API JobGroups Create v4

> ⓘ **NOTE:** This feature is in Beta release.

Contents:

- *Required Permissions*
- *Request*
- *Response*
- *Reference*

Create a jobGroup, which launches the specified job as the authenticated user.

The request specification depends on one of the following conditions:

- Dataset has already had a job run against it and just needs to be re-run.
- Dataset has not had a job run, or the job definition needs to be re-specified.

> ⓘ **NOTE:** Override values applied to a job are not validated. Invalid overrides may cause your job to fail.

> ⓘ **NOTE:** In this release, you cannot execute jobs sourced from datasets in Redshift or SQL DW or publish to these locations via the API. This known issue will be fixed in a future release.

Version: `v4`

Required Permissions

> ⓘ **NOTE:** Each request to the Trifacta® platform must include authentication credentials. See *API Authentication*.

Request

Request Type: `POST`

Endpoint:

```
/v4/jobGroups
```

Request Body - Re-run job:

If you are re-running a job that has already executed and do not need to modify any job settings, you can use the following simplified body to launch it:

```
{
  "wrangledDataset": {
    "id": 7
  }
}
```

Request Body - Specify job:

If you are specifying a new job or must re-run a job with new settings, you must include a version of the following request body. Required parameters are listed below:

```
{
  "wrangledDataset": {
    "id": 1
  },
  "overrides": {
    "execution": "photon",
    "profiler": false,
    "writesettings": [
      {
        "path": "hdfs://hadoop:50070/trifacta/queryResults/admin@trifacta.local/cdr_txt.csv",
        "action": "create",
        "format": "csv",
        "compression": "none",
        "header": false,
        "asSingleFile": false
      }
    ]
  },
  "ranfrom": "cli"
}
```

Request Body - Specify job for dataset with parameters:

If you have created a dataset with parameters, you can specify overrides for parameter values during execution through the APIs. Through this method, you can iterate job executions across all matching sources of a parameterized dataset. For more information on creating datasets with parameters, see
Overview of Parameterization.

In the following, the `runParameters` overrides have been specified for the `varRegion` and the `varReference Type` variables.

```
{
    "wrangledDataset": {
        "id": 33
    },
    "overrides": {
      "writesettings": [
       {
        "path": "<path_to_output_file>",
        "action": "create",
        "format": "csv",
        "compression": "none",
        "header": false,
        "asSingleFile": false
       }
      ]
    },
    "runParameters": {
      "overrides": {
        "data": [{
          "key": "Region",
          "value": "02"
        },
        {
          "key": "varReferenceType",
          "value": "PROD"
        }
       ]
      }
    },
  "ranfrom": "ui"
}
```

Response

Response Status Code - Success: `201 - Created`

Response Body Example:

```
{
    "reason": "JobStarted",
    "sessionId": "eb3e98e0-02e3-11e8-a819-25c9559a2a2c",
    "id": 9,
    "jobs": {
        "data": [
            {
                "id": 12
            },
            {
                "id": 13
            }
        ]
    }
}
```

Reference

Request Reference:

Property	Description
wrangledDataset	(required) Internal identifier for the object whose results you wish to generate. The recipes of all preceding datasets on which this dataset depends are executed as part of the job.

overrides. execution	(required, if first time running the job) Indicates the running environment on which the job is executed. Accepted values:
	
	   • `photon`
	   • spark
	
	For more information, see *Running Environment Options*.

overrides. profiler	(required, if first time running the job) When set to `true`, a visual profile of the job is generated as specified by the profiling options for the platform. See *Profiling Options*.
overrides. writesettings	(required, if first time running the job) These settings define the publishing options for the job. See below.
ranfrom	(optional) If this value is set to `null`, then the job does not show up in the Job Results page. If set to `cli`, the job appears as a CLI job. See *Job Results Page*.

writesettings Reference:

The `writesettings` values allow you to specify aspects of the publication of results to the specified `path` location.

> ⓘ **NOTE:** `writesettings` values are required if you are running this specified job for the dataset for the first time.

> ⓘ **NOTE:** To specify multiple outputs, you can include additional `writesettings` objects in the request. For example, if you want to generate output to `csv` and `json`, you can duplicate the `writesettings` object for `csv` and change the `format` value in the second one to `json`.

These settings correspond to values that you can apply through the UI or through the command line interface.

   • For UI information, see *Run Job Page*.
   • For CLI information, see *CLI for Jobs*.

Property	Description
path	(required) The fully qualified path to the output location where to write the results

action	(required) If the output file or directory exists, you can specify one of the following actions:
	
	   • `create` - Create a new, parallel location, preserving the old results.
	   • `append` - Add the new results to the old results.
	   • `overwrite` - Replace the old results with the new results.

`format`	(required) Output format for the results. Specify one of the following values:
	
	• `csv`
	• `json`
	• `avro`
	• `pqt`
	
	  ⓘ **NOTE:** Parquet format requires execution on a Hadoop running environment (`overrides.execution`
	      must be `spark`).
	
	  ⓘ **NOTE:** To specify multiple output formats, create additional `writesettings` object for each output format.

`compression`	(optional) For `csv` and `json` results, you can optionally compress them using `bzip2` or `gzip` compression. Default is `none`.
	
	  ⓘ **NOTE:** If compression is applied, the filename in the `path` value must end with the appropriate extension for the
	      type of compression:
	
	      • `.gz` for GZIP
	      • `.bz2` for BZIP2

`header`	(optional) For `csv` results with `action` set to `create` or `append`, this value determines if a header row with column names is inserted at the top of the results. Default is `false`.
`asSingleFile`	(optional) For `csv` and `json` results, this value determines if the results are concatenated into a single file or stored as multiple files. Default is `false`.

Response reference:

Property	Description
reason	Current state of the job group at time of API call. Since this call creates the job group, this value is always `Job started` in the response to this call.
sessionId	Session identifier for the job group.
id	Internal identiifer of the job group.
jobs.data.id	Internal identifier of the individual job or jobs created as part of this job group.