# Data Warehouse Basics Interactive Exercises

[Data Warehouse Basics Interactive Exercises](#_y3wuuxz21c42)

[Lab #1](#_ilgqeb1fx625)

[Lab #2](#_adv1hx6qngf9)

## Lab #1

* Connect to ClickHouse with your favorite query tool
  + clickhouse-client
  + ClickHouse play interface
  + Any other tool you have available, such as DBeaver
* Find out how many databases there are
* List the tables in the default database
* Describe the columns of the ontime table

Hints:

Cloud ClickHouse URL: <https://github.demo.trial.altinity.cloud:8443/play>

User: demo

Password: demo

|  |
| --- |
| **show databases** |
| **use** *name* |
| **show tables [from** *name***]** |
| **describe table** *name* |
| **select count() from** *name* |

## Lab #2

1. Create a report showing the number of flights each year.
2. What were the busiest airports in 2017 measured by number of departing flights?
   1. Can you also print the airport name?
3. Which airport had the highest departure delay in 2017?
   1. Same question as above but restrict to airports with more than 100K flights.
4. Which airline carrier has the record for most flights in a single day?
5. Can you show the airport name, number of flights, latitude, and longitude of every airport in 2020 with flights from San Francisco International Airport (SFO) or Portland (PDX)?

Hints:

**-- Most cancelled flights in 2017.**

**SELECT Carrier,**

**toYear(FlightDate) AS Year,**

**sum(Cancelled)/count(\*) \* 100. AS cancelled**

**FROM ontime**

**WHERE Year = 2017**

**GROUP BY Carrier, Year**

**HAVING cancelled > 1.0**

**ORDER BY cancelled DESC**

**LIMIT 10**

**-- Where clause examples.**

**SELECT**

**toStartOfMonth(FlightDate) AS Month,**

**Carrier, count() AS Flights**

**FROM ontime**

**WHERE**

**FlightDate BETWEEN toDate('2015-01-01')**

**AND toDate('2015-06-30')**

**AND Carrier != 'AA'**

**AND Dest IN ('SFO', 'ORD', 'JFK')**

**GROUP BY Month, Carrier ORDER BY Month, Carrier**

**-- Filtering out measurements.**

**SELECT**

**Year,**

**Origin,**

**COUNT() AS Flights,**

**AVG(DepDelayMinutes) AS Delay**

**FROM ontime**

**GROUP BY Year, Origin HAVING Flights >= 100000 ORDER BY Year, Flights DESC**

**-- Join that adds airport name.**

**SELECT o.Dest,**

**any(a.Name) AS AirportName,**

**count(Dest) AS Flights**

**FROM ontime o**

**RIGHT JOIN airports a ON a.IATA = toString(o.Dest)**

**GROUP BY Dest**

**ORDER BY Flights**

**DESC LIMIT 10**