

# Technical interview questions

Technical interview questions are where I fail so work hard here!

This document details questions and answers to technical interview questions, both real and for myself.

## Interview questions

### LLMs

1. What is an embedding?
2. What are

### Software development

1. What are unit tests?
2. What are integration tests?

### Niche Python features

### Data engineering

1. What is infrastructure as code?
2. What is a data lake?
3. What is a data warehouse?

### SQL

## Real interview questions

### Pre-interview Python Developer interview questions

1. Can you add a set within a set in python? why? if not, how could you do it?

A set can only be within a set if it is a frozenset. This is because all items within a set must be hashable.

2. You need to download 20 files from a website at the same time in python, can you do it? what do you use?

Python is typically single threaded, but you could use multithreading or run multiple python scripts to do this. Also asynchronous requests might achieve an analogous goal.

3. what are the differences between a static method and a class method?

A class method takes the class as an argument and the static method doesn't. The class method can operate on the class/object it's been instantiated by.

4. Can you make a class attribute actually private in python?

Because of how Python works you can't truly make something private.

5. How can you redefine an operator in python for a given class, ie class A: so  $A() + A()$  is doable?

You implement the operator as a method within the class.

6. How do you make a class hashable in python?

Implement the hash method

7. For function arguments in python, are they passed by value or by reference?

Reference

8. when do you use a function and when do you use a class with a method?

It's dependent on whether you're performing an operation with a class/classes or not.

## **Quantitative Developer Phone Screen**

### **Python**

1. What are unit tests?

- Pandas has a testing suite
- You can use unittest and pytest to write unit tests in Python.

### **Pandas**

1. What are some issues with working with dataframes?

- Referential integrity, can be managed with copying dataframes

2. What is the difference between a merge and a concat?

- concat adds both dataframes together
- merge performs joins -> SQL manipulation (Left, Right, Inner, Outer)

### **Brainteasers**

1. There are two painters, p1 and p2. If p1 paints a room in 2 hrs and p2 paints a room in 4 hrs, how long would they both take to paint the same room?

- Start by thinking about how much they could paint in 1 hour.
- p1 and p2 can paint 3/4 of the room in 1 hour, therefore they can paint 1/4 of the room in 20 minutes
- The total time is 1 hr and 20 minutes

2. How many zeros in 100 factorial?

- Number of factors of 10 and 5 AS well as number of squares of 5
- 24

## **Data Engineer Phone Screen**

1. What is the difference between OLAP and OLTP?

- OnLine Analytical Processing (OLAP)
  1. Built to run large analytics queries and is typically inefficient at handling lookups of individual records.
  2. Modern column databases are optimised to scan large volumes of data, dispensing with indexes to improve scalability and scan performance.
- OnLine Transaction Processing (OLTP) system
  1. A database that reads/writes records at a high rate.

2. Often referred to as transactional databases, but does not imply that the system in question supports *atomic transactions*.
  3. Generally speaking OLTP databases support low latency and high concurrency.
  4. OLTP databases work well for application backends
  5. Small companies might run analytics on OLTP systems but as it scales runs into resource issues
2. What does ACID mean?
- Support for atomic transactions - Atomicity, Consistency, Isolation and Durability
  - Atomicity: an atomic transaction is a set of several changes that are committed as a unit
  - Consistency: any database read will return the last written version of the retrieved item
  - Isolation: if two updates are in flight concurrently for the same thing, the end database state will be consistent with the sequential execution of these updates in the order they were submitted.
  - Durability: indicates that the data will never be lost, even in the event of power loss
3. What is the difference between NoSQL and SQL?
- NoSQL
    - Non relational databases -> refers to multiple databases that abandon the relational paradigm.
    - key-value: non relational database that retrieves records using a key
    - document stores: a document store is a specialized key-value store. A document is a nested json object
    - wide-column: optimised for storing massive amounts of data with high transaction rates and extremely low latency
    - graph databases: store data with a mathematical graph structure
    - search databases: used to search datas complex and straightforwar semantic and structural characteristics
    - time serires: optimised for retrieving and statistical processing of time series data
  - SQL
    - data is stored in a table of relations (rows) and each row contains multiple fields (columns)
    - each relation in the table has the same schema
    - indexed by primary key
    - foreign keys relate tables to each other
4. What is the difference between document databases and data lakes?
- Data lake is a location where all data can be stored regardless of schema
  - e.g. HDFS