

Troubleshooting DAS Installation 1

Troubleshooting DAS Installation

Date of Publish: 2019-03-20



<https://docs.hortonworks.com>

Contents

Troubleshooting DAS installation.....	3
Problem area: Queries page.....	3
Your queries are not appearing on the Queries page.....	3
Query column is empty, yet you can see the DAG ID and Application ID.....	4
Query column is not empty, but you cannot see the DAG ID and Application ID.....	4
You cannot view queries from other users.....	4
Problem area: Compose page.....	4
You cannot see your databases or the query editor is missing.....	4
You cannot view new databases and tables, or cannot see changes to existing databases or tables.....	5
Replication failure in the DAS Event Processor.....	5
Problem area: Reports page.....	6
You cannot view the Reports tab.....	6
DAS service installation fails with the “python files missing” message.....	6
DAS does not log me out as expected, or I stay logged in longer than the time specified in the Ambari configuration.....	6

Troubleshooting DAS installation

This section provides you a list of possible causes and solutions to debug and resolve issues that you might face while using DAS.

Problem area: Queries page

Issues related to queries can arise mostly because of misconfigured Hive or Tez hooks. You may also experience trouble viewing your queries if you are a non-admin user. Non-admin users can view only their queries.

Your queries are not appearing on the Queries page

Check the role that you have been assigned. If you are not an administrator, then you can only view the queries that you have run. Only administrators can view queries and data of other users.

- To enlist yourself as the admin user, go to **Ambari > Data Analytics Studio > Config > Advanced data_analytics_studio-security-site**, and add your user name in the admin_users field. For example, hive,admin1,admin2.
- Check the configuration settings for Hive and Tez in the Ambari UI. See [Check Cluster Configuration for Hive and Tez](#).
- Verify whether the logs are added to the paths specified in the hive.hook.proto.base-directory and the tez.history.logging.proto-base-dir directory configurations.

Use the following commands to get the paths to the logs for Hive and Tez. If you have not run any queries, then the log directories may be empty.

For Hive:

```
hdfs dfs -ls /warehouse/tablespace/external/hive/sys.db/query_data
```

For Tez:

```
hdfs dfs -ls /warehouse/tablespace/external/hive/sys.db/dag_data
```

```
hdfs dfs -ls /warehouse/tablespace/external/hive/sys.db/dag_meta
```

```
hdfs dfs -ls /warehouse/tablespace/external/hive/sys.db/app_data
```

- Verify whether the following proto directory configurations are present in the /etc/das/conf/das-event-processor.json file: hive.hook.proto.base-directory and tez.history.logging.proto-base-dir. If these configurations are not present in the /etc/das/conf/das-event-processor.json file, then do not edit the das-event-processor.json file - because they will be overwritten by the configurations specified in Ambari.

At this stage, restart the DAS service and recheck. If this doesn't fix the problem, then you might be using multiple configuration groups in Hive and Tez.

If you are using different or multiple configuration groups in Hive or Tez, then make sure that you specify the same paths in all the configuration groups. Presently, DAS does not support multiple configuration groups.

- Verify whether you have the required access and permission to directories specified in the hive.hook.proto.base-directory and tez.history.logging.proto-base-dir directory configurations by checking errors from EventProcessorPipeline in the /var/log/das/event-processor.log file.
- Make sure that the DAS services are running as the hive service user.
- If you have a large number of users running Hive queries simultaneously using DAS, the connections from Hive can get exhausted. This sends the query to a queue in Hive which takes a long time to start running. This can result in the following:
 - If query is fired from DAS, it does not return any log or indicate that the query has been queued.
 - The query does not appear on the UI, because no event is received by DAS.

Evaluate your cluster usage. You may need to add more machines to the clusters to resolve this issue.

