Simplifying Data Management

With DataStax Enterprise
(DSE) OpsCenter

DATASTAX

CONTENTS

ABSTRACT

Managing and optimizing the performance of a widely distributed data management platform that is comprised of many individual servers can be very challenging. This white paper discusses how DataStax Enterprise (DSE) OpsCenter can be used to greatly simplify the task of administering one or many DSE clusters, whether those clusters are deployed on-premises or in the cloud.

INTRODUCTION

Today's online applications require a future-proof data management platform that provides continuous availability, predictable and reliable scalability, high performance, security, and support to handle all types of data. The current era of cloud applications has produced a new set of data management requirements that are beyond the capabilities of legacy relational database management systems (RDBMS).

Today, DSE is deployed to handle modern cloud applications across industry verticals such as retail, financial services, healthcare, media and entertainment, and many others. Many companies are choosing DSE, an always-on, secure, and operationally simple data management platform built on the best distribution of Apache Cassandra™, to power the exploding number of cloud applications that require data distribution across data centers and clouds. A masterless architecture, with an ability to easily distribute data across many data centers and cloud availability zones makes DSE ideal for production installations.

Companies that deploy DSE need an efficient way of managing one or many distributed DSE clusters that sometimes consist of hundreds of nodes and oftentimes span multiple geographies. To handle the administration and monitoring of the DSE infrastructure, DataStax provides DSE OpsCenter.

DSE OPSCENTER

DSE OpsCenter is a web-based, visual management and monitoring solution for DSE. OpsCenter has integrated functionality for real-time and historical monitoring, tuning, provisioning, backup, and security management. OpsCenter provides architects, database administrators (DBA), and operations staff with the tools necessary to intelligently and proactively to manage and monitor their cloud applications and simplify the administrative tasks so they can concentrate on business-critical tasks rather than babysitting their data management systems.

How Does DSE OpsCenter Work?

DSE OpsCenter can be installed on any server – on-premises or in the cloud – that has connectivity to clusters running DSE. Each node in a DSE cluster contains a DataStax agent, which communicates with the central OpsCenter service. The DataStax agent and OpsCenter service work together to monitor and handle tasks on every managed cluster.

OpsCenter provides a web-based console for central management of the data management system along with a visual point-and-click environment which makes administration, monitoring and maintenance easier. Currently, a single installation of OpsCenter is able to efficiently manage and monitor up to a 1,000-node cluster.
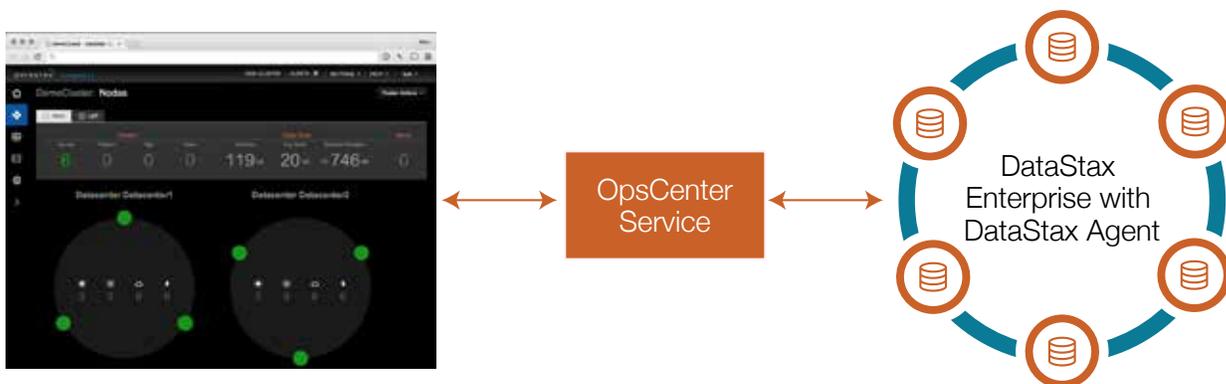

Figure 1 – Overview of DSE OpsCenter Architecture.

The OpsCenter Interface

DSE OpsCenter comes with an enterprise dashboard that makes managing many clusters at once easy. OpsCenter also provides individual cluster dashboards that provide at-a-glance information, which makes it simple to understand the performance and uptime of any running cluster.

