

# Oracle® Fusion Middleware

## Oracle Stream Analytics Install Guide for Hadoop 2.7 and Higher



E93122-03  
April 2019



Oracle Fusion Middleware Oracle Stream Analytics Install Guide for Hadoop 2.7 and Higher,

E93122-03

Copyright © 2018, 2019, Oracle and/or its affiliates. All rights reserved.

Primary Author: Oracle Corporation

This software and related documentation are provided under a license agreement containing restrictions on use and disclosure and are protected by intellectual property laws. Except as expressly permitted in your license agreement or allowed by law, you may not use, copy, reproduce, translate, broadcast, modify, license, transmit, distribute, exhibit, perform, publish, or display any part, in any form, or by any means. Reverse engineering, disassembly, or decompilation of this software, unless required by law for interoperability, is prohibited.

The information contained herein is subject to change without notice and is not warranted to be error-free. If you find any errors, please report them to us in writing.

If this is software or related documentation that is delivered to the U.S. Government or anyone licensing it on behalf of the U.S. Government, then the following notice is applicable:

U.S. GOVERNMENT END USERS: Oracle programs, including any operating system, integrated software, any programs installed on the hardware, and/or documentation, delivered to U.S. Government end users are "commercial computer software" pursuant to the applicable Federal Acquisition Regulation and agency-specific supplemental regulations. As such, use, duplication, disclosure, modification, and adaptation of the programs, including any operating system, integrated software, any programs installed on the hardware, and/or documentation, shall be subject to license terms and license restrictions applicable to the programs. No other rights are granted to the U.S. Government.

This software or hardware is developed for general use in a variety of information management applications. It is not developed or intended for use in any inherently dangerous applications, including applications that may create a risk of personal injury. If you use this software or hardware in dangerous applications, then you shall be responsible to take all appropriate fail-safe, backup, redundancy, and other measures to ensure its safe use. Oracle Corporation and its affiliates disclaim any liability for any damages caused by use of this software or hardware in dangerous applications.

Oracle and Java are registered trademarks of Oracle and/or its affiliates. Other names may be trademarks of their respective owners.

Intel and Intel Xeon are trademarks or registered trademarks of Intel Corporation. All SPARC trademarks are used under license and are trademarks or registered trademarks of SPARC International, Inc. AMD, Opteron, the AMD logo, and the AMD Opteron logo are trademarks or registered trademarks of Advanced Micro Devices. UNIX is a registered trademark of The Open Group.

This software or hardware and documentation may provide access to or information about content, products, and services from third parties. Oracle Corporation and its affiliates are not responsible for and expressly disclaim all warranties of any kind with respect to third-party content, products, and services unless otherwise set forth in an applicable agreement between you and Oracle. Oracle Corporation and its affiliates will not be responsible for any loss, costs, or damages incurred due to your access to or use of third-party content, products, or services, except as set forth in an applicable agreement between you and Oracle.

# Contents

	<b>Preface</b>	
	<hr/>	
	Audience	iv
	Documentation Accessibility	iv
	Conventions	iv
	Related Documents	v
<b>1</b>	<b>How to Install Oracle Stream Analytics?</b>	
	<hr/>	
	Planning Your Installation	1-1
	Installing Oracle Stream Analytics	1-2
	Configuring the Default User	1-3
	Upgrading Oracle Stream Analytics	1-3
		1-4
<b>2</b>	<b>Post Installation Procedures</b>	
	<hr/>	
	Initializing Metadata Store	2-1
	Changing the Admin Password	2-2
<b>3</b>	<b>Launching and Terminating Oracle Stream Analytics</b>	
	<hr/>	
	Launching Oracle Stream Analytics	3-1
	Setting up Runtime for Oracle Stream Analytics Server	3-2
	Validating Data Flow to Oracle Stream Analytics	3-3
	Terminating Oracle Stream Analytics	3-6

# Preface

Installing Oracle Stream Analytics for Hadoop 2.7 and Higher describes the prerequisites for installing Oracle Stream Analytics, steps to install it, configure it, launch it, and terminate it.

## Topics

- [Audience](#)
- [Documentation Accessibility](#)
- [Conventions](#)
- [Related Documents](#)

## Audience

This document is intended for developers and users who need to install and configure Oracle Stream Analytics.

## Documentation Accessibility

For information about Oracle's commitment to accessibility, visit the Oracle Accessibility Program website at <http://www.oracle.com/pls/topic/lookup?ctx=acc&id=docacc>.

