

trilioVault

User Guide

Table of Contents

1 Introduction 3

1.1 Key Concepts 3

1.1.1 Workload 3

1.1.2 Workload Types 3

1.1.3 Workload Configuration 4

1.1.4 Virtual Application Snapshot Technology (VAST™) 5

1.1.5 trilioVault Controller Nodes 5

1.1.6 Storage 5

1.1.7 VAST for VMware® environment 6

1.2 Installation & Configuration 6

2 trilioVault Dashboard 6

2.1 Node Metrics 7

2.2 Backup Storage for Workloads 8

2.3 Workload Snapshot-job Summary 9

2.4 Audit Log 9

2.5 Settings 10

3 Create a Workload 11

3.1 Create Serial/Parallel Workload 13

3.2 Cassandra workload 16

3.3 MongoDB Workload 17

4 Workload Topology 20

4.1 VM View 20

4.2 DB View 20

5 Modify Workload Name and Credentials 21

6 Workload Snapshot 23

6.1 Snapshot View 23

6.2 Create Workload Snapshot 25

7 Workload Restore 28

8 Known Issues 34

9 Appendix 34

9.1 Pre-requisites for using trilioVault for Cassandra workload 34

9.2 Pre-requisites for using trilioVault for PostgreSQL 35

9.3 Pre-requisites for using TrilioVault for MySQL 36

9.3.1 MySQL Backup Option I 36

9.3.2 MySQL Backup Option II 37

# Introduction

Trilio Data provides trilioVault™, a Business Assurance platform. Enterprise IT and Cloud Service Providers can now leverage Backup & Disaster Recovery as a Service for Cloud solutions in both VMware and OpenStack. This multi-tenant, self-service, policy-based solution is designed to protect NoSQL Application Workloads from data corruption or data loss, providing point-in-time snapshots, configuration and change awareness and seamless 1-click recovery. The benefits are faster and reliable recovery, and easier migration of NoSQL databases between Clouds.

## Key Concepts

Before we dive into the trilioVault User Interface, it is important to understand a few key concepts.

### Workload

Traditionally, a workload is defined as the amount of processing that the computer has been given to do at a given time. The workload consists of some amount of application programming running in the computer and usually some number of users connected to and interacting with the computer's applications. However the cloud computing and the new generation of applications changed the definition of workload and the workload is not limited to amount of processing to a single computer. Trilio Data defines workload as a collection of:

* Computing,
* Networking,
* Firewall Settings,
* Networking Configurations, and
* Storage Resources

This collection is participating in processing the workload.

A workload is expressed as a collection of virtual machines (or “VMs”), the interconnectivity between the machines and the virtual disks mapped to those machines.

### Workload Types

Workloads can be generic workload types where few VMs are grouped together. Interdependencies between the VMs can be expressed explicitly.

* For example, a set of virtual machines have no “write” order dependencies but are grouped together because they are all part of one application deployment. These types of workloads are called Parallel Workloads, as there are no inherent dependencies between virtual machines.
* In another example, the virtual machines are dependent on each other. For example data flow goes from VM1 to VM2 in which case VM2 is dependent on VM1. These relations are expressed as Serial Workloads where the dependencies between VMs can be explicitly specified and workload processing will make use of these relations.

There are other types of workloads where relations between VMs are expressed by the applications that these VMs are part of.

* For example NoSQL databases such as MongoDB™, Apache Cassandra™ (or “Cassandra”) and others distribute the data between multiple virtual machines. Serial or Parallel Workloads cannot describe the dependencies between VMs of these applications and require in-depth understanding of application configuration to learn the interdependency between these VMs. To support these types of applications, Trilio Data defines a type for each of these applications.
* There is another type of workload called Composite Workload. For example an application deployment may contain a NoSQL database such as MongoDB and a SQL database such as MySQL. You can define a workload for each of these applications but the interdependencies between these two applications cannot be described by about types. Trilio Data defines a composite workload type to express the data flow between two different workloads.

With these types of workloads, user can express their applications in terms of workloads and define protection policies for their applications.

### Workload Configuration

Workload configuration can be divided into two areas:

#### Application Configuration

Application Configuration deals with the configuration that the application keeps in the VM. Usually such configuration is kept in application specific files under /etc directory for Linux® and in the Registry for Windows®. Application configuration may include:

* Port numbers for various services,
* Database file path names, and
* Peer Nodes if the application is a distributed application.

NOTE: If the application is a distributed application such as MongoDB or Cassandra, the configuration includes number of nodes, replicas and sharding keys etc.

#### Infrastructure Configuration

Infrastructure configuration of an application includes:

* Backend storage,
* Number of volumes,
* Number of VMs, and
* Network connectivity between VMs, IP addresses, and firewall settings.

### Virtual Application Snapshot Technology (VAST™)

VAST is the trilioVault technology that captures and makes application consistent snapshots of workloads. VAST includes application aware workflows that discover application topologies, built application specific snapshot workflows and an execution engine that execute the workflows to get application consistent snapshots. VAST can automatically detect application and infrastructure configuration changes between snapshots and perform snapshot operations accordingly.

* For example, NoSQL class of databases are highly distributed among multiple VMs. It is common to add additional VMs or storage to either increase the capacity or enhance the IO performance. When these changes happen, VAST automatically discovers the changes and performs snapshot operation accordingly. In this example, the workflow takes a full snapshot of the newly added VM or storage device and takes an incremental snapshot of the VMs and devices that existed since the last snapshot operation.
* NoSQL databases are built to be highly available even in the face of one of more node failures. VAST will automatically detect surviving nodes and perform the snapshot of surviving nodes.

### trilioVault Controller Nodes

These are also known as nodes in the trilioVault user interface. A trilioVault node is a virtual machine that runs the trilioVault software. This software is responsible for managing backup and recovery of the workloads that are protected with the software. trilioVault is a distributed appliance. That means you can “spin up” as many VMs as you would like and trilioVault will still function as one service. A multiple VM solution can be used for high availability and load distribution. However, first you must start with a controller node, which is a primary node and then add additional nodes as needed.

### Storage

In this document, storage is referred to the Backup Storage repository. In this current version of trilioVault, there is only one repository. This Backup Storage repository is exposed to the trilioVault controller nodes for storing workload snapshots. This backup storage can be local storage or a NFS File system.

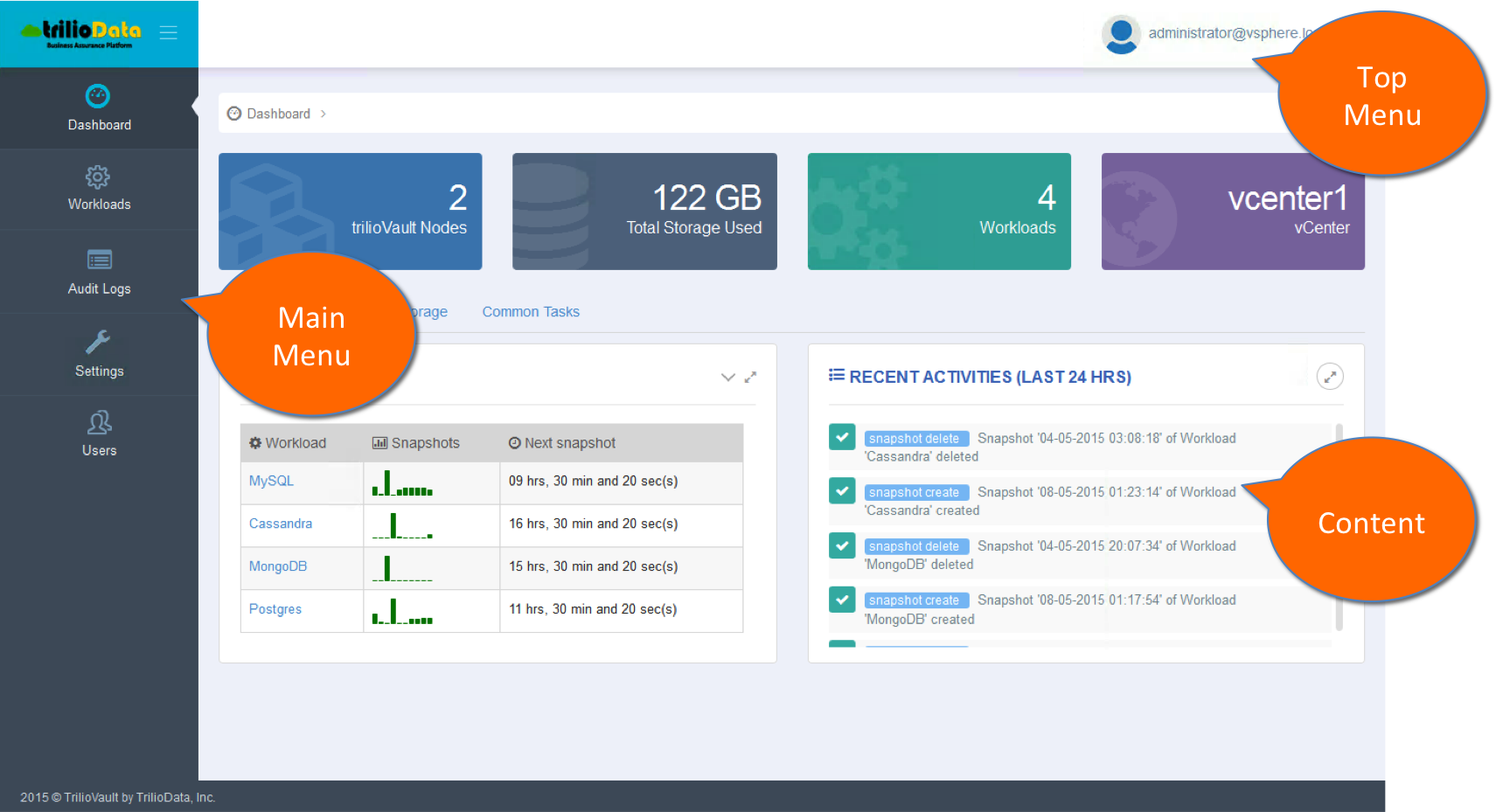
### VAST for VMware® environment

VAST implementation for VMware-based environments leverage the changed block-tracking feature of [VADP](https://pubs.vmware.com/vsphere-55/index.jsp#com.vmware.vddk.pg.doc/vddkBkupVadp.9.1.html) to take efficient snapshots of workloads.

## Installation & Configuration

To install and configure the trilioVault software, please refer to the Quick Start Deployment Guide.

# trilioVault Dashboard



The dashboard for trilioVault can be accessed via url, http://<tvaultgui-ip-address>:3000 from a standard Internet browser. Where <tvault-ip-address> is the IP address of the TrilioVault controller node. Use Vcenter credentials that you have provided in configuration step.

The user interface consists of 3 areas. To the left of user interface is the navigation bar in the form of the main menu. At the top of the user interface, you have the Top Menu and the remainder of the real-estate is for the Content.

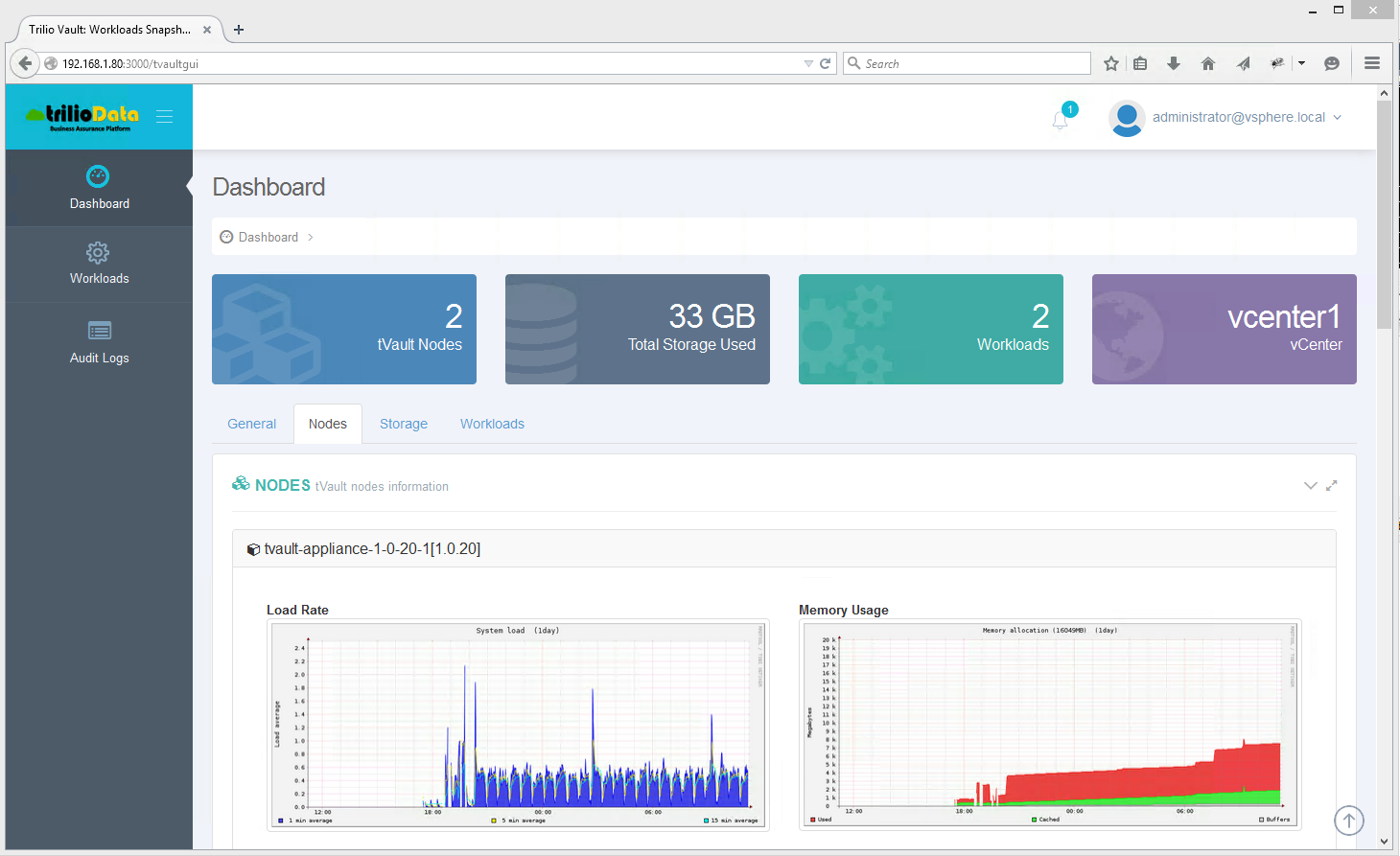
The Main Menu enables a user to navigate through the user-interface, while the Top Menu includes the user that is logged in.