The enterprise dashboard displays a summary of all monitored clusters/nodes along with uptime status, includes an aggregated listing of all alerts that have been triggered across all clusters/nodes, and shows a summary of each cluster's performance and status.
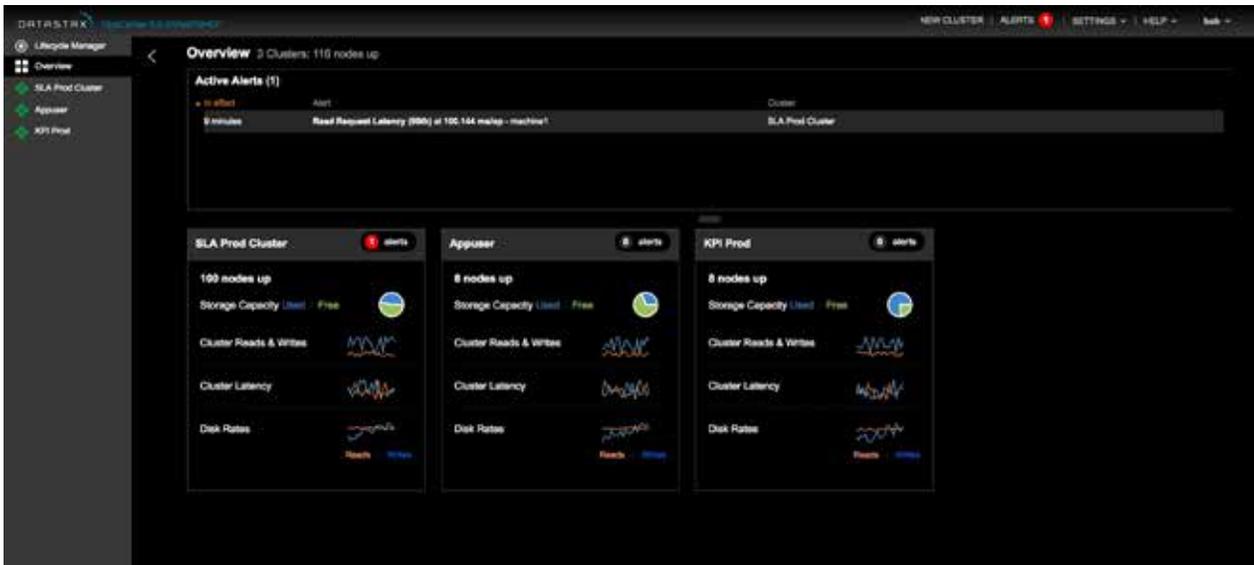
Figure 2 – The Enterprise Dashboard of OpsCenter.

From the enterprise dashboard, you can drill down into individual clusters, which have overview dashboards of their own. Critical performance information, up/down status, alerts, and the overall health of a cluster can be quickly ascertained from these overview dashboards.

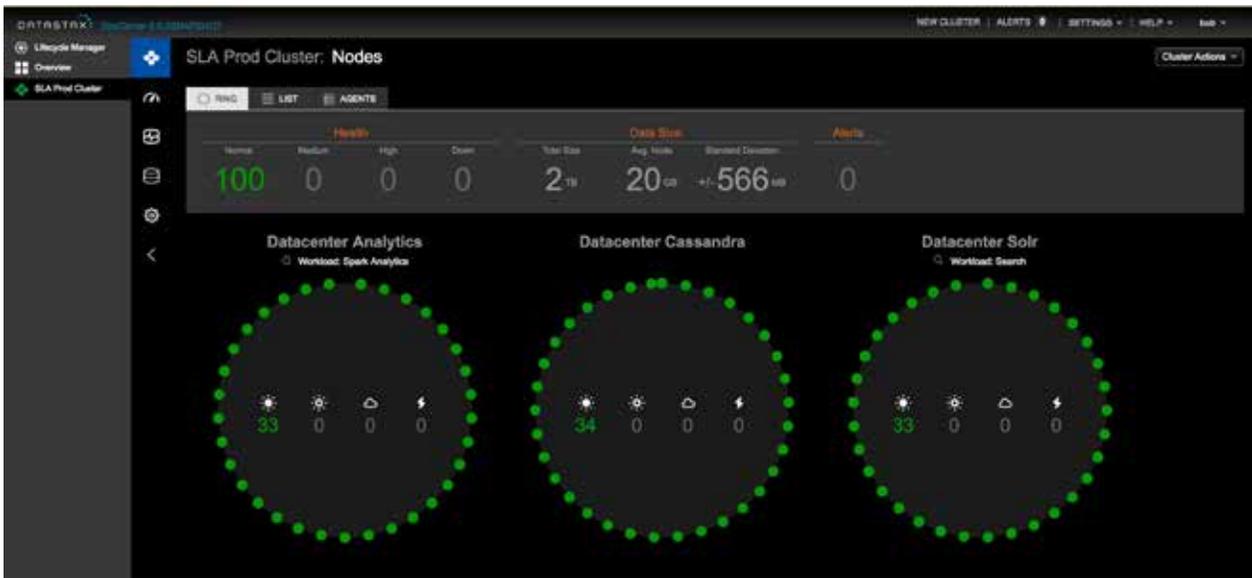*Figure 3 – Individual cluster core dashboard.*

