to use Uila to analyze the root causes of an particular application issue whether it is within the virtual infrastructure; i.e. CPU, memory, storage, or might be caused by another application (which we call Dependent Service), or network issues either in virtual or physical layers.

Here is an example of a multi-tiered application with may sub components deployed cross multiple virtual machines.

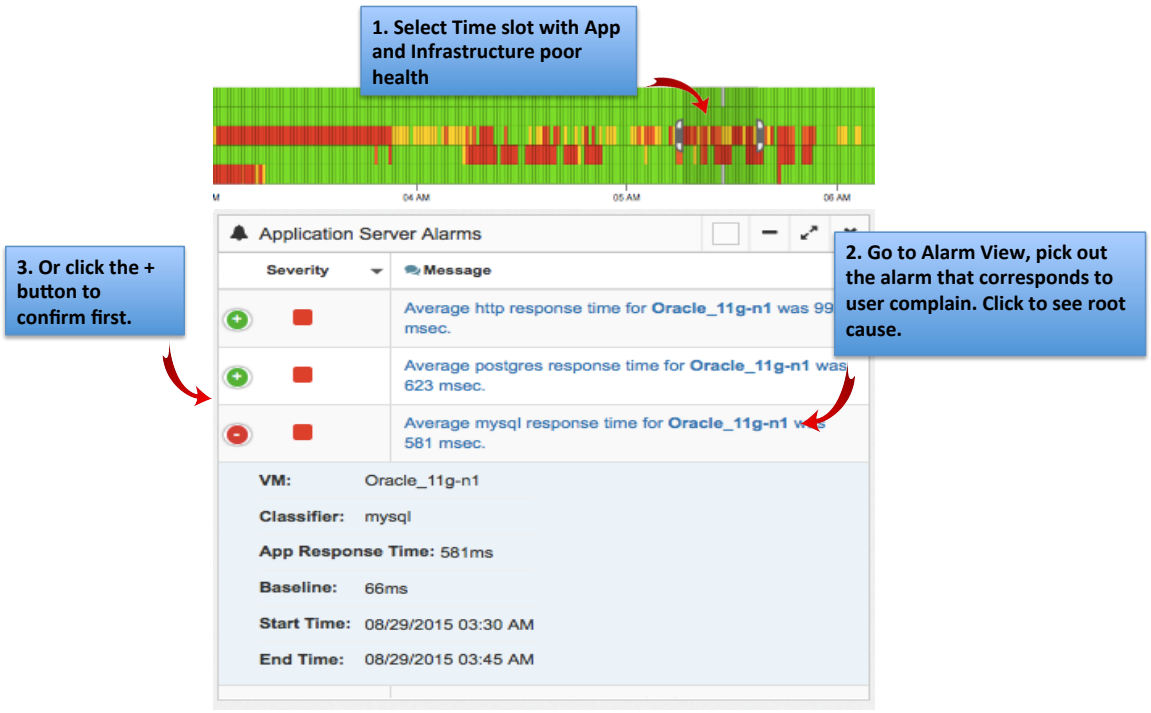


13.1 Rapid Troubleshooting of Application Performance Degradation Root Cause

The scenario below describes the steps of finding the root cause of application performance degradations from the Alarm View.

Step 1. Alert on app performance issue (or receive complaint!)

1. Find a time slot where you see applications and/or infrastructure health are in critical health score (usually associated with red color).
2. Go to Alarm view, Look for the application; e.g. '*mysql*' in VM '*Oracle_11g-n1*, that has response time is high above baseline. Click the Alarm to drill to Root Cause view. See Step 2.
3. Optionally, you can click + circle to see more alarm details.
4. You can also click  the to expand Alarm view to full screen.