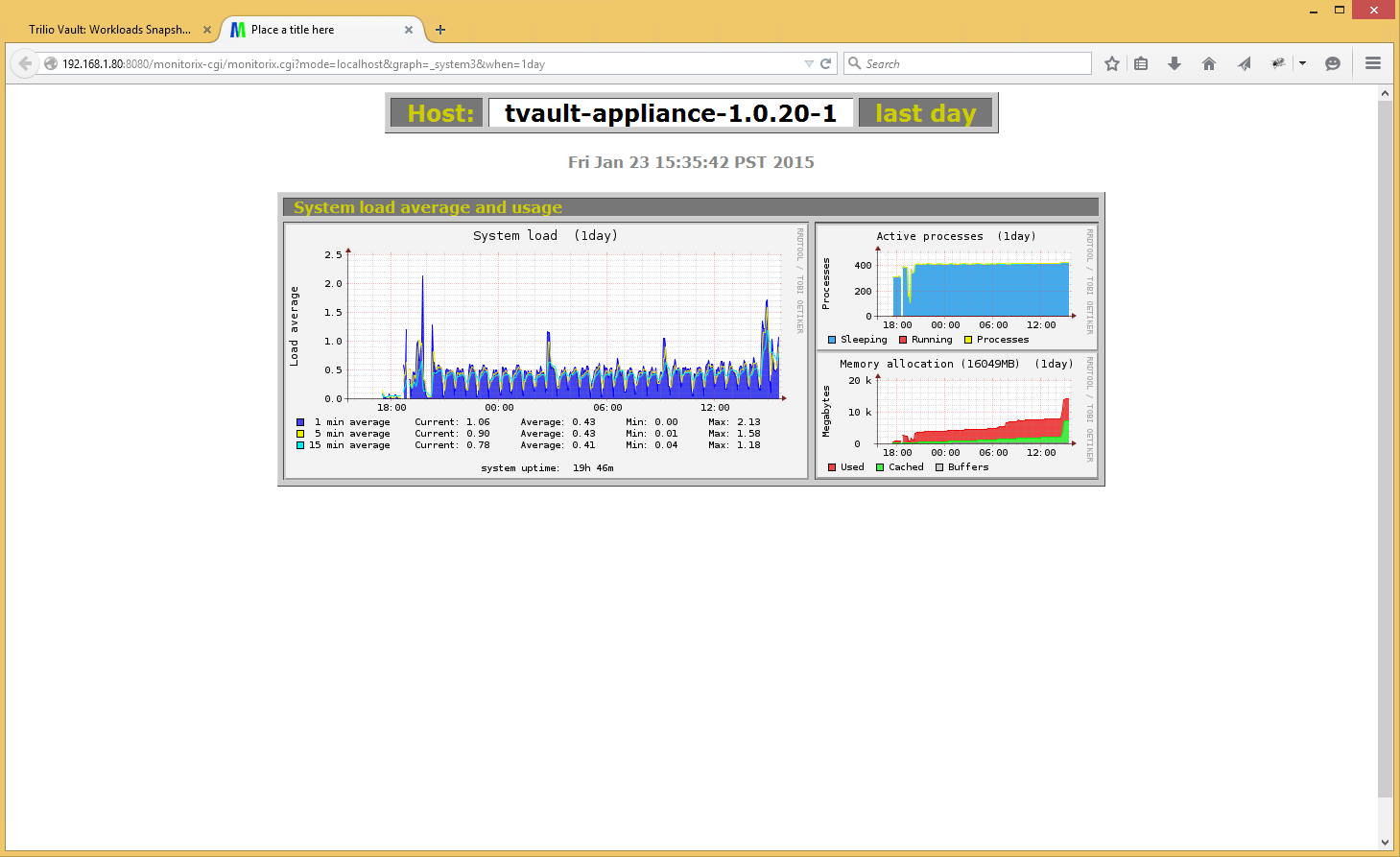
The top-level home is the dashboard view that gives a high level management view of the TrilioVault environment. This consists of the status, general tasks and list of recent activities. The dashboard provides the status for all TrilioVault nodes, amount of Backup Storage used, # of Workloads defined and the VMware vCenter name. In the General tab section, one can view the snapshot summary of the workloads and all the recent activities from the last 24 hours.

## Node Metrics

TrilioVault provides node metrics for System load, memory usage, network, disk activity and File System usage. In order to view the metrics, follow the steps as listed below:

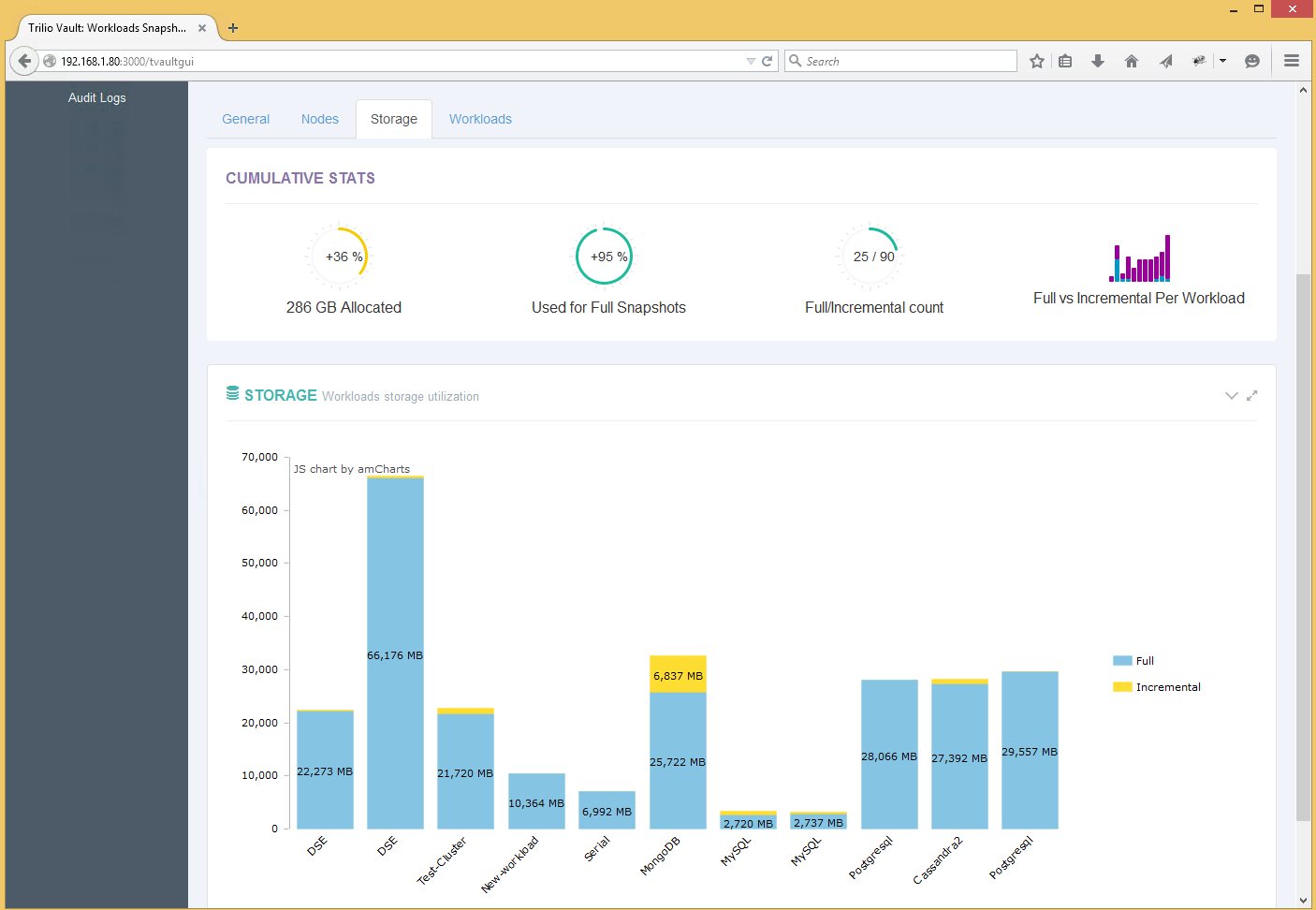
* Navigate to the Dashboard by clicking on the Dashboard icon in the main menu.
* Click on the Node tab. This will list all the TrilioVault controller nodes in the system.
* Click on a particular node. This will give you all the metrics for the node selected.



* You can get further details by clicking on the metrics. An example of System load is given below.

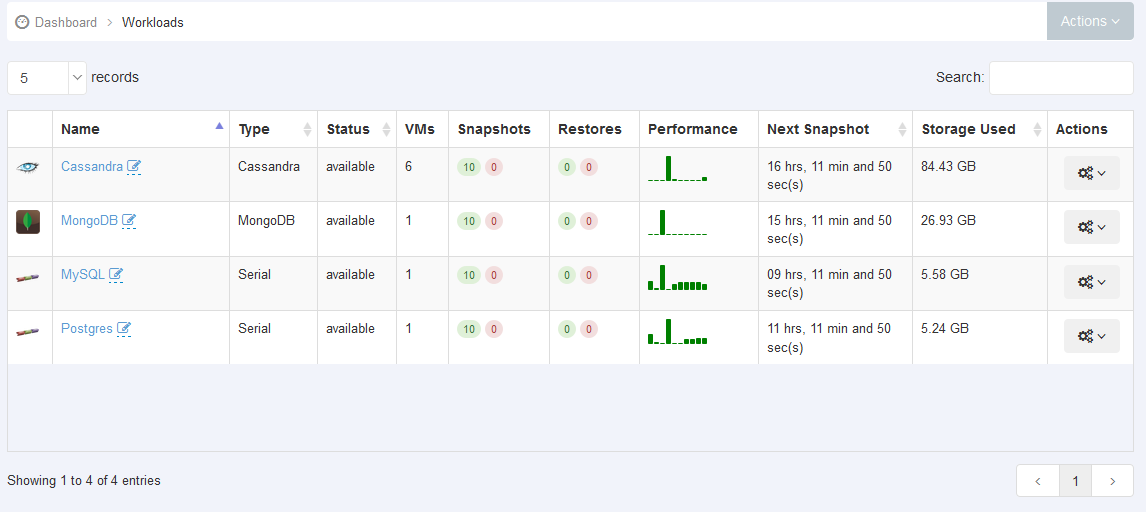
## Backup Storage for Workloads

You can view the Backup Storage used by each and for all workloads by selecting the storage tab in the content area of the Dashboard. In the screenshot below, there are 2 sections. The first sections gives you backup storage consumed by each workload. The system provides a split of storage required for Full vs. Incremental Snapshots Per Workload. The second section provides cumulative stats on % of Storage Used vs. Available Storage, % Storage for Full vs. Incremental Snapshots, # of Full vs. Incremental Snapshot Count and stats of Full vs. Incremental Per Workload.



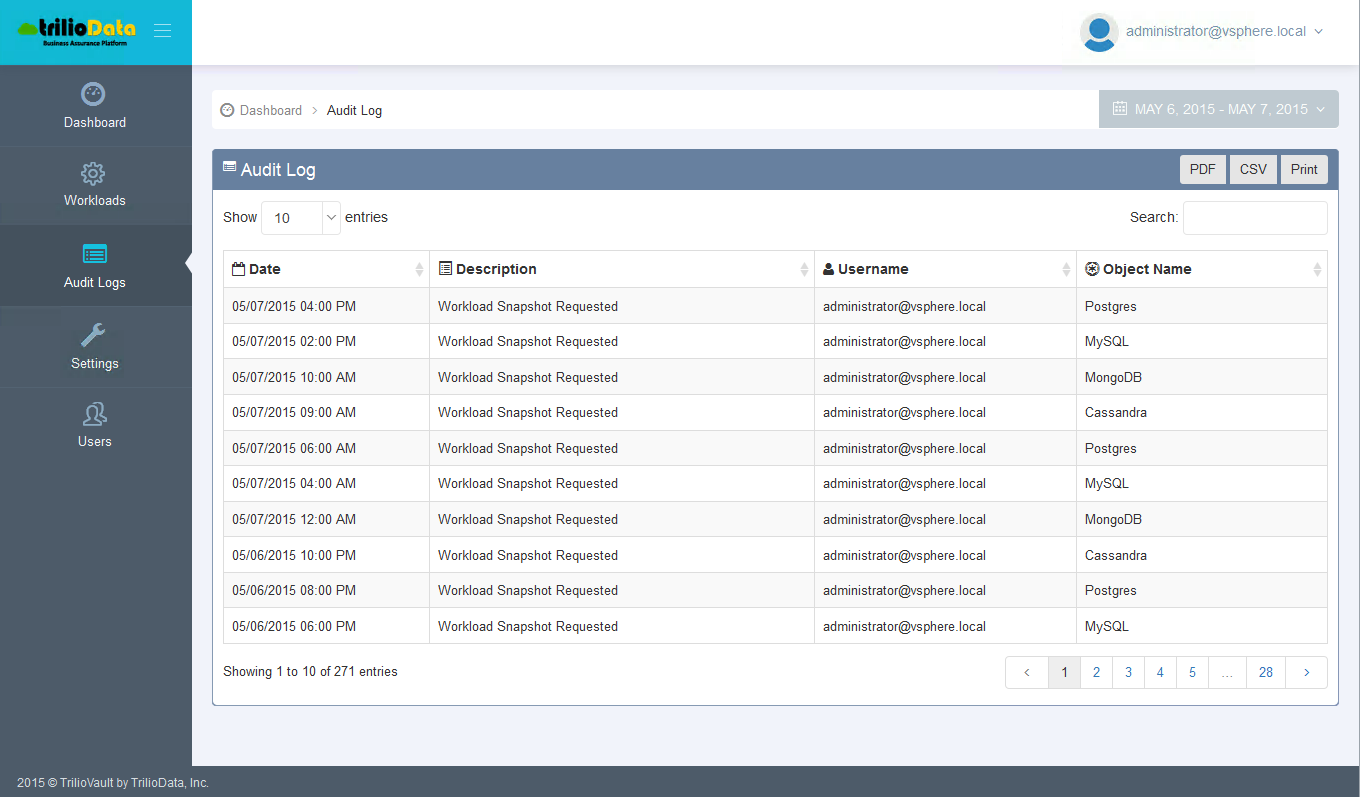
## Workload Snapshot-job Summary

You can get a summary of the Workload snapshot-job by clicking on Workload tab that is the Dashboard’s content area. The workload job summary consists of bar graph that depicts time to Backup full and incremental snapshots and when the next scheduled snapshot will be taken as shown below.



