Altinity

Seeing Is Believing: Popular BI Tools for ClickHouse

Robert Hodges and Dima Borovstov -- Altinity

Presenter Bios

Robert Hodges - CEO at Altinity

30+ years on DBMS plus virtualization and security. ClickHouse is DBMS #20

Dima Borovstov - Implementation Engineer at Altinity

25 years on database apps and visualization. Clickhouse user since 2017

Tableau

What is Tableau?

Most advanced and popular data visualization tool

Powerful for data discovery and exploration

No prior programming knowledge is needed--drag and drop via UI and you see the result right away

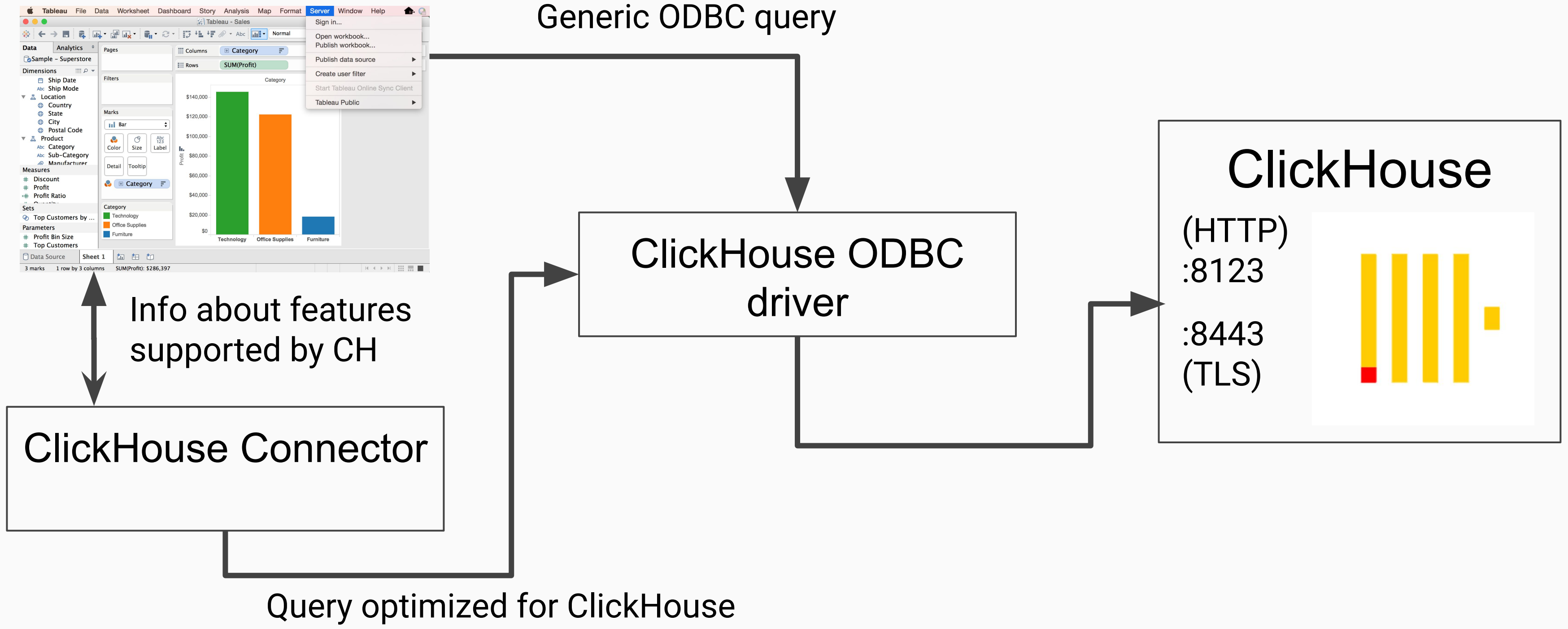
Supports huge list of different data sources (>80) + extendable (odbc, conn)

Can connect to multiple data sources and join data from different datasets

Complete Analytics Ecosystem (Server, Cloud, Reader, Mobile, Prep)

How Tableau connects to ClickHouse

Tableau Desktop/Server



Altinity

Tableau connector for ClickHouse

https://github.com/Altinity/clickhouse-tableau-connector-odbc
Apache 2.0
Developed by Dima Borovstov at Altinity

- Allows users to utilize ClickHouse query syntax from Tableau
- Works in Tableau Desktop (Win + Mac) and Server (Linux)
- Uses ClickHouse ODBC driver
- No need to configure ODBC DSN
- Open source (Apache 2.0) with current release on GitHub
- Easy to install (documentation here)
- Available via Tableau Gallery (coming soon)
- Maintained by Altinity

Demo time!

Let’s have a look at Tableau in action

Tableau as a BI tool

Strengths

- Intuitive drag-and-drop interface
- No scripting or programming knowledge needed
- Wide selection of graphs and charts
- Ability to create interactive dashboards (filters, actions, parameters)
- Calculated fields act similarly to Excel formulas
- Works with Live data sources and extracts
- A wide range of connectivity

Possible Weaknesses

- High cost and inflexible pricing
- Not able to show real time updated data
- Very limited support of scripting and custom visualization
- Lack data modeling and data dictionary capabilities
- Lack of version control

More information

Tableau Documentation

- Official Tableau Help
- Integrating Tableau with ClickHouse
- ClickHouse ODBC Driver installation and configuration

ClickHouse Connector

- ClickHouse Tableau connector Github project

Grafana

Altinity

What is Grafana?

Understands time series data

Simple installation

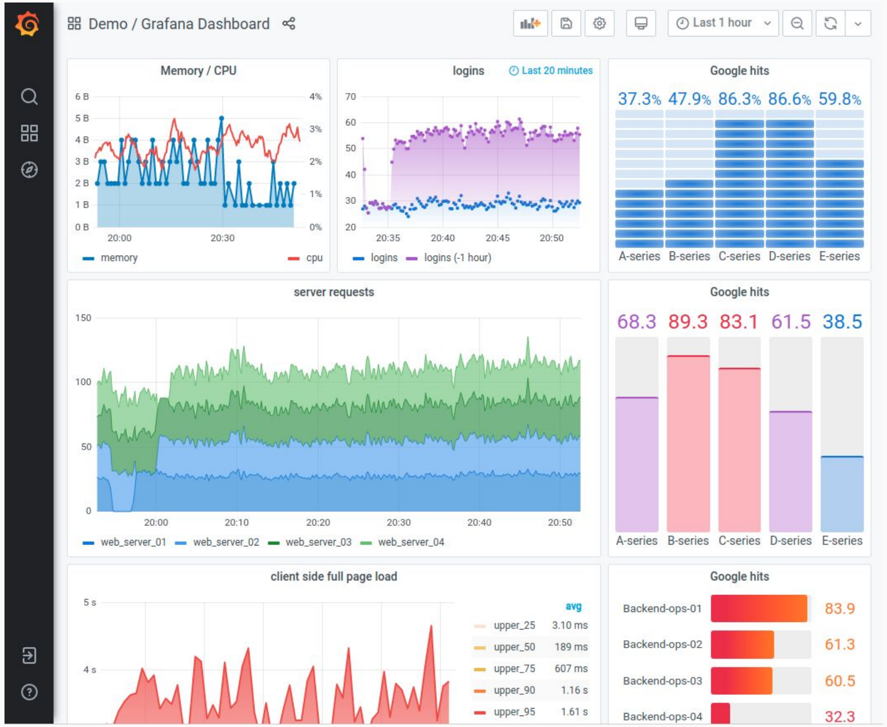
Supports many data sources

Lots of display plugins

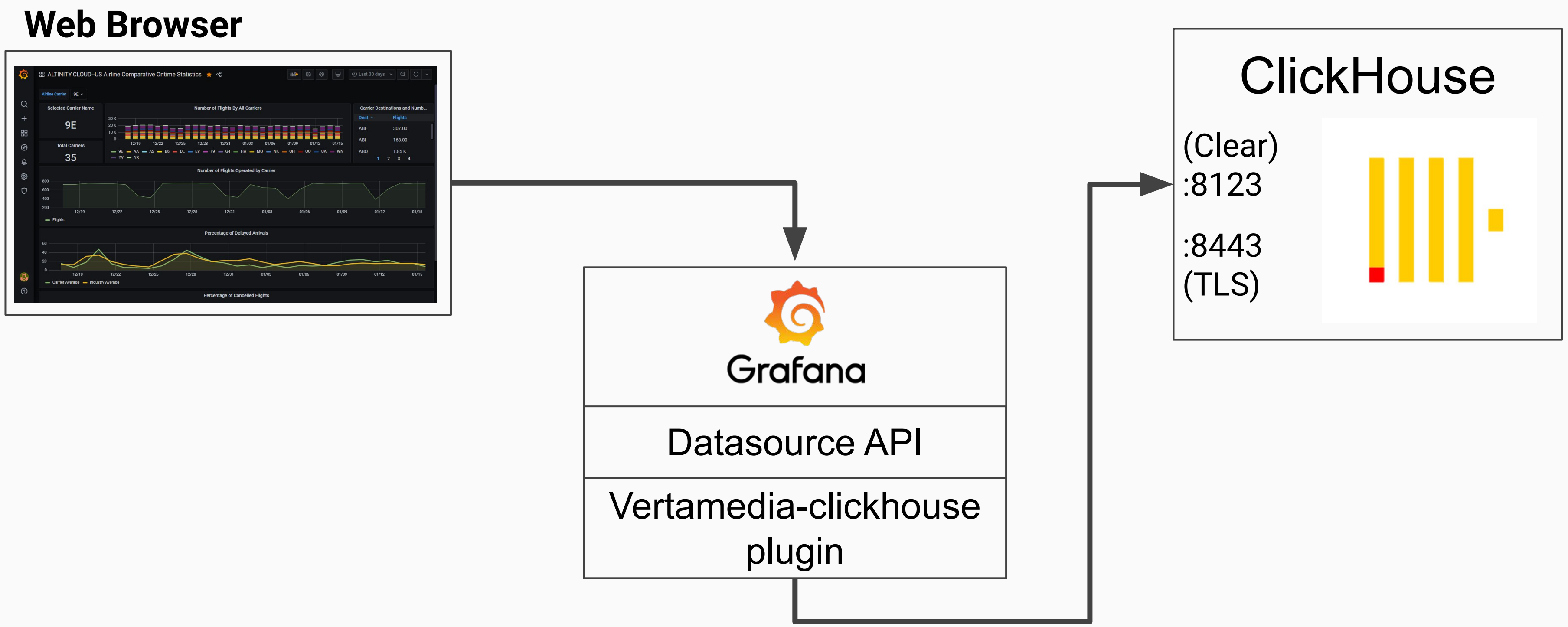
Highly interactive

Great for operational dashboards

Is open source (AGPL v3)



