



## Data Pipeline Run Parameters Setup

*(updated 03 Nov 2023)*

IOblend Designer creates its data pipelines as a series of data pipeline playbook components, where a component is either a:

- **Source**, which defines from where the data is to be pulled
- **Transformation**: either a SQL or Python transformation performed on any data that has been created or read in
- **Sink**, which defines where the resulting data of the data pipeline should be put, otherwise known as the data pipeline target

IOblend Designer organises the set of data pipelines playbook components into a data pipeline Project. It stores the data pipeline as series of JSON metadata files that the IOblend Engine will read and convert into Spark and Python code to be run. IOblend Engine also requires specific project Run Parameters that helps it to convert the IOblend Designer's data pipeline metadata.

In this section, we will be focusing on the **Run Parameters**, but will also cover the other two groups in the context of setting up your data pipeline projects.

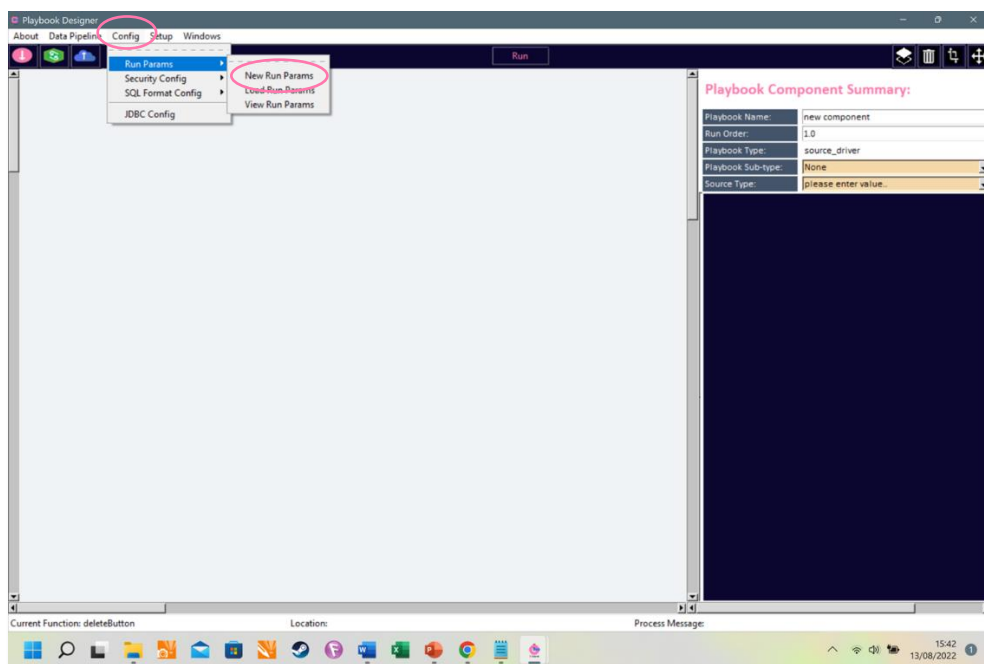
Each IOblend data pipeline requires the user to specify Run Parameters that describe the characteristics of the given pipeline, such as its name, location, user security credentials for sources and sinks, and other parameters. IOblend will use these parameters to store and run any given data pipeline.

As mentioned above, run parameters are specific to the data pipeline project, as such they should really be created first when starting a new data pipeline and before any data pipeline playbook components are dropped onto the Designer window. If these Run Parameters are not set up properly you will not be able to execute your data pipeline.

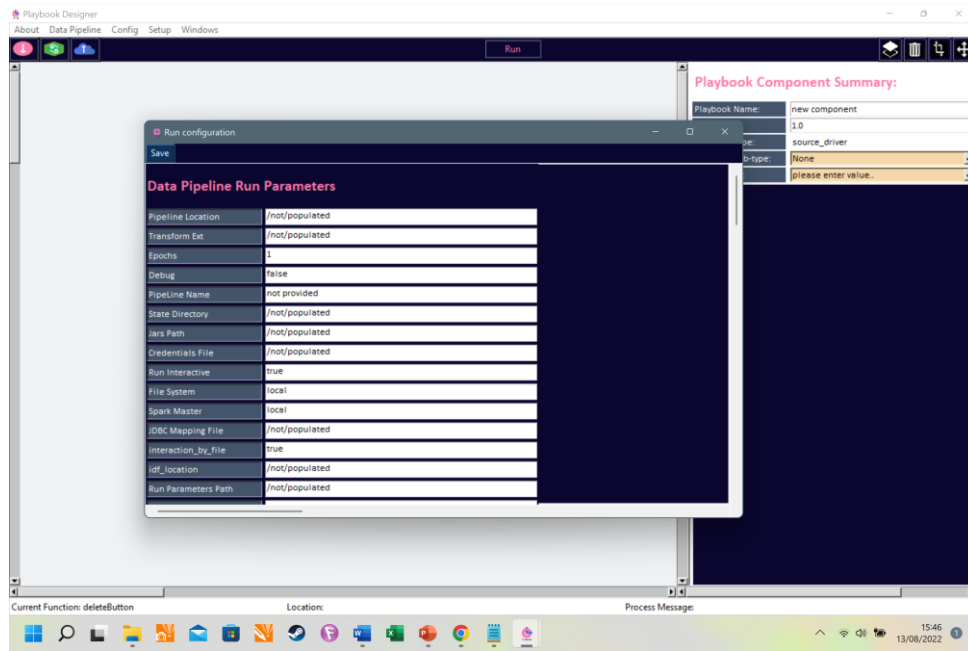
**You must define the Run Parameters for each *new* data pipeline you design for the system to understand what it needs to do.**