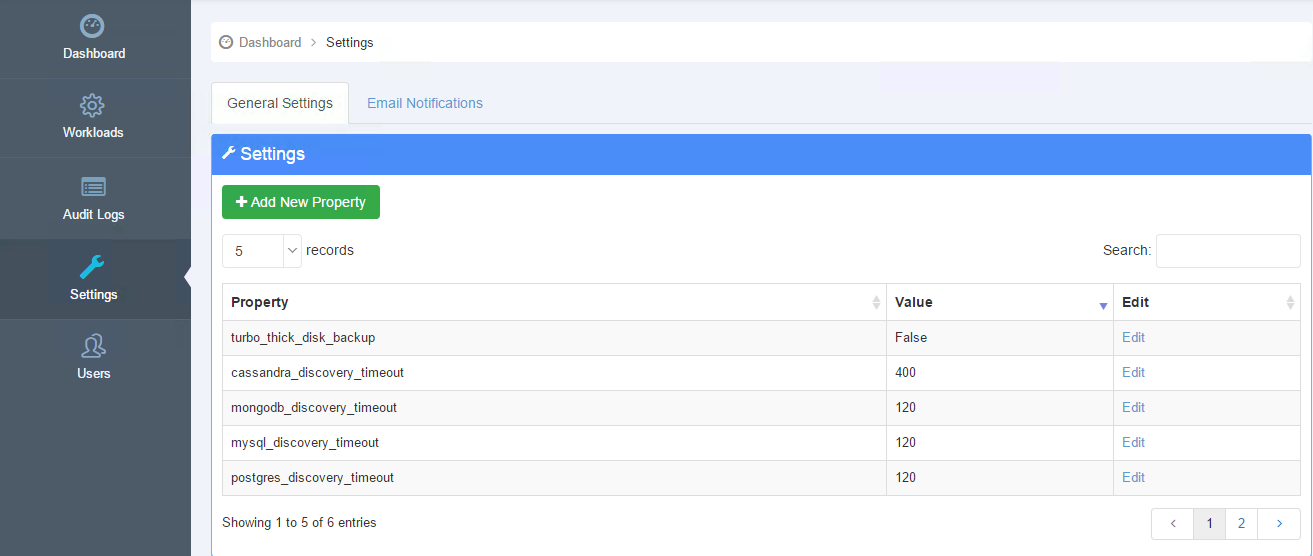
## Audit Log

trilioVault generates the Audit Log that consists of user initiated or scheduled actions. This list can be viewed by clicking on the ‘Audit Log’ from the main menu. Please see an example of the Audit Log that was generated. This Audit Log can also be saved in PDF & CSV formats.

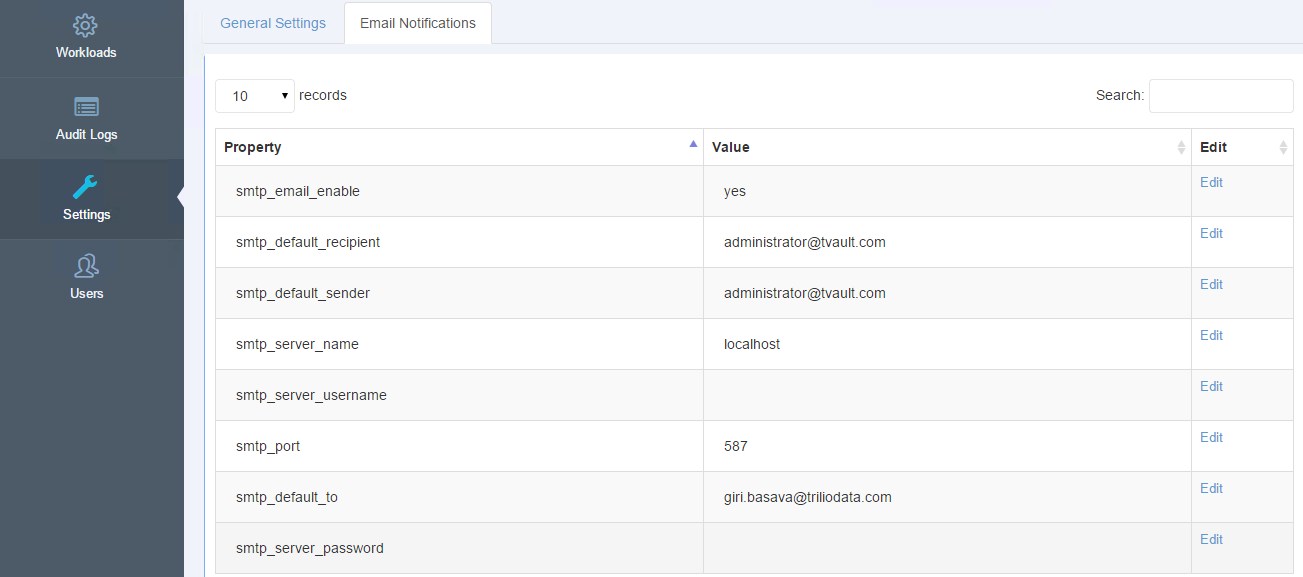


## Settings

In the settings section, one can set tunable parameters for the trilioVault software. There are default values for the parameters that should only be changed for troubleshooting perspectives by clicking on the edit field for the parameter as shown below. The new value has to be saved for it to be enforced.

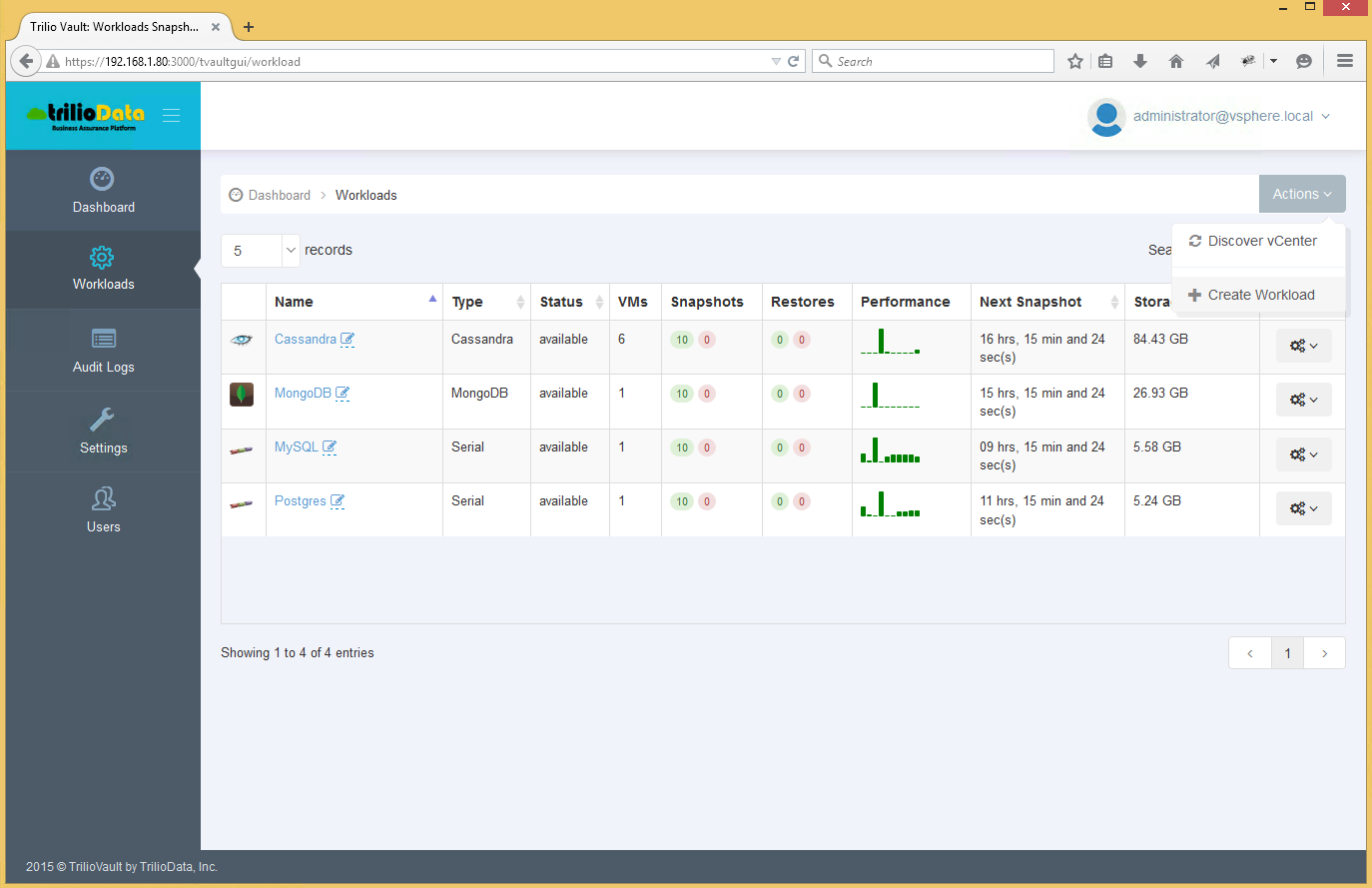


In this section, one can also set the parameters for email notification as shown in the figure on the next page under “Email notification”

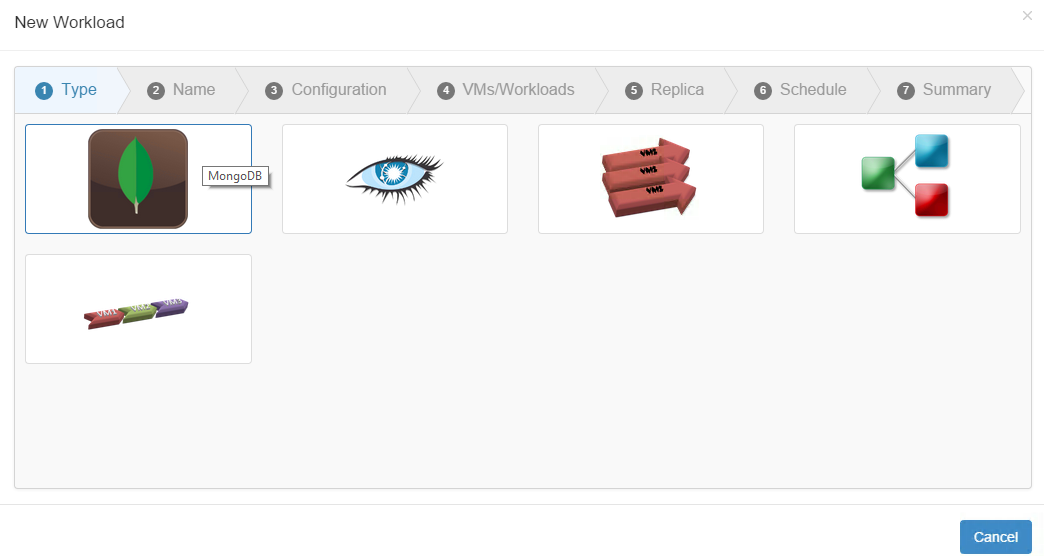


# Create a Workload

You can start the workload creation by clicking “Create Workload” in the dashboard under the General Task area or by clicking Workload in the main menu and then selecting Create Workload under the Actions menu as shown in the figure below.

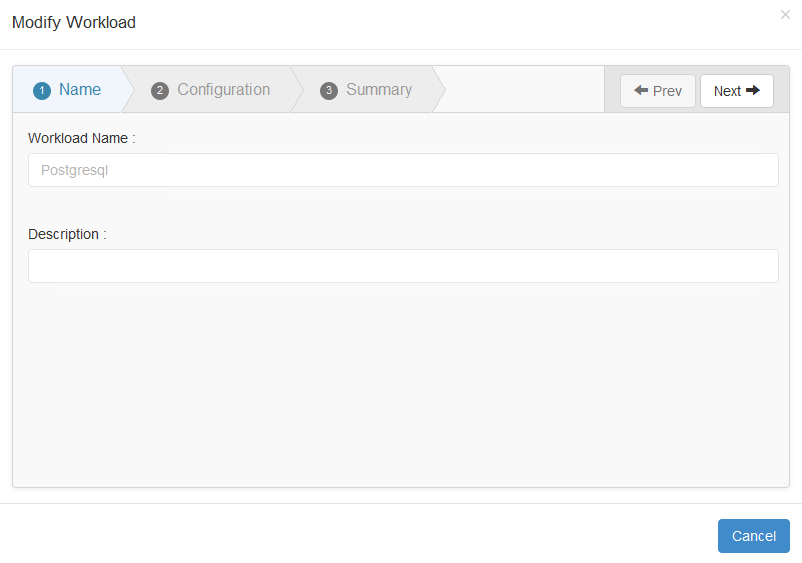


NOTE: Before creating any workloads, the trilioVault should first discover the vCenter configuration. You can start the discover process by clicking “Discover vCenter” in the dashboard under the General task area.

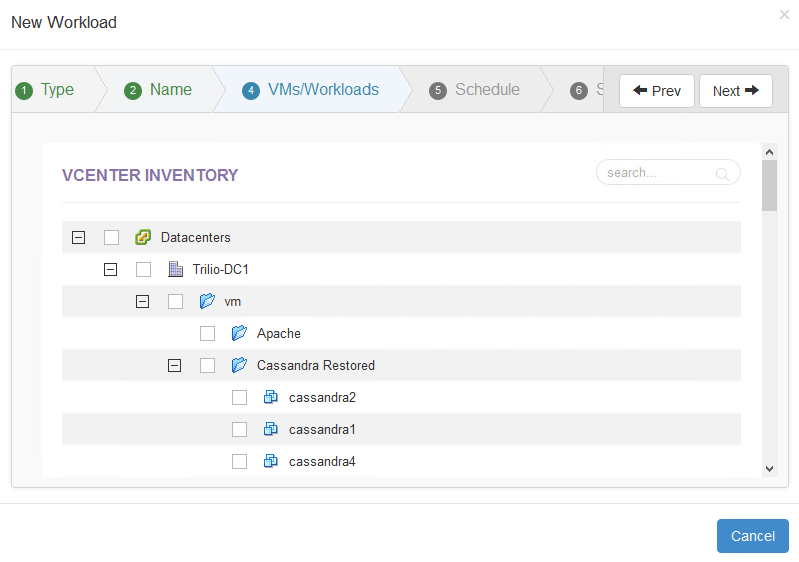


The Workload creation wizard starts with a list of workload icon types available for this deployment. These icons easily identify each workload type. Among these are application specific icons for Apache™ Hadoop®, Cassandra and MongoDB. Once a user chooses a particular workload type, rest of the wizard steps are determined based on the workload type.

## Create Serial/Parallel Workload

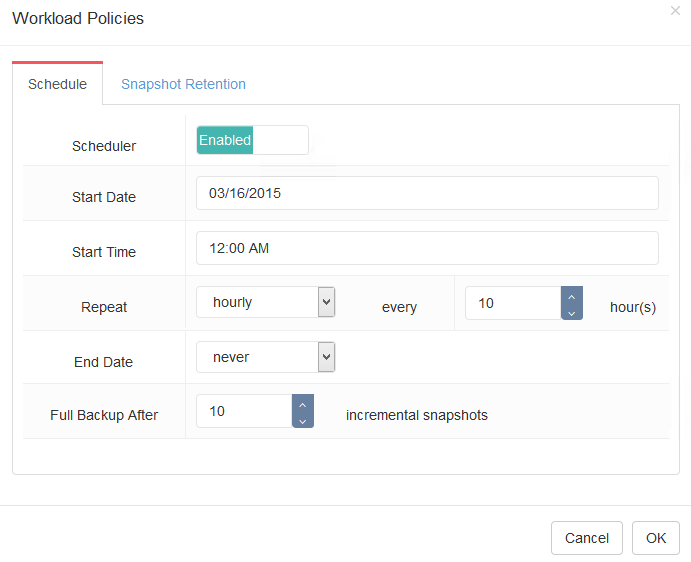


Provide the name and description of the workload and click next to select the virtual machines from VM inventory.



