Postgres (Databases - RDMS):

1. Database components/layout (transport/network layer, interpretation layer, execution engine, storage)
    a. http://www.nhu.edu.tw/~chun/CS-ch14-Databases.pdf
    b. B Tree
    c. Example storage layer: https://www.cs.umb.edu/~poneil/lsmtree.pdf

2. Postgres vs Mysql: https://blog.panoply.io/postgresql-vs.-mysql

3. Schema/Table Definition, Constraints (indices [primary, secondaries]), Access/Permissioning
    a. Use shop; /c people(ssn, fname, lname)
    b. Alter shop.people add primary_key etc
4. Different types of dbs: RDMS, NoSQL, OLTP, OLAP

Python:

1. Python vs Java vs C++ vs Go vs Rust - why do data engineers prefer python?
2. Comprehension : list, dictionaries

```
vals = [1,2,3,4,5]

sums = [1,3,6,10,15]
sums = [sum(vals[:i+1]) for i, x in enumerate(vals)]

{1:1, 2:3, 3:6, 4:10, 5:15}
sums_dict = {x:sum(vals[:i+1]) for i, x in enumerate(vals)}
```

Airflow:

What is airflow? Why would I use airflow and not standard sql?
cron vs Airflow vs Luigi vs Matillion vs SparkContext vs AWS Step Functions…
What is a DAG in ETL? What is a DAG in airflow?

users, products -> clean data -> user+products+mapping -> sales_report

Spark:

SQL:

Product tbl [p_name(P), weight]

People tbl [ssn(P), f_name(S), l_name]

people_product_map tbl [ssn(P), p_name(P)]

1,	abcd
1,	bcde
1,	cdef
2,	abcd
3,	bcde
4,	cdef

 # products bought by people

```
select
  *
from people as p
join people_product_map ppm
        on p.ssn=ppm.ssn
join product t
        on ppm.pname = t.pname
```