Query column is empty, yet you can see the DAG ID and Application ID

- Check the configuration settings for Hive and Tez in the Ambari UI. See [Check Cluster Configuration for Hive and Tez](#).
- Check whether Hive has access to write in to the directory specified by the following proto directory configuration: `hive.hook.proto.base-directory`.
- Check whether Hive can write in to the directory specified by the following proto directory configuration: `hive.hook.proto.base-directory`.
- Check whether Hive has access to read from the directory specified by the following proto directory configuration: `hive.hook.proto.base-directory`.

Query column is not empty, but you cannot see the DAG ID and Application ID

- Check the configuration settings for Hive and Tez in the Ambari UI. See [Check Cluster Configuration for Hive and Tez](#).
- Check whether the Tez job has access to write in to the directory specified by the following proto directory configuration: `tez.history.logging.proto-base-dir`.
- Check whether the data is being written in to the following subdirectories of the directory specified in the `tez.history.logging.proto-base-dir` directory configuration: `app_data`, `dag_data`, and `dag_meta`.
- Check whether das has access to read from the following subdirectories of the directory specified in the `tez.history.logging.proto-base-dir` directory configuration: `app_data`, `dag_data`, and `dag_meta`.

You cannot view queries from other users

If you need to view queries and data of other users, then assign admin role to the user by going to **Ambari > Data Analytics Studio > Config > Advanced data_analytics_studio-security-site**, and specifying the user in the `admin_users` field. For example, `hive,admin1,admin2`.

Problem area: Compose page

Misconfigured repl in Hive can affect the query compose page on the DAS UI. Issues can also occur if HSI or HS2 is not running.

You cannot see your databases or the query editor is missing

- Verify whether HSI or HS2 services are running by going to **Ambari > Hive > Summary**.
- If you are not running llap, then go to **Ambari > Data Analytics Studio > Config > Advanced data_analytics_studio-properties > Data Analytics Studio config file template** and set the `use.hive.interactive.mode` parameter to `false`.
- Verify whether the Hive configuration parameter `hive.metastore.transactional.event.listeners` is set to `org.apache.hive.hcatalog.listener.DbNotificationListener` by going to **Ambari > Hive > Configs > Advanced > Advanced hive-site**.
- DAS event processor runs the `repl dump` command as a hive service user. The hive service user can run the `repl dump` command if the user has `repl admin` privileges.
- Verify whether DAS can read the `repl root` directory permissions - if das and hive are running as different service users.

There may be exceptions in the `/var/log/das/event-processor.log` file with `repl dump` failures because of read errors.



Note: If the das services cannot read the `repl root` directory permissions, then it can cause a number of `repl` directories to be created, but not deleted.

- Lastly, this can be caused due to incorrect hive service name in ZooKeeper. This is a known issue as of the DAS 1.2.0 release.

You cannot view new databases and tables, or cannot see changes to existing databases or tables

- Verify whether the hive configuration parameter `hive.metastore.transactional.event.listeners` is set to `org.apache.hive.hcatalog.listener.DbNotificationListener` by going to **Ambari > Hive > Configs > Advanced > Advanced hive-site**.
- Verify whether HSI or HS2 services are running by going to **Ambari > Hive > Summary**.
- This can also happen if the Hive metadata replication fails. To troubleshoot replication-related failures, see [Replication failure in the DAS Event Processor](#).

Replication failure in the DAS Event Processor

DAS uses replication as a way to copy database and table metadata information from Hive to DAS Postgres database. If the replication fails, then you may not be able to see database or table information in DAS.

The Hive metadata replication process occurs in the following two phases:

- Bootstrap dump
- Incremental dump

When the DAS Event Processor is started for the first time, the entire Hive database and table metadata is copied in to DAS. This is known as the bootstrap dump. After this phase, only the differences are copied in to DAS at one-minute intervals, from the time the last successful dump was run. This is known as an incremental dump.

If the bootstrap dump never succeeded, then you will not see any database or table information in DAS. If the bootstrap dump fails, then the information regarding the failure is captured in the most recent Event Processor log.

