Dataset

Represents a dataset in Dremio. All datasets returned by the REST API have an `entityType` of `dataset`.

Dataset Parameters

The JSON representation of a dataset looks like this:

```
{
    "entityType": "dataset" [immutable after creation],
    "id": String [immutable, generated by Dremio],
    "path": [String] [immutable after creation],
    "tag": String [immutable, generated by Dremio],
    "type": String ["PHYSICAL_DATASET", "VIRTUAL_DATASET"] [immutable],
    "fields": [DatasetField] [immutable],
    "createdAt": String (RFC3339 date) [immutable, generated by Dremio],
    "accelerationRefreshPolicy": DatasetAccelerationRefreshPolicy [optional, only for physical datasets in a source],
    "sql": String [optional, required for virtual datasets],
    "sqlContext": [String] [optional, only for virtual datasets],
    "format": DatasetFormat [optional, required for promoted datasets],
    "approximateStatisticsAllowed": Boolean [optional, introduced in Dremio 2.1.0]
}
```

Name	Type	Description
id	String	Dataset ID. Generated by Dremio, immutable.
path	[String]	Dataset path. Immutable after creation.
tag	String	Identifies the instance, changed each time it is modified. Generated by Dremio, immutable.
type	String	The dataset type, must be either `PHYSICAL_DATASET` or `VIRTUAL_DATASET`. Immutable after creation.
fields	[DatasetField]	The dataset fields representing the schema of the dataset. Immutable.
createdAt	String	RFC3339 date (example: `2017-10-27T21:08:22.858Z`) representing the creation datetime. Immutable.
accelerationRefreshPolicy	DatasetAccelerationRefreshPolicy	Represents the acceleration refresh policy for the dataset. Applies only to physical datasets that exist in a source.
sql	String	The sql for the dataset, applies only to virtual datasets and required for them.
sqlContext	[String]	The context for the sql, applies only to virtual datasets and is optional.
format	DatasetFormat	The dataset format configuration, applies only to promoted physical datasets and is required.
approximateStatisticsAllowed	Boolean	When set, count distinct queries will return approximate results.

Fields Parameter

Represents a dataset field's schema in Dremio.

The JSON representation of a field looks like this:

```
{
    "name": String – the field name,
    "type": {
        "name": String ["STRUCT", "LIST", "UNION", "INTEGER", "BIGINT", "FLOAT", "DOUBLE", "VARCHAR", "VARBINARY",
"BOOLEAN", "DECIMAL", "TIME", "DATE", "TIMESTAMP", "INTERVAL DAY TO SECOND", "INTERVAL YEAR TO MONTH"],
        "subSchema": [DatasetField] [optional],
        "precision": Number [optional],
        "scale": Number [optional]
    }
}
```

For complex types (`LIST`, `STRUCT`, `UNION`), `subSchema` will provide a list of `DatasetField` representing the composition. For example, `UNION` will have a `subSchema` which represents all the primitive types that have been detected.

For `DECIMAL` type, `precision` / `scale` are provided.

AccelerationRefreshPolicy Parameter

Represents the dataset acceleration refresh policy for a dataset.

```
{
    "refreshPeriodMs": Number,
    "gracePeriodMs": Number,
    "method": String ["FULL", "INCREMENTAL"],
    "refreshField": String [optional],
    "accelerationNeverExpire": Boolean,
    "accelerationNeverRefresh": Boolean
}
```

Name	Type	Description
refreshPeriodMs	Number	How often (in milliseconds) to refresh all reflections on the dataset.
gracePeriodMs	Number	How old (in milliseconds) data in a reflection can be and still be used for accelerating queries.
method	String	For every refresh, either a `FULL` or an `INCREMENTAL` (only works in certain cases, please see docs) update.
refreshField	String	For certain datasets, a `refreshField` can be set if `method` is `INCREMENTAL`.
accelerationNeverExpire	Boolean	Controls whether the reflection is able to expire.
accelerationNeverRefresh	Boolean	Controls whether the reflection regularly refreshes.

Format Parameter

Folders/files can be promoted to physical datasets by applying a format. When applying a dataset format to a folder, all files in that folder should conform to the selected type.

Text (delimited)

Applies to text files with delimiters (CSV, TSV, etc).

```
{
  "type": "Text",
  "fieldDelimiter": String,
  "lineDelimiter": String,
  "quote": String,
  "comment": String,
  "escape": String,
  "skipFirstLine": Boolean,
  "extractHeader": Boolean,
  "trimHeader": Boolean,
  "autoGenerateColumnNames": Boolean
}
```

JSON

```
{
  "type": "JSON"
}
```

Parquet

```
{
  "type": "Parquet"
}
```

Excel

```
{
  "type": "Excel",
  "sheetName": String,
  "extractHeader": Boolean,
  "hasMergedCells": Boolean
}
```

XLS

```
{
  "type": "XLS",
  "sheetName": String,
  "extractHeader": Boolean,
  "hasMergedCells": Boolean
}
```

Was this page helpful?   👍 Yes   👎 No

```
{
  "type": "Text",
  "fieldDelimiter": String,
  "lineDelimiter": String,
  "quote": String,
  "comment": String,
  "escape": String,
  "skipFirstLine": Boolean,
  "extractHeader": Boolean,
  "trimHeader": Boolean,
  "autoGenerateColumnNames": Boolean
}
```