- How would you design the backend for instagram?

design Instagram
design Instagram
设计图片共享app
system design, insgram upload photo.
设计ins。

- 设计一个fb的功能，用户发个帖子，可以设置权限，比如自己可见，好友可见，好友的好友可见。如何实现这种功能，fb的用户有2billion，最多好友有4000个。

第三轮system design，问在FB里，一个人A发表了一个post,他的所有在线的朋友（B，C，D）都要收到这条post,并且如果有朋友（e.g, B) comment了，此条comment也要实时推送到所有在线的A的朋友（C，D）。怎么实现。
Archetecture Design, 如何design一个在facebook上make comment的method，一个人发comment，同时在网页上浏览的人都能看到。

- 设计脸书的留言系统，主要讨论后台怎么处理post, delete以及有了新留言后如何推送通知给所有正在浏览页面的用户

- design题是设计news feed手机客户端

设计 newsfeed 但是对schema设计这块问的非常详细

- Design题. 设计耶尔普 (纠结在了geoHashing的细节上)

system design 设计个类似yelp的东西

- 系统设计，设计Twitter search, 后续内容需要自己问：QPS, DAU, text, 图片, audio均有可能，多个词的时候去and 的logic (比如 Facebook interview feedback 那么返回三个词都存在的)

- 系统设计。 设计联系人状态列表。写了api, db schema, 简单讨论scalability，

系统设计， 联系人app，只要client part, 问了很多关于caching 好友列表和好友头像的问题，还要设设用户无网络时怎么办。

- FB 的 privacy setting 设计，怎么设计api，code大概怎么写，使用哪些数据，数据怎么存，要支持下列功能

1. Fiend 可见， Friend of Friends 可见
2. 每个user可以自己设些group，把一些user加进每个Group，其他用户不知道他设的group是什么，怎么让group里面的人看到我发的状态？怎么存group？

design, privacy setting, 就是被everyone可见，只被朋友的朋友可见，group可见，问底层SQL怎么存（每一column是什么，怎么Join)，问存friend mapping怎么存，用SQL存和用noSQ存friend mapping的trade off. 然后问friend of friend怎么找，之前看过面经，直接说出最基本的n^2和两种O(n)方法，结果没一种是面试官想要

设计脸熟隐私设定。给一个用户ID和一个post ID。 需要返回这个post是否能被该user看见。先实现全体可见和朋友可见。再实现group可见(white list, black list)。最后实现朋友的朋友可见。这个系统是脱离news feed的一个独立系统，和news feed没有什么关系。主要考察schema如何设计，table如何join。

https://www.facebook.com/notes/facebook-engineering/tao-the-power-of-the-graph/10151525983993920/
Considering the TAO paper by FB. “Facebook’s application servers would query this event’s underlying nodes and edges every time it is rendered. Fine-grained privacy controls mean that each user may see a different view of the checkin: the individual nodes and edges that encode the activity can be reused for all of these views, but the aggregated content and the results of privacy checks cannot.

- 设计点餐系统

主要讨论了 DB shema 和 API 设计， 加上一点点 单点失败

- 设计订机票app

- System Design: Typehead

Design：typeahead 抓着trie的实现问，注意这的trie不是一般LC的简单trie，是真正的production typeahead系统的trie，包括存储优化，trie construction，MapReduce + trie update，等

typeAhead，虽然之前准备过，大的架构答得还行，但是问了几个具体的问题，答得很不好所以估计要挂。这些具体的问题还是值得思考一下的，我给大家列一下，希望能帮到后面的人：1. top n hot key word怎么生成，问了下map reduce的东西 2. typeAhead这里的hot key words考虑多久的时效性，比如你是按照1 month，1 week，1 day 还是1 hour的数据给出hot key words。3. 大家都知道要用Trie去存数据，并且Trie是放在cache里的，那么这个cache什么时候去更新？怎么更新？要不要加TTL？你更新的这个cache的频率会对用户query的时效性产生很大的影响，并且你更新也会对数据库和服务器造成额外的负担，你怎么去平衡。最后加了一个问题说如果这个服务是面向多个国家的，过了一段时间你发现你的推荐在某些国家点击率很高，有些国家点击率很低，你要怎么优化。总之都和你之前的一系列答案有关。问得相当的细。