How Grafana connects to ClickHouse

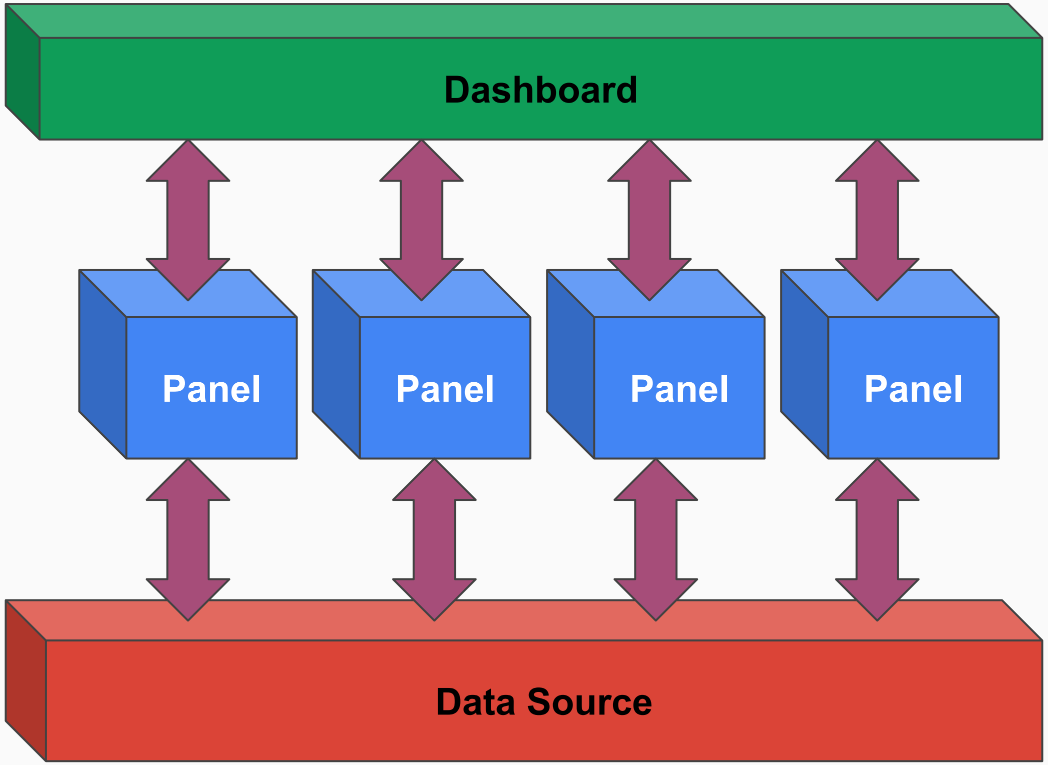


ClickHouse Grafana Plugin (Data Source)

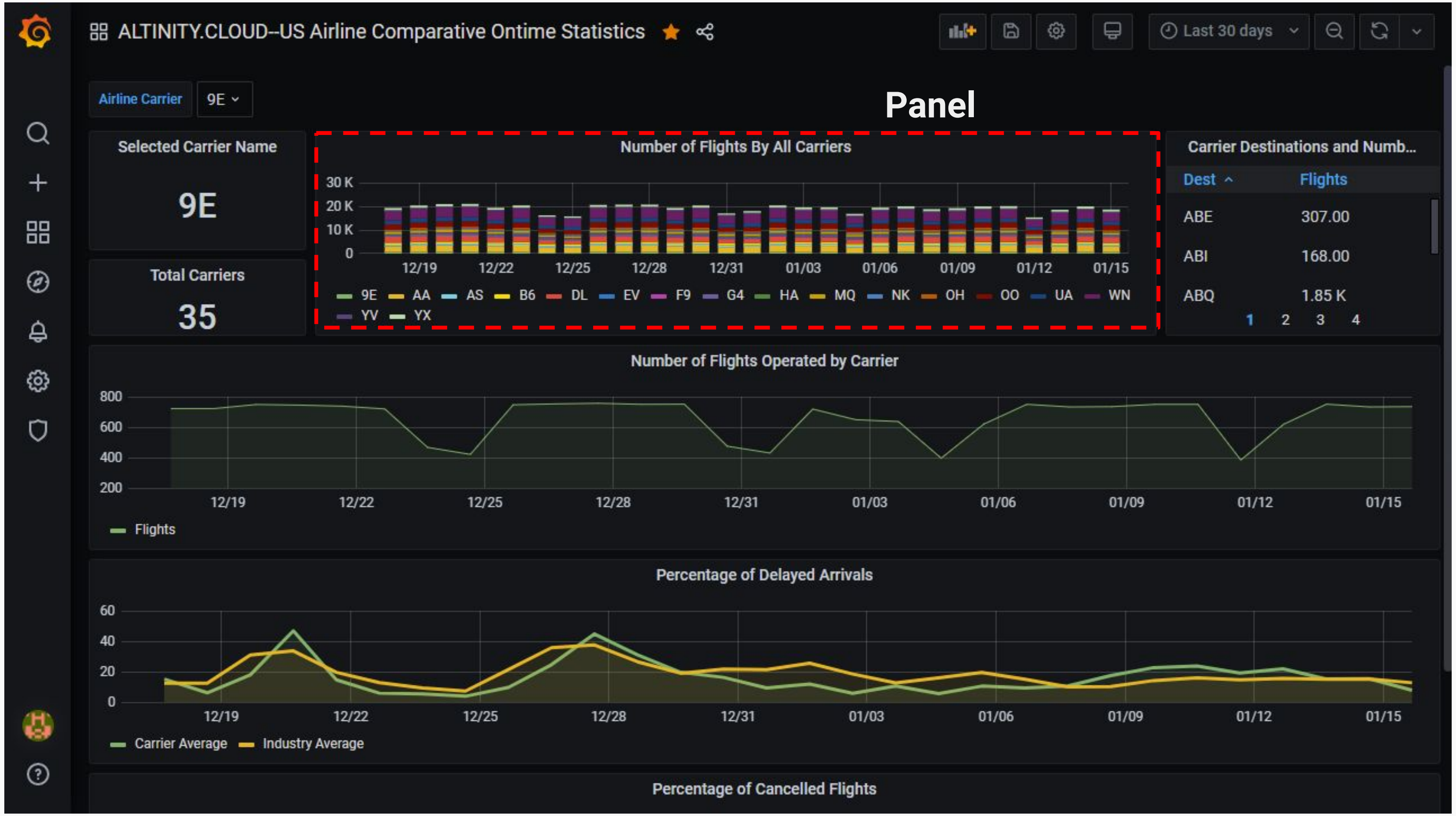
https://github.com/Vertamedia/clickhouse-grafana
Apache 2.0
Developed by Roman Khavronenko

- Distributed as vertamedia-clickhouse-datasource on grafana.com
- Uses ClickHouse HTTP Interface
- TLS Support
- Current release on https://grafana.com: 2.3.1
- Maintained by Altinity

Grafana dashboard organization



Typical Grafana Dashboard


ALTINITY.CLOUD--US Airline Comparative Ontime Statistics
Last 30 days
Airline Carrier
9E
Panel
Selected Carrier Name
9E
Total Carriers
35
Number of Flights By All Carriers
9E AA AS B6 DL EV F9 G4 HA MQ NK OH OO UA WN YV YX
Carrier Destinations and Numb...
Number of Flights Operated by Carrier
Flights
Percentage of Delayed Arrivals
Carrier Average
Industry Average
Percentage of Cancelled Flights

Dest	Flights
ABE	307.00
ABI	168.00
ABQ	1.85 K

1 2 3 4

Altinity

Altinity.Cloud provides a test endpoint

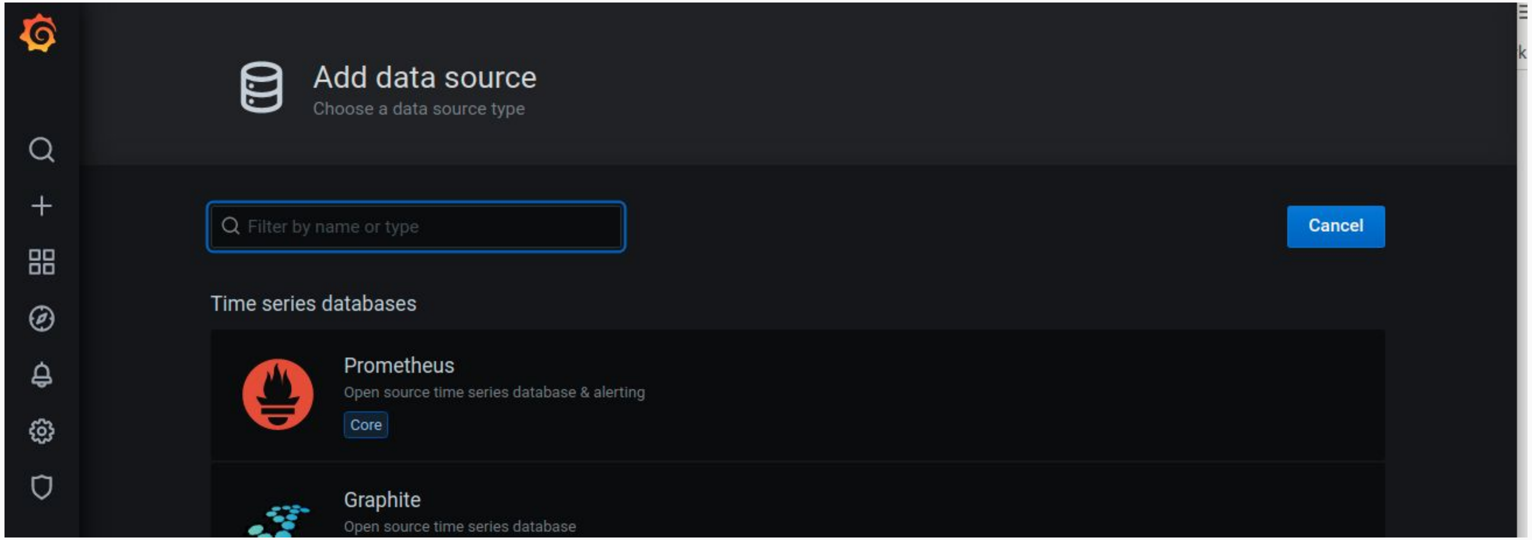
Connection parameters

URL	https://github.demo.trial.altinity.cloud:8443
User	demo
PW	demo

Datasets

airports	Airport names and locations
github_events	Full event history from Github (3.1B rows)
ontime	Airline ontime data (196M rows)
tripdata	NYC taxi commission ride data (1.3B rows)