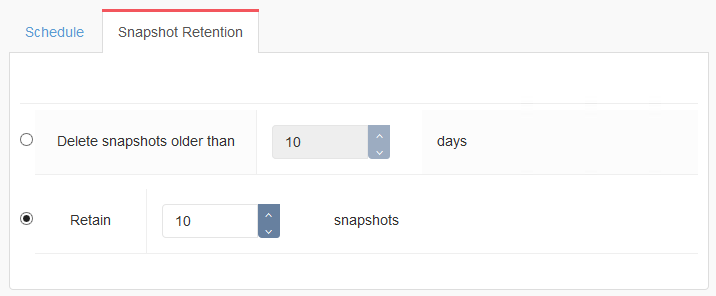
In this example, select the Postgresql™ virtual machine from the tree-view. In this tree-view, you can also search for the virtual machine you may be looking for. A user can select one or more virtual machines as required for a given workload.

Then the next step is to define a protection policy for the workload in the form of a snapshot scheduler 🡪 details and the retention policy as shown below.



The options for the scheduler are as follows:

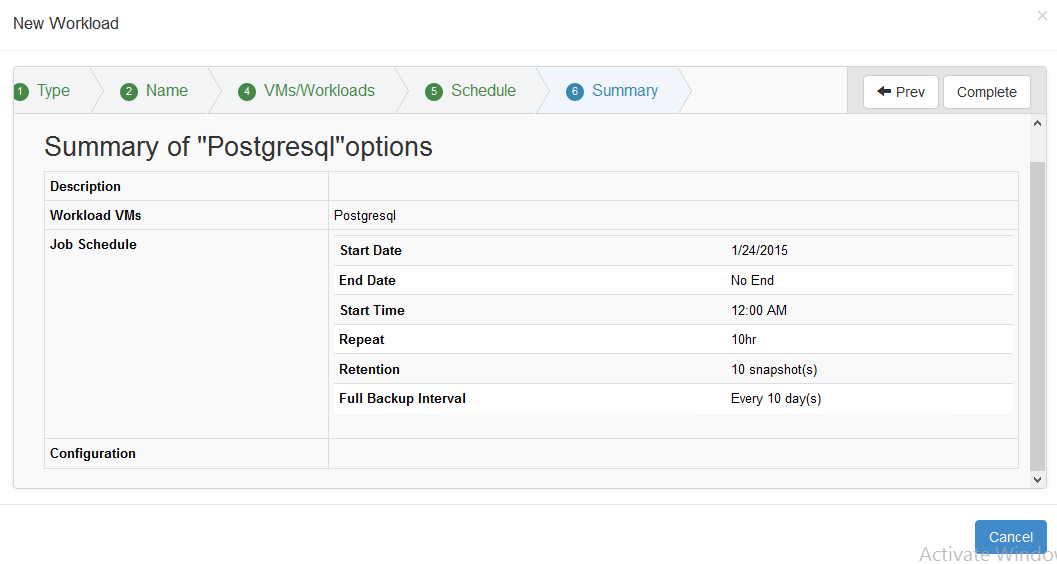
* Start date and time
* Frequency of the snapshots
* End date if applicable
* After how many incremental snapshot should the system take a full snapshot



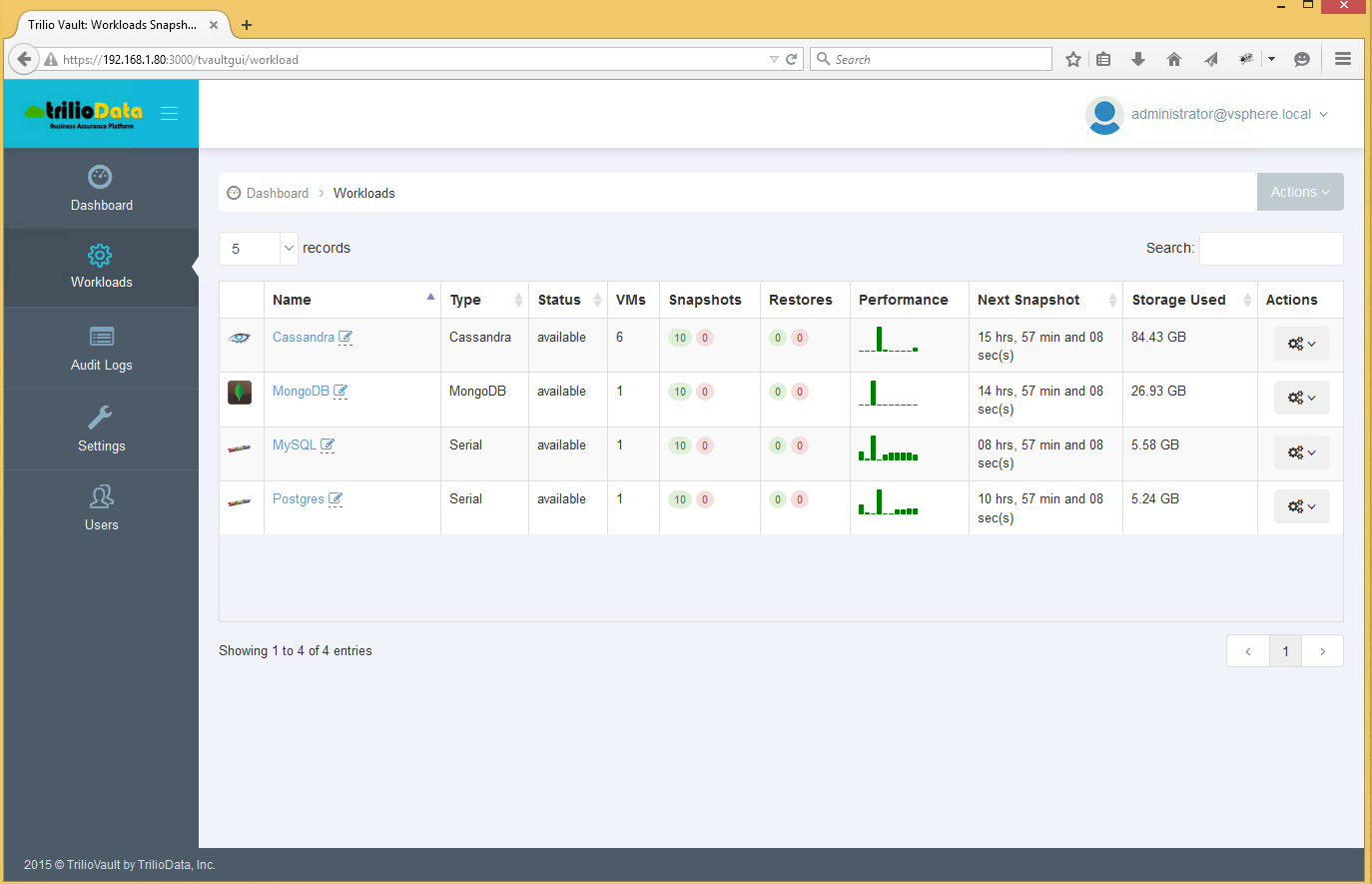
In order to select the snapshot retention, the options are as follows:

* How many snapshots are to be retained.
* Options to delete older snapshots beyond certain days.

Once the protection policy is defined, one can view the summary for the workload before the completion step.



The following screenshot was taken after the trilioVault had taken several snapshots of the Postgresql workload.



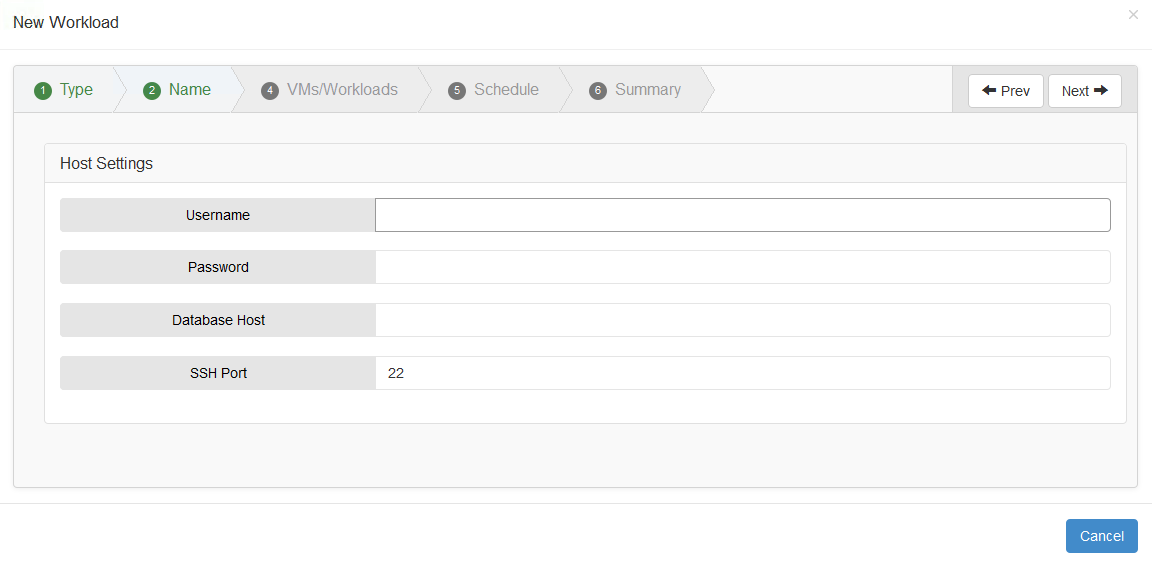
The above view shows that PostgreSQL workload consists of 1 virtual machine, has 10 snapshots.

## Cassandra workload

Before creating the Cassandra workload, make sure that the Cassandra Cluster nodes and the database is running and check the pre-requisite section in the Appendix section.

The steps for creating the Cassandra workload are as follows:

* Choose the Cassandra workload wizard.
* Provide a name and description for the Cassandra workload.
* Provide the IP address to one of the Cassandra nodes along with the credentials (username and password) and the trilioVault system will discover all the nodes in the Cassandra cluster.



* Choose a snapshot schedule and retention policy.
* View the summary before workload creation.

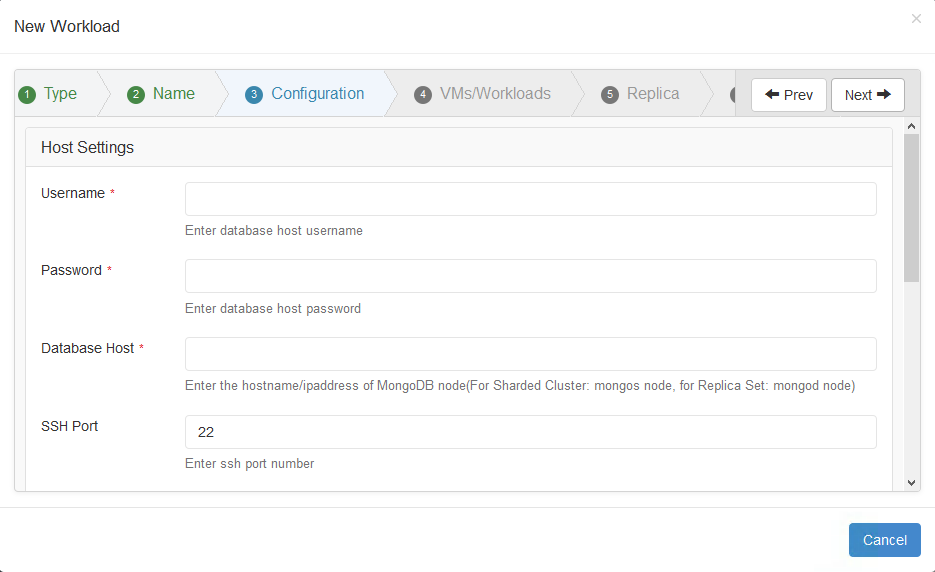
## MongoDB Workload

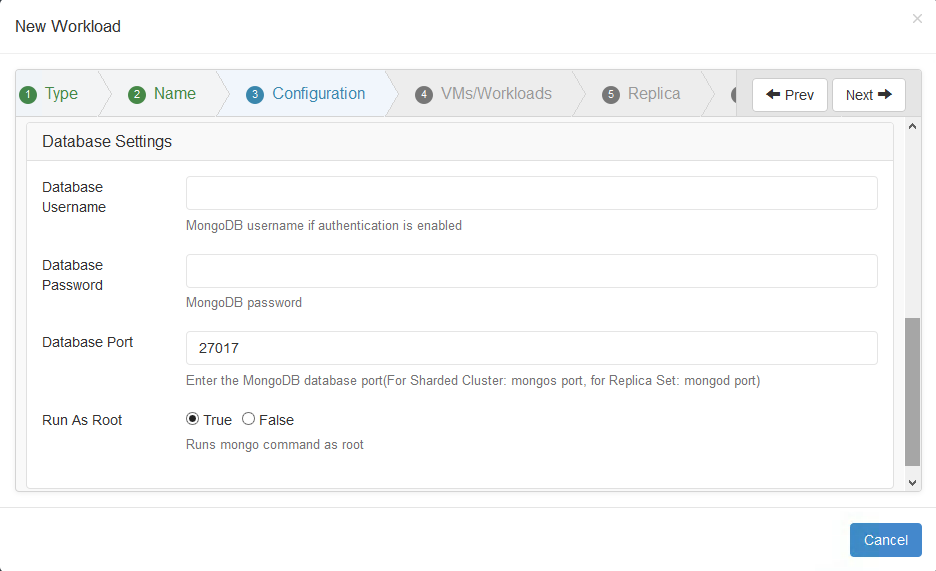
Before creating the MongoDB workload, make sure that the MongoDB Cluster nodes and the database is running.

