

Reports.....

The Reports page lets you create reports about the usage of HDFS in your cluster—data size and file count by user,

group, or directory. It also lets you report on the MapReduce activity in your cluster, by u

To display the Reports page, select ClustersCluster nameReports.

For users with the Administrator role, the Search Files and Manage Directories button on the file browser for searching files, managing directories, and setting quotas.

If you are managing multiple clusters, or have multiple nameservices configured (if high availability is configured) there will be separate reports for each cluster and nameservice.

See the following pages for more details:

Directory Usage Report

The directory usage report allows you to browse the HDFS filesystem in a way that is similar to the File Browser. However, the Directory Usage Report also allows you to sort the listings and select actions to perform actions on them.

Minimum Required Role: BDR Administrator (also provided by Full Administrator and Cluster Administrator). This feature is not available when using Cloudera Manager to manage Data Hub clusters.

You can also view the last access time, the last modified time of any file in a directory, and the size of the directory. This usage information is updated on an hourly basis.

You can customize the report by adding filters. A number of preconfigured filters are available, and you can create a custom filter.

Related Information

The File Browser

Accessing the Directory Usage Report

You can access the directory usage report through the Clusters menu or the HDFS File Browser.

Procedure

1. Click Clusters Cluster Name Reports Directory Usage.
2. Optionally, you can access the report through the HDFS File Browser.
 - a) Click Clusters HDFS service File Browser.
 - b) Click the Directory Usage link located in the lower-right portion of the File Browser.

Using the Directory Usage Report

In the directory usage report, you can sort the display, view files and subdirectories in the directory. When you first open the report, the top level of the HDFS filesystem displays:

Click the Filters link to select options that limit the output by various criteria.

Click any column header to sort the display.

Cloudera Manager Reports

Click a directory name to view the files and subdirectories in the directory.

Click Download CSV to download a CSV version of the report. Select one of the following options:

- Top 1K Rows
- Top 10K Rows
- Top 100K Rows Select one or more rows by checking the boxes on the left and then choose an action to perform on the selection from the Actions for Selected drop-down menu. You can select the following actions:
- Manage Quota – A dialog box opens in which you can set a quota for the number of files or disk space. These values are displayed in columns in the file listing.
- Include selected directories in disk usage reports – The selected directories appear in the disk usage reports.
- Exclude selected directories from disk usage reports – The selected directories do not appear in the disk usage reports. Related Information Disk Usage Reports

Filters

You can use filters to limit the display and to search for files.

Procedure

1. To apply filters to the directory usage report, click the Filters drop-down menu near the top of the page and select one of the following preconfigured filters:
 - Large Files
 - Large Directories
 - By Specific Owner
 - By Specific Group
 - Old Files
 - Old Directories
 - Files with Low Replication
 - Overpopulated Directories
 - Directories with Quotas
 - Directories Watched
2. To modify any of these filters, click the Customize link and select new criteria. Click Clear to revert to the preconfigured criteria for the filter.
3. Click the Search button to display the report with the filters applied.
4. You can also select Custom from the Filters drop-down menu to create a report in which you define the criteria. To create a custom report:
 - a) Select any of the following criteria from the drop-down menu on the left:
 - Filename
 - Owner
 - Group
 - Path
 - Last Modified
 - Size
 - Diskspace Quota

- Namespace Quota
 - Last Access
 - File and Directory Count
 - Replication
 - Parent
 - Raw Size
 - Size With Snapshot
 - Total Size
- b) Select an operator from the drop-down menu.
 - c) Enter a value and units of measure for the comparison.
 - d) Select the units of measure for the comparison from the drop-down menu. (Some criteria do not require units of measure.)
 - e) Click the icon to add additional criteria.
 - f) Click the Search button to display the directory usage report with the custom filter applied.

The report changes to display the result of applying the filter. A new column, Parent is added to the report, showing the full path to each file or subdirectory.

Disk Usage Reports

There are two types of disk usage reports: Current Disk Usage By Directory and Historical Disk Usage By Directory.

Procedure

1. To use these reports, select one or more directories to watch by clicking the icon for the directory.
2. Alternatively, you can select multiple directories, and then click Actions for Selected to include selected directories in disk usage reports.

Disk Usage Reports

The disk usage reports show HDFS disk usage statistics, either current or historical, by user or by directory. The By Directory reports display information about the directories in the Watched list, so if no directories are selected, there will be no results found for these reports. You can also specify which directories to include in the report by selecting them from the Directory Usage Report.

Related Information

Designating Directories to Include in Disk Usage Reports

Directory Usage Report

Viewing Current Disk Usage by User, Group, or Directory

The current disk usage reports show "current" disk usage in both chart and tabular form. The data for these reports comes from the fsimage kept on the NameNode, so the data in a report is current as when the last checkpoint was performed. Typically the checkpoint interval is (by default) 12 hours, but if checkpoints are not being performed as frequently, the disk usage report may not be up to date. The report displays the current usage and does not account for deleted files that only exist in the NameNode. Deleted files are included in the usage information when you run the du command.

To create a disk usage report:

- Click the report name (link) to produce the resulting report. Each of these reports show:

Bytes The logical number of bytes in the files, aggregated by user, group, or directory. This report shows the files sizes, not taking replication into account.

Raw Bytes The physical number of bytes (total disk space in HDFS) used by the files aggregated by user, group, or directory. This does include replication, and so is actually Bytes times the number of repli

File and Directory Count The number of files aggregated by user, group, or directory.

Bytes and Raw Bytes are shown in IEC binary prefix notation (1 GiB = $1 * 2^{30}$).