Adding the ClickHouse data source



Add data source

Choose a data source type

Time series databases

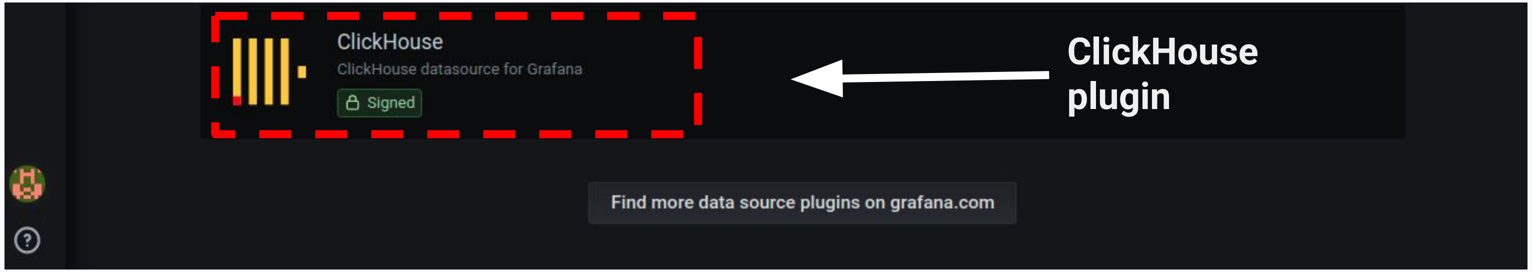
Prometheus

Open source time series database & alerting

Graphite

Open source time series database

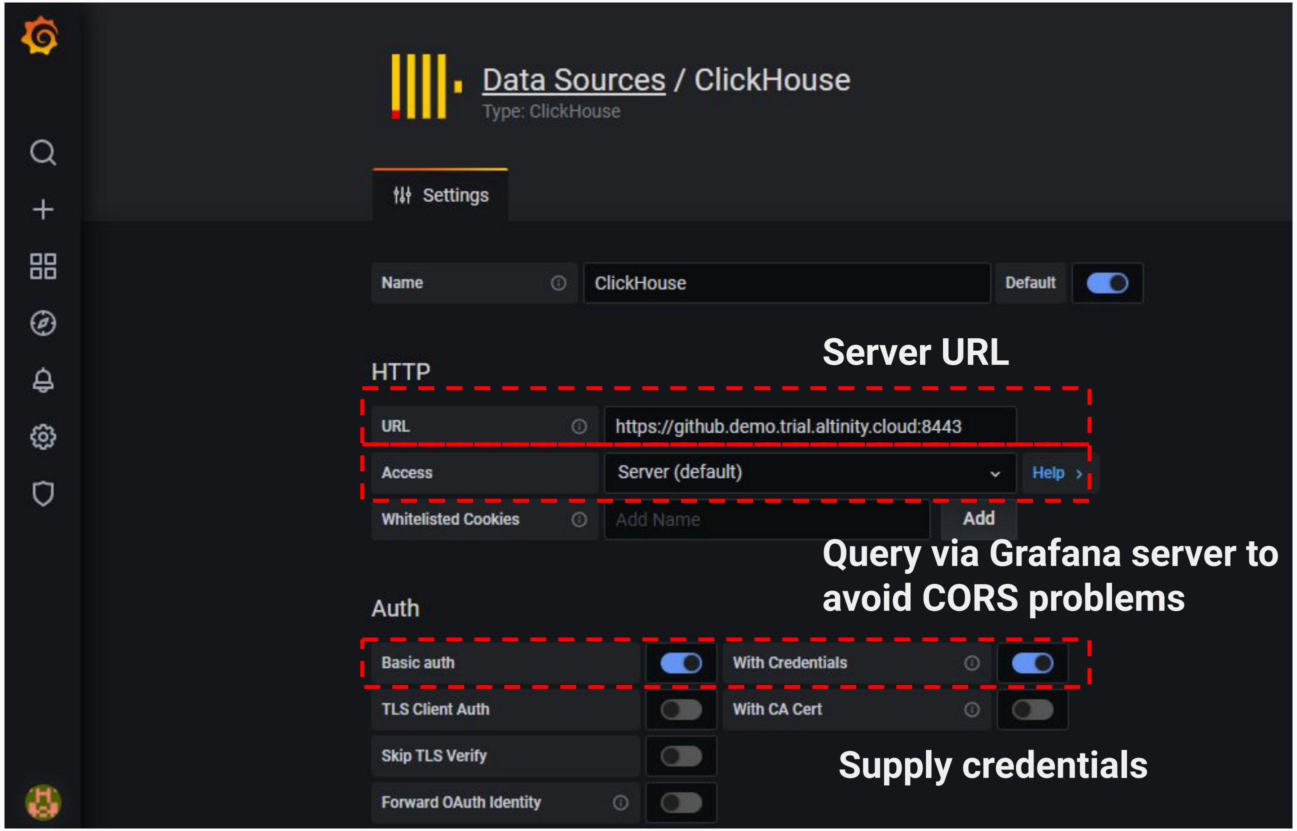
• • •



ClickHouse

ClickHouse datasource for Grafana

Defining data source (1)



Data Sources / ClickHouse
Type: ClickHouse

Settings

Name	ClickHouse	Default	●

HTTP

URL	https://github.demo.trial.altinity.cloud:8443	
Access	Server (default)	Help
Whitelisted Cookies	Add Name	Add

Auth

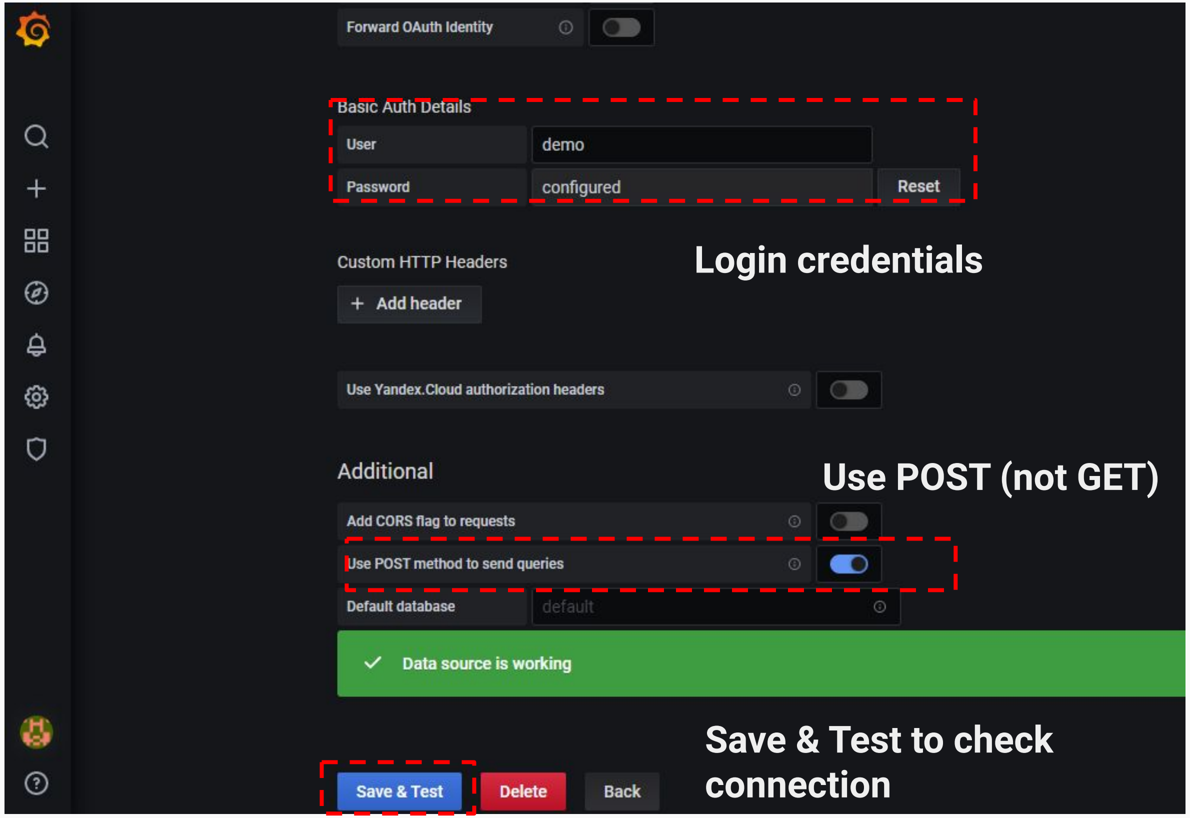
Basic auth	●	With Credentials	●
TLS Client Auth	○	With CA Cert	○
Skip TLS Verify	○		
Forward OAuth Identity	○		

Server URL

Query via Grafana server to avoid CORS problems

Supply credentials

Defining data source (2)