Figure 4 – Ring view of a cluster's status and operations.

All administration and monitoring functions can be easily navigated via either the tabs across the top of the interface or the left-hand side navigation pane.

OpsCenter Security

OpsCenter has built-in security that controls the clusters a user can manage and the tasks they can perform. OpsCenter defaults to having no security enabled, but activating certain security parameters can easily change that.

OpsCenter's security uses a familiar user/role-based paradigm where roles and users are created by a super user account, with various permissions being assigned to roles that are then assigned to users. OpsCenter supports LDAP/ Active Directory (AD), which allows you to have a centralized repository for user and group management and leverages this for authentication and authorization purposes. Once established, a user can sign in and manage the clusters that have been assigned to them and only run the operations they have been granted to perform.

In addition, OpsCenter also supports Off Server Key Management using KMIP (Key Management Interoperability Protocol), which protects encryption keys and only transmits keys to DataStax servers that have been explicitly authorized by the Key Manager. This increases the protection for data at rest. OpsCenter also includes session based login with expire sessions and logout options.

CREATING NEW CLUSTERS

DSE OpsCenter's visual Lifecycle Manager Deploy feature allows administrators to create new DSE clusters either on premise or in the cloud. A wizard guides you through the process of specifying how many nodes will make up a cluster, what the nodes will consist of (e.g. just database, a mixture of database, analytics, graph and search nodes, etc.), and the cluster's configuration.

Once you are satisfied with what the new cluster will look like, OpsCenter will download the specified software onto the target machines, install the software, configure all agents, and add the new cluster to OpsCenter's managed systems - all with the click of a button. OpsCenter's management framework makes it easy to add existing clusters using this service.

Figure 5 – Process of creating a new database cluster with OpsCenter.

MANAGING DATABASE CLUSTERS

DSE OpsCenter makes it easy to administer widely distributed clusters that may involve multiple data centers, geographies, and the cloud.

Adding and Removing Capacity

DSE OpsCenter makes it simple to add new capacity to an existing cluster in an online fashion. New nodes can be visually added through OpsCenter, which makes it easy to scale a cluster and ensure that it can handle increasing workloads. In a similar fashion, nodes can be visually removed from a cluster through OpsCenter.
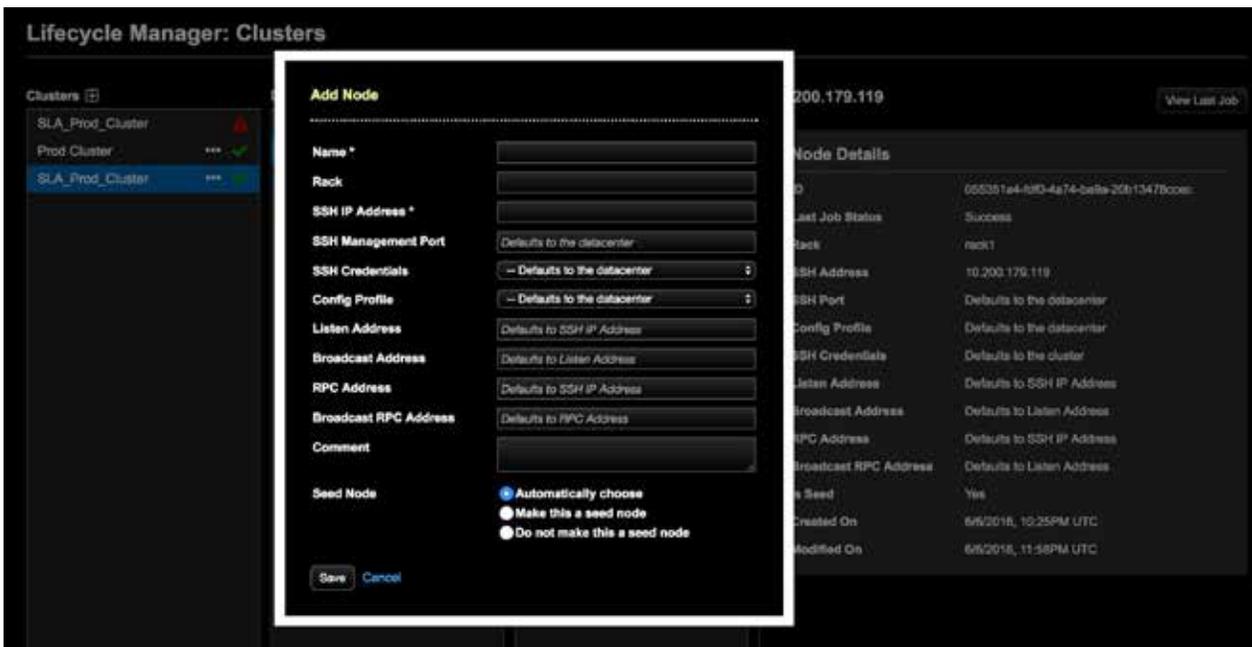