实现谷歌搜索的自动补全功能。 我首先设计一个最简单的version， 设计了trie node结构，然后demo了下，继续问我user的数据怎么和服务器的数据交互，我分析了push 和pull mode，说使用服务器pull mode更好。然我我继续分析，当我的用户继续增加的时候，采用根据地理位置的不同设置不同的data center。 然后他问我如果用户太大了，还是使用trie么？我说继续使用trie，但是一台电脑肯定放不下。然后他问我怎么办，我说多台电脑。然后他问我多台电脑那怎么搜索，然后我说因为我们需要统计高频的几个搜索记录，所以用mapreduce多台主机一起搜索，并且写了MapReduce的输入输出。最后让我估算一共需要多少台电脑存所有的搜索记录。 估算结果如果全部使用内存的话120台。最后他貌似满意。

- 设计一个“脸书上消耗时间”的功能，记录每天每个用户在脸书上消耗的时间，要求尽可能的在客户端计算量少，要求设计相应数据库以及api的输入以及输出。
  计算每隔多久向服务器发送数据；以及如果用户数据错误怎么办；
-

- Desgin an Advertisement (AD) statistic system. 每次用户登录的时候，系统都会show几个广告给他。广告总共有K种类别，可以认为K<=10。用户看到广告可能会点击，Click Through

Ratio (CTR) = 用户点的广告数量/给用户看的广告数量。注意若同一个广告被用户点击了多次，只算一个click。设计一个系统，answer the following two types of queries:

1. Given a user, return his CTR for all types of ADs.
2. Given an AD type, return its average CTR ovre all the users.

Follow ups:

a. What if K becomes very large? for example, we consider each product as one type, thus K can be as large as 10000.

b. New query type: Given an AD type, return the top-X users with highest CTR. 1< X < 100.

个人感觉这个题似乎更偏向数据库设计？我当时其实有点懵，花了很多时间解释如何设计relationalDB, 处理这些query,以及可能的优化，e.g., 这个application明显是write-heavy的，所以caching(delayed write)会很有用

- 设计一个照片 app, 但是主要只讨论了 client, 服务器部分面试官不感兴趣
- 还是设计一个照片 app, 和前一个不太一样，有offline看照片功能，client和服务器都讨论了
- 设计题，poi

系统：烙印小哥， POI，问的挺细

设计题：POI

3. 设计 POI

设计皮哦哎，用了四叉树。问了如何sharding。

POI system design，用了quad tree，问了万一location如果在边界怎么办

system design, 给一个location找1mile里的所有XX

POI: a point of interest.

Q1. Given the current location, how to find the most closest k points.

Q2. Given the current location, how to find all the points within k miles.

There kinds of answers: GeoHash, K-D tree and Space-filling Curve.

- system design， design places nearby

- 设计一个游戏下载平台
- 设计download api
- 设计一个在线游戏的排名系统。游戏是回合制的，每一回合结束后都会获得一个分数。玩家可以在游戏中添加好友，每个玩家有任意个好友。在每一回合结束后，游戏界面需要弹出两个排序的表格：（1）玩家及其所有好友的最高分数排序top 10；（2）玩家的分数在所有用户（million级别）最高分数中的排名，以及该排名的前十位、后十位的玩家及分数

- Design：objectionable content的reporting系统，就是fb上每个post用户都可以report，设计一个reporting系统，要和news feed连起来，保证被report的内容不要show出来。除了各种系统设计面试官要求讨论怎么scale up，即fb每天有billions of新帖子，而后台真正去看帖子到底应不应该被禁的人工劳力不多，怎么尽可能高效地利用人力资源，尽量用算法去automate这个过程，让它公平公正。

- 系统设计， 脸搜 post系统。

end to end facebook post search， 包括如何添加新的post到如何search post

- 系统设计，我面的是叫product design，主要是api层面的。然后是一个abc小哥，让我来设计一个facebook sharing url时的preview。就从前端怎么抓取url然后怎么valid再到后端用爬虫抓取url内容的整个流程都说了一下，还有怎么用到cache啊，然后post之后其它用户获取preview的时候流程是怎么样的。