Forward OAuth Identity

Basic Auth Details

User	demo
Password	configured

Custom HTTP Headers

Use Yandex.Cloud authorization headers

Additional

Add CORS flag to requests

Use POST method to send queries

Default database	default

✓ Data source is working

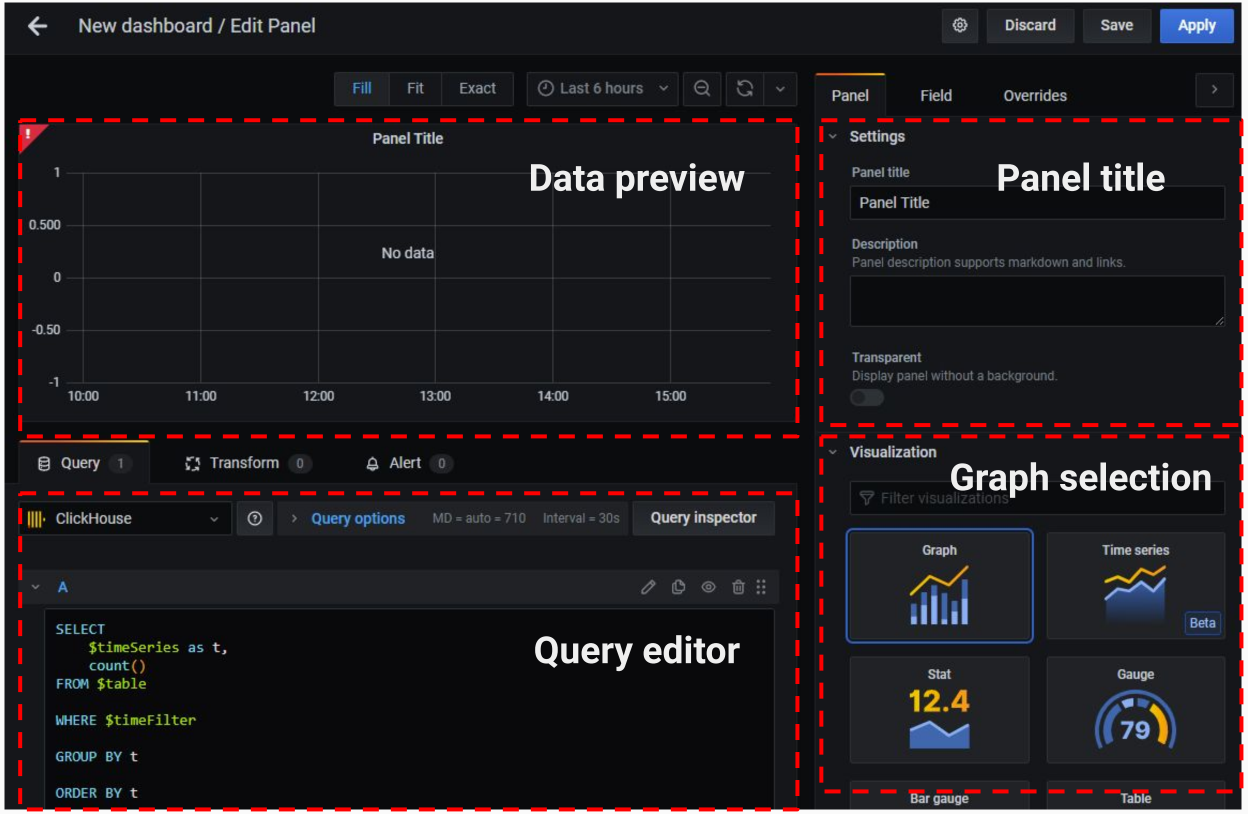
Login credentials

Use POST (not GET)

Save & Test to check connection

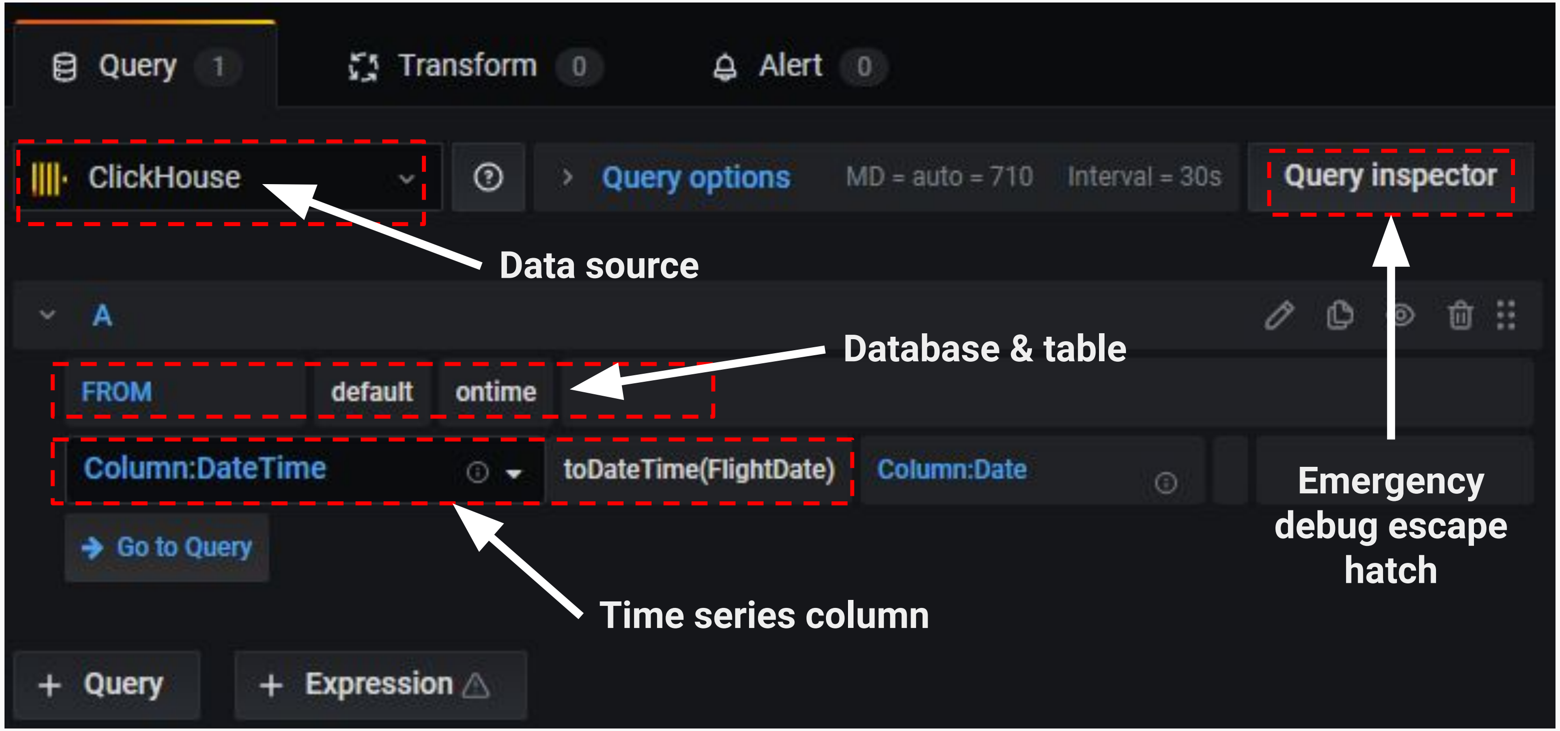
Altinity

Creating a time series graph

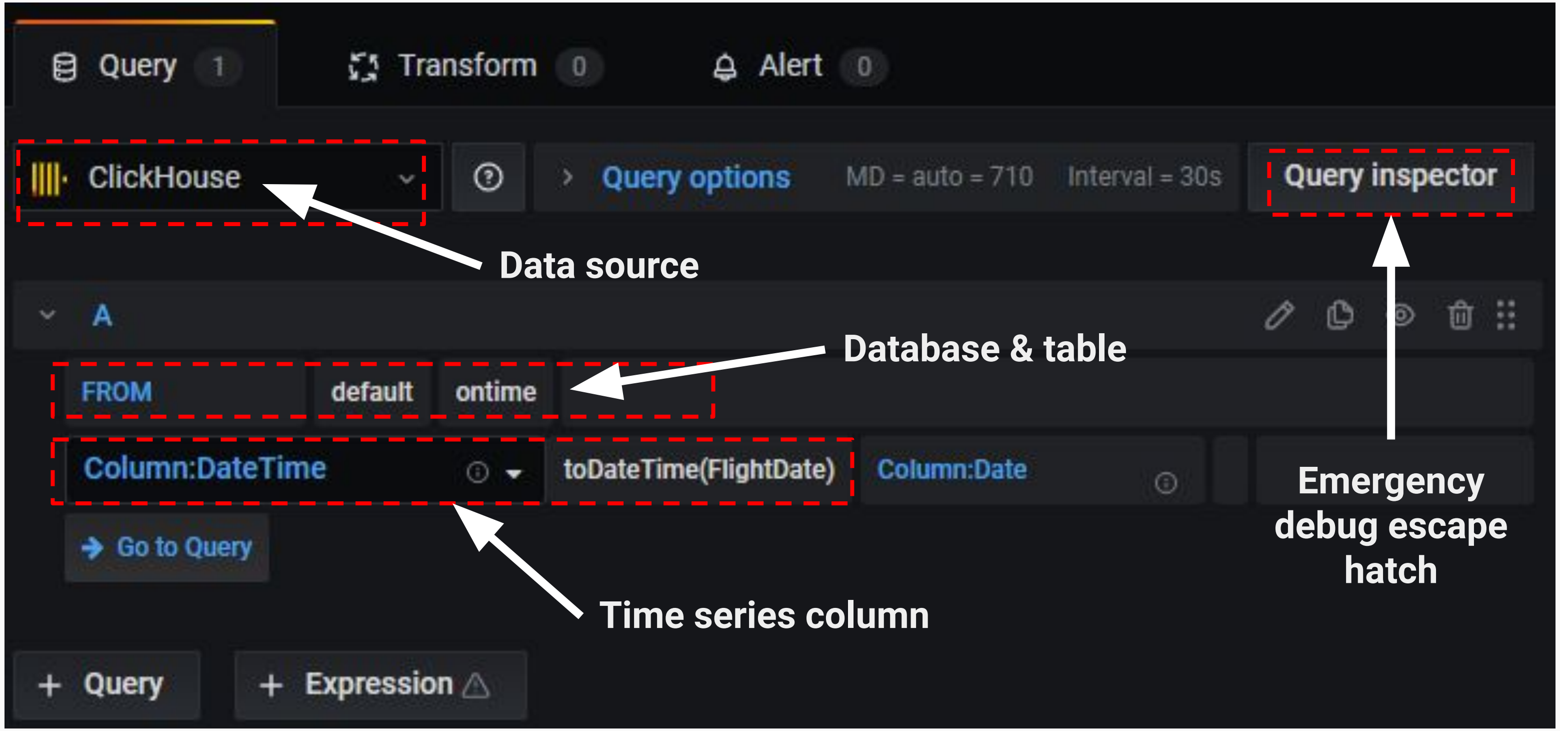


```
SELECT
    $timeSeries as t,
    count()
FROM $table

WHERE $timeFilter

GROUP BY t

ORDER BY t
```