[Figure x.x]

Step 2. Drill down directly to infrastructure root cause

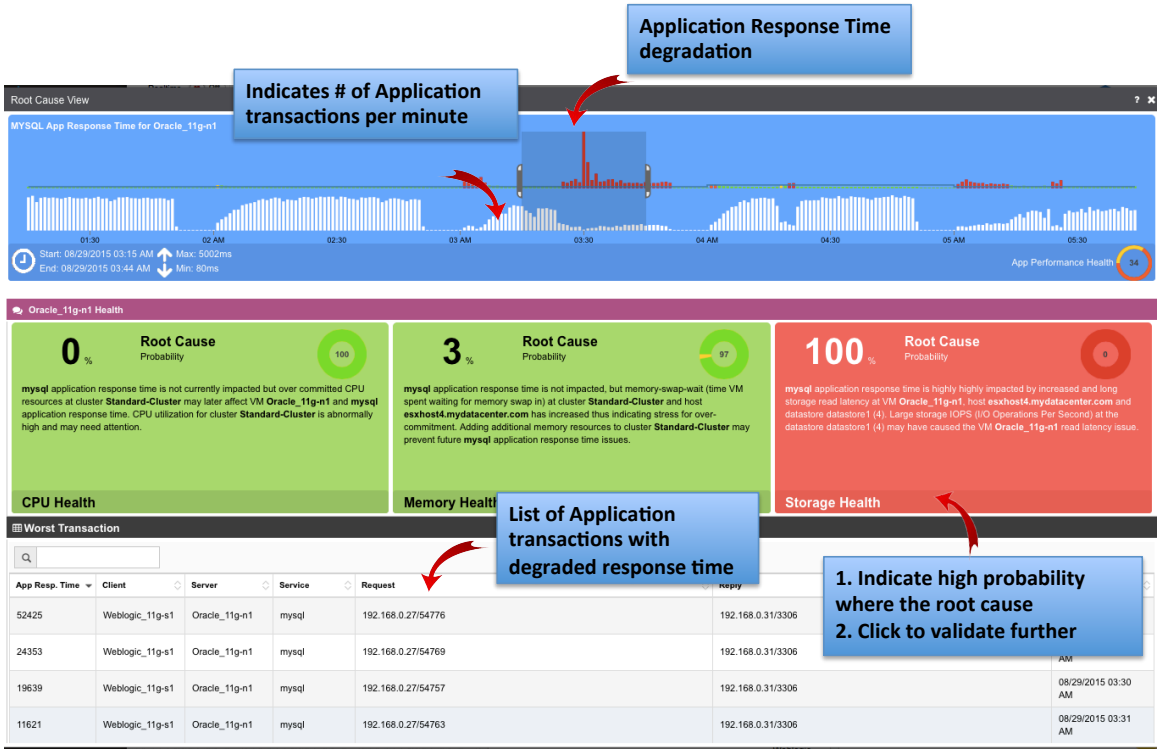
See Illustration in Figure X.X

1. *mysql* ART for VM '*Oracle_11g-n1'* are slowed to above critical threshold for about consecutive 30 minutes trend.
2. Uila root cause analytics is able to confirm the root cause is due to '***mysql*** *application response time is highly impacted by increased and long storage read latency at VM* ***Oracle_11g-n1****, host* ***esxhost4.mydatacenter.com*** *and datastore datastore1 (4). Large storage IOPS (I/O Operations Per Second) at the datastore*

*datastore1 (4) may have caused the VM **Oracle_11g-n1** read latency issue*. Click any where in the Storage Health sub-window for further investigation and validation.

Note: Bottom of the screen lists each of the *mysql* send and receive transactions between 2 VM services that corresponding to slow ART.

3. See Figure X.X. VM, Host and Cluster all show CPU Ready Wait value above critical threshold (Red).
4. However, VM, Host and Cluster all show CPU Usage within normal threshold (Green), which mean root cause to slow HTTP ART may not be CPU related. Click  to return to the first Root Cause view.
5. Next step is to investigate the Storage Health; which is the second worse performer.

