Database section

**What are file storage options?**

**Usually blob storage** here, Amazon S3, can also use CDN with this as well, to serve static image, distrbiute image across the globe so it can be faster for storage

What are the 3 top choices for database?

1. Mysql, sql
2. Nosql : couchbase, Mongodb
3. Ever increasing nosql (Columnar DB)   
   Cassandra Hbase

What are the 3 types of consistency? Can you give examples and let me know which one is suited for your particular database?

Strong consistency, eventual consistency and weak consistency

How does nosql query data for multiple tables?

* Embedded documents (a document within a document)

What are the 2 types of locking

1. Optimisitc and pessimistic

Optimistic is used with version here

Can look at hotel booking system for the applicatino of this?

How does full text search like Elastic search gets integrated with postgresql?

- an example can be found in the ticketMaster example, we must keep our elastic search in sync with the database here

-

Redis section

1. How do we configure the data in redis with a ttl

Check ticket master example in the answers sheet

**Why is consistent hashing so important?**

It’s the technique to achieve even distrubtion of data in horizontal scaling across servers or database serverss, this is why this is important.

What is it basically?

* Map servers and keys on to the ring using a uniformly distributed hash function.
* To find out which server a key is mapped to, go clockwise from the key position until the first server on the ring is found.
* When new server is added, only one key needs to be remapped here

**If we have a non-unifomr key distribution on the ring, how do we solve this problem?**

This is done using vNodes on each server node

**Genearl microservice section**

How to achieve a low latency here?

1. Index, using caching, load balancing

How to achieve high availability

- using load balancing, using redundancy and replication, failover clustering for database here

- Health monitoring and alerts here

How to decide which shard to call?

1. We can use consistent hashing here

cacheHostNumber = hash\_function(key) % numberOfCacheHosts

How to make sure transcatino is correct when dealing with multiple dbs in a microservice environment?

1. Using saga or 2 phase commit

What to do when you want to store metric-based database?

* Use a time-based database like OpenTSDB for metric related system design like

Ad click aggregator

What’s the pros and cons of using a cron job?

Scalability issues in large systems:

Managing cron jobs across a large distributed system can become cumbersome due to the centralized nature of cron.

**What are some alternatives to using a cron job?**

A distributed job scheduler

**How Slack Built a Distributed Cron Execution System for Scale**

What’s apache spark?

Used for stream processing with apache kafka

**What are some real use cases for this?**

Big data, data warehousing and etc.

Websocket section

1. How does server sent event work in sprint?

Webflux

2. What are the disadvantages and advantages of using server sent event in spring?

**Case studies**

Uber

How do we prevent multiple ride requests from being sent to the same driver simultaneously? (Uber)

1. Lock with TTL with in-memory database

The Ride Matching Service attempts to acquire a lock on the driverId in Redis. If the lock is successfully acquired, it means no other service instance can send a ride request to the same driver until the lock expires or is released.

**Ad click aggregator example**

How do we make sure that advertisers can query metrics at low latency? (how many times an ad is clicked for a time window frame

* Using stream processing with Apache spark
* Using batch processing here

**Url shortener**

How to deal with having a centralized counter? (Url shortener)

Using redis to store a single counter that can be accessed by everything

**Kafka section**

How do we scale kafka event-driven microservice here?

How to increase # of partition in a group?

**Hotel reservation section**

**How to prevent the 2 users from taking 1 transaction?(Can check the hotel reservation system)**

1. Pessimistic locking
2. Optimistic locking (using version)

**Chat system**

How do we make sure each client will get a message when sender sends message?

Each client will be subscribed to a particular pub/sub and the pub/sub is connected to a websocket server and when things happen it will comes to th

**Ticketmaster case study:**

**How to ensure that the ticket is locked for the user whilie they are checking out?**

1. Use distributed locking with redis ttl here

**How do we make sure that the booking map is always up to date?**

**1. We can** use SSE for real time seat update

**2.** using virtual waiting q for popular events for websocket connection

1. **How can you improve text search to ensure we meet our low latency requirements?**

1. Create indexes on event table,

2. Use full text indexes in the database such as mysql or postgresql here:

**How does elastic search sync with Postgresql?**

we can use change data capture (CDC) for real-time or near-real-time data synchronization from PostgreSQL to Elasticsearch. This setup captures changes in the PostgreSQL database, such as inserts, updates, and deletes, and replicates them to the Elasticsearch index.