Defining data source and time series

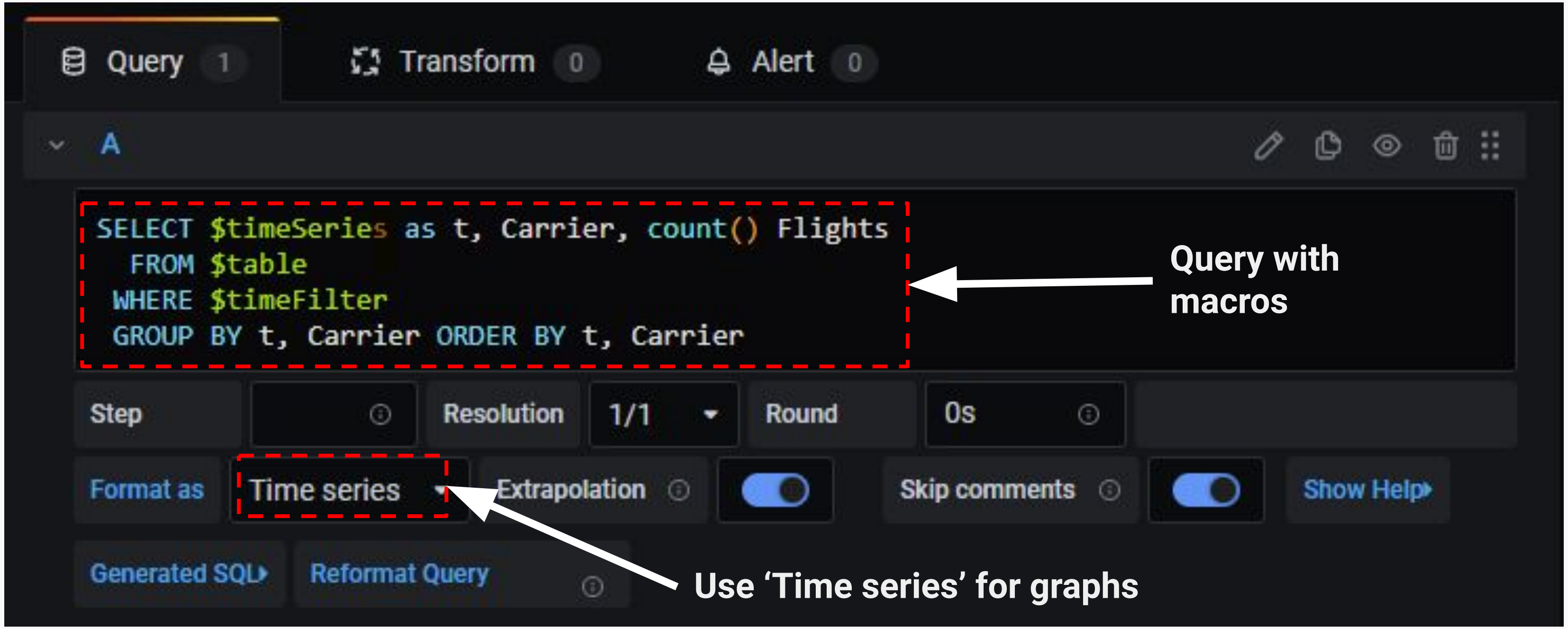


Defining data source and time series



Altinity

Defining the query



```
SELECT $timeSeries as t, Carrier, count() Flights
  FROM $table
 WHERE $timeFilter
 GROUP BY t, Carrier ORDER BY t, Carrier
```

Step:
Resolution: 1/1
Round: 0s
Format as: Time series
Extrapolation: on
Skip comments: on

Query with macros

Use 'Time series' for graphs

Grafana as a BI tool

Strengths

- Outstanding time series support
- Tight integration with non-SQL data sources like Prometheus
- Interactive drill-down on data
- Supports alerting
- Great for operational analytics

Possible Weaknesses

- Requires advanced SQL coding skills
- No reuse of code or panel definitions
- Every query hits the server (no caching)
- Limited number of visualization types

More information

Grafana Website and Documentation

- Grafana documentation
- ClickHouse data source documentation

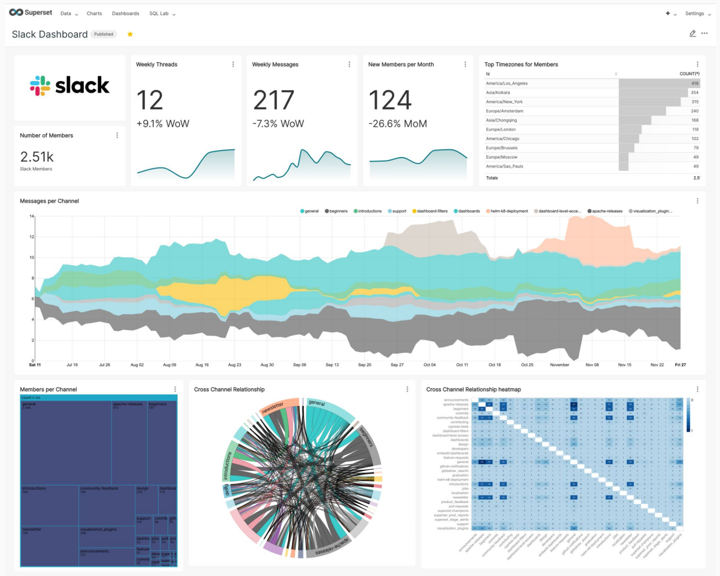
Installing and using Grafana with ClickHouse

- Creating Beautiful Grafana Dashboards on ClickHouse: a Tutorial
- US Airline Comparative Ontime Statistics Dashboard

Superset

Altinity

What is Apache Superset?



Top Timezones for Members

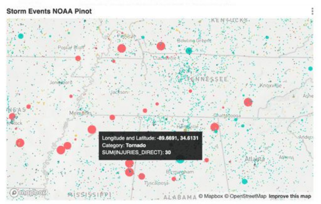
tz	COUNT(*)
America/Los_Angeles	416
Asia/Kolkata	354
America/New_York	315
Europe/Amsterdam	240
Asia/Chongqing	168
Europe/London	118
America/Chicago	102
Europe/Brussels	79
Europe/Moscow	49
America/Sao_Paulo	49
Totals	2.5

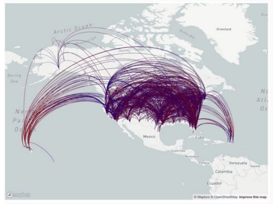
Modern open source BI platform

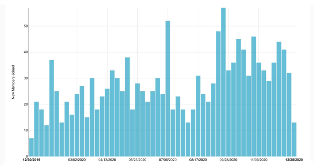
(Thanks to Srini Kadamati @ Preset)

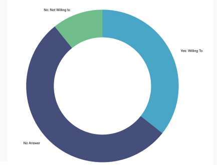
snowflake
Google BigQuery
amazon REDSHIFT
dremio
druid
presto
ROCKSET
elasticsearch
Amazon Athena
trino
VERTICA
PostgreSQL
SQL Server
MySQL
MariaDB
ClickHouse
Exasol
Greenplum
IBM DB2
monetdb
SQLite
SYBASE
ORACLE

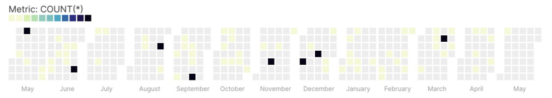
Works with nearly any SQL speaking data engine