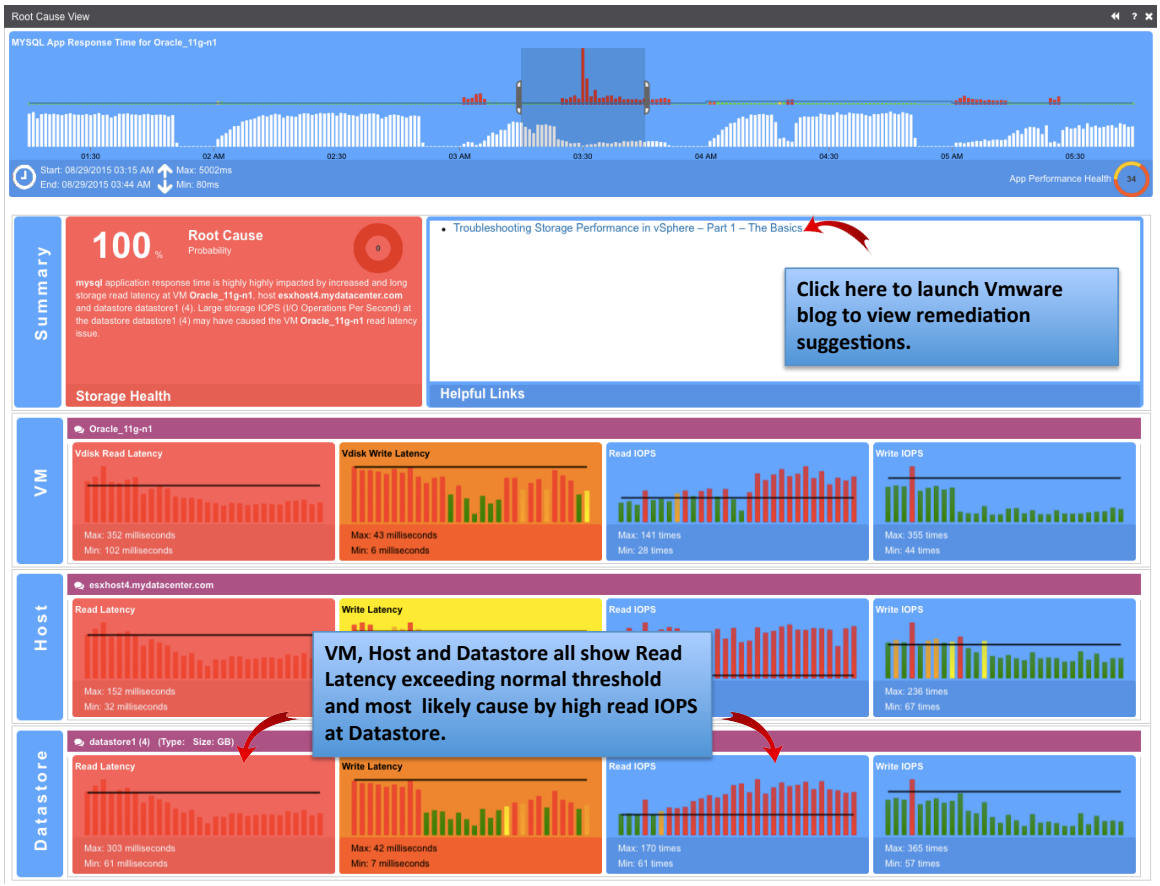


[Figure x.x]

Step 3. Validate Root Cause and Get Remediation

See Figure X.X below.

1. VM, Host and Datastore all show Read Latency exceeding normal threshold and most likely caused by high read IOPS at Datastore.
2. Next step, Click "Troubleshooting Storage Performance in vSphere – Part 1 – The Basics" link to review troubleshooting steps to resolve issue.



[Figure x.x]

14 Appendices

[To Add]

14.1 Reference Documents

VMware vCenter Installation and Setup Guide:
http://pubs.vmware.com/vsphere-55/topic/com.vmware.ICbase/PDF/vsphere-esxi-vcenter-server-552-installation-setup-guide.pdf

VMware Server and Host Management:
http://pubs.vmware.com/vsphere-55/topic/com.vmware.ICbase/PDF/vsphere-esxi-vcenter-server-552-host-management-guide.pdf

Virtual Machine Administration Guide:
http://pubs.vmware.com/vsphere-55/topic/com.vmware.ICbase/PDF/vsphere-esxi-vcenter-server-552-virtual-machine-admin-guide.pdf

VMware vSphere Monitoring and Performance:

http://pubs.vmware.com/vsphere-55/topic/com.vmware.ICbase/PDF/vsphere-esxi-vcenter-server-55-monitoring-performance-guide.pdf