

RapidMiner to DataVirtuality JDBC

Once connected, you will be able to instantly begin making models and predictions with the test data sources already set up within this test instance.

Connecting DataVirtuality Server to RapidMiner

1. Download the JDBC connector, by accessing the server via your favorite browser.
 - a. <http://localhost:8080/downloads.html>
2. Once downloaded, move the jar file into the appropriate RapidMiner repository.
 - a. For example, on Mac this is: `/Applications/RapidMiner Studio.app/Contents/Resources/RapidMiner-Studio/lib/jdbc/datavirtuality-jdbc.jar`
3. In RapidMiner Studio, select **Connections > Manage Database Drivers**, add the new driver, and populate the driver properties with the values below:
 - a. Name: **DataVirtuality**
 - b. URL: **jdbc:datavirtuality:datavirtuality@mm://**
 - c. Jar File: **/Applications/RapidMiner Studio.app/Contents/Resources/RapidMiner-Studio/lib/jdbc/datavirtuality-jdbc.jar**
 - d. (The driver class will populate automatically as: **com.datavirtuality.dv.jdbc.Driver**)
4. Click the save button and exit the manager.
5. Select **Connections > Manage Database Connections** and add a new connection.
 - a. Enter a name of your choice.
 - b. Select **DataVirtuality** from the dropdown driver selection.
 - c. Under host use the address of the server: **localhost**, with port **31000**.
 - d. In the "Advanced" panel, add the parameter "ON", to the SHOWPLAN option.
 - i. The virtual database option here should also be: **datavirtuality**.
6. Click "OK" to add the new Database Connection.
7. DataVirtuality should now successfully be integrated with RapidMiner.
 - a. To view your new connection, refresh the **DB** object, within the Repository panel.

Setting Up DataVirtuality Studio

DataVirtuality Studio is the client tool used to interface with the engine running DataVirtuality Server.

1. Open the downloads page, for your DataVirtuality Server instance.
 - a. <http://localhost:8080/downloads.html>
 - b. Under **DataVirtuality Server**, select Download and choose your host operating system to begin the download.
2. Follow the instructions to set up a default install of DataVirtuality Studio.

3. Once successfully installed, launch DV Studio from your Applications.
 - a. In windows this is: **Start > DV Studio**
 - b. In Mac: **Command + Space, DV Studio**
4. When the Connection Windows appears, insert the following values into the login module:
 - a. Title: Something descriptive of your instance (eg. "My DataVirtuality Test")
 - b. Host: **localhost**
 - c. Port: **31000** (or 31001 for SSL)
 - d. Database: **datavirtuality**
5. Select connect to... connect!

Pulling Data into RapidMiner

All new data sources can be quickly added through two channels, by selecting **Data Sources > Add Data Source** in the top navigation, or by right-clicking on the Data Sources folder object, located in the left-most panel and selecting **Add Data Source**.

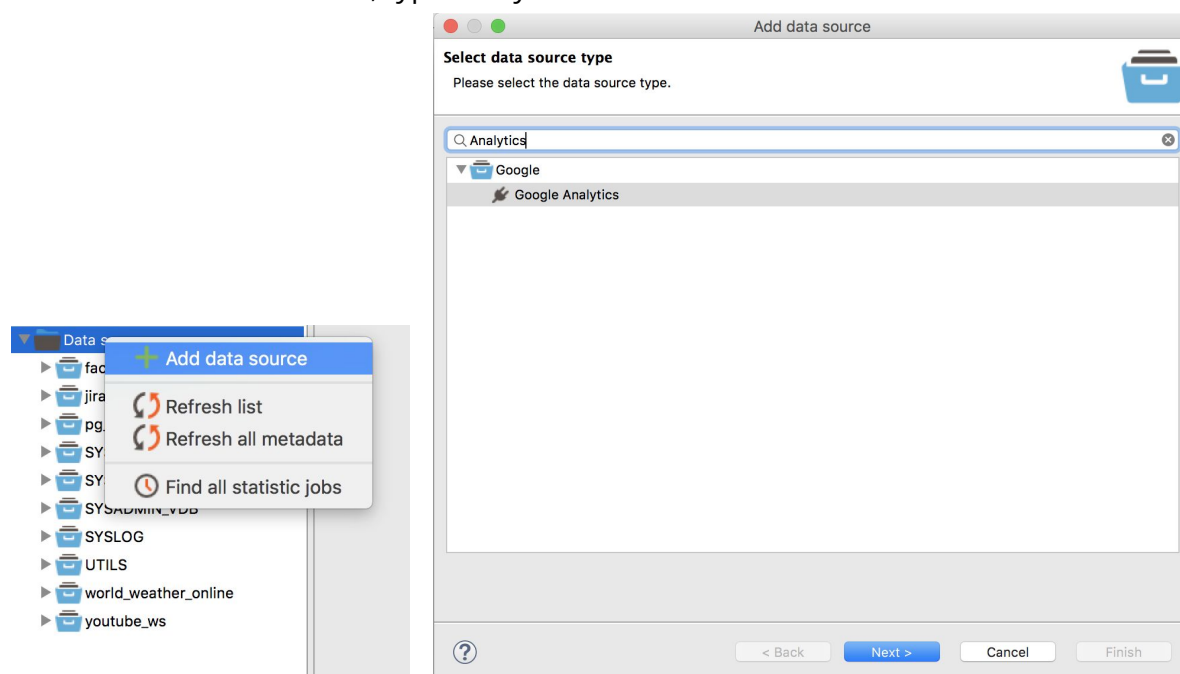
Here you will see a list of traditional data sources, such as MS SQL and Oracle, as well as web services like social media APIs, flat files, big data connectors, etc.