The steps for creating the MongoDB workload are as follows:

* Choose the MongoDB workload wizard
* Provide a name and description for the MongoDB workload
* Provide the IP address to one of the MongoDB nodes (Mongos node for a sharded cluster otherwise one of the mongod nodes for a Replicaset) along with the credentials (username and password) and the trilioVault system will discover all the nodes in the MongoDB cluster.

|  |  |  |
| --- | --- | --- |
|  | Attribute | Description |
| Host Settings | Database Host | Provide the mongos host, if it is sharded cluster otherwise provide the host for one of the mongod nodes in a Replicaset trilioVault should be able to lookup the host name. Make sure your DNS settings are entered correctly during configuration. |
| Username | Username of the MongoDB host |
| Password | Username of the MongoDB host |
| SSH Port | Usually set to 22. Make sure all MongoDB hosts are enabled for password-based authentication. Future releases of trilioVault will be integrated with password management systems. |
| Database Settings | Database Username | Provide DB user if database authentication is enabled. Otherwise leave this field empty. |
| Database Password | Provide DB password, if database authentication is enabled. |
| Database Port | Provide the mongos port, if it is sharded cluster otherwise provide the port for one of the mongod nodes in a Replicaset |



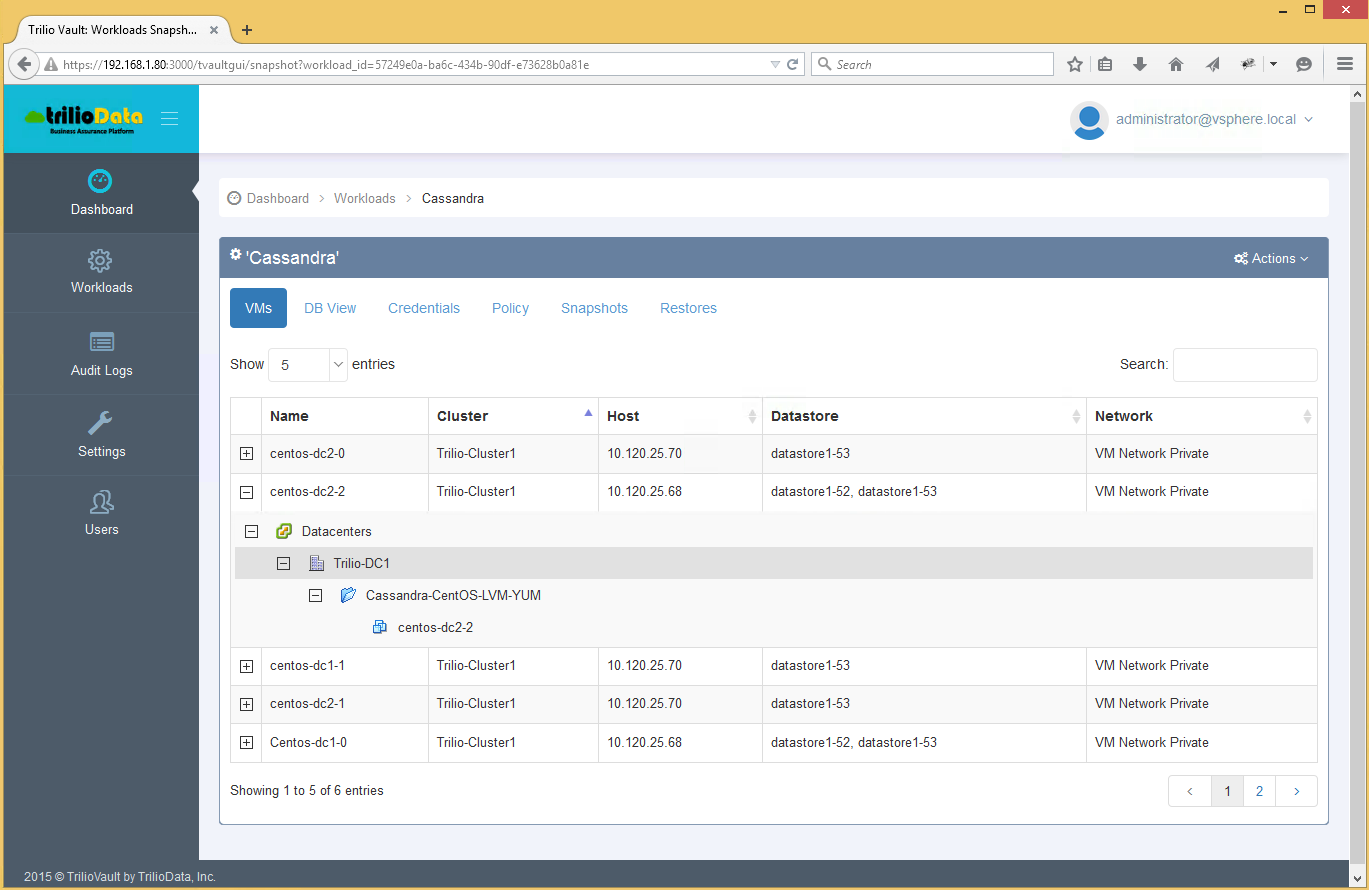


* Choose a snapshot schedule and retention policy.
* View the summary before workload creation.

# Workload Topology

## VM View

To view the topology for a given workload, click on the workload hyperlink (for e.g. Cassandra) in the workload panel, the software will take you to the topology view for the given workload as shown below. In this “VM View”, you can see the details of all the virtual machines instances in the workload.



## DB View

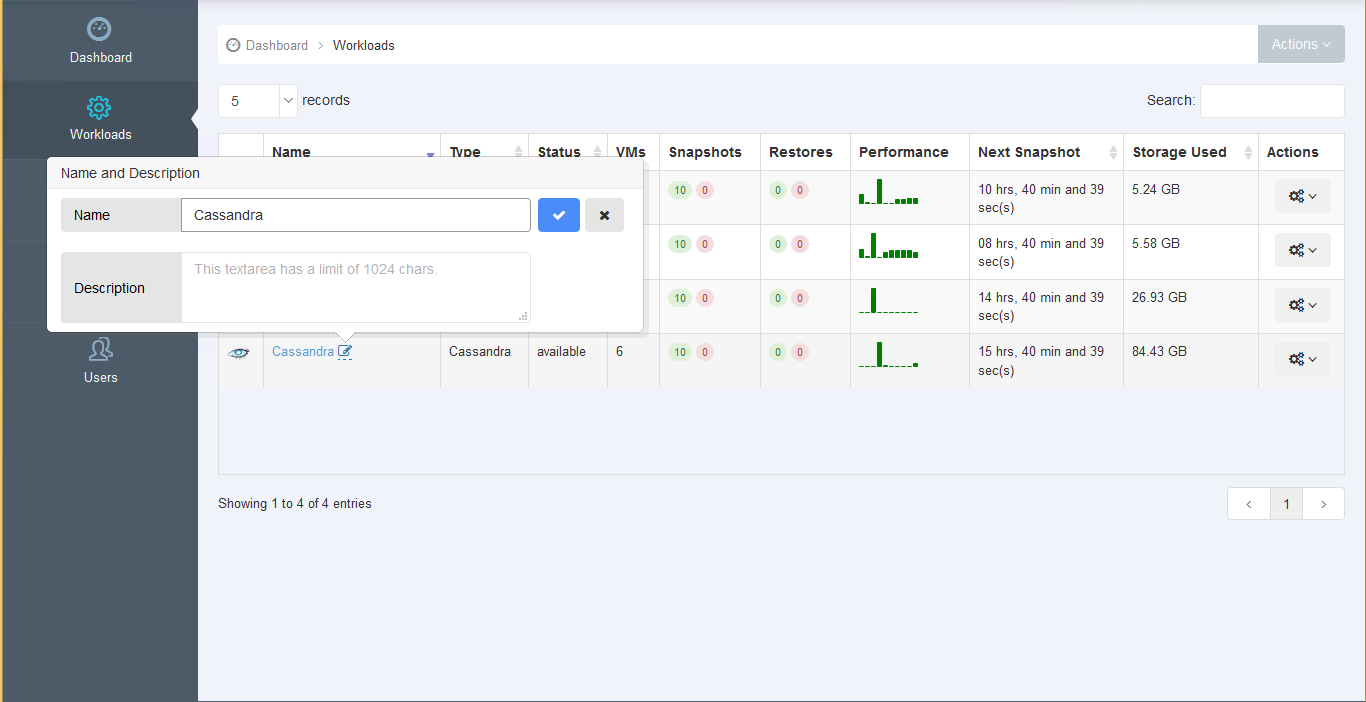
For Cassandra and MongoDB, one can also view the Database topology as known to database administrators of these databases, by clicking on the “DB View”. For Cassandra, trilioVault will give you the Datacenter view for the Cassandra workload as shown below.



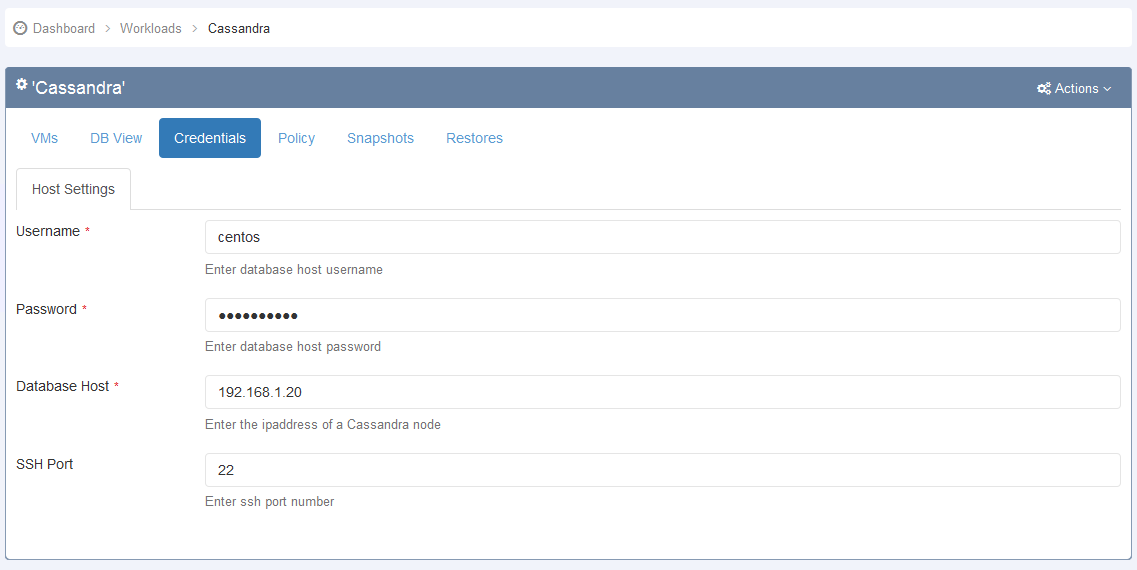
One can also view the “Keyspaces” along with replication strategy options. This can enable one to select the Cassandra Data Center(s) that need to be backed up. By default, one will backup all the Data Centers, but can pick and choose the ones that needs to be backed up by dragging the datacenter name in question into the box as shown above.

# Modify Workload Name and Credentials

By clicking on “Edit” icon as shown below in the workload panel, the software will enable one to modify the workload name and description.

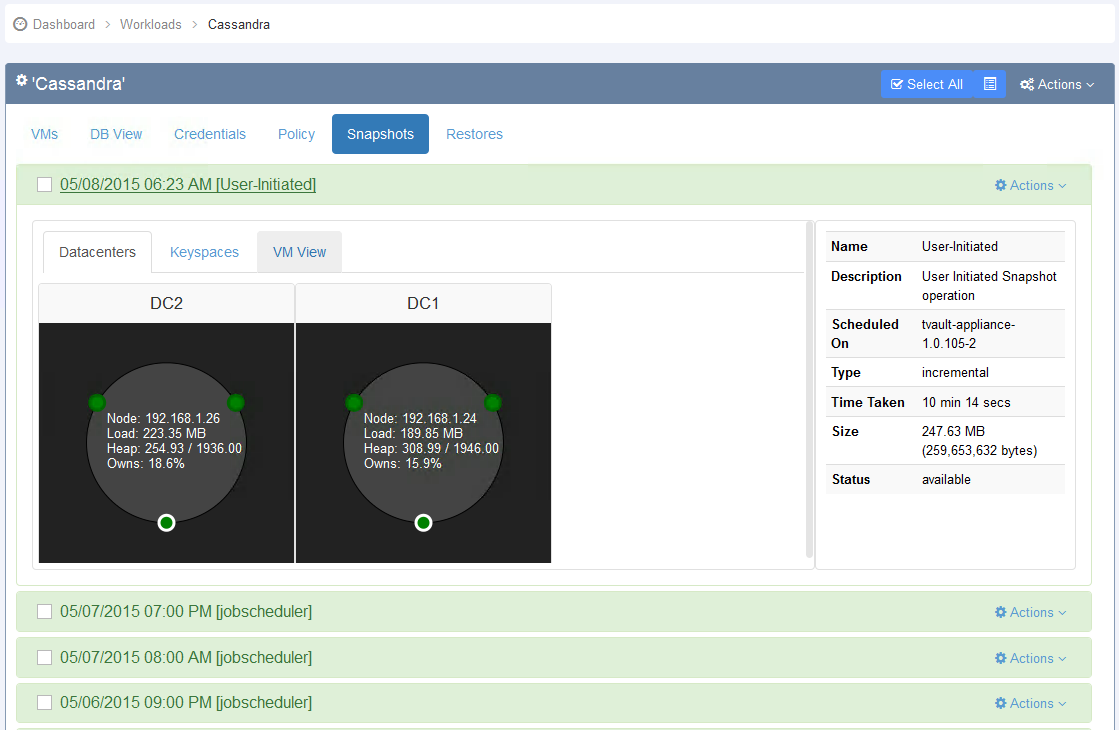


To modify the access credentials, click on the Credentials tab for the given workload as shown below.



# Workload Snapshot

## Snapshot View



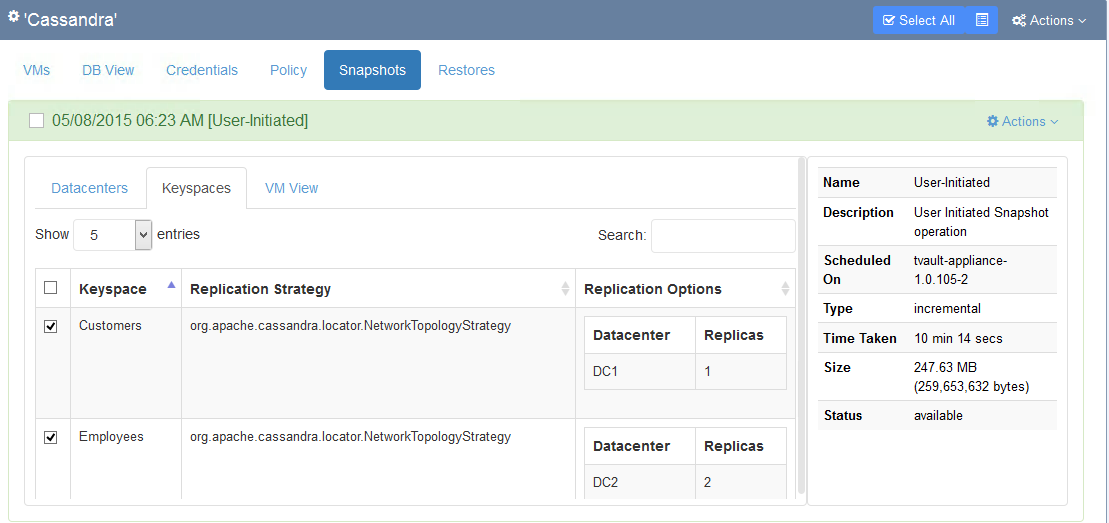
1. By clicking on “Snapshots” tab for a given workload, the software will take you to the snapshot view of the workload. The snapshot view lists all the snapshots that are tagged by the timestamp at which they were taken. Each snapshot is tagged as either “User-Initiated” or “jobscheduler”.

* User-initiated snapshots, as name implies is executed by user on-demand basis.
* “Jobscheduler” snapshot operations are initiated by job scheduler.