### Accessible Access to Oracle Support

Oracle customers who have purchased support have access to electronic support through My Oracle Support. For information, visit <http://www.oracle.com/pls/topic/lookup?ctx=acc&id=info> or visit <http://www.oracle.com/pls/topic/lookup?ctx=acc&id=trs> if you are hearing impaired.

## Conventions

The following text conventions are used in this document:

Convention	Meaning
<b>boldface</b>	Boldface type indicates graphical user interface elements associated with an action, or terms defined in text or the glossary.
<i>italic</i>	Italic type indicates book titles, emphasis, or placeholder variables for which you supply particular values.
monospace	Monospace type indicates commands within a paragraph, URLs, code in examples, text that appears on the screen, or text that you enter.

## Related Documents

Documentation for Oracle Stream Analytics is available on [Oracle Help Center](#).

Also see the following documents for reference:

- *Understanding Oracle Stream Analytics*
- *Developing Custom Stages and Functions*
- *Known Issues in Oracle Stream Analytics*
- *Spark Extensibility for CQL in Oracle Stream Analytics*
- *Using Oracle Stream Analytics*

# 1

## How to Install Oracle Stream Analytics?

Oracle Stream Analytics is a very advanced, scalable, reliable, and flexible platform that helps you in designing event-processing applications very efficiently. It caters to a wide variety of industries like Telecom, Financial Services, Transportation, Retail, Insurance, and so on. You can use Oracle Stream Analytics for Real-time Fraud Detection, Real-time Location-based Marketing, Real-time Smart Inventory Management, Real-time Dynamic Pricing, Real-time Asset Tracking, and so on. This document helps you to plan and execute the installation of Oracle Stream Analytics in a seamless manner.

### Topics:

- [Planning Your Installation](#)
- [Installing Oracle Stream Analytics](#)
- [Upgrading Oracle Stream Analytics](#)

## Planning Your Installation

To plan the installation of Oracle Stream Analytics 18.1.0.0.1 efficiently, ensure that you have the required hardware and software. You should also perform the prerequisite procedures before starting the installation process.

You can use the information in the [certification matrix](#) before installing Oracle Stream Analytics 18.1.0.0.1. The certification matrix provides you useful links to support pages, supported software, and system requirements in general. The following software is required for operation of Oracle Stream Analytics:

- Oracle JDK 8 Update 131 and higher versions
- Repository Database
  - Oracle Database versions 12.2.0.1 or higher, 12.1.0.1 or higher, and 11.2.0.4 or higher
  - Else, you can use MySQL version 5.6 or 5.7
- A running Hadoop cluster with version 2.6 or 2.7
- A running Kafka cluster with version 0.10.2.0
- Locally installed Spark with version 2.2.1 built for Hadoop 2.7

### Note:

Install Spark and JDK in the same node on which you plan to install Oracle Stream Analytics. See [Installing Oracle Stream Analytics](#).

- Google Chrome browser with version 60 or higher

# Installing Oracle Stream Analytics



This topic applies only to Oracle user-managed services.

After you ensure that you have procured the required software and hardware and completed the procedures to set up the required environment for installation, use the following steps to install Oracle Stream Analytics 18.1.0.0.1:

1. Create a directory, for example, `spark-downloads`, and [download](#) Apache Spark into the newly created folder as specified below:

- Spark release: 2.2.1
- Package type: Pre-built for Apache Hadoop 2.7 and later
- Download Spark: `spark-2.2.1-bin-hadoop2.7.tgz`

2. Extract the Spark archive to a local directory.

You can see a subfolder, `spark-2.2.1-bin-hadoop2.7`.

3. Create a new directory, for example, `OSA-downloads` and [download](#) `OSA-18.1.0.0.1.zip` from Oracle [eDelivery](#) and extract it into the newly created folder.

You can find the `OSA-18.1.0.0.1-README.txt` file in the `OSA-18.1.0.0.1` zip file.

4. Review the file `OSA-18.1.0.0.1-README.txt`.

5. Set the environment variables:

- a. Set the `SPARK_HOME` environment variable in the `OSA-18.1.0.0.0/osa-base/etc/osa-env.sh` file to point to the directory where you have extracted the Spark archive. For example:

```
SPARK_HOME=/products/spark-downloads/spark-2.2.1-bin-hadoop2.7
```

- a. Set the `JAVA_HOME` environment variable in the `OSA-18.1.0.0.0/osa-base/etc/osa-env.sh` file to point to the directory where you have extracted the JDK archive. For example:

```
JAVA_HOME=/products/java-downloads/jdk1.8.0_131
```

6. Configure your data source in `OSA-18.1.0.0.1/osa-base/etc/jetty-osa-datasource.xml` as per instructions in [Initializing Metadata Store](#).