Figure 6 – Adding a node to an existing cluster in OpsCenter

Maintaining Cluster Balance

When an existing cluster is expanded via the addition of new nodes, unless virtual nodes are used, the cluster's data will need to be rebalanced. Performing this operation manually can be complex and time consuming. Fortunately, OpsCenter supplies the ability to rebalance a cluster with the single click of a button.

When you rebalance a cluster through OpsCenter, it will examine a cluster, determine its current state of balance, and then display a forecast of how the cluster will look when rebalanced. If you are satisfied with what OpsCenter intends to do, you can then proceed with the rebalance operation. OpsCenter will keep you fully informed of the rebalance task's progress.
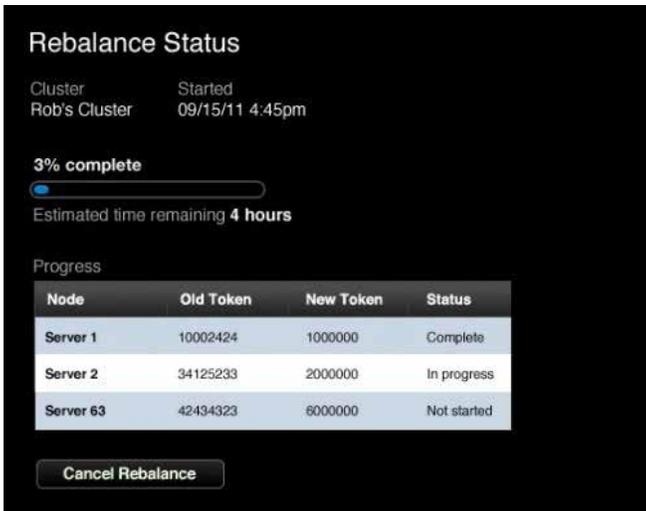
Figure 7 – OpsCenter conducting a rebalance operations

Maintaining Cluster Data Consistency

In a widely distributed and replicated DSE cluster, there are situations that can cause some nodes to have inconsistencies with respect to the replicated data they hold. Detecting when this happens and running what is called a repair operation can be challenging for a busy operations staff.

For clusters running DSE versions 6.0 or greater, DSE's built-in NodeSync feature eliminates the need to run repair operations. OpsCenter is able to visually monitor and toggle NodeSync on a per keyspace or table basis, alleviating operators from needing to run any repair related operations on their clusters.

For clusters running DSE versions below 6.0, OpsCenter's Repair Service allows administrators to transparently handle repair operations with no manual intervention and constantly keeps a cluster's data consistent with little to no performance impact on the cluster.

