**Question**

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**

1. User can Login or be Anonymous
2. Once we paste some data and give options like post name, visibility of the post (can everyone see it), Duration of post (How long it should be active), Specific Folder structure to maintain.
3. Assumption has to be made on how large the database has to be. Let’s consider that huge traffic is expected and size of the database has to be large.
4. Since it is assumed it is going to be a huge application we need to implement cache for user created URL’s in the browser so that it can easily accessed next time. (We might need to periodically synchronize it to keep it updated)
5. Database indexing is needed in our case to find out the details efficiently
6. While Insertion of a new row if there is need for re-indexing again each time then we have to look other hashing concept URL as the key
7. For finding out if the URL is unique we need a mechanism. If we indexed based on the URL then we can search and retrieve in unit time Big O (1) and by this way we could check after generating a new URL for the data and check if there is a duplicate if so then create a new URL randomly and find a unique and store the details
8. In order to by-pass checking each and every time whether the duplicate is present before inserting the record (Creating a new URL for the content and storing the details) we could pre-process and generate around 1000 Keys and cache it and generate new ones once certain threshold is reached, say 250 keys remain.

**Indexed Database (URL based Indexing)**

If there is huge requirement of database values to be stored we can have hashed alphabetically URL wise

**Aa**

**Bb**

**Zz**

**or**

**Generate URL and check in DB if duplicates are present if not insert**

**Randomly generate and keep in cache to avoid repeated calls frequently**

**or**

**Browser Cache is implemented to reduce the calls to application server often**

**Modification (Optimal Solution)**

Here the URL along with data is assumed to be stored in an online server. If it is an offline application it can reside as a file system.

In addition to it we could do the analytics on how many times the URL has been accessed and keep track of it to remove the unused URL based on a certain mechanism.

Using 10 character alphanumeric value could be an effective mechanism for generating URL