- a. Uncomment and edit one of the two Data source configurations, either for Oracle Database or MySQL depending on the database you want to use as metadata store. The uncommented fragment for Oracle database is shown below:

```
<New id="osads" class="org.eclipse.jetty.plus.jndi.Resource">  
  <Arg>  
    <Ref refid="wac"/>  
  </Arg>  
  <Arg>jdbc/OSADatasource</Arg>  
  <Arg>  
    <New class="oracle.jdbc.pool.OracleDataSource">  
      <Set name="URL">jdbc:oracle:thin:@myhost.example.com:
```

```

1521:OSADB</Set>
  <Set name="User">{OSA_USER}</Set>
  <Set name="Password">{OSA_USER_PWD}</Set>
  <Set name="connectionCachingEnabled">true</Set>
  <Set name="connectionCacheProperties">
    <New class="java.util.Properties">
      <Call name="setProperty"><Arg>MinLimit</Arg><Arg>1</
Arg></Call>
      <Call name="setProperty"><Arg>MaxLimit</Arg><Arg>15</
Arg></Call>
      <Call name="setProperty"><Arg><InitialLimit</
Arg><Arg>1</Arg></Call>
    </New>
  </Set>
</New>
</Arg>
</New>

```

- b. Change the database host, port, SID, Oracle Stream Analytics schema user name and password fields marked in bold in the code in step 6a.

## Configuring the Default User

osaadmin user is predefined user with administrative privileges in Oracle Stream Analytics. You cannot delete the default user.

The installation directory has two scripts: `<installRoot>/osa-base/bin/start-osa.sh` and `<installRoot>/osa-base/bin/configure.sh`. The first script invokes `configure.sh` and then starts the Jetty server. If you do not need to start the server, but only need to configure the default user, then run the `configure.sh` script. The list of available parameters for both the scripts is given below:

- `help` - show help content
- `dbroot` - data base root user, or any user who has previligies to create db schema, db user
- `dbroot_password` - password of the dbroot user
- `osaadmin_password` - password of the predefined admin user: `osaadmin`

You do not need to specify `osaadmin_password` while creating a new schema. The script automatically detects that the `osaadmin` user is not created and prompts for `osaadmin_password`.

While installing Oracle Stream Analytics, configure the `osaadmin` user with an appropriate password by specifying a value for the `osaadmin_password` parameter. You can change the password later using the Oracle Stream Analytics application user interface.

## Upgrading Oracle Stream Analytics



This topic applies only to Oracle user-managed services.

If you have an existing Oracle Stream Analytics 18.1.0.0.0 installation, use the following steps to upgrade to Oracle Stream Analytics 18.1.0.0.1:

### Metadata Upgrade

Use these steps if there is no additional configuration apart from the data source configuration and environment variables:

1. Backup the existing Oracle Stream Analytics 18.1.0.0.0 metadata database, for example, use Oracle Database Backup tool.
2. Stop Oracle Stream Analytics 18.1.0.0.0.
3. Unzip Oracle Stream Analytics 18.1.0.0.1 installer.
4. Copy the data source configuration from existing installation to the new installation:

```
cp OSA-18.1.0.0.0/osa-base/etc/jetty-osa-datasource.xml OSA-18.1.0.0.1/
osa-base/etc/jetty-osa-datasource.xml
```

5. Copy the environment variable `osa-env.sh` from existing installation to the new installation:

```
cp OSA-18.1.0.0.0/osa-base/etc/osa-env.sh OSA-18.1.0.0.1/osa-base/etc/
osa-env.sh
```

### Upgrade Using the Existing Jetty Configurations

Use these steps if you want to upgrade using the existing Jetty configurations:

1. Backup the existing Oracle Stream Analytics 18.1.0.0.0 metadata database, for example, use Oracle Database Backup tool.
2. Stop Oracle Stream Analytics 18.1.0.0.0.
3. Unzip Oracle Stream Analytics 18.1.0.0.1 installer.
4. Use the following set of commands:

```
#!/bin/bash
OSA_V18.1.0.0.0=<refers to the existing installation that you would
like to update>
OSA_V18.1.0.0.1=<refers to the unzipped OSA 18.1.0.0.1 installer>

cp ${OSA_V18.1.0.0.1}/osa-base/bin/osa.installer.jar ${OSA_V18.1.0.0.0}/
osa-base/bin/osa.installer.jar
cp ${OSA_V18.1.0.0.1}/osa-base/extensibility-api/osa.spark-
cql.extensibility.api.jar ${OSA_V18.1.0.0.0}/osa-base/extensibility-api/
osa.spark-cql.extensibility.api.jar
cp ${OSA_V18.1.0.0.1}/osa-base/lib/ext/osa.web.jetty-session.jar $
{OSA_V18.1.0.0.0}/osa-base/lib/ext/osa.web.jetty-session.jar

cp ${OSA_V18.1.0.0.1}/osa-base/resources/libs/wlthint3client.jar $
{OSA_V18.1.0.0.0}/osa-base/resources/libs/wlthint3client.jar
cp ${OSA_V18.1.0.0.1}/osa-base/resources/libs/spark-archive.jar $
```

```
{OSA_V18.1.0.0.0}/osa-base/resources/libs/ #spark-acrhive.jar is a new
file added in Oracle Stream Analytics 18.1.0.0.1 installer

#comment below line if you want to keep existing logging configuration
cp ${OSA_V18.1.0.0.1}/osa-base/resources/log4j2.xml ${OSA_V18.1.0.0.0}/
osa-base/resources/log4j2.xml

cp ${OSA_V18.1.0.0.1}/osa-base/resources/modules/spark-osa.jar $
{OSA_V18.1.0.0.0}/osa-base/resources/modules/spark-osa.jar
cp ${OSA_V18.1.0.0.1}/osa-base/sql/*.sql ${OSA_V18.1.0.0.0}/osa-base/
sql/

rm -f ${OSA_V18.1.0.0.0}/osa-base/start.d/logging-log4j2.ini

cp ${OSA_V18.1.0.0.1}/osa-base/webapps/osa.web.*.war ${OSA_V18.1.0.0.0}/
osa-base/webapps/
cp ${OSA_V18.1.0.0.1}/osa-base/webapps/osa.web.admin.xml $
{OSA_V18.1.0.0.0}/osa-base/webapps/
#osa.web.admin.xml is a new file added in Oracle Stream Analytics
18.1.0.0.1 installer

cp ${OSA_V18.1.0.0.1}/osa-base/version.txt ${OSA_V18.1.0.0.0}/osa-base/
version.txt
```

# 2

## Post Installation Procedures

You are done with the installation of Oracle Stream Analytics. Now, you must complete some post installation tasks before you start using Oracle Stream Analytics.

### Topics

- [Initializing Metadata Store](#)
- [Changing the Admin Password](#)

## Initializing Metadata Store



This topic applies only to Oracle user-managed services.

After installing Oracle Stream Analytics, you need to configure the metadata store with the database admin credential details and the version of Oracle Stream Analytics as required.

### Note:

If you do not have the database admin credentials, ask your database administrator to create a Oracle Stream Analytics database user and initialize the content under the user by using the SQL scripts available in the `OSA-18.1.0.0.1/osa-base/sql` folder. The Oracle Stream Analytics database username must match the one configured in `jetty-osa-datasource.xml`.

To initialize the metadata store, you need database admin credentials with sysdba privileges:

1. Change directory to `OSA-18.1.0.0.1/osa-base/bin`.
2. Execute the following command:

```
./start-osa.sh dbroot=<root user> dbroot_password=<db root password>
```

### Note:

Ensure that you replace the database root username and password as required.

The following console messages indicates that the Oracle Stream Analytics schema is created and the metadata store is successfully initialized:

```
OSA DB user created: <User>  
OSA schema version: 18.1.0.0.1
```

 **Note:**

If you don't see the above messages, check the `OSA-18.1.0.0.1/osa-base/logs` folder to identify the cause and potential solution.

3. Press enter on the console and run the `OSA-18.1.0.0.1/osa-base/bin/stop-osa.sh` file to complete the metadata store initialization.

## Changing the Admin Password



This topic applies only to Oracle user-managed services.

Oracle Stream Analytics comes with a default user, `osaadmin`. You must create your own password to login by using that user ID.

Use the following steps to change the password:

1. Change directory to top-level folder `OSA-18.1.0.0.1` and execute the following command:

```
java -cp ./lib/jetty-util-9.4.7.v20170914.jar  
org.eclipse.jetty.util.security.Password osaadmin <your new password>
```

You can see the following codes in the console:

```
2018-03-01 15:37:55.800:INFO::main: Logging initialized @117ms to  
org.eclipse.jetty.util.log.StdErrLog  
alphago  
OBf:1u9d1uv81uha1u9rlugg1uuy1ua5  
MD5:34d0a556209df571d311b3f41c8200f3  
CRYPT:osX/8jafUvLwA
```

2. Make a note of the obfuscated password string, marked in bold in the code in step 1, and copy it to a notepad.
3. Connect to the Oracle Stream Analytics metadata store database by using your SQL tool such as SQL Plus or SQL Developer.

