

Using SmartSense 1

Using SmartSense

Date of Publish: 2018-07-12

<http://docs.hortonworks.com>

Contents

Roles required for using SmartSense.....	3
Capturing bundles.....	4
Capture bundles for proactive analysis.....	4
Capture bundles for troubleshooting.....	4
Automatically capturing and uploading bundles via SmartSense gateway.....	5
Create a new capture schedule.....	5
Update the capture schedule.....	6
Download and upload bundles.....	6
Review SmartSense recommendations.....	7
Reviewing a recommendation.....	8
Apply a recommendation.....	9
Review and revert previously applied recommendations.....	10
Export recommendations as excel spreadsheet.....	11
Access the activity explorer.....	11
Chargeback dashboard.....	12
HDFS dashboard.....	12
Job Comparison dashboard.....	13
User Summary dashboard.....	13
Workload dashboard (MR, Pig, Tez, Hive).....	13
Workload Trend dashboard.....	14
YARN dashboard.....	14

Roles required for using SmartSense

The following table describes bundle capture-related actions and roles required to perform them.

"Other Users" include Cluster Administrator, Cluster User, Service Operator, Service Administrator, and Cluster Operator as defined in Ambari documentation related to understanding cluster roles.

Table 1: Roles required for using SmartSense

Action	Ambari Administrator	Other Users*
Access Ambari View		
Initiate SmartSense capture		
Initiate support capture		
View "Bundles" page		
View bundle		
Upload a bundle		
Download encrypted bundles		
Download unencrypted bundles		
Delete bundles		
View capture schedule		
Update capture schedule		
Pause capture schedule		
Activate capture schedule		

Action	Ambari Administrator	Other Users*
Delete capture schedule		
View recommendations		
Apply recommendations		
Revert recommendations		

Related Information

[Understanding cluster roles](#)

Capturing bundles

After you install the SmartSense service and view, data collection can begin.

To trigger an ad hoc capture, access the **SmartSense View** by clicking the



icon and selecting **SmartSense View**, and then follow steps depending on your use case:

- Capturing for proactive analysis: Use this if you would like to prevent issues, improve security, and/or increase availability and performance of your cluster.
- Capturing for troubleshooting: Use this if you are working with support to troubleshoot a support case.

Capture bundles for proactive analysis

To capture bundles for analysis, follow these steps.

Procedure

1. To trigger an ad hoc capture, access the **SmartSense View** by clicking the



icon and selecting **SmartSense View**.

2. Click the Capture link in the top-right corner to access the Data Capture page.
3. On the Data Capture page, under Select the intent for data capture, select **Proactive Analysis**.
4. Click the **Capture** button.
SmartSense will analyze cluster configuration and metrics for all cluster nodes, and will produce recommendations to prevent issues, improve security, availability and performance of your cluster.

Capture bundles for troubleshooting

To capture bundles for troubleshooting, follow these steps.

Procedure

1. To trigger an ad hoc capture, access the **SmartSense View** by clicking the



icon and selecting **SmartSense View**

2. Click the Capture link in the top-right corner to access the Data Capture page.
3. On the Data Capture page, under Select the intent for data capture, select **Support Case Troubleshooting**.
4. Enter your Case Number.
5. Select the type of diagnosis:

- Cluster Service

a) Select services for diagnosis. The list services that can be captured is available in the documentation related to services available for capture.

b) Select hosts for diagnosis: All Hosts or choose Only Specific Hosts and select specific hosts.

- YARN Application:

Enter Application ID. The YARN application details, application master logs and a subset of container logs will be captured.

- Hive Query:

Note:

This option is only available when DAS is installed on the cluster.

Enter one of the following: Tez DAG ID, YARN App ID, MR Job ID, or Hive Query ID. The SQL query, execution plan, application logs will be captured.

6. Click the Capture button.

This triggers Ambari agents on each node to invoke the HST agent to capture specific data.

