
intake_hbase Documentation

Release 0.0.1

Joseph Crail

Jul 25, 2018

Contents:

1 Quickstart	1
1.1 Installation	1
1.2 Usage	1
2 API Reference	3
3 Indices and tables	5

`intake-hbase` provides quick and easy access to tabular data stored in Apache [HBase](#). This plugin reads hbase query results without random access: there is only ever a single partition.

1.1 Installation

To use this plugin for `intake`, install with the following command:

```
conda install -c intake intake-hbase
```

1.2 Usage

1.2.1 Ad-hoc

After installation, the function `intake.open_hbase` will become available. It can be used to execute queries on the hbase server, and download the results as a list of dictionaries.

Three parameters are of interest when defining a data source:

- `query`: the query to execute, which can be defined either using [Lucene](#) or `'JSON'` syntax, both of which are to be provided as a string.

1.2.2 Creating Catalog Entries

To include in a catalog, the plugin must be listed in the plugins of the catalog:

```
plugins:
  source:
    - module: intake_hbase
```

and entries must specify `driver: hbase`.

1.2.3 Using a Catalog

`intake_hbase.source.HBaseSource`

class `intake_hbase.source.HBaseSource` (*table*, *connection*, *divisions=None*, *qargs=None*, *metadata=None*)

Execute a query on HBASE

The data are returned as tuples of (ID, data) where the data is a dict of field-value pairs.

Parameters

table: **str** HBase table to query. If within a project/namespace, either use the full table name, e.g., `myproject_mytable` or use `table_prefix='myproject'` in the connection parameters.

connection: **str or dict** See happybase connection arguments <https://happybase.readthedocs.io/en/latest/api.html#happybase.Connection>

divisions: **list or None** Partition key boundaries. If None, will have one partition for the whole table. The number of partitions will be `len(divisions) - 1`.

qargs: **dict or None** Further arguments to `table.scan`, see <https://happybase.readthedocs.io/en/latest/api.html#happybase.Table.scan>

Attributes

datashape

description

hvplot Returns a hvPlot object to provide a high-level plotting API.

plot Returns a hvPlot object to provide a high-level plotting API.

Methods

<code>close()</code>	Close open resources corresponding to this data source.
<code>discover()</code>	Open resource and populate the source attributes.
<code>read()</code>	Return all results
<code>read_chunked()</code>	Return iterator over container fragments of data source
<code>read_partition(i)</code>	Return a (offset_tuple, container) corresponding to i-th partition.
<code>to_dask()</code>	Return a dask-bag of results
<code>yaml()</code>	Return YAML representation of this data-source

read()

Return all results

to_dask()

Return a dask-bag of results

CHAPTER 3

Indices and tables

- `genindex`
- `modindex`
- `search`

H

HBaseSource (class in intake_hbase.source), 3

R

read() (intake_hbase.source.HBaseSource method), 4

T

to_dask() (intake_hbase.source.HBaseSource method), 4