DSE OpsCenter provides a visual way of enabling the Repair Service and monitoring its progress. In addition, OpsCenter also supports incremental repairs that can vastly improve the efficiency of the repair process.
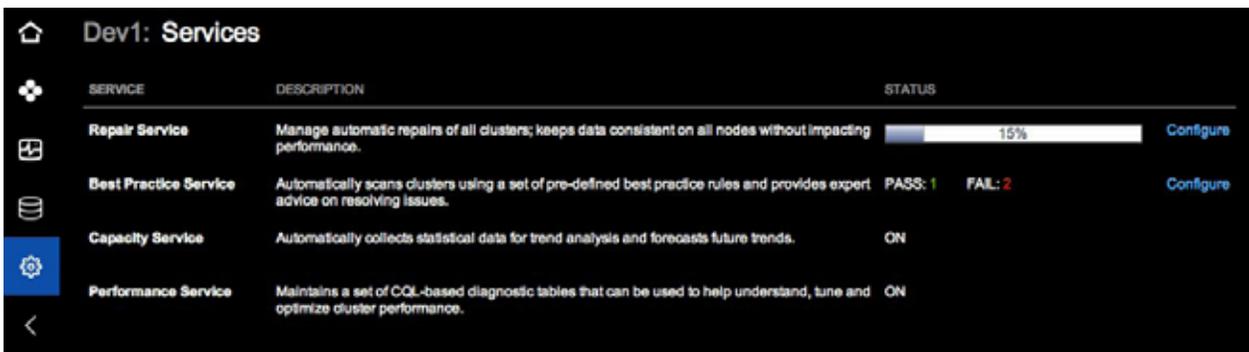

Figure 8 – The Repair Service managed visually via OpsCenter.

Bulk Operations

Oftentimes a task may need to be performed on more than one node. If a cluster consists of hundreds of nodes, it can become difficult and time-consuming to carry out the specific task. DSE OpsCenter supplies a bulk operations interface, which makes performing operations on more than one node very easy. Simply select all or a subset of nodes in a cluster, select the operation you want to perform, and OpsCenter carries out the task for you across all the nodes.
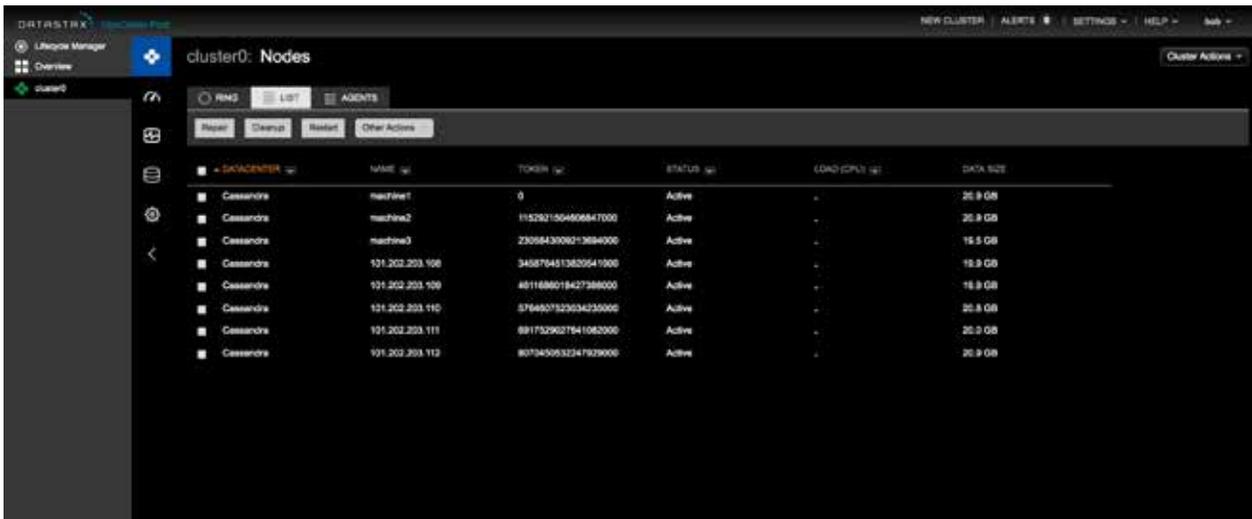

Figure 9 – Managing bulk operations in OpsCenter.

Backup Service

One of the most critical capabilities of a data management platform is backup and restore. Even with a NoSQL database like Cassandra, which can easily replicate data and keep redundant copies of information on multiple servers in different physical locations, data loss can still occur through the accidental dropping of database objects or the unintended deletion of data.

The Backup Service in OpsCenter delivers robust data protection and provides organizations peace of mind in the event of data loss. It makes creating and carrying out a solid backup and restore plan easy. You can schedule backups of a cluster through OpsCenter and monitor backup activity via a comprehensive reporting section. Backups can be customized to do things like move backup files to the cloud (Amazon S3 and Microsoft Azure Blob Storage), to S3 API compatible object storage, or to other local destinations such as NFS. It also comes with a full and efficient incremental backup, including Commit log backups that enable point-in-time recovery.

These backup files can also be automatically compressed to save disk space.

Figure 10 – The backups interface in OpsCenter.