In the left-most panel, is helpful information for the various objects within DV Studio and can also be accessed through the **Help** option in the top navigation.

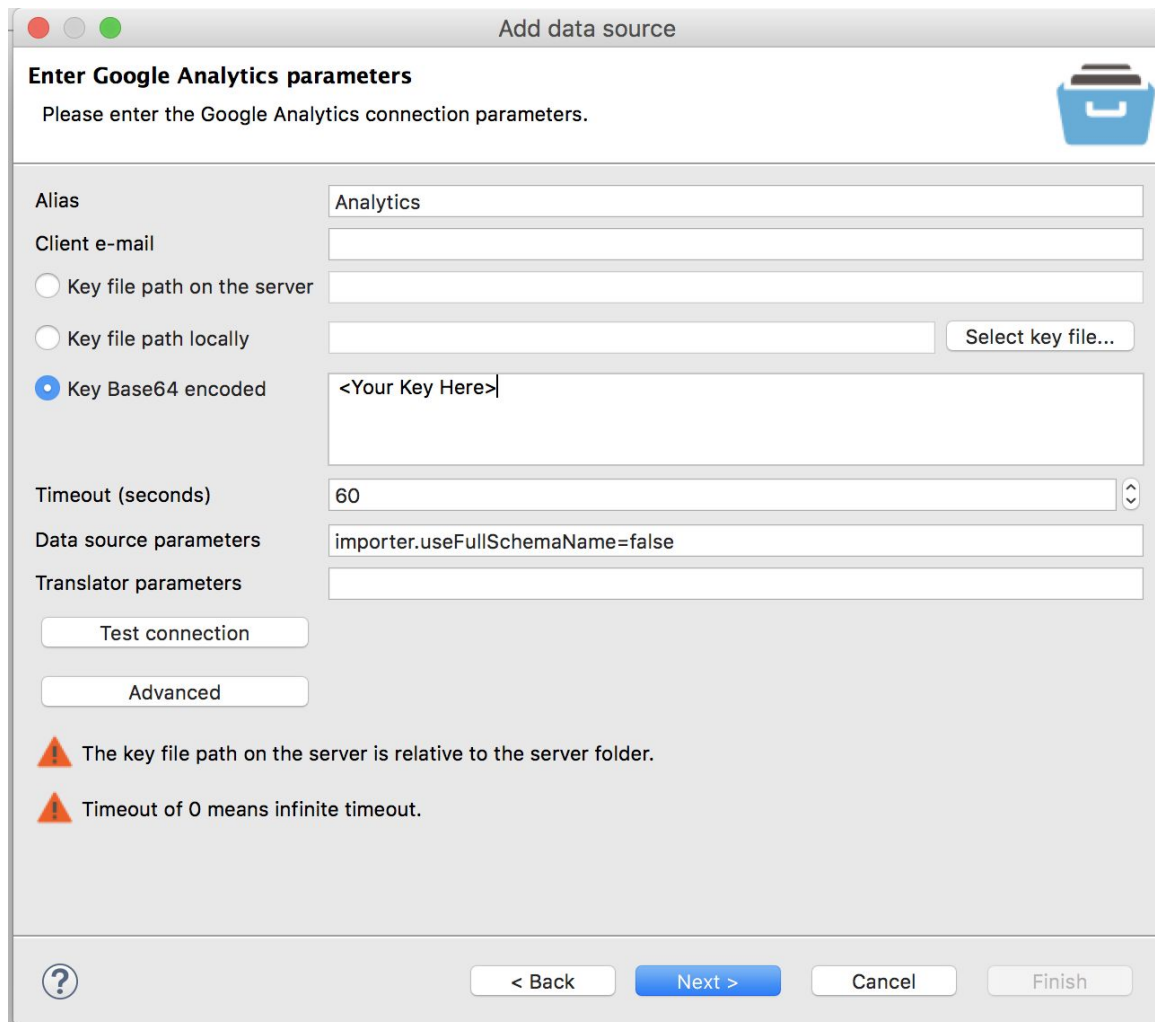
Adding an example data connection:

Let's add a new connection to our DataVirtuality Server! We are going to be connecting to Google Analytics and investigating our user data and user visits to gain some deeper knowledge about how our advertisements are performing.