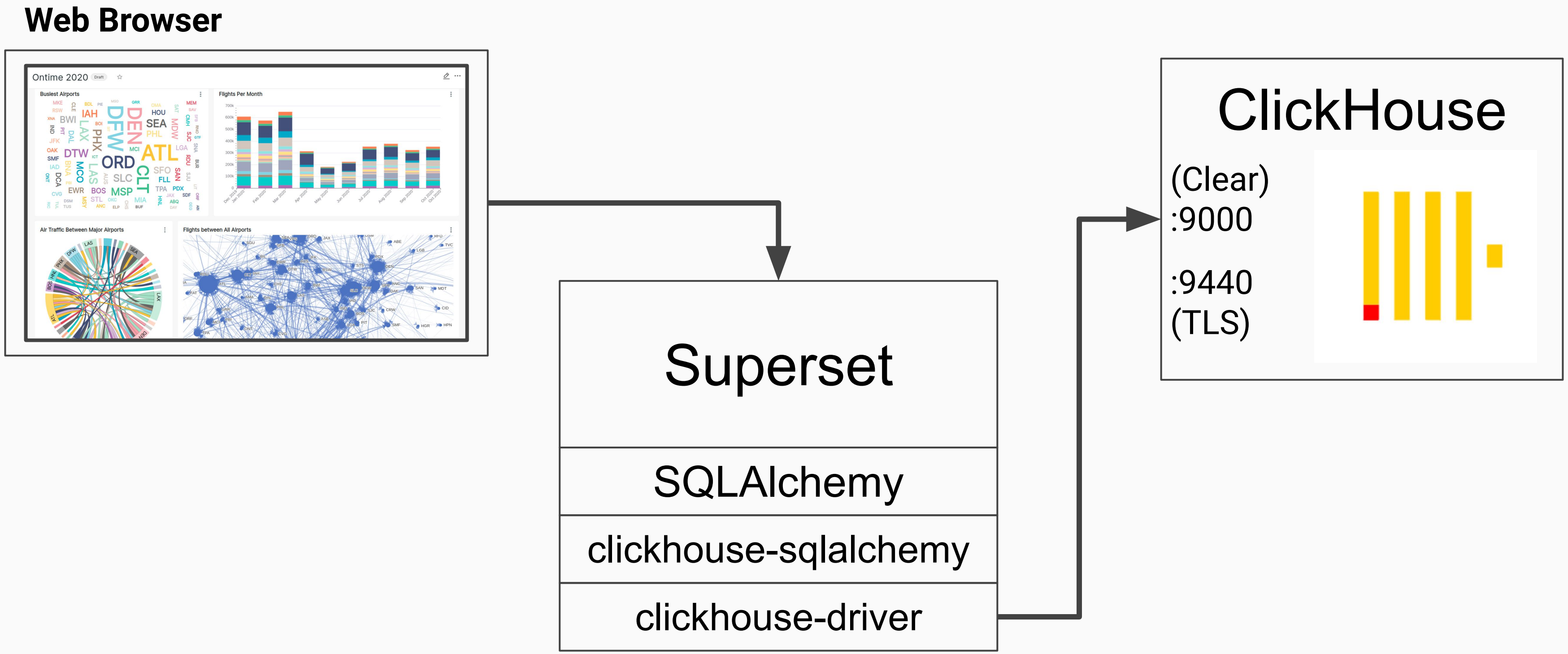


Large diversity of charts

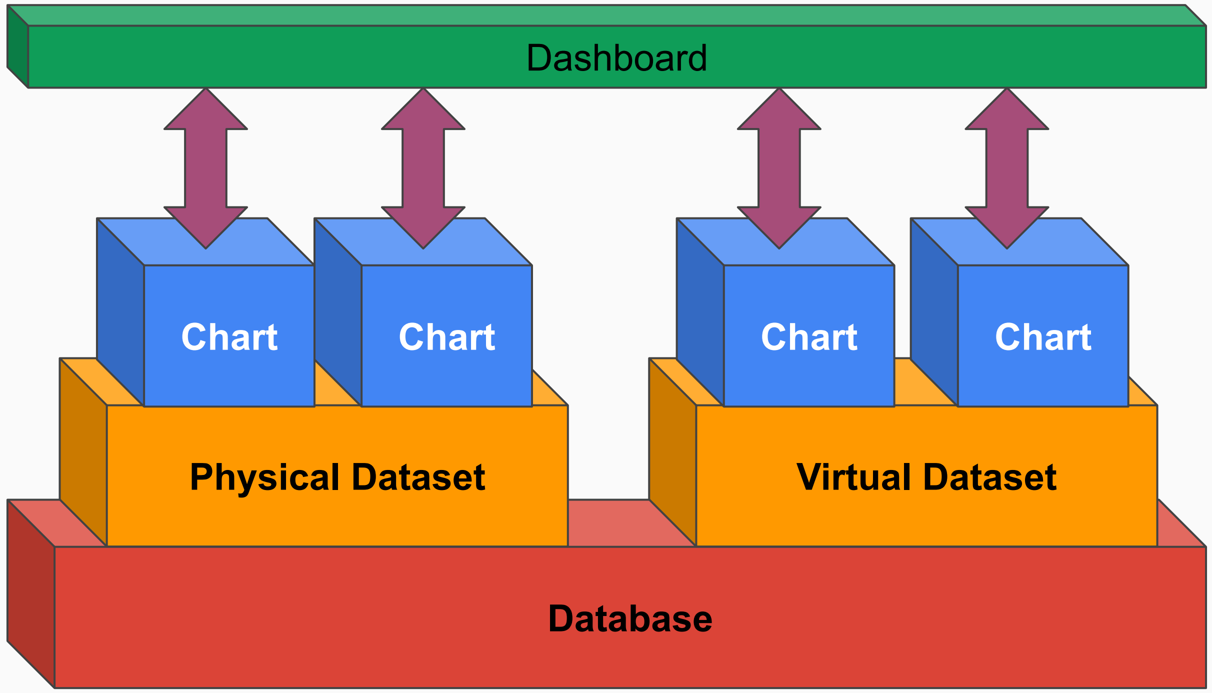
Altinity

preset

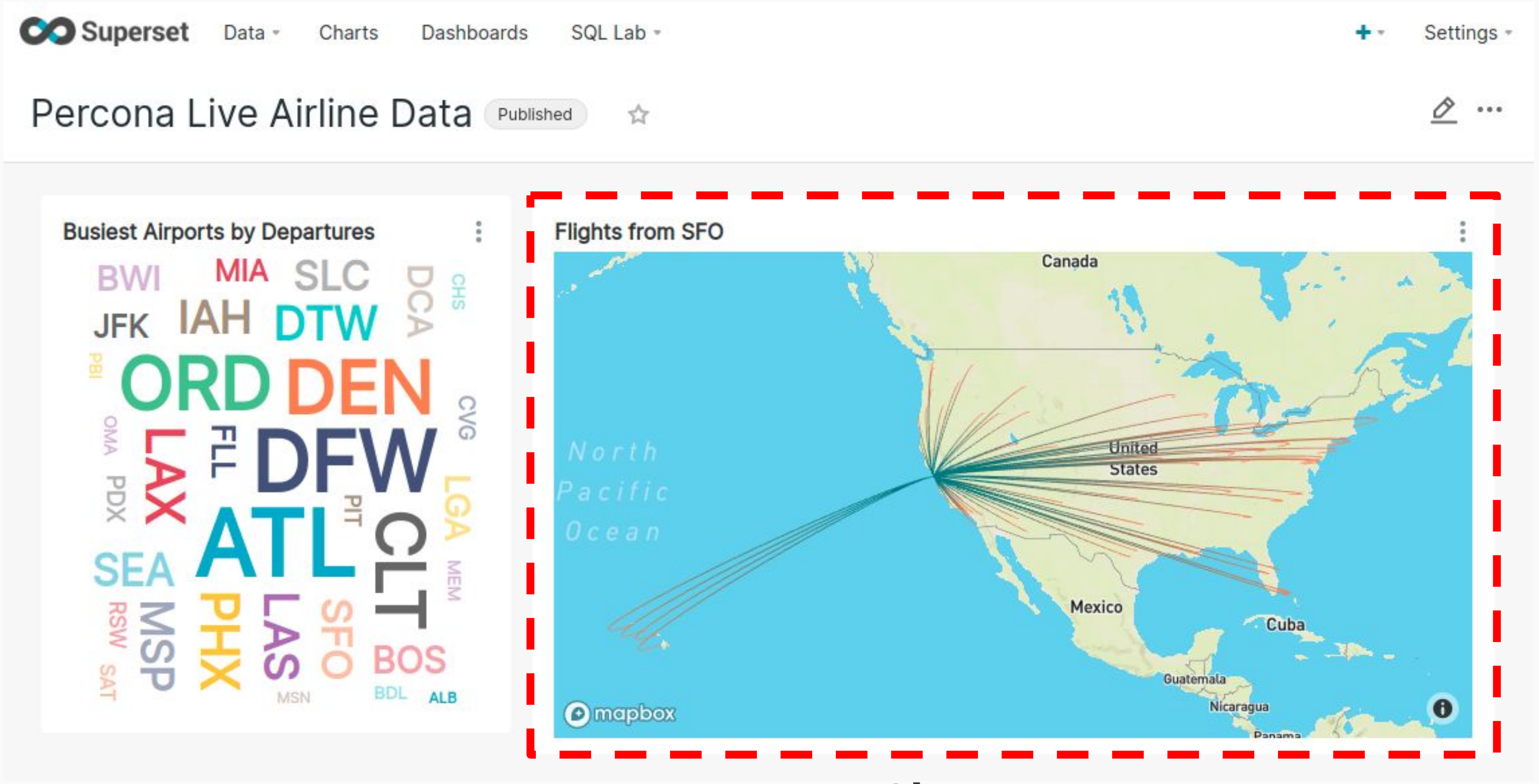
Superset connection to ClickHouse



Superset dashboard organization

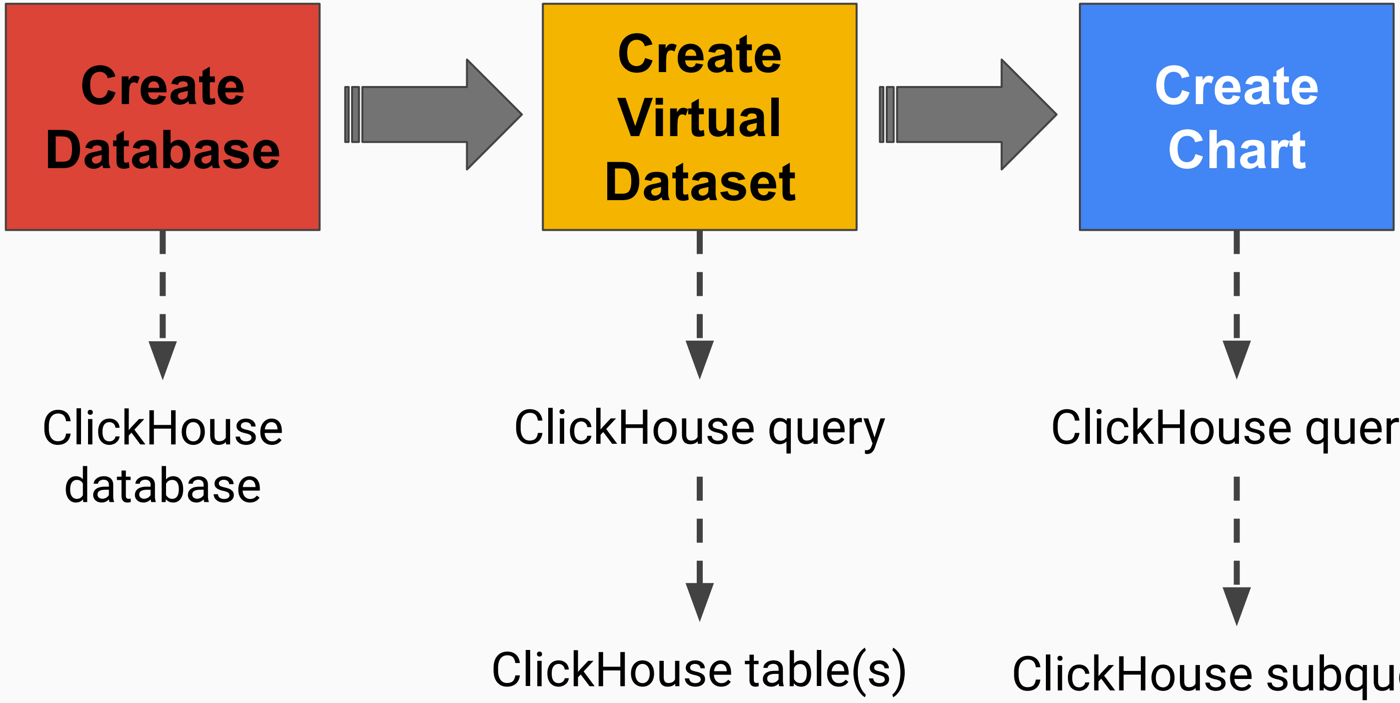


A simple dashboard in Superset

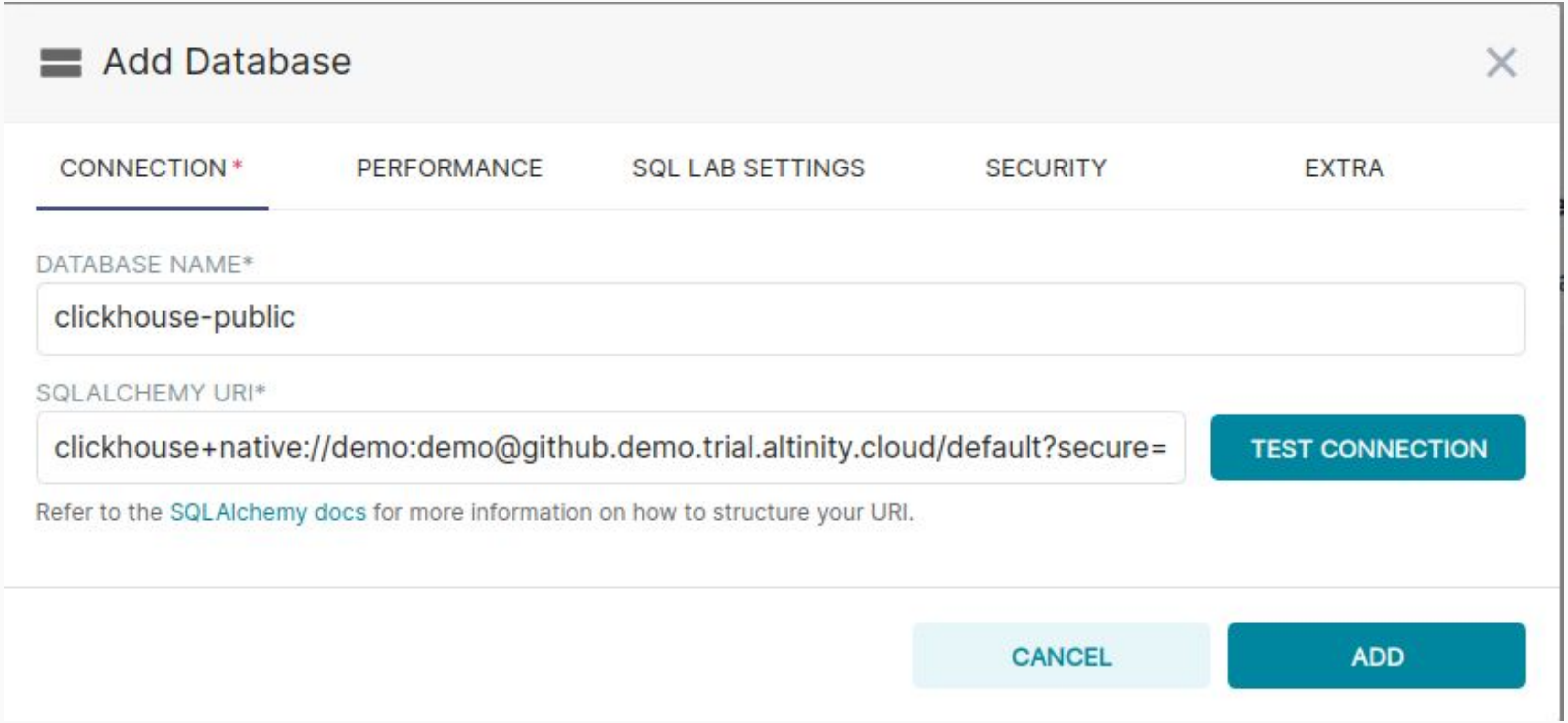

Superset
Data
Charts
Dashboards
SQL Lab
Settings
Percona Live Airline Data
Published
Busiest Airports by Departures
BWI MIA SLC DCA CHS JFK IAH DTW PBI ORD DEN CVG OMA LAX FLL DFW PDX PIT LGA SEA ATL CLT MEM RSW MSP PHX LAS SFO BOS SAT MSN BDL ALB
Flights from SFO
Canada
North Pacific Ocean
United States
Mexico
Cuba
Guatemala
Nicaragua
Panama
mapbox

Chart

Creating a chart on a virtual dataset



Database connection page


Add Database

DATABASE NAME*

clickhouse-public

SQLALCHEMY URI*

clickhouse+native://demo:demo@github.demo.trial.altinity.cloud/default?secure=

Refer to the SQLAlchemy docs for more information on how to structure your URI.

ClickHouse SQLAlchemy Drivers

