

Configuring Data Collector

Data Collector is the component that accepts reports from testing and code analysis tools. You can change Data Collector's default storage and upload limits to meet data requirements for your organization.

In this section:

- [Configuring the Report Storage Threshold](#)
- [Configuring the Maximum Size of Report Uploads](#)
- [Configuring Data Retention Settings](#)
- [Configuring Test Failures Threshold](#)
- [Controlling Coverage Data Processing for DTP Enterprise Pack Artifacts](#)

Configuring the Report Storage Threshold

When Data Collector receives and processes files from analyzers, a report file is processed and moved to the `[DTP_HOME]/data/_DEFAULT_/dc/stored` folder. This folder contains subfolders that are structured according to a YEAR/MONTH/DAY/XXX/YYY format where XXX and YYY are integer values. The values are calculated based on file name.

You can set a maximum size for the stored folder and DTP will automatically prune older entries.

1. Open the `DCConfig.xml` file located in the `[DTP_HOME]/grs/config/` directory and locate the following entries:

```
<stored-folder>
  <max-size>20</max-size>
  <size-to-clean>5</size-to-clean>
</stored-folder>
```

2. Change the value for the `<max-size>` property (in GB) to an acceptable maximum storage capacity; the default is 20 GB.
3. Change the value for the `<size-to-clean>` property (in GB) to the amount of data you want pruned when the maximum capacity is reached; the default is 5 GB.
4. Save the file.

If these entries do not appear in the `DCConfig.xml` file, then the defaults will be used.

Configuring the Maximum Size of Report Uploads

By default, the maximum size of the XML report file that Data Collector can accept is 512M. You can change the limit by adding the following Data Collector JVM argument:

```
-Dcom.parasoft.sdm.api.rawstorage.datacollector.uploadMaxSize=512
```

For Linux installations, you can modify the following string located in the `[DTP_HOME]/bin/variables` file:

```
JAVA_DC_CONFIG_ARGS=" -Dcom.parasoft.sdm.api.rawstorage.datacollector.uploadMaxSize=512"
```

Configuring Data Retention Settings

You can configure the amount of unit testing, coverage, resource coverage, and metrics data that DTP stores. Test, coverage, and metrics data are collectively referred to as "test details" because together they provide critical testing information about the build. The data from these practices consume substantial storage, so DTP preserves space by regularly pruning the data associated with older builds in the filter.

About Resource Coverage

DTP collects coverage at the method level and at the resource level. Method-level coverage data is required to view coverage in the [Coverage Explorer](#). Unless explicitly stated, "coverage" in this documentation means method-level coverage.

Resource coverage data refers to the code coverage data associated with resource groups. A resource group is a collection of files and/or folders defined by a set of one or more ANT file patterns. Resource groups are subsets of a project that you can associate with a filter. They enable you to view software quality information for specific parts of the code. See [Adding Resource Groups to Projects](#) for additional information about resource groups.

Default Data Retention Settings

The amount of data stored is based on the number of historical builds, as opposed to a hard limit, such as gigabytes. By default, DTP keeps the following data:

- Unit test data: two (2) historical builds
- Coverage data: two (2) historical builds
- Resource coverage data: ten (10) historical builds
- Metrics data: eight (8) historical builds.

Configuration

To configure the data retention settings:

1. Stop the Data Collector service. See [Stopping DTP Services](#).
2. Open the DTP_HOME/grs/config/DCConfig.xml configuration file and locate the <details-retention-builds-count> group of settings:

```
<!-- <details-retention-builds-count>
  <tests>2</tests>
  <coverage>2</coverage>
    <resource-coverage>10</resource-coverage>
  <metrics>8</metrics>
</details-retention-builds-count-->
```

3. Uncomment the settings and specify how much data should be stored for each practice. The values refer to number of builds. Specifying 1, for example, means storing one build worth of data for the practice.

```
<details-retention-builds-count>
  <tests>3</tests>
  <coverage>4</coverage>
    <resource-coverage>12</resource-coverage>
  <metrics>9</metrics>
</details-retention-builds-count>
```

Additional Details About Retention Settings

If the settings are commented out or no value is specified, then the default value is used.

If the value is set to 0, then there will be no limit on the number of builds saved. Use caution when saving an unlimited number of builds to the database. Retaining a large number of builds will reduce performance.

4. Save the file and restart Data Collector.

Configuring Test Failures Threshold

By default, Data Collector will reject reports that contain more than 5000 reported test failures. You can change the limit by adding the following Data Collector JVM argument:

```
-Dcom.parasoft.sdm.rawstorage.failures.limit=5000
```

For Linux installations, you can modify the following string located in the [DTP_HOME]/bin/variables file:

```
JAVA_DC_CONFIG_ARGS=" -Dcom.parasoft.sdm.rawstorage.failures.limit=5000"
```

If a negative value is provided, Data Collector will not reject reports due to number of test failures.

Controlling Coverage Data Processing for DTP Enterprise Pack Artifacts

By default, Data Collector does not process coverage data for the [Change Based Testing](#) (CBT) or [Risky Code Changes](#) (RCC) DTP Enterprise Pack artifact. This is to reduce the potential performance impact associated with the running the artifacts. If you want to use these artifacts, you must enable coverage data processing in the ReducedCoverageConfig.xml configuration file located in the DTP_HOME/conf/ directory. This file enables you to specify a list of coverage images that you want Data Collector to process.

Coverage Report Size and Volume Affect Processing Time

If your coverage report files exceed 1 GB and you send multiple reports per build, you may experience an extra five minutes per GB in Data Collector processing time.

1. Open the file and set the value of the `type` parameter in the `<images>` element to either `accept` or `reject`:

```
<images type="accept">
```

Setting the parameter to `accept` means that Data Collector will *only* accept coverage data for the coverage images in the list. Setting the parameter to `reject` means Data Collector will accept coverage data for *all* coverage images *except* for those in the list.

2. Specify the coverage images to accept or reject by adding them in `<image>` elements within the `<images>` element:

```
<images type="accept">
  <image>Example Coverage</image>
  <image>Test Coverage</image>
</images>
```

3. Save the file.

See the XML file for more examples of `accept` or `reject` settings.