1. Add the datasource, type Analytics to locate the connection and click Next.



2. Enter your Analytics account Key, email, and a name for the connection, then click Next.



Add data source

Enter Google Analytics parameters

Please enter the Google Analytics connection parameters.

Alias: Analytics

Client e-mail:

☐ Key file path on the server:

☐ Key file path locally: Select key file...

☒ Key Base64 encoded: <Your Key Here>

Timeout (seconds): 60

Data source parameters: importer.useFullSchemaName=false

Translator parameters:

Test connection

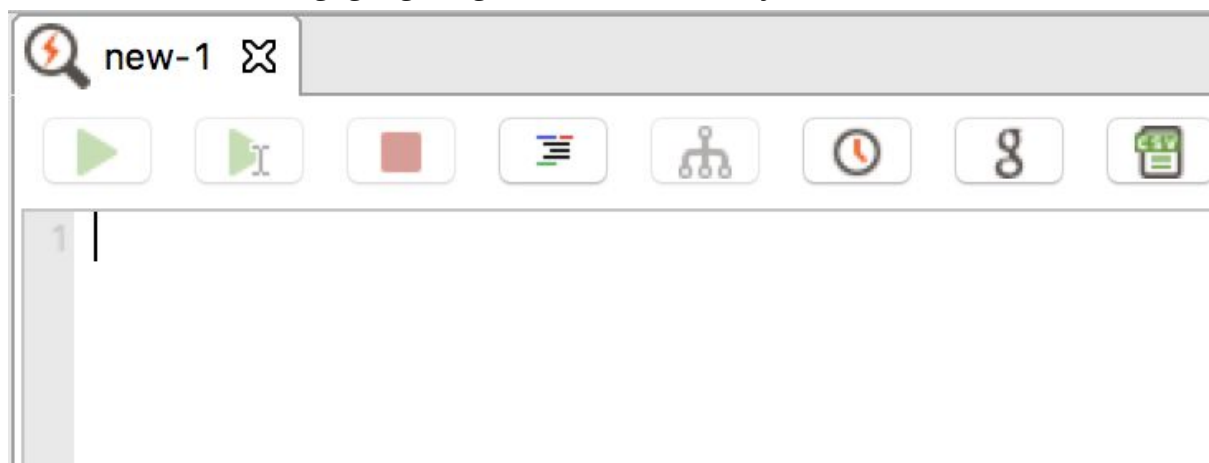
Advanced

⚠ The key file path on the server is relative to the server folder.

⚠ Timeout of 0 means infinite timeout.

? < Back Next > Cancel Finish

3. Click finish to simultaneously test the connection and add it to your data sources.
4. Select **Window > Open SQL Editor**
5. Click the small "g" google logo to launch the Analytics Wizard.



6. Under Metrics, select **Users** and **NewUsers**.
7. Under Dimensions, Time, select **date**.

8. Click **OK**.

9. A SQL query will be automatically generated in order to retrieve the data from this data source. In order to keep a lasting copy of this data, above the **SELECT**, copy and paste this line:

CREATE VIEW "views.analytics_example" AS

Your window should look like so:

```

1 CREATE VIEW "views.analytics_example" AS
2 SELECT
3   a.users, a.newUsers, a.date
4 FROM
5   (exec
6     Analytics.get("profile"=>'', "startDate"=>'2018-06-21', "endDate"=>'2018-06-21', "metrics"=>'users,newUsers', "dimensions"=>'da
7     te', "samplingLevel" =>'Default')) a;|

```

10. Now hit the green “play” button located above and slightly to the left of this query.

11. On the left hand side of the studio, open the **views** folder, by double clicking on the folder icon.



12. Now select the refresh button, to load the newly created google analytics view.



Name	Description	Cardinality
analytics_example		

13. To use this table inside of RapidMiner, we need to use the **Read Database** Operator, and call the view like so:

```
SELECT "date", users, newUsers FROM "views.analytics_example" ORDER BY "date" DESC;;
```

14. Run the process by clicking the play button to view your data.