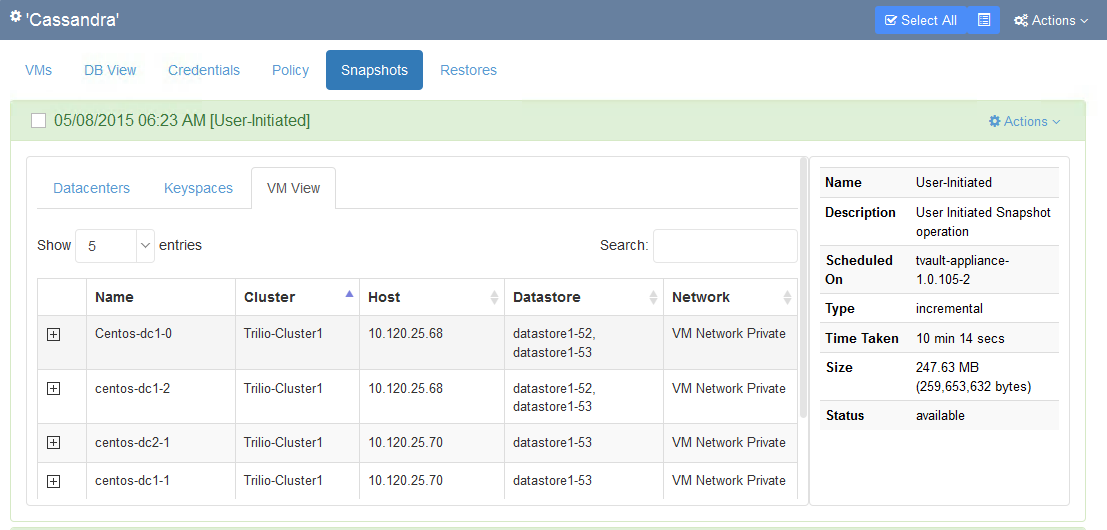
1. Clicking a snapshot hyperlink expands the view to display the details of the snapshot. The left side contains the topology of the snapshot at the time of snapshot operation. The right hand side includes additional details such as size of the snapshot. Though the snapshot does not include whether the snapshot is full or incremental, such data can be inferred by the position of the snapshot. The bottom most is the full snapshot is usually the largest in size.

NOTE: In some cases, an incremental snapshot size can be larger than full snapshot. This happens when the under storage is thin provisioned and there was tremendous IO activity between two snapshot operations.

1. One can also visualize the Database view (DM View) for Cassandra workload as shown above and verify the Keyspaces that are part of the snapshot as shown below.

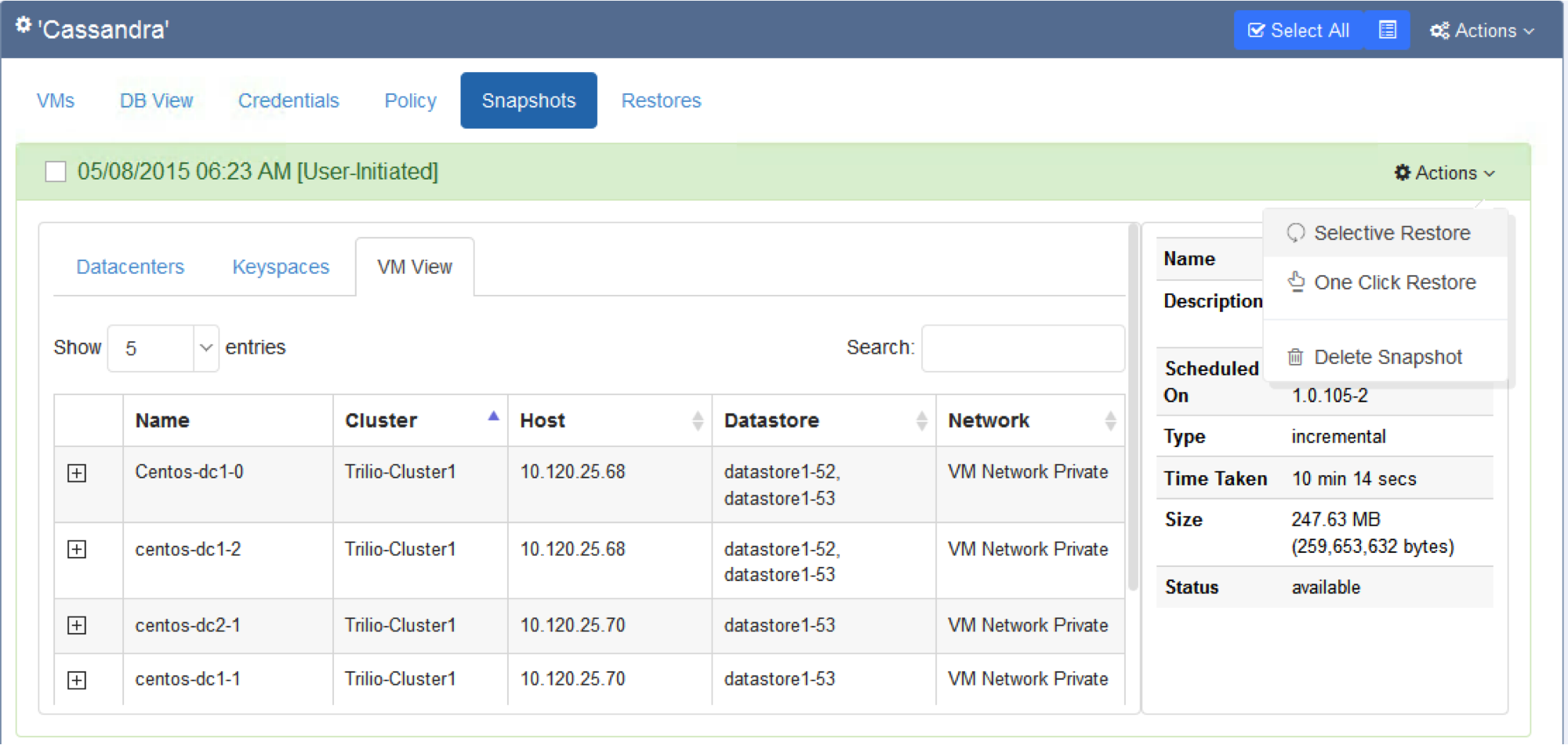


One can also view the “VM view” for the given snapshot as shown below.



1. Three operations can be performed on a snapshot by clicking on the Actions pull down menu.

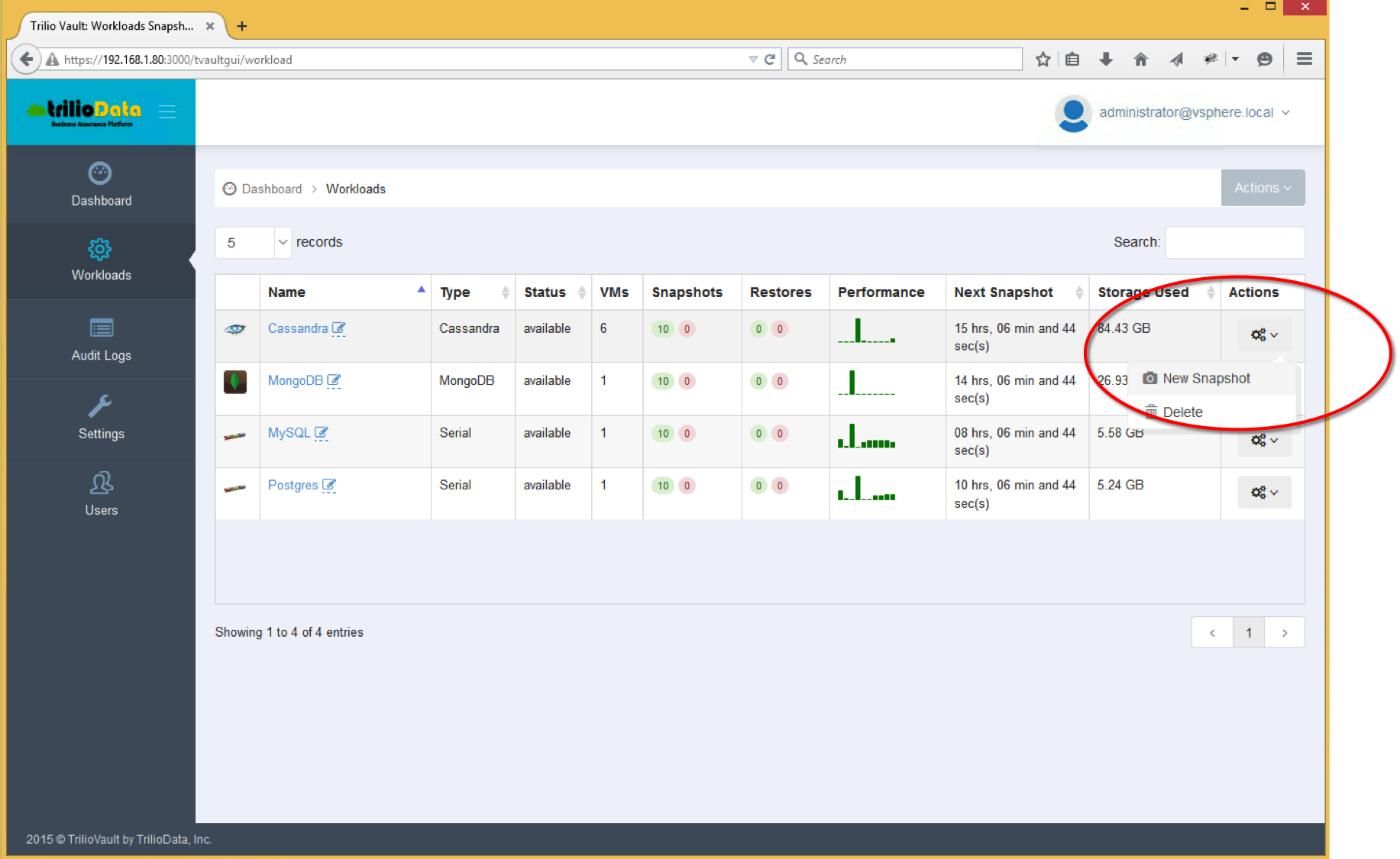
* With the delete operation, you can delete the snapshot.
* Selective Restore operation launches the restore wizard for that snapshot, enabling one select the nodes that need to be restored.
* One-Click Restore operation, enables one to restore the workload to a production environment with single click.



## Create Workload Snapshot

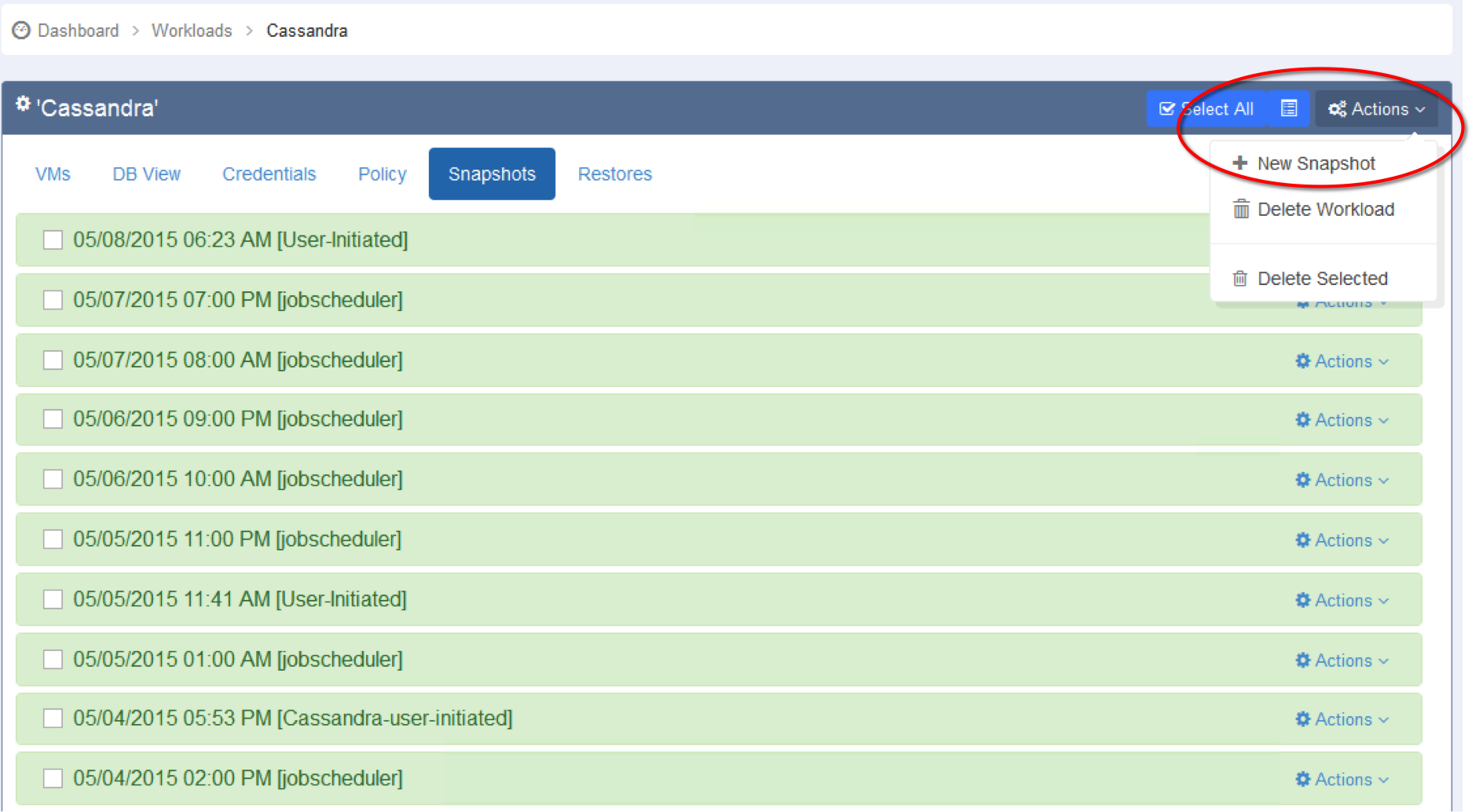
trilioVault will create the snapshots automatically based on the protection policy. You can also create an on-demand snapshot of the workload. To do so, follow the steps as articulated below:

* Select the Workload from the Main Menu.
* Select the “New snapshot” from the Action pull down menu for the given workload as shown below.

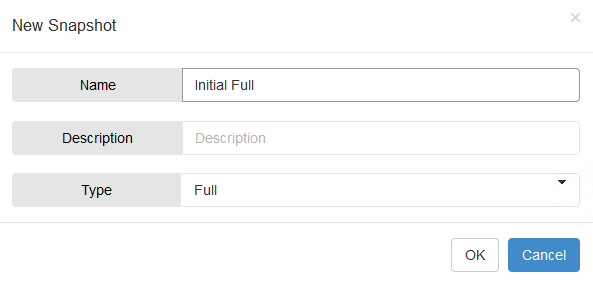


In the screenshot above it shows that there are 10 good snapshots and there were no error in creating any of them.