The credentials must match the ones provided in the `OSA-18.1.0.0.1/osa-base/etc/jetty-osa-datasource.xml` file.

4. Update the `osa_users` table by using the command below and replace the obfuscated password with the string that you copied in step 2:

```
update osa_users set pwd='<CopiedObfuscatedPasswordString>' where  
username='osaadmin'
```

 **Note:**

This on-premise version of Oracle Stream Analytics doesn't support role based access. All users have admin privileges. To create additional users with obfuscated passwords, see [Adding Users](#).

# 3

## Launching and Terminating Oracle Stream Analytics

Once you have completed the post installation tasks, you are ready to launch Oracle Stream Analytics, and start using it. Launching and terminating Oracle Stream Analytics is easy and you just need to run simple commands to do them.

### Topics

- [Launching Oracle Stream Analytics](#)
- [Setting up Runtime for Oracle Stream Analytics Server](#)
- [Validating Data Flow to Oracle Stream Analytics](#)
- [Terminating Oracle Stream Analytics](#)

## Launching Oracle Stream Analytics

After you have installed Oracle Stream Analytics, the next step is to start the Oracle Stream Analytics server that will launch the application.

To launch Oracle Stream Analytics and start the server:

1. Change directory to `OSA-18.1.0.0.1/osa-base/bin` and run `./start-osa.sh`.

You can see the following message on the console:

```
OSA schema version: 18.1.0.0.1
The schema is preconfigured and current. No changes or updates are
required.
```



### Note:

If you don't see this message, check the log file in the `OSA-18.1.0.0.1/osa-base/logs` folder.

2. To ensure that the server is up and running, go to Chrome browser and type `localhost:9080/osa` in the address bar to launch Oracle Stream Analytics.
3. Login using your default admin credentials.

If you do not see the home page, check the logs in the `<OSA-INSTALL-DIR>/osa-base/logs` folder.

# Setting up Runtime for Oracle Stream Analytics Server

Before you start using Oracle Stream Analytics, you need to specify the runtime server, environment, and node details. You must do this procedure right after you launch Oracle Stream Analytics for the first time.

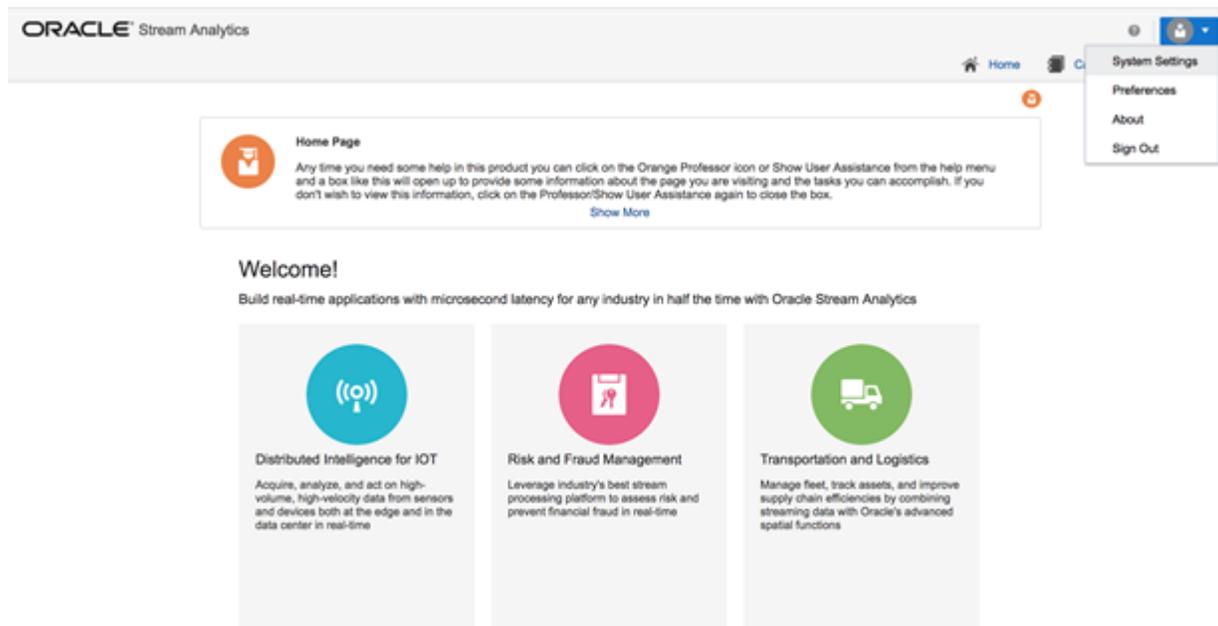
Do the following to set up runtime for Oracle Stream Analytics:

1. In the Chrome browser, enter `localhost:9080/osa` to access Oracle Stream Analytics' login page, and login using your credentials.

