

LinkedIn System Design

1. 已知一个函数，输入用户 ID，可以返回该用户的所有好友(degree 1 friends)，并且能够实现按照好友 ID 从小到大排序。现要求实现算法和相关的系统，返回一个用户的所有好友的好友(degree 2 friends)，以及 degree 3 friends。

Note: 通过临时加一些限制条件，能够进行时间或空间的优化。

2. 已知一个文件系统，可以 create files, delete files, sequentially scan file content, read file content randomly, append file content. 对于 key - value pairs，在给定的文件系统中实现 put, get, delete 方法。其中 key 比较小，全部 key 可以放在内存中，value 会比较大。

Answer:

- a. In the main memory, maintain a hashmap, key is the given input key, value is the file ptr and line ptr (key-value pair location in the file system).
- b. Each time we need to update the value, write a new key value pair at the end of the file, and also update the related key value pairs in the main memory. Then when main memory cracks, we can reconstruct the main memory hashmap from the files content.
- c. Make file modifications to save space: scan the file sequentially for each key to find the latest update, then put the latest key-value pairs into another file. After we finished doing this for all the keys in a file, delete that file.
- d. Deal with synchronization: when we do the file modification, put all the newest updates into another file, then replay it.
- e. Each file should have a limited amount of different keys, so that file modification is not that hard.

3. Design Tiny URL / Note: 如何配置 Memcache。

4. Design LinkedIn

- a. Search 功能里 inverted index 和 data of user , data of company 怎么存，分别用 NoSQL 还是 SQL? -> 列举了 NoSQL 和 SQL 区别，最后选择 SQL。
- b. 前端模块的 service / application 区别。
- c. 设计 timeline, push / pull model。

5. How to design LinkedIn profile page which can support rich media such as video, picture and audio?

6. LinkedIn 有很多 host server 在做很多不同的 service，这些 service 发出的 exception 都会被写到 log 里。设计一个系统，监测 24 小时之内 top 500 的 exceptions。

7. How to design the API for searching similar people? Details about the frontend / backend.

8. Design a system to calculate the revenues generated by ad click.

9. 设计系统来实现 5 分钟, 1 小时, 24 小时内分享前 10 多的文章 / Note: Distributed System

10. 社交网站上的文章转发，如何设计系统可以得到实时的转发量榜单和 weekly digest，要求数据库的设计。有人转发一个文章时 request 是什么样的，如何快速得到实时的转发量榜单，如何得到 weekly digest 等。

Note:

- a. weekly digest 就把每转发一次的信息存入到数据库中，然后直接读取就行了。
- b. 实时转发我说可以保存到一个 in-memory database 里面，可以快速读取。

11. Design IP Black List

这个考虑起来很有意思，先要问问是 IPv4 还是 IPv6。前者所有地址加起来共有 2^{32} 个，如果内存足够大，一个 BITMAP 在 $O(1)$ 就搞定了。如果内存不够或者是 IPv6，那就考虑分块，多级索引，二分查找，反正复杂度立刻上升为是 $O(\log)$ ，当然可以详细讨论这个 LOG 具体怎么做。

12. Design a recommendation system

Data collection / Training algorithm / Evaluate recommendation results / Evolve with C

13. 假设现在我们有 N 个服务器，每个服务器有一个 check sum，如果黑客攻击其中某一个服务器，那么该服务器的 check sum 就会改变，一旦改变，就是被黑客攻击了。设计一个系统，如何 detect hacker attack。

14. 设计 Amazon Buy 页面

- a. 如何设计 database 来存信息，然后 render page? 用 SQL 来存各个 table。
- b. 如果数据很大怎么办?
可以把 shared data 存在各个 slave 上，靠 master 去 query，然后 OOD server interface，怎样读取各个 table 的信息，然后写了一个 server class 来用这些 interface 来 render page。
- c. 如果很多客户同时读怎么办，我的解决方法有 cache / multi-threading。
- d. How to suggest product, 建一个 Product weighted graph, 然后用 BFS。

15. Design a dashboard system for administrator to monitor the top URL in last 5 minutes.

Note: high traffic, database schema design, user experience, performance issue, communication protocol, etc.

16. Design a notebook application like evernote or onenote, it should support search and collaboration operations.

17. 一个 user enter company or university information while signing on, how to help the user better locate his company / university? 我先说用 hashtable 存 company / university 和 id pairs，然后用 Trie 实现 auto-completion 帮助 user 完成 look-up。如果 High-level 怎么快速 locate? 可以 combine location, rank of the top selected company / university, connection's company / university, logo 来推荐。如果用户填资料的时候没有选对应的选项，自己填了一个怎么办? 可以暂时存 null，之后再让用户 complete information。