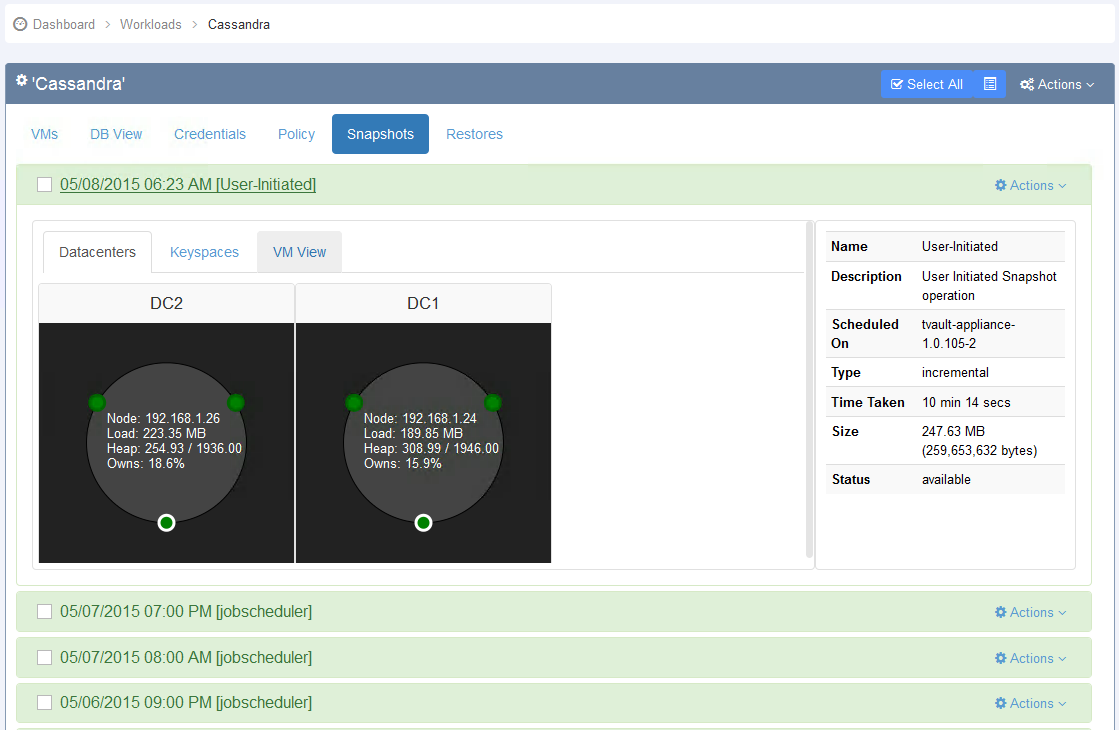
* Alternatively, one can also create a snapshot by first clicking on the Snapshots tab for a given workload and then selecting “New Snapshot” from the Actions pull down menu as shown below.



* Give a name and select whether a full or incremental snapshot needs to be created. On selecting “OK”, the system will create the snapshot as shown below.



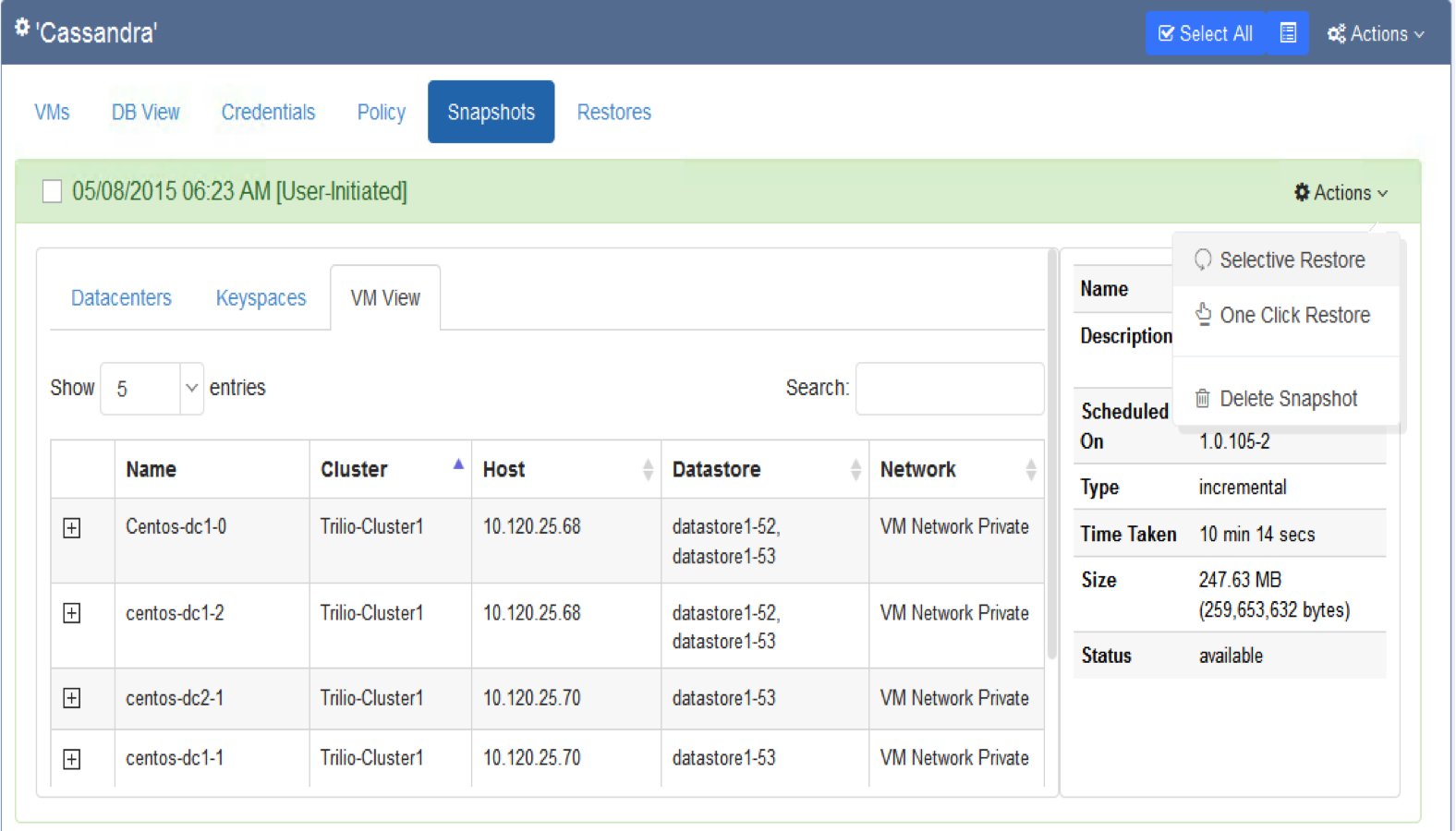
Once the snapshot is created, it will show the status as follows:



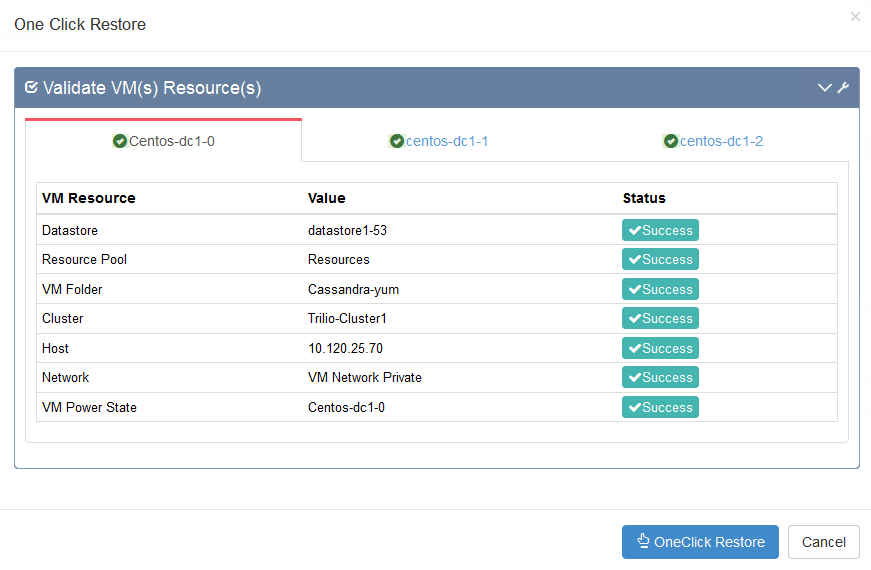
In this above screenshot, the system informs the user that the snapshot was user initiated, it took ‘x’ amount of time to create a full snapshot and the status is marked as available for restore if required.

# Workload Restore

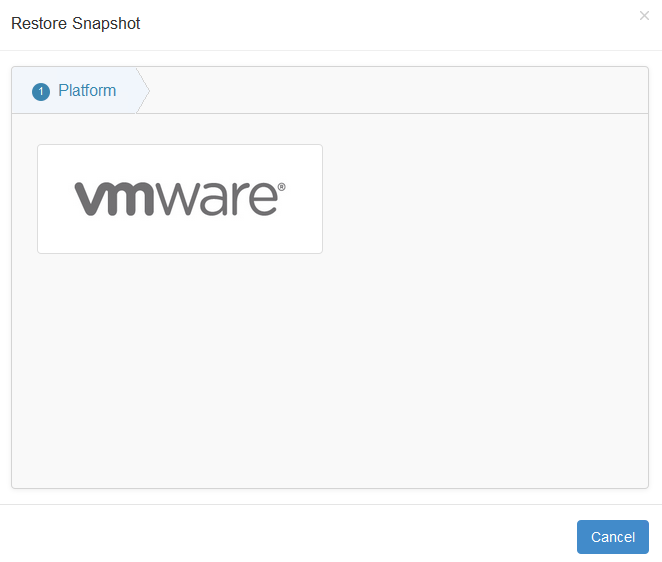
In order to restore a workload, one can select a snapshot and select the One-Click Restore or the Selective option as shown below.



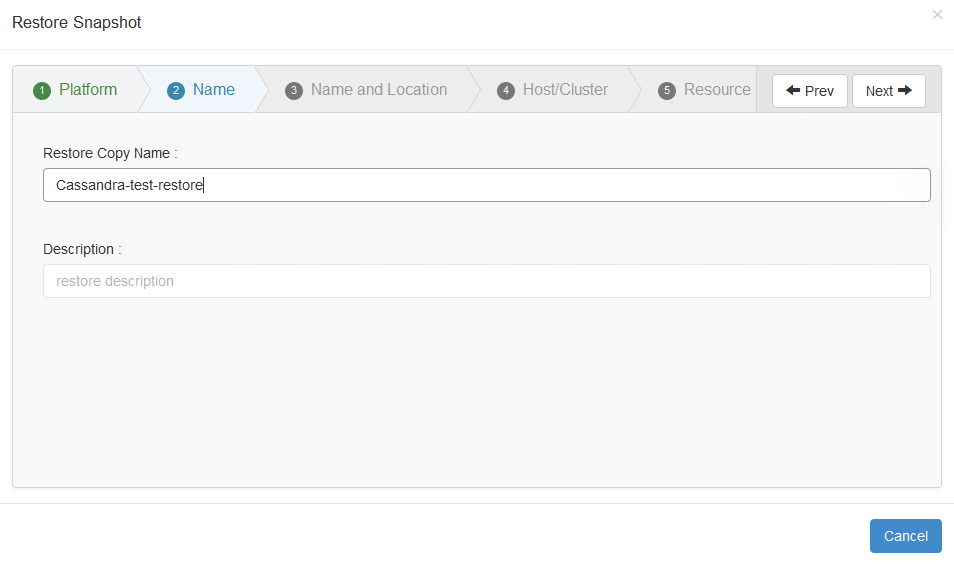
If one chooses the single click restore then the s/w will check if environment is ready to perform the restore. Make sure the virtual machines are either turned down or deleted before performing the restore.



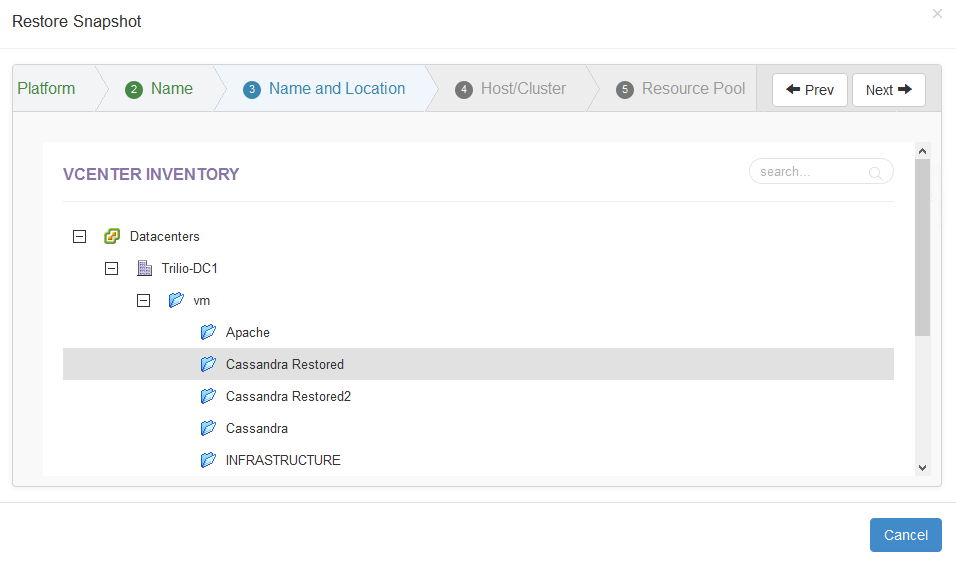
If on the other hand, one chooses the Selective Restore operation then the Restore Snapshot Wizard is invoked. Since trilioVault captures entire application stack, it can restore a workload to a completely different environment than the environment that of the original workload snapshot. As mentioned earlier, the workload configuration includes infrastructure and application configurations. This wizard lets you modify the workload snapshot configuration parameters so the workload snapshot can be restored to a different environment.



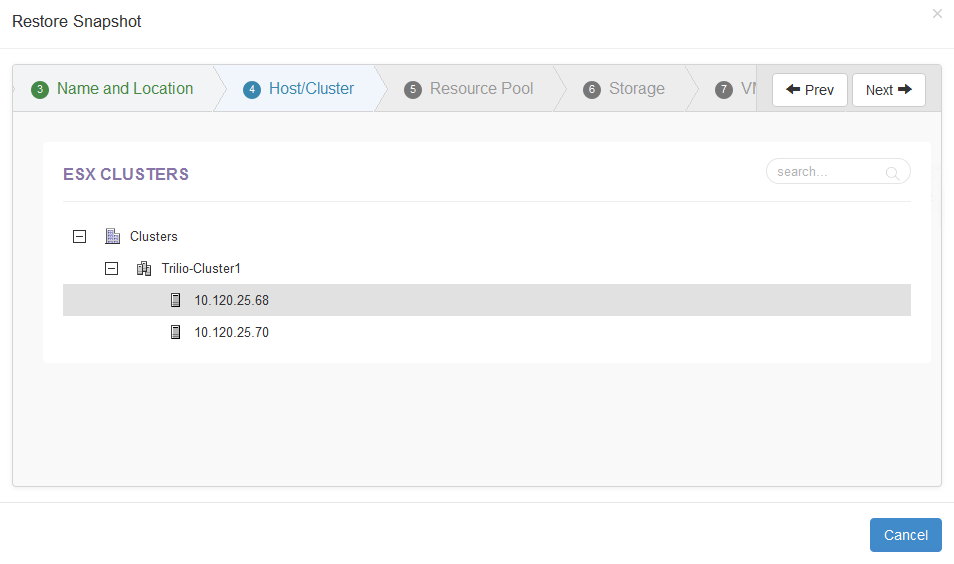
* First step of the restore snapshot wizard is choosing the target platform as shown above. Currently, we support VMware and OpenStack. In the future, Trilio Data will support additional cloud platforms to include popular public clouds.
* Suppose we chose VMware as the target platform for the Cassandra restore as an example. The wizard will then ask the user to provide a name and description for the restore copy as shown below.



