1. What is latency?

Ans:

Latency is generally considered to be the amount of time it takes from when a request is made by the user to the time it takes for the response to get back to that user.

2. Explain the term caching.

Ans:

Caching is the process of storing copies of files in a cache, or temporary storage location, so that they can be accessed more quickly

It is a temporary storage.

Ex: User Request Req1 to server, when its first time it will hit the database and Respond with Res1, Next time if user request same Request Req1 this time it wont hit database it will fetch the result from cache.

3. What is cache invalidation and cache eviction?

Ans:

**Cache Invalidation**

Ex: At particular time for Req1 response will be Resp1 and it is stored in cache but after some time database has been update now for same request Req1 response is Resp2.

When Req1 is made we will get Resp1 as this is stored in cache but this is invalid.

Therefore, process of updating cache is known as cache invalidation.

Ways of updating

1. Update Response
2. Cache expiry method (TTL)

**Cache Eviction**

Ex: In cache I restrict the value for at most 100 Request & Response.

If I want to store 101 Request in cache at that time, I have to remove one Request / Response from cache.

There are ways to remove

1. FIFO (First In First Out)
2. LRU (Least Recently Used)
3. LFU (Least Frequently Used)

4. Explain the different patterns/strategy of cache.

Ans:

**Cache aside pattern / strategy**

DB

Server

Client

Cache

Data modelling of cache & database may be different

Cache is talking with server & never with Database

Support heavy read operation

Code to make sure that the cache & Database is consistent

System will work if cache is not working

**Cache read/write through pattern/strategy**

DB

Client

Server

Cache

Caching is directly talking with server & database

If the cache goes down the system goes down

Supports heavy read.

Data modelling must be same.

**Cache write around pattern / strategy**

DB

Client

Server

Cache

Cache is directly connected with both server & database.

Server is directly connected with both cache & database.

No failure if cache goes down.

Partial system failure if database fails.

**Writeback pattern**

Cache

DB

Client

Server

It will store bunch of requests once in database.

Database fails we can handle it for some time.

Supports write heavy.

5. What is the REST API? Mention the four most common methods.

Ans:

A REST API is an API implementation that adheres to the REST architectural constraints.

It acts as an interface.

The communication between the client and the server happens over HTTP.

A REST API takes advantage of the HTTP methodologies to establish communication between the client and the server.

REST also enables servers to cache the response that improves the application’s performance.

Methods:

1. GET
2. POST
3. PUT
4. DELETE