Ibis

Python Data Analysis Productivity Framework

about me

Ivan Ogasawara

- Ibis core contributor
- SciPy Latin America Bolivia Ambassador (telegram: @scipyla)
- Software Engineering at Quansight (twitter: @quansightai)

Introduction

- Ibis has a similar **API** provided by **Pandas**
- Created by Wes McKinney (~ Nov/2014)
- Handling multiple types of database (**backends**) using the same **API**
- No need to handle **SQL**
- Handling large volumes of data with just **expressions**
- Organizes expressions using **graphs**

Installation

```
# with pip
pip install ibis-framework[all]

# with pip getting source from github
pip install git+https://github.com/ibis-project/ibis.git#egg=ibis-framework[all]

# with conda
conda install -c conda-forge ibis-framework
```

Backends

- Apache Impala
- Apache Kudu
- Hadoop Distributed File System
- PostgreSQL / PostGIS
- SQLite
- Google BigQuery
- Yandex Clickhouse
- Pandas (Experimental)
- OmniSciDB (Experimental)
- MySQL (Experimental)
- Apache Spark SQL / PySpark (Experimental)

Expressions

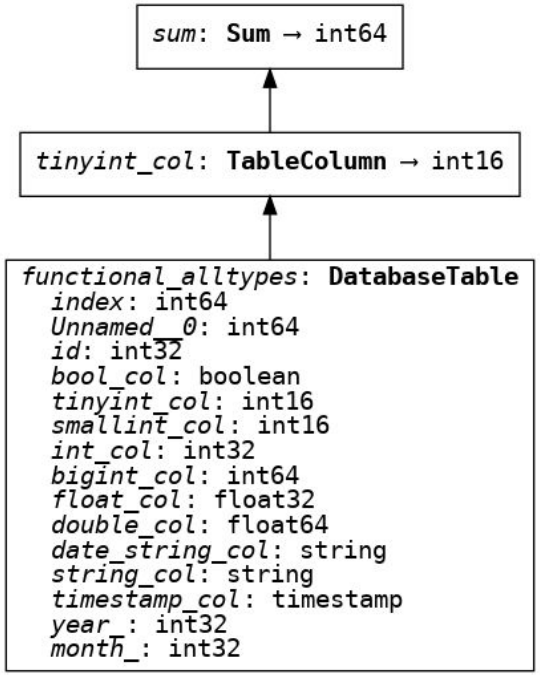
```
import ibis

# connection using OmniSciDB backend
con = ibis.omniscidb.connect(**con_data)

# con.list_tables()
t = con.table('functional_alltypes')

# on jupyter-lab
display(t.tinyint_col.sum())

# or: display(t['tinyint_col'].sum())
```



Expressions

```
print(t[[t.id, t.int_col]].head().execute())
```

```
     id  int_col
0  6690        0
1  6691        1
2  6692        2
3  6693        3
4  6694        4
```

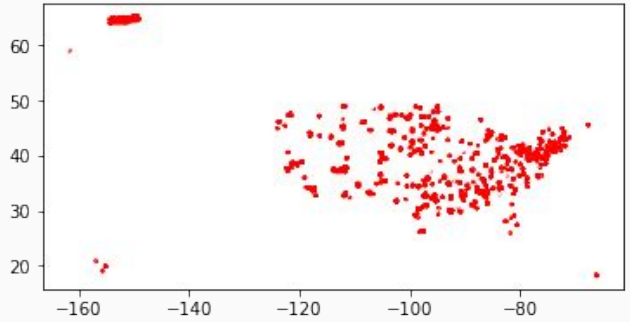
Expressions

```
t = omniscidb.table('zipcodes_2017')
display(t)

style_kwds = { ... }
expr = t[t.omnisci_geo].head(400)

expr.execute().plot(**style_kwds)
```

```
zipcodes_2017: DatabaseTable
  ZCTA5CE10: string
  AFFGEOID10: string
  GEOID10: string
  ALAND10: int64
  AWATER10: int64
  omnisci_geo: multipolygon
```



Features

- Support for the main functions available in **Pandas**
  - **sum, mean, max, min, all, any, std, var, corr, group_by, sort_by, etc**
- Window operations support (**SQL OVER**)
- **Geo Spatial** data and operations support
- **User-defined function (UDF)** support
- **Common table expressions (CTE)** support

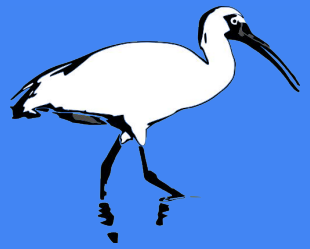
New features

- Support for more operations for **PySpark**
- Improve documentation
- Support for more window operations for **OmniSciDB**
- Support for more **Geo Spatial** operations for **OmniSciDB** and **PostGIS**

A complete list of the new features and bugfixes:

https://github.com/ibis-project/ibis/pulls?page=3&q=is%3Apr+is%3Aclosed+merged%3A%3E2019-06-24&utf8=%E2%9C%93

¡Thanks!



Contacts

- GitHub/Twitter: xmnlab
- ivan.ogasawara@gmail.com
- ivan.ogasawara@quansight.com

QUANSIGHT