The directories shown in the Current Disk Usage by Directory report are the HDFS directories that are currently watched directories. You can add or remove directories to or from the watch list from this report. Click the Files and Manage Directories button at the top right of the set of reports for the cluster of interest. The report data is also shown in chart format:

- Move the cursor over the graph to highlight a specific period on the graph and see the actual value (data size) for that period.
- You can also move the cursor over the user, group, or directory name (in the graph legend) to highlight the portion of the graph for that name.
- You can right-click within the chart area to save the whole chart display as a single image (a .PNG file) or as a PDF file. You can also print to the printer configured for your browser. Related Information Designating Directories to Include in Disk Usage Reports

Viewing Historical Disk Usage by User, Group, or Directory

You can use the historical disk usage reports to view disk usage over a time range you define. The reports show statistics reported per hour, day, week, month, or year.

Procedure

1. To create the report, click the report name (link) to produce the initial report. This generates a report that shows Raw Bytes for the past month, aggregated daily.
2. To change the report parameters, select the Start Date and End Date to define the time range of the report.
3. Select the Graph Metric you want to graph: bytes, raw bytes, or files and directories count.
4. In the Report Period field, select the period over which you want the metrics aggregated. The default is Daily. This affects both the number of rows in the results table, and the granularity of the data points on the graph.
5. Click Generate Report to produce a new report. As with the current reports, the report data is also presented in chart format, and you can use the cursor to view the data shown on the charts, as well as save and print them. For weekly or monthly reports, the Date indicates the date on

which disk usage was measured. The directories shown in the Historical Disk Usage by Directory report are the HDFS directories you have set as watched directories. Related Information Designating Directories to Include in Disk Usage Reports

Downloading Reports as CSV and XLS Files

Any report can be downloaded to your local system as an XLS file (Microsoft Excel 97–2003 worksheet) or a CSV (comma-separated value) text file.

About this task

To download a report, do one of the following:

Procedure

- From the main page of the Report tab, click CSV or XLS link next to in the column to the right of the report name
- From any report page, click the Download CSV or Download XLS buttons. Either of these opens the Open file dialog box where you can open or save the file locally.

Activity, Application, and Query Reports

The Reports page contains links for displaying metrics on the following types of activities

About this task

The Reports page contains links for displaying metrics on the following types of activities

- Disk usage
- MapReduce jobs
- YARN applications
- Impala queries
- HBase tables and namespaces

Procedure

1. To view the Reports page, click ClustersClusterNameReports. You can generate a report to view aggregate job activity per hour, day, week, month, or year, by user or for all users.
2. Click the Start Date and End Date fields and choose a date from the date control.
3. In the Report Period drop-down, select the period over which you want the metrics aggregated. Default is Daily.
4. Click Generate Report. For weekly reports, the Date column indicates the year and week number (for example, 2013-01 through 2013-52). For monthly reports, the Date column indicates the year and month by number (2013-01 through 2013-12).

The File Browser

The File Browser tab on the HDFS service page lets you browse and search the HDFS namespace files and directories.

Cloudera Manager Reports

Minimum Required Role: BDR Administrator (also provided by Full Administrator and Cluster Administrator). This feature is not available when using Cloudera Manager to manage Data Hub clusters.

The File Browser page initially displays the root directory of the HDFS file system in the grid. The grid displays the immediate subdirectories below. Click any directory to drill down into the contents of that directory for available actions.

Searching Within the File System

When you search within the file system, you can select from custom search criteria such as file size, file type, and more.

To search the file system, click Custom report in the Reports section. The file and directory listings are based on the fsimage stored on the NameNode, so the listings will be only as current as the last checkpoint interval is (by default) once per hour, but if checkpoints are not being performed, the listings may not be up to date.

To search the file system:

1. From the HDFS service page, select the File Browser tab.
2. Click Choose and do one of the following:
 - Select a predefined query. Depending on what you select, you may be presented with different fields to fill in or different views of the file system. For example, selecting Size will provide a choice of arithmetic operators and fields where you provide the size to be used as the search criteria.
 - a. Select a property in the Choose... drop-down.
 - b. Select an operator.
 - c. Specify a value.
 - d. Click to add another criteria (all of which must be satisfied for a file to be considered a match) and repeat the preceding steps.
3. Click the Generate Report button to generate a custom report containing the search results. If you search within a directory, only files within that directory will be found. For example, if you browse /user and do a search, you might find /user/foo/file, but you will not find /bar/baz.

Enabling Snapshots

To enable snapshots for an HDFS directory and its contents, see the topic Managing HDFS Snapshots.

Setting Quotas

To set quotas for an HDFS directory and its contents, see Setting HDFS Quotas.

Designating Directories to Include in Disk Usage Reports

You can designate directories to include in a disk usage report.

Procedure

1. To add or remove directories from the directory-based Disk Usage reports, navigate through the file system to see the directory you want to add. You can include a directory at any level without including its parent.
2. Check the checkbox Include this directory in Disk Usage reports. As long as the checkbox is checked, the directory appears in the usage reports. To discontinue inclusion of the directory in Disk Usage reports, clear the checkbox.

Downloading HDFS Directory Access Permission Reports

For each HDFS service, you can download a report that details the HDFS directories a group has access to.

Before you begin

Minimum Required Role: Cluster Administrator (also provided by Full Administrator)

Procedure

1. In the Cloudera Manager Admin Console, click ClustersClusterNameReports.
2. In the Directory Access by Group row, click CSV or XLS. The Download User Access Report pop-up displays.
3. In the pop-up, type a group and directory.
4. Click Download. A report of the selected type will be generated containing the following information – path, owner, permissions, and size – for each directory contained in the specified directory that the specified group has access to.