We will explain what each parameter means and how it works below.

1. Launch IOblend. In the Designer window select “**Config ---> Project Run Params --> New Run Params**”



2. A pop-up box will appear. You will need to populate each field with paths and values. All fields contain description – just hover the cursor over the field titles for a small pop-up box to appear.
  - o Let's go over each Run Parameter in more detail



## Pipeline location

- This parameter defines the location of the data pipeline. You can specify any convenient location either on Local (e.g. *Developer Edition*) or Cloud environment (*Enterprise Edition*). The path could look something like this:

**C:/Users/abc/Documents/ioblend\_designer/your\_project\_folder\_name/playbooks**

---

## Transform Ext

- This parameter defines the location of the external transform. IOblend uses two types of transforms, Internal, which is SQL, and External, which is written in Python. You can specify the location of the Python scripts here. The path could look something like this:

**C:/Users/abc/Documents/ioblend\_designer/your\_project\_folder\_name/transforms**

---

## Epochs

- This parameter specifies the number of cycles you want the pipeline to run. **0** (zero) means the pipeline will run continuously unless manually terminated -

used for real time streaming data. Otherwise, type a specific number of iterations you want your pipeline to execute.

---

### Debug

- This parameter needs to be set to **True** (*just type true in the window*) if you are developing a data pipeline. If set to True, IOblend will allow you to interactively debug the pipeline as you build it, step by step. In production, this parameter should be set to **False** – this will enable IOblend to run the pipeline automatically without pausing for each step. Case-sensitive.
  - **Note:** set *Run Interactive* and *Interaction\_by\_file* fields to **True** as well when developing the pipelines to allow the step-by-step viewing of the results of the pipeline outputs.
- 

### Pipeline Name

- Give your pipeline a unique name
- 

### State Directory

- This parameter defines the location of the pipeline's state directory. IOblend generates a small amount of data about its current internal state for a pipeline. This path describes where that data should be written. Typically, it is a folder on the machine the IOblend engine is running from. The path should look something like this:

**C:/Users/abc/Documents/ioblend\_designer/your\_project\_folder\_name/state**

---

### Jars Path

- This parameter defines the absolute or relative path to **jars** used in IOblend, such as JDBC jars. All JDBC drivers should be placed in this location. The path should look something like this:

**C:/Users/abc/Documents/ioblend\_designer/your\_project\_folder\_name/jars**

---

## Credentials File

- This parameter defines the location of the user credentials for all sources and sinks in your pipeline. Security credentials can be “Master”, meaning you can have all your access credentials saved in one file for ease of work when developing projects, or you can have specific permissions set for each project, where you only specify credentials pertinent to a specific data pipeline. You can use the **Config->Security config->New** credentials menu selection in IOblend Designer to populate this file. When you first create it from this menu option, it will populate your credentials file with template credentials for each source/sink type. You can use these templates to set your credentials for your own sources and sinks. How to do this will be covered in another section.
- The path should look something like this:

**C:/Users/abc/Documents/ioblend\_designer/your\_project\_folder\_name /config**

- You will find the examples of the formats in the supplied P3-test pipeline, which can be found in the IOblend designer under P3-test/config.

The **sources** and **sinks** will look like the below:

```
{
  "sample_local_stream_jdbc": {
    "dbtable": "ABC",
    "user": "root",
    "password": "XYZ"
  },
  "sample_static_local_jdbc": {
    "dbtable": "DEF",
    "user": "root",
    "password": "YYY"
  },
  "watch_azure_static": {
    "client_id": "enter your client id",
    "tenant_id": "enter your tenant id",
    "subscription_id": "enter your subscription id",
    "storage_account_name": "enter storage acct name",
    "client_secret": "client secret",
    "container": "container name",
    "conn_type": "serviceprinciple",
    "conn_env_setup": true,
```

```
"gen2": true },

"watch_azure_stream": {
  "client_id": "enter your client id",
  "tenant_id": "enter your tenant id",
  "subscription_id": "enter your subscription id",
  "storage_account_name": "enter storage acct name",
  "client_secret": "client secret",
  "container": "container name",
  "conn_type": "serviceprinciple",
  "conn_env_setup": true,
  "gen2": true },

"watch_bigquery_stream_pjdbc_large": {
  "dbtable": "ABC",
  "user": null,
  "password": null },

"watch_sink_bq_pjdbc_iw": {
  "dbtable": "ABC",
  "user": null,
  "password": null},

"ABC": {
  "OAuthServiceAcctEmail": "bigquery-
sa@your_acct.iam.gserviceaccount.com",
  "OAuthPvtKeyPath": "add your auth pvt key path"},

"gcp_config_files": {
  "OAuthServiceAcctEmail": "bigquery-
sa@your_acct.iam.gserviceaccount.com",
  "OAuthPvtKeyPath": "add your auth pvt key path",
  "OAuthPvtKeyPath_json": "add your auth pvt key path.json"}
```