Preferred

clickhouse-sqlalchemy
Apache 2.0
Developed by Konstantin Lebedev

- Uses ClickHouse Native TCP
- TLS support
- Bug fixes for Superset
- Current pypi.org release: 0.1.6, Mar 15 2021
- Supported by Altinity

Deprecated

sqlalchemy-clickhouse
Apache 2.0
Developed by Marek Vavrusa

- Currently documented in Superset
- Uses ClickHouse HTTP Interface
- No TLS support
- Current pypi.org release: 0.1.5.post0, Aug 9 2018

ClickHouse connection strings

SQLAlchemy URL format:

```
clickhouse+native://[user:pw]@host[:port]/database[?options…]
```

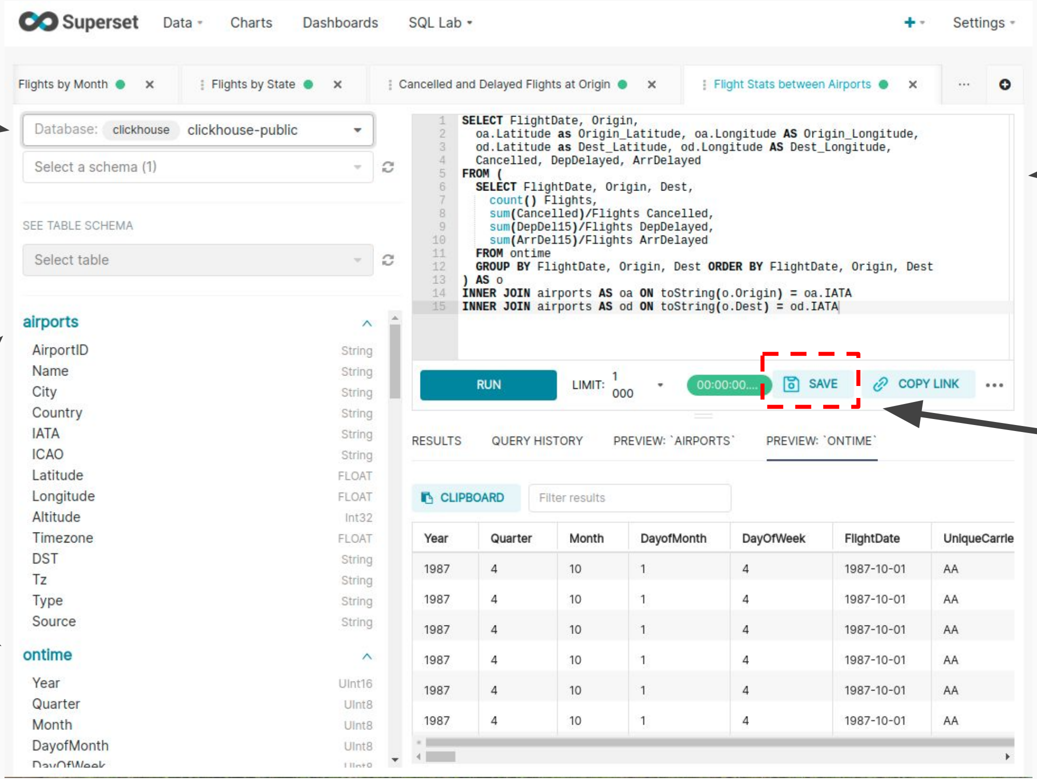
ClickHouse on localhost (e.g., your laptop)

```
clickhouse+native://localhost/default
```

ClickHouse public endpoint:

```
clickhouse+native://demo:demo@github.demo.trial.altinity.clou
d/default?secure=true
```

Build, run, and save query in SQL Lab



airports

AirportID	String
Name	String
City	String
Country	String
IATA	String
ICAO	String
Latitude	FLOAT
Longitude	FLOAT
Altitude	Int32
Timezone	FLOAT
DST	String
Tz	String
Type	String
Source	String