Results

After HST agents complete their captures and report data to the HST server, the completed bundle is available in the bundles list for download, or it is automatically uploaded to the SmartSense Gateway, if configured. Refer to the documentation related to downloading and uploading bundles.

Related Information

[List of services available for capture](#)

[Download and upload bundles](#)

Automatically capturing and uploading bundles via SmartSense gateway

When enabled, the gateway automatically uploads completed bundles to Hortonworks when a capture is completed. This includes SmartSense analysis as well as support case troubleshooting bundles. You can also schedule SmartSense analysis bundles for capture and automatic upload in the SmartSense Ambari View.

Create a new capture schedule

If you have deleted the default capture schedule, you can create a new one by following these steps:

Procedure

1. Access **SmartSense View** by clicking  and selecting **SmartSense View**.
2. Click the **Schedule** link in the top right corner to access the scheduler settings.
3. Select the scheduling period (weekly or monthly) and the day of the week and time of day that you want the capture to take place. Click **Set Capture Schedule**.

Note:

Scheduler changes take up to one hour to take effect.

Update the capture schedule

The SmartSense view provides a way to easily create, update, pause, resume, and remove the schedules used for automated bundle capture and upload. When you deploy it, SmartSense creates a default capture schedule. To view this default capture schedule and update it, follow these steps:

Procedure

1. Access **SmartSense View** by clicking  and selecting **SmartSense View**.
2. Click the **Schedule** link in the top right corner to access the scheduler settings.
3. Remove, pause, or resume existing schedules.

You can also update the capture schedule by selecting a new scheduling period (weekly or monthly) or changing the day of the week and time of day that you want the capture to take place.

Note:

Scheduler changes take up to one hour to take effect.

Download and upload bundles

Completed bundles can be manually downloaded and uploaded. To view and download bundles, follow these steps:

Procedure

1. Access the **SmartSense View** by clicking  and selecting **SmartSense View**.
2. Click the **Bundles** link in the top right corner.
This page shows all bundles that have been captured and their status. If data is still being captured, the UI automatically updates itself with the capture progress until completed.
3. After the bundle is in a completed state, you can:
 - Download it manually by clicking **Download** and selecting either **Download Encrypted** or **Download Unencrypted**.
 - Upload it manually by clicking **Upload**.

What to do next

You can also automate and schedule this process by using the SmartSense gateway. When using the SmartSense gateway, all bundles are uploaded to Hortonworks. When support case troubleshooting bundles are received, they trigger a case notification. This case notification uses the case number provided during the capture initiation process. For more information about available ways to upload support bundles, refer to documentation related to bundle transport.

Related Information

[Bundle transport](#)

Review SmartSense recommendations

You can access your SmartSense recommendations from SmartSense View in Ambari web UI, as described in the following steps.

Prerequisites

In order for recommendations to be generated for your cluster, you first need to capture a bundle and then upload it through HTTPS gateway for analysis. Alternatively, you can schedule automatic capture and upload using the SmartSense gateway.

Steps

1. Access the **SmartSense View** by clicking



and selecting **SmartSense View**.

2. Click the **Recommendations** link in the top right corner.
3. From the recommendations page, you can view open recommendations, and view previously deferred and ignored recommendations.
4. To make sure that the recommendations are up-to-date, from the



menu select **Get Latest**.

5. You can search and filter the recommendations:
 - Click on a column name to sort the recommendations accordingly.
 - Use the search box in the top right corner to filter recommendations.
 - To review and apply, ignore, or defer a recommendation, click on its corresponding row.

The following information is available for each item on the recommendations list:

Table 2: Recommendations list

Column	Description
Priority	One of:    
Service	The HDP service to which the recommendation applies.
Recommendation	The summary of the recommendation.
Category	Broad category (such as "Operations", "Performance", or "Security") to which the recommendation belongs.
Avg. Rating	Average customer rating for the recommendation.
Classifiers	These markers indicate actions available for any recommendation:  means that the recommendation can be applied automatically through Ambari.  means that the configuration is not managed by Ambari and you must apply the recommendation manually.  means that in order to apply the recommendation you must also apply dependent recommendations.  means that the recommendation has previously been applied and then reverted.