DSE OpsCenter makes it simple and straightforward to restore the data with granular options to do full and object-level restore operations (local or from cloud).

It also supports lightweight database cloning where an option is provided to restore backup files from cloud (Amazon S3 and Microsoft Azure Blob Storage), S3 API compatible object storage, or local destination such as NFS to the same cluster or to a different cluster managed by OpsCenter.

The restore functionality is further enhanced with the addition of true point-in-time restore capability, which allows businesses to restore to any particular date/time in the past without needing a backup, after configuring this feature. This is helpful in situations where a user mistakenly drops a database object and would like to roll back to the point immediately before this incident.

Figure 11 – A point in time restore interface in OpsCenter.

Other Administration Capabilities

DataStax OpsCenter visually handles many other administration functions including:

- Starting/stopping of nodes

- Editing cluster configuration

- Rolling cluster restarts

- Viewing data objects

- Handling Cassandra tasks such as cleanup, flush, and more

- Notifications of software updates

- Generating reports detailing cluster and individual node configuration

OpsCenter provides a REST API that customers can use to integrate the metrics into their existing monitoring environment.

BUILT-IN AUTO FAILOVER

Monitoring and managing critical data continuously without any interruption is a requirement for many organizations due to their strict High Availability (HA) policies. OpsCenter comes with a built-in auto failover option. This enables customers to have multiple instances of OpsCenter with an active-passive setup (only one OpsCenter can be active at a time), which addresses the HA needs by automatically failing over to the standby OpsCenter when the primary fails. The failover will be transparent to the user with zero interruption in most of the services.

MONITORING PERFORMANCE

DSE OpsCenter makes it easy to monitor the uptime and performance of your DSE clusters with a number of different performance interfaces.

The enterprise dashboard provides an at-a-glance snapshot view of the uptime and performance of all the managed clusters.

For each individual cluster, a set of customizable core performance metrics / graphs are provided on a per-cluster dashboard that provides insight into the storage and I/O operations of the system.

Figure 12 – Monitoring the performance of a cluster

Each cluster's dashboard allows you to easily see what is currently happening on a cluster as well as examine the same statistics historically using either the historical tab presets or the custom timeline controls. OpsCenter also allows you to manage, monitor, and track usage of in-memory tables. Moreover, different tabs may be added to the dashboard that allows certain related metrics to be organized together for fast correlated analysis. Charts may be easily added and removed from any tab, and saved / recalled for subsequent uses of OpsCenter.

Capacity Planning

OpsCenter lets you easily perform capacity planning for your DSE clusters. OpsCenter continuously collects key metrics about your managed clusters into its repository and does so with little to no overhead being imposed on the system.

OpsCenter provides the ability to do historical trend analysis across custom time periods for any statistic so you can quickly determine when your clusters experience the most stress and if resource consumption is static or increasing with key metrics such as disk usage. Finally, OpsCenter lets you take historical information and perform forecasting so you can predict future needs (e.g. when to buy more disks or RAM) and answer questions such as "When will this cluster reach 50TB in size?"
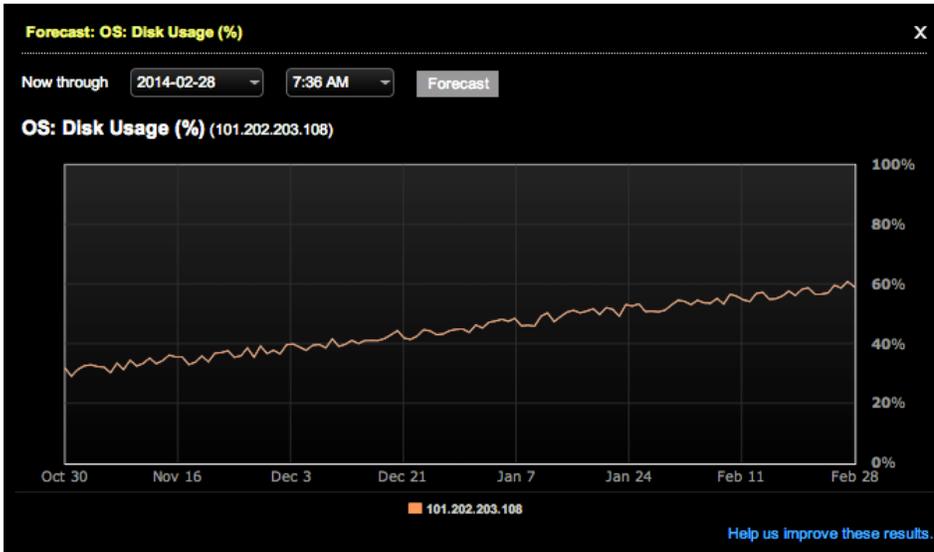
Figure 13 – Forecasting future capacity needs in OpsCenter.