 **Note:**

The password is a plain-text password.

2. Click the user profile at the top right corner, and then click **System Settings**, as shown in the following figure:



3. In the Environment tab of the System Settings dialog box, ensure the following:

- a. In the **Kafka ZooKeeper Connection** field, specify the host and the port where Kafka Zookeeper is running.  
The Kafka Zookeeper connection is required to see live results while designing pipelines.
- b. Select **Yarn** as the **Runtime Server** and in the **YARN Resource Manager URL** field, specify the host name and port of the Hadoop cluster where the YARN resource manager is running.
- c. Select **WebHDFS** as **Storage** and in the **path webhdfs://** field, specify the host name of Hadoop cluster where the Name Node is running along with a root folder, for example spark-deploy (as shown in the figure above).  
If the folder does not exist, it will automatically be created, but the user specified in the Hadoop authentication below must have write permissions.
- d. Select **Simple** as **Hadoop Authentication** because **Kerberos** is not supported in this version of Oracle Stream Analytics.  
The specified **Username** must have write permissions for the folder specified in the webhdfs:// URL.

## Validating Data Flow to Oracle Stream Analytics

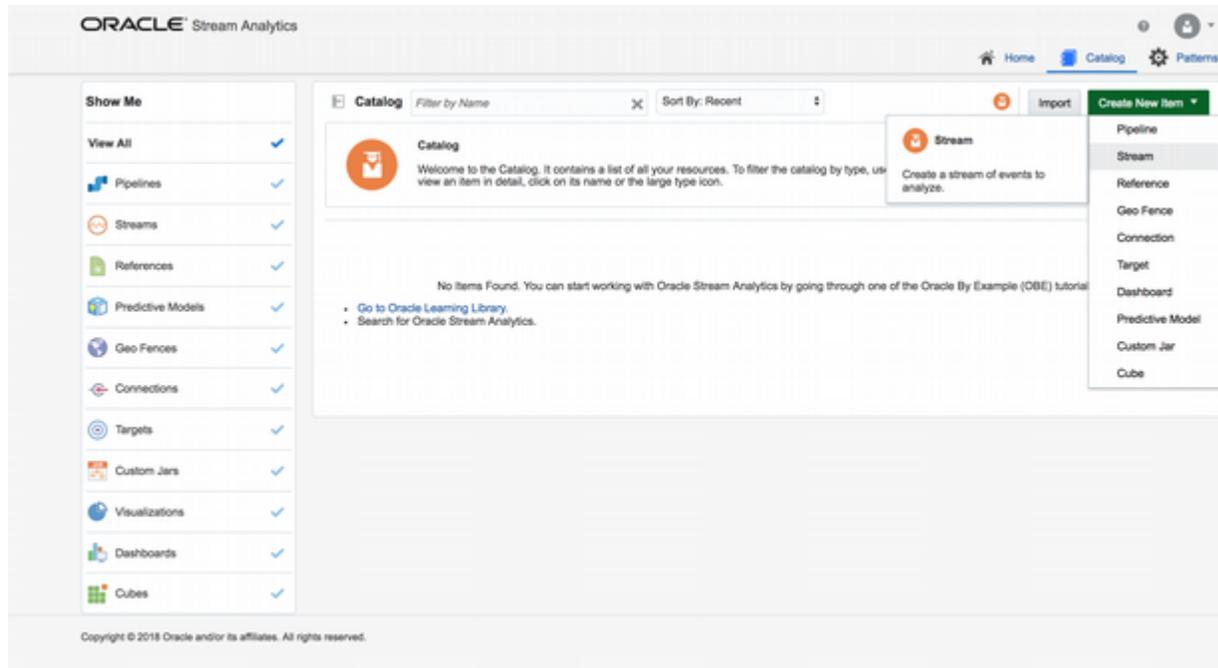
After you have configured Oracle Stream Analytics with the runtime details, you need to ensure that sample data is being detected and correctly read by Oracle Stream Analytics