Related Information[Capturing bundles](#)[Download and upload bundles](#)[Automatically capturing and uploading bundles via SmartSense gateway](#)[Reviewing a recommendation](#)**Reviewing a recommendation**

To review a recommendation, click on its corresponding row.

The following actions are available for each recommendation:

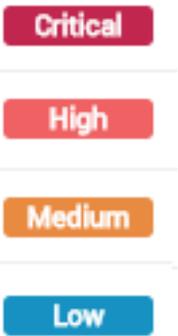
- Ignore - Let us know that you do not want to apply this recommendation and help us understand why.
- Defer - Let us know that you will be applying this at a later date, but not right now.

- Mark As Applied (for recommendations that have to be applied manually), or Proceed to Apply (for recommendations that can be applied automatically)

If you ignore or defer a recommendation, you can still apply it later.

After clicking on a specific recommendation, the following information is available for each recommendation:

Recommendation details

Column	Description
Priority	One of: 
Status	One of: <ul style="list-style-type: none"> • Open • Applied • Ignored • Deferred • Reverted • Reopened
Recommendation Category	Broad category (such as "Operations", "Performance", or "Security") to which the recommendation belongs.
Affects	Describes to which specific software component the recommendation is related.
Rule Id	Unique ID that identifies the SmartSense rule related to the SmartSense recommendation.
Description	Background and context related to the recommendation.
Findings	Description of how your cluster deviates from the recommended configuration.
Recommendation	An outline of specific changes that need to be made to your cluster to apply the recommendation.
Configurations	Lists affected configuration properties, including: <ul style="list-style-type: none"> • Config File - The specific file that needs to be changed • Property Name - The specific property that needs to be changed • Captured Value - Configured value at the time of bundle capture • Current Value - Current configured value in Ambari • Recommended Value - Recommended value for this cluster
Affected hosts	Hosts on which the configuration change is required.

Apply a recommendation

While some recommendations can be applied automatically, others have to be applied manually.

There are two ways to tell how a recommendation can be applied:

- When reviewing open recommendations, you can see in the Classifiers column, what options are available for which recommendation.

- When reviewing a specific recommendation, you can see one of the two options: Mark As Applied (for recommendations that must be applied manually) or Proceed to Apply (for recommendations that must be applied automatically).

Automatically apply a recommendation

Applying a recommendation automatically involves the following tasks:

Steps

1. From the Recommendations page, click on the table row corresponding to the recommendation that you want to review.
2. Click on Proceed to Apply.
3. Review recommended changes.
4. Enter a comment in the Change Notes field. This comment will later allow you to track the Ambari configuration version created after applying a configuration.
5. Click on Apply.
6. You can optionally provide feedback for this recommendation and then click on Submit Feedback. Or you can opt out and click I will provide later. You can still provide feedback later, from the History page.
7. You can view the configuration change in Ambari configuration history.

Manually apply a recommendation

Applying a recommendation manually involves the following tasks:

Steps

1. From the Recommendations page, click on the table row corresponding to the recommendation that you want to review.
2. Apply the recommendation manually.
3. Click on Mark As Applied.
4. Click on I have to confirm that you've applied the changes.

Note:

You can revert previously applied recommendations. This option is available on the History page.

Related Information

[Review SmartSense recommendations](#)

[Reviewing a recommendation](#)

[Review and revert previously applied recommendations](#)

Review and revert previously applied recommendations

To view the history of previously applied recommendations, follow these steps:

Procedure

1. Click on the



menu and select Show History.

The History tab allows you to review previously applied, ignored, deferred, and reverted recommendations, and, if needed, revert applied recommendations and review and reopen deferred and ignored recommendations.

2. To get more details about a specific recommendation, click



3. You have an option to Review and Revert. If a recommendation has previously been reverted, this option is grayed out.

Export recommendations as excel spreadsheet

You can export SmartSense recommendations to an Excel spreadsheet (XLSX file format).