Proactive Alerts

DSE OpsCenter provides the capability to create proactive alerts for things like node outages, I/O bottlenecks, and more in DSE. You can create alerts based on a wide variety of events and statistics, customize the message and choose to be notified both through the OpsCenter interface and outside of OpsCenter (e.g. via an email or SNMP Traps) when a threshold is reached.

Figure 14 – The Alerts interface in OpsCenter.

Visual Performance Service

OpsCenter's Visual Performance Service lets you quickly diagnose performance issues within the cluster, thereby eliminating a lot of manual effort required during troubleshooting. By enabling this service, OpsCenter proactively identifies problem areas and provides context-specific recommendations on possible causes and suggests potential ways to fix the issues.

Operational maturity OpsCenter Performance Service empowers administrators to:

- Visually enable metrics they want to monitor and eliminate the need for custom scripting

- Track metrics at different times of the day for exploration/analysis purposes and compare the results
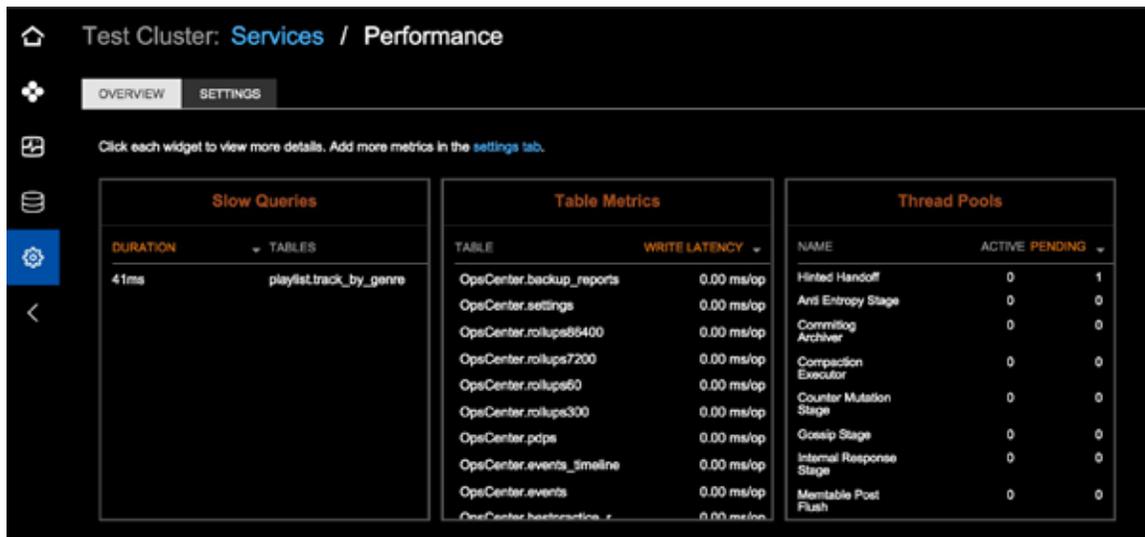

Figure 15- OpsCenter Performance Service in OpsCenter.

Best Practice Service

To help adhere to the best practices for configuring, securing, and optimizing your DSE clusters for maximum performance, OpsCenter provides Best Practice Service. The Best Practice Service is designed to periodically scan clusters and automatically detect issues that threaten a cluster's viability.

The Best Practice Service utilizes a set of rules that span a variety of categories (e.g. security, replication) that analyze different aspects of each cluster that the OpsCenter manages. Any best practice deviations are reported back along with expert advice on how to fix each found issue.