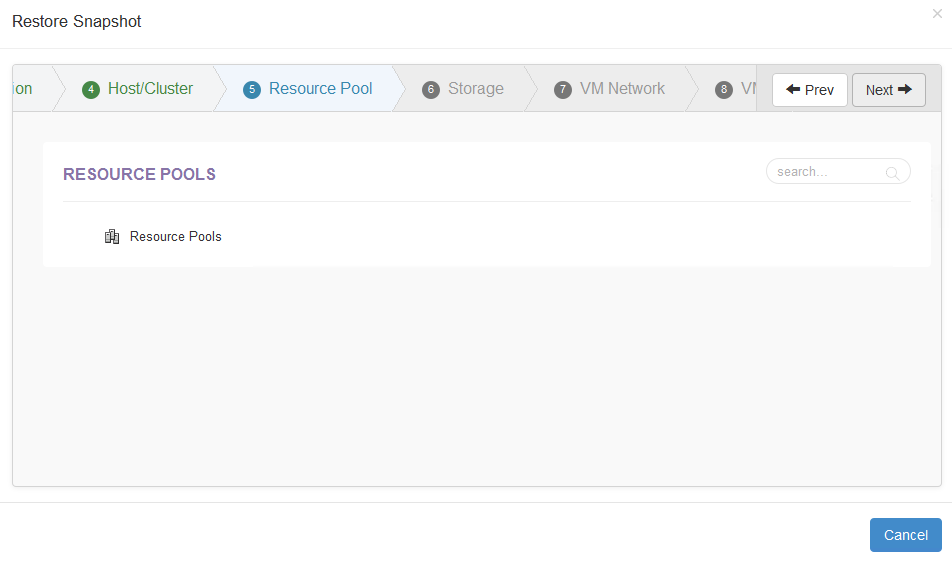
* Similar to VMware vSphere’s create virtual machine wizard, the next three steps will identify Name and Location, Host/Cluster and Resource group of the workload. However in order to provide even more flexibility with the recovery, at the end of the wizard, user is given a change to change any particular VM placement.
* Select the location folder as shown below.



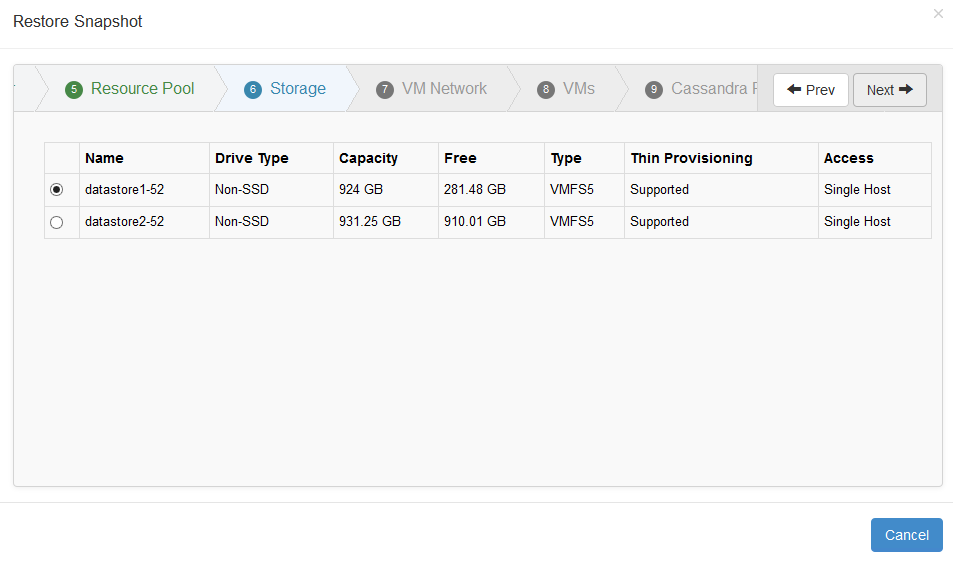
* Then select the ESX host in the example as shown below.



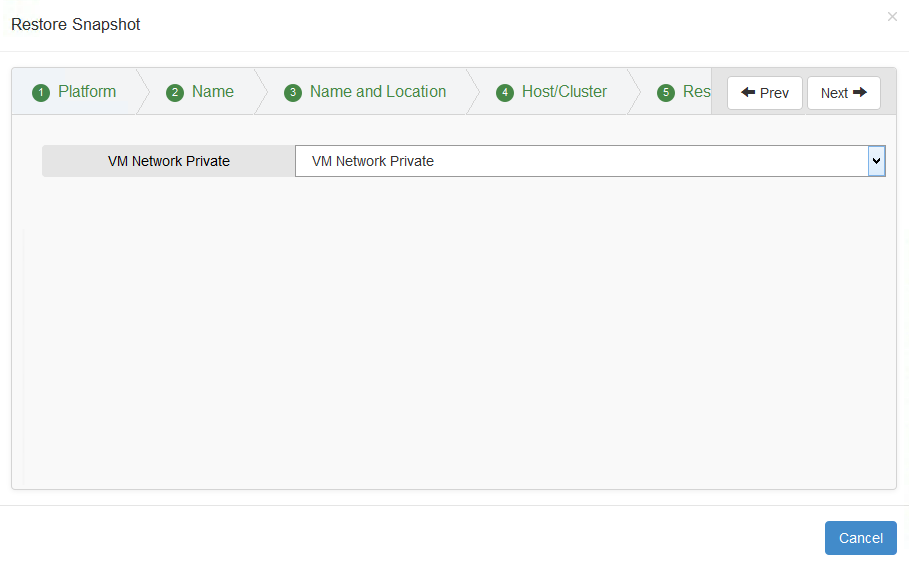
* Identify the resource group as shown below. This selection is optional.



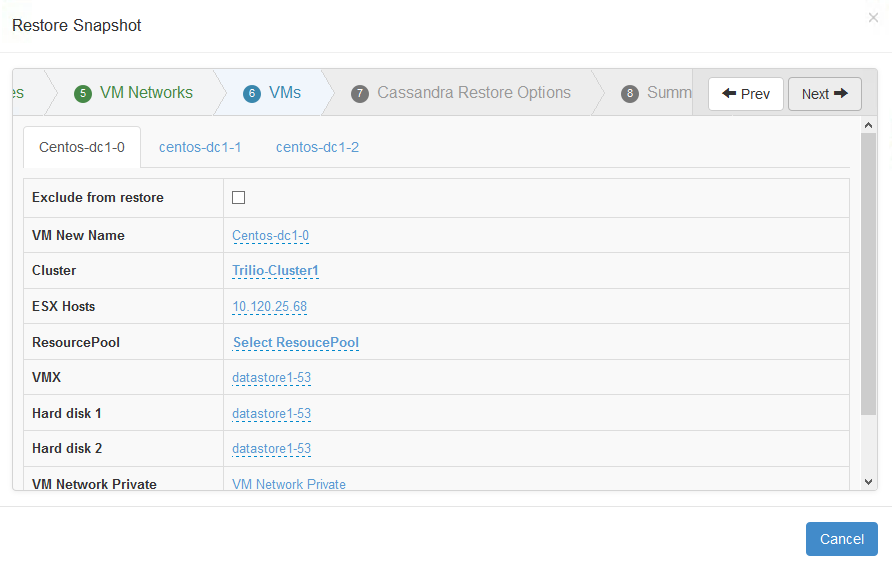
* Select the Datastore for the restore copy of the workload.



* Select a VM network for virtual machines for recovered workload.

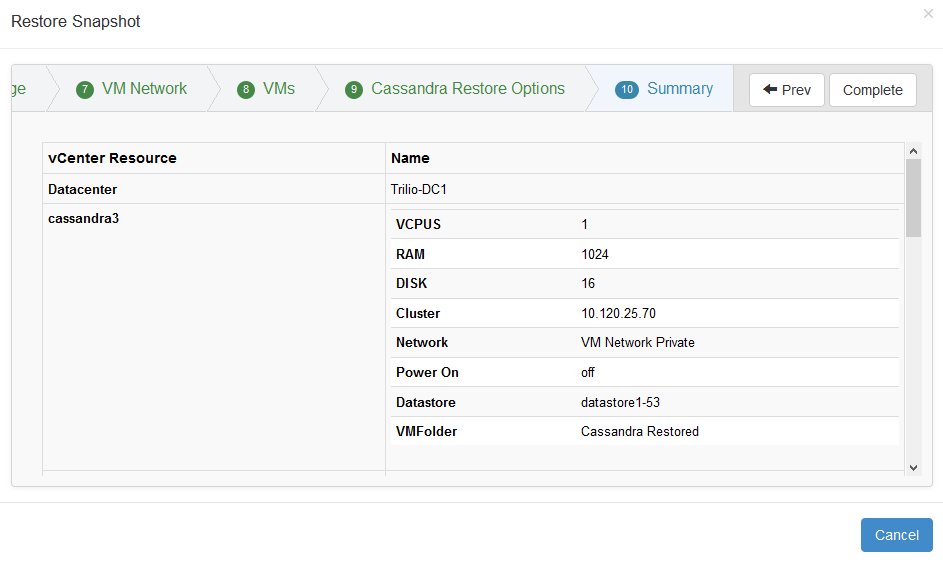


* The restore wizard until now identified configuration settings for entire workload. However, a user may choose to deploy VMs on to different clusters/resource groups or datastores. This page lets you change individual VM settings. A user can choose a different datastore, host/cluster or resource group. Also, by default all VMs are created powered down. However, the user can choose to power on the VM at the end of the recovery by choosing the “Power-on” check box. Lastly, a user can also specify the order in which these VMs need to be powered on.



* If the operating system for the workload instances were originally created on a non-LVM environment then trilioVault will allow one to change the IP address of the virtual machine instance during the restore process.

* This page provides the summary of restore options that are chosen for this recovery. After completing the wizard, trilioVault opens the restore view where all recoveries that are performed for this snapshot are displayed with the currently executing recovery process at the top. Remember that a snapshot can be restored any number of times as long as the restore option will not interfere with other instances of the workload VMs.



# Known Issues

1. If an expected dialog box does not open, reload the page and try again.
2. If the delete workload operation does not work, click on the snapshots and get back to workloads view and then delete the workload.

# Appendix

## Pre-requisites for using trilioVault for Cassandra workload

trilioVault software leverages the Cassandra “nodetool” utility for discovering the Cassandra cluster. It is important that the software can access this utility for a given user. And depending on how Cassandra is installed, the “nodetool” may or may not be accessible. The best practice is to run the command “which nodetool” for a given node. If this command cannot find the path to the nodetool then one has to create a file called trilio.sh file in the /etc/profile.d directory as an example for centos or Ubuntu linux distributions. To this file one would add the directory of the “nodetool” utility to the PATH variable as shown below:

PATH=$PATH:<nodetool-directory>

export PATH

## Pre-requisites for using trilioVault for PostgreSQL

In order to take a consistent snapshot of PostgreSQL database without stopping the database, please follow these steps:

* Make sure that the postgres service is running.
* Create <archives> directory under the same directory where the PostgreSQL data files are located. You can create this in the Linux shell by executing “mkdir <location of the postgresql data directory>/<archives>
* WAL archiving mode needs to be enabled in the postgresql.conf file. Please enter the following in the postgresql.conf file.
  + wal\_level = archive
  + archive\_mode = on
  + archive\_command = ‘cp %p <location of postgresql data files>/<archives>/%f’
* Make sure that either VMware vmtools or a 3rd party service such as **open-vm-tools package is installed and running.**
  + **To Install Open VM Tools**

Run the commands:

* $ sudo apt-get update
* $ sudo apt-get install open-vm-tools
* **Create /usr/sbin/pre-freeze-script file with the following contents using vi or emacs editor**

#!/bin/bash

sudo –u postgres /usr/bin/psql –c “SELECT pg\_start\_backup(‘triliodata’, true);”

#note – In the above command you can use the appropriate database user that can execute the above SQL command

* **Create /usr/sbin/post-thaw-script file with the following contents using vi or emacs editor**

#!/bin/bash

sudo –u postgres /usr/bin/psql –c “SELECT pg\_stop\_backup();”

sudo rm –rf <location of the postgresql data directory>/<archives>/\*

#note – In the above command you can use the appropriate database user that can execute the above SQL command

* Execute the following commands within the Linux command shell:
  + chmod 755 /usr/sbin/pre-freeze-script
  + chmod 755 /usr/sbin/post-thaw-script

## Pre-requisites for using TrilioVault for MySQL

In order to take a consistent snapshot of MySQL database, please follow these steps:

* Make sure that the MySQL database service is running.
* Make sure that either VMware vmtools or a 3rd party service such as **open-vm-tools package is installed and running.**
  + **To Install Open VM Tools**

Run the commands:

* $ sudo apt-get update
* $ sudo apt-get install open-vm-tools

### MySQL Backup Option I

To quiesce MySQL, we will do a mysqldump to create a transaction consistent backup of MySQL without downtime.

Note the disadvantage of this method is that it takes more space since it creates a copy of the database. Alternatively, if you can stop/start mysql during an off hour backup that is a simple way to quiesce ( no transactions can occur if the database is stopped ) as well. You would stop MySQL in the pre-freeze-script, and start MySQL in the post-thaw-script in that case

* **Create /usr/sbin/pre-freeze-script file with the following contents using vi or emacs editor**

#!/bin/bash

sudo mysqldump --single-transaction --all-databases > /var/mysqlbackup

* **Create /usr/sbin/post-thaw-script file with the following contents using vi or emacs editor**

#!/bin/bash

sudo rm –rf /var/mysqlbackup

* Execute the following commands within the linux command shell:
  + chmod 755 /usr/sbin/pre-freeze-script
  + chmod 755 /usr/sbin/post-thaw-script

### MySQL Backup Option II

In this option, the MySQL service is stopped for a few seconds, while the snapshot is taken. The pros for this method is it does not require additional space as required in the previous option. However for a few seconds, the database is unavailable. This method is suitable if there is a window when the database can be stopped for a few seconds. The MySQL service is stopped for a few seconds while the snapshot is created then  
started again and is described in Method 1.

* **Create /usr/sbin/pre-freeze-script file with the following contents using vi or emacs editor**

#!/bin/bash

sudo service mysql stop

* **Create /usr/sbin/post-thaw-script file with the following contents using vi or emacs editor**

#!/bin/bash

sudo service mysql start

* Execute the following commands within the linux command shell:
  + chmod 755 /usr/sbin/pre-freeze-script
  + chmod 755 /usr/sbin/post-thaw-script