To validate data flow into Oracle Stream Analytics, use the following steps:

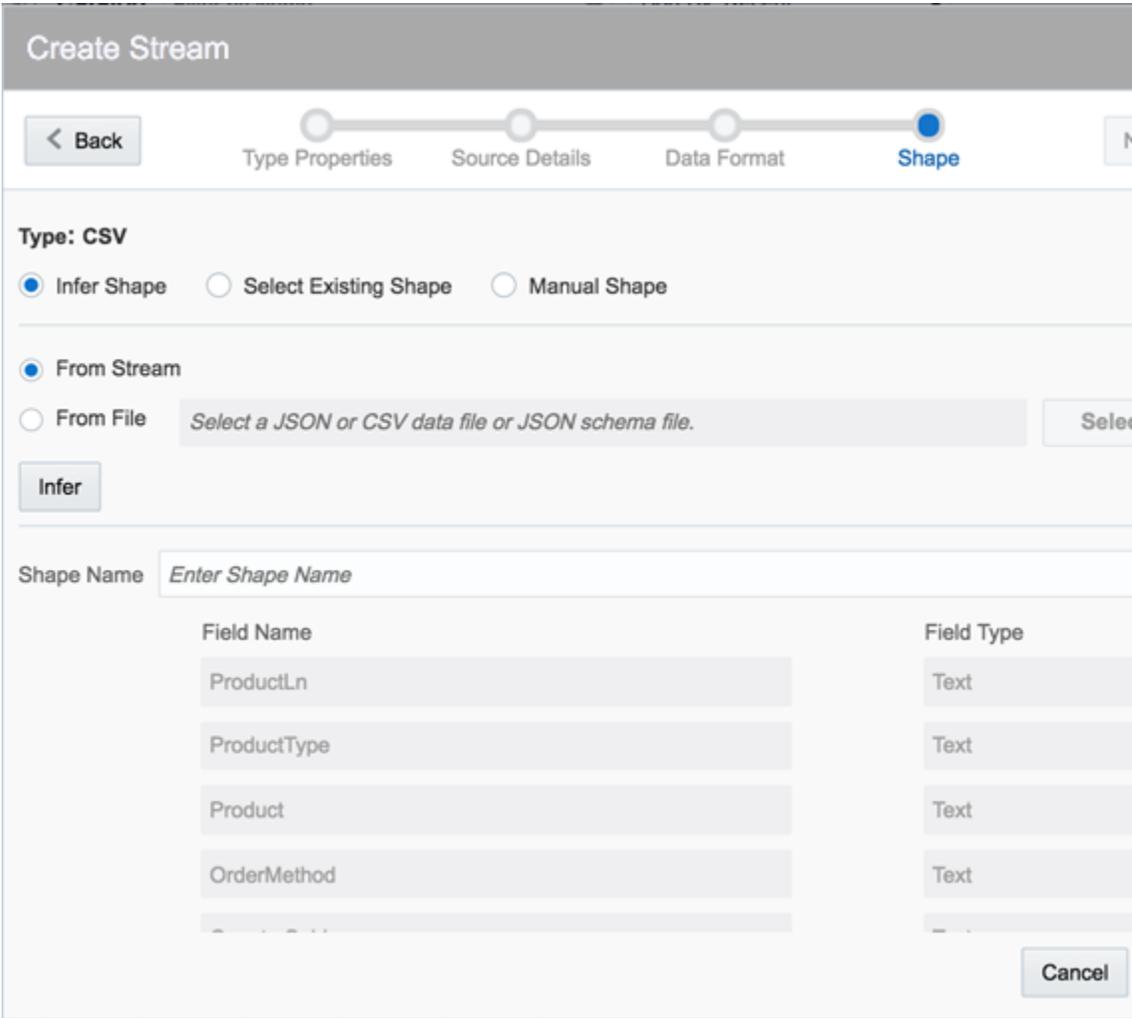
1. Copy the six lines below into a CSV file, for example `sample.csv`.

```
ProductLn,ProductType,Product,OrderMethod,CountrySold,QuantitySold,UnitsS
alePrice
Personal Accessories,Watches,Legend,Special,Brazil,1,240
Outdoor Protection,First Aid,Aloe Relief,E-mail,United States,3,5.23
Camping Equipment,Lanterns,Flicker Lantern,Telephone,Italy,3,35.09
Camping Equipment,Lanterns,Flicker Lantern,Fax,United States,4,35.09
Golf Equipment,Irons,Hailstorm Steel Irons,Telephone,Spain,5,461
```

- In the **Catalog**, as shown in the image below, click **Create New Item**, and then click **Stream**. create a stream of type File.