If an incremental dump fails, then you will not see any new changes to the databases and tables in DAS. The incremental dump relies on events stored in the Hive metastore, because these events take up a lot of space and are only used for replicating data. The events are removed from Hive metastore daily, which can affect DAS.

Fixing incremental dump failures

If you see the message “Notification events are missing in the meta store”, then:

1. Stop the DAS Event Processor.
2. Log in to the DAS Postgres database.
3. Re-initiate the bootstrap dump by running the following commands:

```
update das.databases set creation_source = 'EVENT_PROCESSOR';
update das.tables set creation_source = 'EVENT_PROCESSOR';
update das.columns set creation_source = 'EVENT_PROCESSOR';
delete from das.db_replication_info;
```

4. Restart the DAS Event Processor.



Note:

The error message "Notification events are missing in the meta store" is a symptom and not the cause of an issue, at most times. You can usually fix this by resetting the database. However, to determine the actual cause of the failure, we need to look at the first repl dump failure. Older Event Processor logs may contain the information about the actual cause, but they are purged from the system at a set frequency. Therefore, if you hit this issue, we recommend that you first copy all the Event Processor logs that are available; else it might be too late to diagnose the actual issue.

If the first exception that you hit is an `SQLException`, then it is a Hive-side failure. Save the HiveServer and the Hive Metastore logs for the time when the exception occurred.

File a bug with Cloudera Support along with the above-mentioned logs.

Problem area: Reports page

The Reports tab is not available for DAS-Lite. It may also not appear if you have not specified the DataPlane Smartsense ID in Ambari.

You cannot view the Reports tab

- Set up the SmartSense ID in DAS configurations.
 1. On the Ambari UI, go to **Data Analytics Studio > Configs**.
 2. Under **Advanced data_analytics_studio-webapp-properties**, specify `das-smartsense-id` in the **Dataplane Smartsense id** field.
 3. Restart the DAS service.
- Verify whether you are running DAS and not DAS-Lite by using the following commands:
 - (On Debian and Ubuntu): `apt list --installed '*data-analytics*'`
 - (On RHEL and Fedora): `rpm -qa | grep 'data-analytics'`

DAS service installation fails with the “python files missing” message

The installation might fail while you are adding the DAS service through the Ambari UI with an error such as the following: Python files are missing.

Procedure

1. SSH in to the Ambari host as a root user.
2. Set the `agent.auto.cache.update` property to true in the `/etc/ambari-server/conf/ambari.properties` file.
3. Restart ambari using the following command:

```
ambari-server restart
```

DAS does not log me out as expected, or I stay logged in longer than the time specified in the Ambari configuration

The login session timeout factor in DAS is governed by the following two properties:

- Session timeout in seconds (`data_analytics_studio_session_timeout`) under **Ambari > Data Analytics Studio > CONFIGS > Advanced data_analytics_studio-webapp-properties**
- `knoxsso.token.ttl` under **Ambari > Knox > CONFIG > Advanced knoxsso-topology**

The `data_analytics_studio_session_timeout` property is a DAS-specific configuration which is used to define the time period for which you want DAS to keep you logged in. This, particularly, comes in handy when you have a long running query. If a user is running a query which runs for a long time, say more than 24 hours, and if the user gets logged out before the query execution finishes, the user may lose the query results. To prevent this issue, you (the DAS admin) must configure the time period in the Session timeout in seconds field accordingly. By default, DAS does not log you out for 24 hours or 86400 seconds after you log in.

The `knoxsso.token.ttl` property is cluster-wide configuration which applies to all the services in the cluster that use Knox SSO. The logout time is given in milliseconds.

If the user is actively accessing DAS, then he will not be logged out. However, if the user is inactive, then the session timeout is equal to the value specified in either the **Session timeout in seconds** field in Ambari or in the `knoxsso.token.ttl` property - whichever is higher.

For example, if you have set 5 minutes in the `knoxsso.token.ttl` property and 86400 seconds (24 hours) in the **Session timeout in seconds** field, then the user will remain logged into DAS for 24 hours, or until the user is actively using DAS.