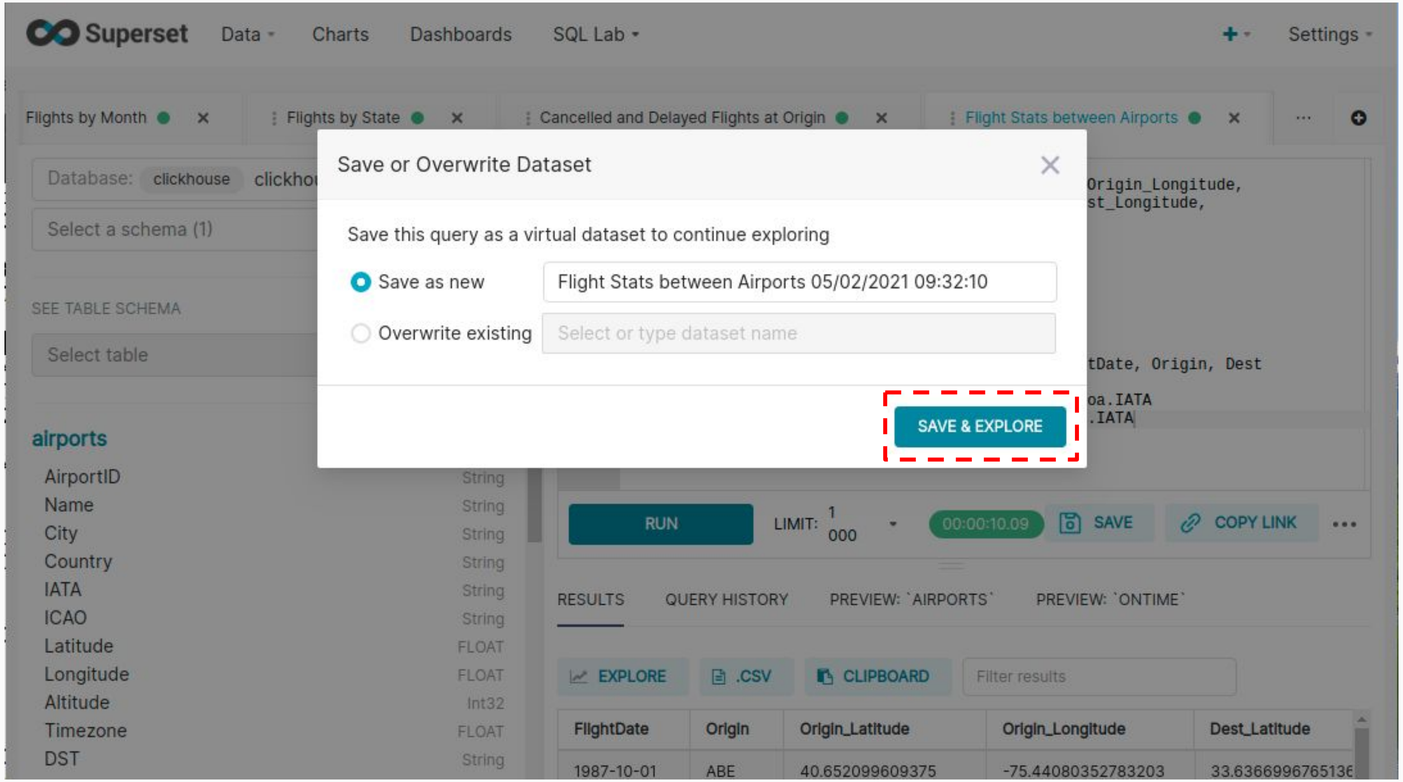
ontime

Year	UInt16
Quarter	UInt8
Month	UInt8
DayofMonth	UInt8

```
SELECT FlightDate, Origin,
  oa.Latitude as Origin_Latitude, oa.Longitude AS Origin_Longitude,
  od.Latitude as Dest_Latitude, od.Longitude AS Dest_Longitude,
  Cancelled, DepDelayed, ArrDelayed
FROM (
  SELECT FlightDate, Origin, Dest,
    count() Flights,
    sum(Cancelled)/Flights Cancelled,
    sum(DepDel15)/Flights DepDelayed,
    sum(ArrDel15)/Flights ArrDelayed
  FROM ontime
  GROUP BY FlightDate, Origin, Dest ORDER BY FlightDate, Origin, Dest
) AS o
INNER JOIN airports AS oa ON toString(o.Origin) = oa.IATA
INNER JOIN airports AS od ON toString(o.Dest) = od.IATA
```

Year	Quarter	Month	DayofMonth	DayOfWeek	FlightDate	UniqueCarrie
1987	4	10	1	4	1987-10-01	AA
1987	4	10	1	4	1987-10-01	AA
1987	4	10	1	4	1987-10-01	AA
1987	4	10	1	4	1987-10-01	AA
1987	4	10	1	4	1987-10-01	AA
1987	4	10	1	4	1987-10-01	AA

Use EXPLORE to save as dataset



Save or Overwrite Dataset

Save this query as a virtual dataset to continue exploring

[✓ Save as new] Flight Stats between Airports 05/02/2021 09:32:10

Overwrite existing Select or type dataset name

```
Origin_Longitude,
st_Longitude,

tDate, Origin, Dest

oa.IATA
.IATA
```

airports

AirportID	String
Name	String
City	String
Country	String
IATA	String
ICAO	String
Latitude	FLOAT
Longitude	FLOAT
Altitude	Int32
Timezone	FLOAT
DST	String

FlightDate	Origin	Origin_Latitude	Origin_Longitude	Dest_Latitude
1987-10-01	ABE	40.652099609375	-75.44080352783203	33.63669967651З

Creating a deck.gl Arc chart

Chart Type*

* Requires a Mapbox token -- See docs

Time Dimension

Lat/Long

Filter

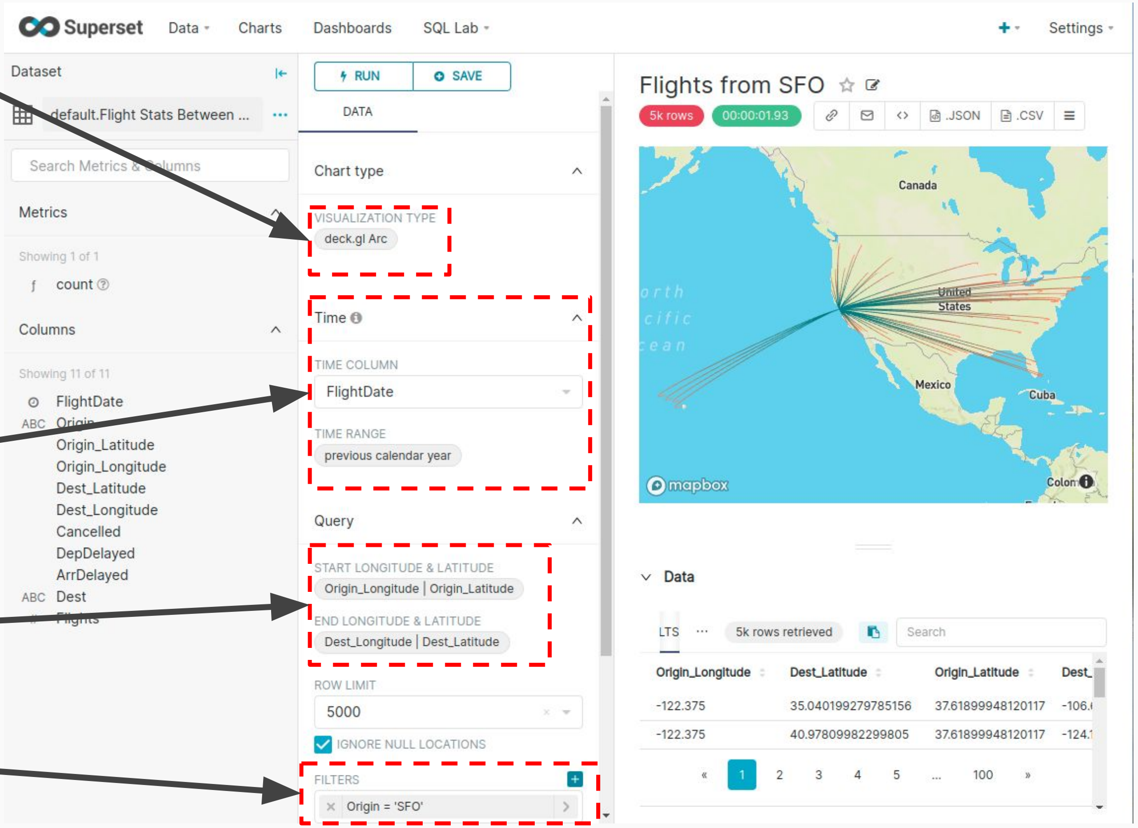


Chart type

VISUALIZATION TYPE: deck.gl Arc

Time

TIME COLUMN: FlightDate

TIME RANGE: previous calendar year

Query

START LONGITUDE & LATITUDE: Origin_Longitude | Origin_Latitude

END LONGITUDE & LATITUDE: Dest_Longitude | Dest_Latitude

ROW LIMIT: 5000

[✓] IGNORE NULL LOCATIONS

FILTERS: Origin = 'SFO'

Origin_Longitude	Dest_Latitude	Origin_Latitude	Dest_
-122.375	35.040199279785156	37.61899948120117	-106.
-122.375	40.97809982299805	37.61899948120117	-124.

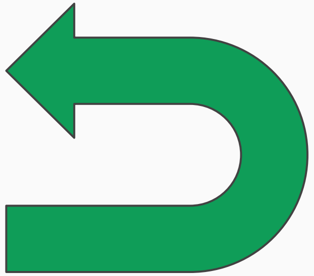
« 1 2 3 4 5 ... 100 »

How Superset queries virtual datasets

```
SELECT "Origin_Longitude" AS "Origin_Longitude",
       "Dest_Latitude" AS "Dest_Latitude",
       "Origin_Latitude" AS "Origin_Latitude",
       "Dest_Longitude" AS "Dest_Longitude"
FROM
  (
```

Dataset subquery

```
  ) AS expr_qry
WHERE "FlightDate" >= toDate('2020-01-01')
  AND "FlightDate" < toDate('2021-01-01')
  AND "Origin" = 'SFO'
  AND "Dest_Latitude" IS NOT NULL AND "Dest_Longitude" IS NOT NULL
  AND "Origin_Latitude" IS NOT NULL AND "Origin_Longitude" IS NOT NULL
LIMIT 5000;
```



Filters pushed down to base table

Superset as a BI tool

Strengths

- Good time series support
- Dozens of interesting charts with more constantly being added
- Low-code/no-code charting
- Clean semantic layering with code reuse
- Caching
- Good for business analytics

Possible Weaknesses

- Map visualizations difficult to use
- SQL data sources
- Limited interactive manipulation/drill-down on time series data

More information

Superset Documentation

- Apache Superset documentation
- Preset documentation on Superset (Supports Superset commercially)

Installing and using Superset with ClickHouse

- Visualizing ClickHouse Data with Apache Superset
    - Part 1: Installation
    - Part 2: Dashboards
- Altinity docs: Integrating Superset and ClickHouse

What else?

Altinity

Other BI tools that support ClickHouse

- Looker - https://looker.com/
- Metabase - https://www.metabase.com/
- Redash - https://redash.io/
- Seektable - https://www.seektable.com

Questions?

Thank you

P.s. We’re hiring!!!

Altinity

https://altinity.com

Grafana

https://grafana.com/

Superset

https://superset.apache.org/

https://preset.io

Tableau

https://tableau.com