System design.当用户粘贴一个URL到Facebook发帖的编辑器里的时候，下面会瞬间自动生成一个那个网页的一些信息，比如图片，标题或前几行文字之类的。预估一下从用户粘贴进去到这个概要生成需要多少时间。一开始觉得这题挺奇怪，后来觉得其实就是要你把系统构架列出，分块预估，然后再把时间汇总起来。 结束后马上发现很多要点我都答漏了，比如DNS，loadbalancer。对一些部件的响应时间预估需要一些工作经验。(前端把URL paste进去后，大概流程是前端立即call FB 的后端API， 这个call发生之前应该检查device的local DNS cache 找域名到IP的转换，local cache如果没有server域名信息，要到最近的ISP DNS server要，要是还没有，再往上一层要，直到国家域名中心之类的总中心。然后提一下根据是通过4G网络还是宽带上的Facebook，这个连接 API的速度差很多。连上server首先会遇到load balancer，把request 分发到一系列distributed backend server中的一个，然后这个server会首先检查cache里有没有这个URL，如果有就把网页HTML读出来，并到S3里把相关的图片信息读出来，打包或分两次发回给前端。如果cache里没有这个URL，就要到访问这个URL去抓取该网页的HTML和图片，访问该URL前也要先走一遍前面提的从域名找IP的那一圈。抓到HTML后要做一些处理，因为不是网页上所有的信息我们都需要，处理完后放到cache里，并且返回前端。还要有一个dedicated server来处理网页上的图片，因为这个可能比处理HTML慢很多。比如图片的分级压缩，因为用户会从不同大小的屏幕上Facebook，一张原始图片得压缩成各种分辨率供不同的屏幕选用。然后把这一系列图片存到S3，并且把在S3的地址存在上一步那个HTML里。基本模块应该就这些。加上数据访问的时候可能存在跨region，比如美东美西或不同大洲间的互访，把这些汇总一下就能估计出个大概时间范围了。中间被问到那个处理图片的service和load balancer 后面紧连得那个service 在不在一台机器上，我说都可以。他就逼问，要是让你做，只能选一个方案，你选在不在一台机器上，我就说分开，他说OK，其实都可以的。然后问了下分不分开的优缺点。感觉重点其实是把部件和流程画出来，预估时间也就自然而然了。)

- design a crawler： 需求是要求尽量减少node之间的通信。一共1万台机器。要求爬取wikipedia的所有网页

加面 design web crawler

第三轮, 系统设计, 爬虫.

system design， 网络爬虫系统

- 第三轮是设计轮，叫设计一个通讯录。。。感觉重点在于服务器和客户端的同步
- 系统设计 设计memcache 只考虑数据结构怎么实现 不用理什么client，server，replication，load balancer啥的
- 设计 memcacheD
- 搜索引擎设计

desig一个search的feature，很general，各个部分都问了问，但是都不怎么深入

- 设计脸书search

设计end to end fb app的搜索 就是设计一个完整的search功能，能存，能读。比如你搜索一个key word，返回post，返回人，返回文章。然后就是问怎么存，怎么读，怎么优化读

思路应该是：如果只是英语的话，应该提取keyword 然后建立invert index，然后对于keyword做shading，然后上面加一个aggeration service来聚合结果。

- 设计一个online musice service的两个API， API1: record user listen to song, API2: return top 10 songs played by any user.
- design a message q system
- system design. Audio mixer
- 设计题：实现 facebook group
- System Design Tiny Url

TinyUrl

- system design, client 给 server 传输文件 的系统。 一个/多个clients <-> 一个 / 多个 server
- 设计轮：网页event售票系统， 随便聊聊画画conponent图和数据表，讲讲哪里容易failure，说说API，信用卡信息谁处理，遇到大流量怎么分流，
- system design: design messengers's online/offline status.

ML design:ranking, ads targeting, recommendation, spam
题目的描述非常简单："You are give a large set of transcripts, use them find the best student".

ML设计：国人大哥，公众号推荐系统，问的比较详细，如何建模、如何抽取样本、采取怎样特征、用户冷启动、Item冷启动都有涉及，还让手推了一下LR的loss func

Design：Newsfeed ranking，着重讲ML，从提取数据，到feature怎么选，到算法，training，online/offline metrics, serving, model update，都有涉及。

设计：设计一个系统来检测fake information（从数据收集一直到train和eval）；设计一个CV系统来帮助在线卖家自动分类所卖物品的类别

ML system design. Newsfeed ranking
news feed 排序
机器学习设计，问设计news feeds什么设计，这种题没有标准答案，主要围绕ML的东西讲，设计feature，设计算法什么

美国大叔， ml设计， 设计怎么capture illegal的post。

第三轮机器学习问了怎么设计newsfeed
第四轮ml design，面经里出现过，news feed ranking，
Instragram newsfeed

ml design: 类似于朋友圈发图的location功能，比如你坐在学校的星巴克里，你想发状态告诉朋友你在这里。但问题是获得的location 坐标信息不准确，你点当前location会返回附近很多 candidates。比如说附近有星巴克，CVS，church。需要设计模型来rank 这些candidates

怎么推荐fb page，涉及到的方便大概有feature engineering，怎么evaluation，如果online offline结果有差距怎么办，candidate page太多了怎么办

ML Design: ads ranking