Read Functions Reference

These are the read functions you can use in the json settings files. This is a reference document and does not show how to use these functions as well as one of the procedures or guides.

Functions

|  |  |
| --- | --- |
| **Function Name** | **Description** |
| read\_table | Performs a batch read of a Delta table. |
| read\_snapshot\_windowed | Retrieves latest record for each key, ordering by ingest timestamp. |
| read\_latest\_ingest | Loads only the records from the most recent ingest time. |
| stream\_read\_cloudfiles | Reads streaming data using Auto Loader (cloudFiles). |
| stream\_read\_table | Creates a streaming DataFrame from an existing Delta table. |

Settings

|  |  |
| --- | --- |
| **Function Name** | **Key(s) Required** |
| read\_table | src\_table\_name |
| read\_snapshot\_windowed | src\_table\_name, surrogate\_key, ingest\_time\_column |
| read\_latest\_ingest | src\_table\_name, ingest\_time\_column |
| stream\_read\_cloudfiles | readStreamOptions, readStream\_load, file\_schema |
| stream\_read\_table | src\_table\_name, readStreamOptions |

Example

Many settings are dynamically generated if you don’t specify them. The file schema should always be given to enforce the approved schema. Note: the file schema is too long to put in this example but will be covered in the databricks-utilities documentation for get-schema.py.

"read\_function": "functions.read.stream\_read\_cloudfiles"

"readStreamOptions": {

"cloudFiles.format": "json",

"encoding": "utf-8",

"pathGlobFilter": "powerPlay.json",

"multiline": "false"

},