The Best Practice Service can be completely customized to only utilize certain rules for certain clusters, be scheduled to run as often or little as you would like for each cluster, and more. No matter whether you are a novice or an expert with DSE, the Best Practice Service ensures your clusters are configured, secured and optimized to run as efficiently as possible.
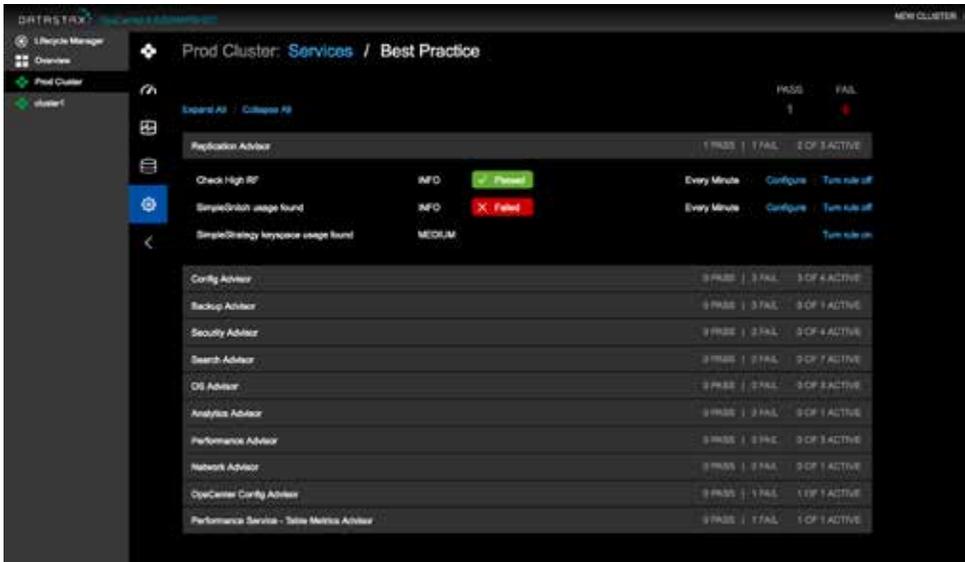
Figure 16 – The Best Practice Service in OpsCenter

Statistical Collection for Troubleshooting

When experiencing performance slowdowns, one of the challenges that the administrators run into is to find out the information needed to diagnose an issue and fix it. DSE OpsCenter makes it easy for users to gather relevant information about a cluster and all its nodes for troubleshooting purposes.

With just a single click, you can easily collect all the pertinent details about your cluster into a single package that can be sent to DataStax Support for analysis. This OpsCenter feature greatly accelerates the troubleshoot process by quickly identifying exactly what went wrong and the needed resolution.

OPSCENTER FEATURES WITH DSE

Monitoring and managing critical data continuously without any interruption is a requirement for many organizations due to their strict High Availability (HA) policies. OpsCenter comes with a built-in auto failover option. This enables customers to have multiple instances of OpsCenter with an active-passive setup (only one OpsCenter can be active at a time), which addresses the HA needs by automatically failing over to the standby OpsCenter when the primary fails. The failover will be transparent to the user with zero interruption in most of the services.

MONITORING	CLUSTER MANAGEMENT
Custom Cassandra and server statistics	Visual cluster provisioning /management for Cassandra
Events dashboard	Perform utility functions on each node or in mass
View replication relationships	Multi-cluster dashboard, support for 1000 node cluster deployment
Proactive alerts	Visual cluster provisioning for analytics, graph and search
External notifications (e.g. email)	Advanced security with role based access controls
Analytics monitoring	Automated node/ring balancing of data
Search monitoring	Visual Backup Service management
In-memory tables monitoring	Visual Performance Service
	Visual restore management
	Visual Repair Service management
	Visual capacity planning w/ forecasting
	Automated software update notifications
	Automated support collection package
	Visual Best Practice Service with built-in expert advisors
	Built-in Auto-Failover
	View Keyspaces
	View Column Families

CONCLUSION

DSE OpsCenter greatly simplifies the task of administering one or many DataStax Enterprise clusters, whether those clusters are deployed on-premises or in the cloud.

For more information about DSE Platform and DSE OpsCenter, visit www.datastax.com. To download a free evaluation version of DSE and DSE OpsCenter, visit http://www.datastax.com/download.

ABOUT DATASTAX

DataStax delivers the always-on, Active Everywhere distributed hybrid cloud database built on Apache Cassandra™. The foundation for personalized, real-time applications at massive scale, DataStax Enterprise makes it easy for enterprises to exploit hybrid and multi-cloud environments via a seamless data layer that eliminates the issues that typically come with deploying applications across multiple on-premises data centers and/or multiple public clouds.

Our product also gives businesses full data visibility, portability, and control, allowing them to retain strategic ownership of their most valuable asset in a hybrid/multi-cloud world. We help more than 400 of the world's leading brands across industries transform their businesses through an enterprise data layer that eliminates data silos and cloud vendor lock-in while powering modern, mission-critical applications. For more information, visit DataStax.com and follow us on @DataStax.