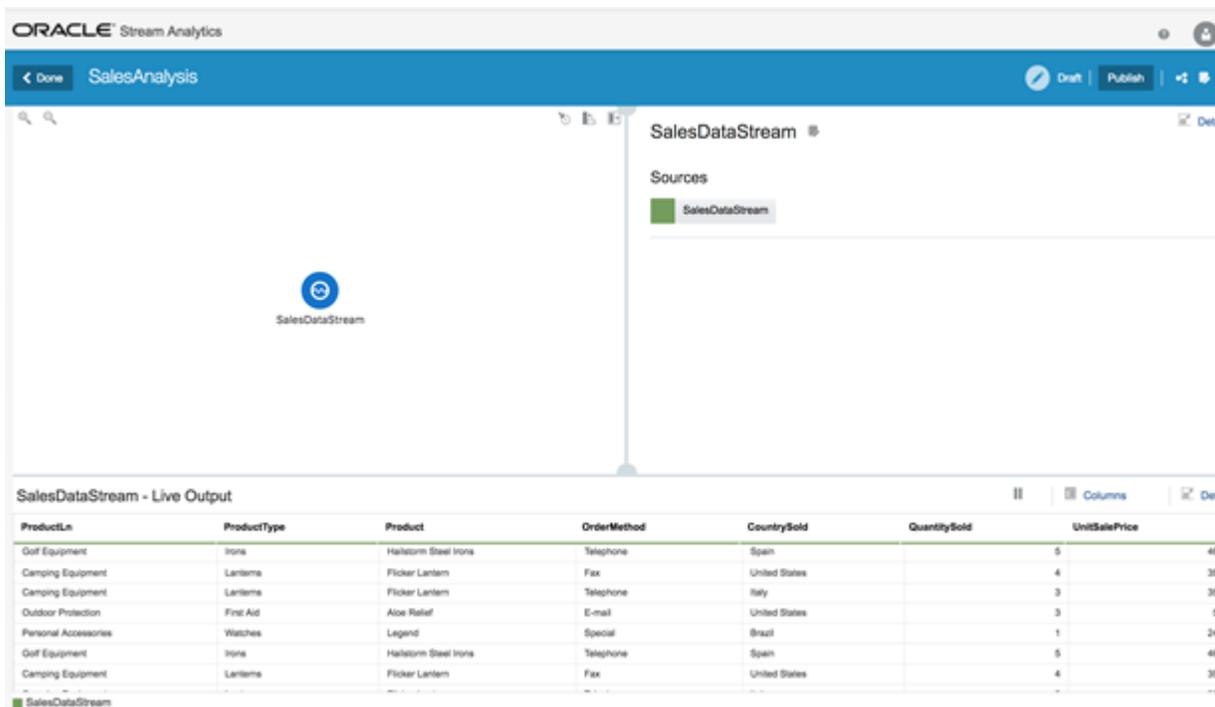
- In the **Type Properties** page of the **Create Stream** dialog box, provide the **Name**, **Description**, and **Tags** for the Stream, select the **Stream Type** as **File**, and then select **Create Pipeline with this source (Launch Pipeline Editor)**.
- Click the **Next** button to navigate to the **Source Details** page of the **Create Stream** dialog box.
- In the **Source Details** page, click **Upload file** to upload the `sample.csv` file, and then click **Next** to navigate to the **Data Format** page.
- In the **Data Format** page, select the **CSV Predefined Format** as **Default** and select the **First record as header**, and then click **Next** to navigate to the **Shape** page.
- In the **Shape** page, verify that the shape of the event is successfully inferred as in the following image, and then click **Save**.



- 8. In the **Create Pipeline** dialog box, enter the **Name**, **Description**, **Tags** of the pipeline, select the **Stream** that you created, and then click **Save**.

You can see the pipeline editor and you can see the message `Starting Pipeline` followed by the message `Listening to Events`.

 **Note:**  
This is the first access of the cluster and it takes time to copy libraries, please be patient. You should eventually see the screenshot below with single node representing the stream source.



To complete your pipeline, see [Creating a Pipeline](#).

## Terminating Oracle Stream Analytics

You can terminate Oracle Stream Analytics by running a simple command.

Use the following command to terminate Oracle Stream Analytics:

```
OSA-18.1.0.0.1/osa-base/bin/stop-osa.sh
```