To do that, click on the



menu and select Export as Excel. The spreadsheet will be downloaded to your default download location.

Access the activity explorer

The activity explorer includes an embedded instance of Apache Zeppelin, which hosts pre-built notebooks that visualize cluster utilization data related to user, queue, job duration, and job resource consumption. To access the activity explorer, follow these steps:

Procedure

1. Navigate to the Ambari **Dashboard** and click the **SmartSense** service.
2. In the **Summary** tab, access **Quick Links > Activity Explorer**.
This launches the activity explorer in a new browser tab.
3. Log in with your activity explorer admin credentials.
4. From the **Notebook** dropdown in the top toolbar, select the name of the notebook that you want to view.

Zeppelin organizes data in notebooks, where each notebook contains rows of paragraphs. Each paragraph visualizes the results of a single SQL statement using either a table, bar chart, pie chart, area chart, line chart, or scatter plot.

The following preconfigured notebooks are available:

- Chargeback dashboard
- HDFS dashboard
- Job Comparison dashboard
- User Summary dashboard
- Workload dashboard (MR, Pig, Tez, Hive)
- Workload Trend dashboard
- YARN dashboard

Once you opened a notebook, be aware of these three operations:

- a. Since the notebooks represent a view of SmartSense utilization data at a specific point in time, they need to be refreshed. In order to refresh all of the data shown in all paragraph of a notebook, you need to:
 - Hover over the row containing the notebook title, and a set of controls will appear.
 - Click on the



button to “Run all paragraphs”. The data for each paragraph in the notebook will be refreshed.

- b. Top N paragraphs show the top 10 entries by default, but you can change this number by entering a new number in the **Top** input field and then typing enter.
- c. Charts have interactive filters that let you select and deselect specific resources by clicking on the circle in the chart legend. For example, if there are four resources being displayed in a chart, and you only want to see four, you can click on a colored circle in the legend to filter it out:

HIVE MR PIG TEMPLETON

Once clicked, the inside of the circle will change to white, and the entry will not be displayed in the chart. For example, if you deselect "Hive", the legend will look like this:

HIVE MR PIG TEMPLETON

Once

Chargeback dashboard

The Chargeback dashboard helps operators understand which resources are being consumed and what costs are associated with these resources. This dashboard exposes five types of resources:

- CPU Hours (in hours) - The amount of CPU used by MapReduce and Tez jobs
- Memory Hours (in gigabytes) - The amount of memory consumed by MapReduce and Tez jobs, and length of consumption
- Storage (in gigabytes) - The amount of HDFS space being consumed
- Data IO (in gigabytes) - The amount of data read and written to HDFS
- Network IO (in gigabytes) - The amount of data sent and received over the cluster's network

The information is contained in the following paragraph:

Paragraph	Description
Chargeback Report	<p>This paragraph lets you associate a financial cost with each of the five resources presented in the previous paragraph. Based on these per unit financial costs, you can see how much should be charged for each resource type.</p> <p>The report also sums up the charge per resource to a per user total, so it's easy to see how much should be charged for that specific user for their total resource consumption.</p> <p>The goal is to show how much money each user should be charged for the cluster resources that they have consumed.</p>

HDFS dashboard

The HDFS dashboard helps operators better understand how HDFS is being used and which users and jobs are consuming the most resources within the file system.

This dashboard includes the following paragraphs:

- File Size Distribution
- Users with Maximum Small Files
- Users with Maximum HDFS Utilization
- HDFS File Size Trend
- HDFS Utilization Trend

Most of these paragraphs have titles that are self-explanatory. A few of them are described below to provide more context:

Paragraph	Description
File Size Distribution	For any large multi-tenant cluster, it's important to identify and keep the proliferation of small files in check. The paragraph displays a pie chart showing the relative distribution of files by file size categorized by Tiny (0-10K), Mini (10K-1M), Medium (30M-128M), and Large (128M+) files. The goal is to show how dominant specific file size categories are within HDFS. If there are many small files, you can easily identify (in the next paragraph) who is contributing to those small files.
Users with Maximum Small Files	Understanding how prevalent files of specific sizes are is helpful, but the next step is understanding who is responsible for creating those files. The goal of this paragraph is to show who is responsible for creating the majority of small files within HDFS.