} - etc, add as many sources/sinks as your pipeline requires.

---

## Run Interactive

- Like the **Debug** field, this parameter needs to be set to **True** (*just type true in the window*) if you are developing a data pipeline. If set to True, IOblend will allow you to interactively view the pipeline output as you build it, step by step. In production, this parameter should be set to **False** – this will enable IOblend to run the pipeline automatically without pausing for each step.
  - **Note:** set **Debug** and **Interaction\_by\_file** fields to **True** as well when developing the pipelines to allow the step-by-step debugging of the results of the pipeline.
- 

## File System

- IOblend can operate with multiple files system formats, such as local OS (**Windows, iOS/Linux**) and Cloud (**Azure, AWS S3, GCP**). This specifies which file system will be used to store all run time parameter items artifacts, including the Project Run parameters file itself.
  - **NOTE: the following run time parameters all need to be stored in the same file system denoted by the File system parameter: Pipeline Location, Transform Ext, Jars Path, Credentials File, JDBC mapping file, Run parameters path**
- 

## Spark Master

- This parameter defines the Spark resource manager options for executing dynamic Spark data pipelines created from your playbooks. You can set one of the following options:
  - **local** (single core local environment)
  - **local [\*]** (multi-core local environment)
  - **YARN**
  - **Mesos**
  - **Kubernetes**

- **Note:** In the Developer Edition, only local features are enabled. You can use all of them in the Enterprise Edition. For the development of your projects, leave the setting on “local” or local [\*]. When you later want to push your project out in production, either consult your data engineer or refer directly to the Spark documentation to determine the best setting for your particular case.
- 

### JDBC Mapping File

- This parameter defines the absolute path to `jdbc_db_sql.json` file. The file maps specific SQL syntax to enable IOblend’s JDBC templates to read and write from the JDBC end points. This is an internal mapping file that should **not** need to be altered by the user, unless adding a new JDBC database that is not already defined in the mapping table. If this is the case, please contact us and we will help you amend this file accordingly. Otherwise just point to the default `jdbc_db_sql.json` file provided in the IOblend download. The default file can be located:

`C:/Users/abc/Documents/ioblend_designer \p3-test\config`

The final JDBC path should look something like this:

`C:/Users/abc/Documents/ioblend_designer/your_project_folder_name/config`

### Interaction\_by\_file

- This parameter needs to be set to **True** (*just type true in the window*) if you are developing a data pipeline. This is an internal IOblend parameter. In production, this parameter should be set to **False** – this will enable IOblend to run the pipeline automatically without pausing for each step. IOblend Designer will use this file to communicate with the IOblend Engine.
  - **Note:** set *Run Interactive* and *Debug* fields to **True** as well when developing the pipelines to allow the step-by-step viewing of the results of the pipeline outputs.
-



## IDF location

*(interactive debug file location)*

- This parameter defines the absolute path to a directory from where the pipeline will be run and interaction files can be written to when running the pipeline interactively from the Designer.

`C:/Users/abc/Documents/ioblend_designer/your_project_folder_name/interaction`

---

## Run Parameters Path

- This parameter defines the directory from where the pipeline run parameters are stored. The path should look something like this:

`C:/Users/abc/Documents/ioblend_designer/your_project_folder_name/config`

---

## SSH

- In the Developer Edition this should be set to **Local**. This means that the IOblend Designer will interact with a local copy of the IOblend Engine.
  - If running the Enterprise Edition, please provide the full SSH command to access the remote server where the IOblend engine has been installed. **Note SSH will not work in the Developer Edition.**
  - Available options are
    - **local**, meaning the pipeline will be executed locally
    - **ssh username@localhost**, an SSH command
- 

## IOblend Directory

- This parameter defines the absolute path to a directory where the IOblend Engine binaries are installed. This feature is only used when deploying via the **SSH**, locally or remotely. If you are unsure how to deploy via the SSH, please consult with your data engineering team or contact us for assistance. The path should look something like this:

`C:/Users/abc/ioblend/ioblend_package/ioblend/src/ioblend_dist`

- **Note:** In the Developer Edition, SSH is not enabled

3. Hit **Save**. You are now all set to begin developing your data pipeline

