**Pastebin:** Design a system like Pastebin, where a user can enter a piece of text and get a randomly generated URL to access it.

**Solution:**

User can login or anonymously paste the notes, select the time for expiration, directory, label and then the random URL will be generated which will be tiny 6 characters on top of the URL length

If the user has created the account then he can manage the notes [create, edit, delete]

Major modules:

* Login Module
* URL generation module
* Retrieval of the content when requested back with the URL or by logging in to the user account
* Database for the user and the anonymous ones are randomly put in the server with the indexed URL for easier retrieval. Users too will be kept in the **sharded** database based on the hash index along with storing these index in the user table. If sharding is done then the database needs to be re-indexed whenever new database server is added
* The system is going to be a read heavy one
* Cache mechanism can be used if needed to reduce the load on read replicas. Cache invalidation has to be in place when the user change the data for anonymous posts it will not be editable

**URL generation module**

Generation of the URL has to be tiny and also unique in comparison to millions of the generated ones. We have to generate a random alpha numeric character of shorter length 10 ^ 10 (10 billion) URL would be sufficient it should be random and a person could not guess. We can generate a URL and check in database and generate again if duplicate exists.

**Storage of the URL and the Notes**

Can use a file system or a database – if we are going to have a disk it will be a little slow due to latency than in-memory cache for frequently read ones. Since users too are going to have a notes creating a disk based file system ensure data is not lost. We could shard the database and then hash and identify the index for the URL and it will be BigO (1)

Follow Up: Analytics of the data

We can store the logs of the data in a flat file and then use it for analytics purpose like following columns URL, timestamp, accessed no of times and then increase the ranking of the URL based on these and promote it to the in-memory cache if necessary