Job Comparison dashboard

The Job Comparison dashboard allows you to compare two executions of the same job.

To compare two executions of a job, provide the job IDs under Job To Compare and Job To Compare With, select the units that you would like to use for the results, and then click on the



button.

The following paragraphs are included in the results:

Paragraph	Description
Job Summary Comparison	Compares specific configuration values.
Configuration Differences	Compares specific configuration values.
Counter Differences	Compares counters.

User Summary dashboard

The User Summary dashboard provides a summary of a specific user's activity (such as HDFS utilization, workload information, and so on) on the cluster.

To obtain this report for a specific user, provide the User Id, select the reporting period for which you would like to see the report, select the units that you would like to use for the results, and then click on the



button.

The following paragraphs are displayed in this report:

- HDFS Utilization
- HDFS File Size Trend
- Workloads Executed
- Data Processed
- CPU Comparison
- Memory Consumption

Workload dashboard (MR, Pig, Tez, Hive)

The Workload dashboard (MR, Pig, Tez, Hive) provides key information about workloads that use MapReduce or Tez for execution.

This dashboard includes the following paragraphs:

- Longest Running Jobs

- Most Resource Intensive Jobs
- Most Resource Wasting Jobs
- Workloads With Highest HDFS Operations
- Workloads Creating Max HDFS Files
- Workloads With Largest HDFS Writes
- Workloads With Highest CPU Consumption
- Workloads With Most Inefficient Data Read
- Workloads With Most Input Data Explosion
- Job Distribution By Type
- Job Submission Trend By Day.Hour

Most of these paragraphs have titles that are self-explanatory. A few of them are described below to provide more context:

Paragraph	Description
Most Resource Wasting Jobs	Resource wasting is calculated by calculating the difference between the memory asked for and the memory that was actually used. For example, if a job asks for 100 8GB containers but only uses 5GB per container, 3GB per container is considered wasted. This is calculated per job, and the top 10 are listed.
Job Submission Trend By Day.Hour	This paragraph shows the number of jobs submitted by day and hour with the notation being <day>.<hour>. For example: <ul style="list-style-type: none"> • Monday.1 - 1am on Monday • Monday.20 - 8pm on Monday The goal of this dashboard is to identify specific job submission hotspots during the week and day. You can use this information to identify the best time to schedule resource intensive jobs to execute.

Workload Trend dashboard

The Workload Trend dashboard provides trends of key execution metrics (time, CPU, memory, amount of data processed, and so on) for all executions of the same job over a period of time.

To obtain this report for a specific workload, provide the workload name, select the reporting period for which you would like to see the report, select the units that you would like to use for the results, and then click on the



button.

This dashboard includes the following trends about your selected workload:

- Execution Duration
- Number of Tasks
- Data Processed
- Memory Consumption
- CPU Consumption

YARN dashboard

The YARN dashboard provides key information for queue, application, container, and NodeManager host metrics.

This dashboard includes the following paragraphs:

- Application Runtime Duration by Queue
- CPU Consumption Per Queue
- Memory Consumption Per Queue
- Applications per Queue

- Applications Requesting Max Containers
- Applications With Max Containers Failures
- Hosts With Max Containers Executed
- Hosts With Max Application Failures
- Hosts With Longest Localization Time
- Hosts With Longest Container Launch Delay

Most of these paragraphs have titles that are self-explanatory. One of them is described below to provide more context:

Paragraph	Description
Applications With Max Containers Failures	This paragraph shows the top jobs with the highest number of failed containers and the reason for each failure, so that you can quickly identify which containers failed and why.