

## 阶段一：预备知识（有基础的话可以不看，或者选看）

0. 深度学习（和大数据没关系，对做研究有帮助，我在研一上学期简单学了一下）：<https://space.bilibili.com/1567748478/channel/seriesdetail?sid=358497>，感觉看不懂也没关系，尽力去理解就行，这个up的视频对理解有帮助：<https://space.bilibili.com/1921388479/channel/lecturedetail?sid=85730>
1. java（只看了上部）：[https://www.bilibili.com/video/BV17F411T7Ao/?spm\\_id\\_from=333.999.0.0&vd\\_source=4c1702a0e944e2290a9681d1aeb164cc](https://www.bilibili.com/video/BV17F411T7Ao/?spm_id_from=333.999.0.0&vd_source=4c1702a0e944e2290a9681d1aeb164cc)
2. python：[https://www.bilibili.com/video/BV1qW4y1a7fU/?spm\\_id\\_from=333.999.0.0&vd\\_source=4c1702a0e944e2290a9681d1aeb164cc](https://www.bilibili.com/video/BV1qW4y1a7fU/?spm_id_from=333.999.0.0&vd_source=4c1702a0e944e2290a9681d1aeb164cc)
3. SSM（大数据用不到，做管理系统会用到，当时为了看博导系统的代码才去看的）：[https://www.bilibili.com/video/BV1Fi4y1S7ix/?spm\\_id\\_from=333.999.0.0&vd\\_source=4c1702a0e944e2290a9681d1aeb164cc](https://www.bilibili.com/video/BV1Fi4y1S7ix/?spm_id_from=333.999.0.0&vd_source=4c1702a0e944e2290a9681d1aeb164cc)
4. linux和shell：[https://www.bilibili.com/video/BV1n84y1i7td/?spm\\_id\\_from=333.999.0.0&vd\\_source=4c1702a0e944e2290a9681d1aeb164cc](https://www.bilibili.com/video/BV1n84y1i7td/?spm_id_from=333.999.0.0&vd_source=4c1702a0e944e2290a9681d1aeb164cc)

## 阶段二：离线数仓部分（师兄说学完这部分可以找实习，我目前只学到这部分）

5. hadoop（源码先不看，找工作时候可以看看）：[https://www.bilibili.com/video/BV1Qp4y1n7EN/?spm\\_id\\_from=333.999.0.0&vd\\_source=4c1702a0e944e2290a9681d1aeb164cc](https://www.bilibili.com/video/BV1Qp4y1n7EN/?spm_id_from=333.999.0.0&vd_source=4c1702a0e944e2290a9681d1aeb164cc)
6. zookeeper+HA高可用：[https://www.bilibili.com/video/BV1to4y1C7gw/?vd\\_source=4c1702a0e944e2290a9681d1aeb164cc](https://www.bilibili.com/video/BV1to4y1C7gw/?vd_source=4c1702a0e944e2290a9681d1aeb164cc) + [https://www.bilibili.com/video/BV1EP4y1j7V1/?spm\\_id\\_from=333.999.0.0&vd\\_source=4c1702a0e944e2290a9681d1aeb164cc](https://www.bilibili.com/video/BV1EP4y1j7V1/?spm_id_from=333.999.0.0&vd_source=4c1702a0e944e2290a9681d1aeb164cc)

7. hive: [https://www.bilibili.com/video/BV1g84y147sX/?spm\\_id\\_from=333.999.0.0&vd\\_source=4c1702a0e944e2290a9681d1aeb164cc](https://www.bilibili.com/video/BV1g84y147sX/?spm_id_from=333.999.0.0&vd_source=4c1702a0e944e2290a9681d1aeb164cc)
8. flume: [https://www.bilibili.com/video/BV1wf4y1G7EQ/?spm\\_id\\_from=333.999.0.0&vd\\_source=4c1702a0e944e2290a9681d1aeb164cc](https://www.bilibili.com/video/BV1wf4y1G7EQ/?spm_id_from=333.999.0.0&vd_source=4c1702a0e944e2290a9681d1aeb164cc)
9. kafka: [https://www.bilibili.com/video/BV1vr4y1677k/?spm\\_id\\_from=333.999.0.0&vd\\_source=4c1702a0e944e2290a9681d1aeb164cc](https://www.bilibili.com/video/BV1vr4y1677k/?spm_id_from=333.999.0.0&vd_source=4c1702a0e944e2290a9681d1aeb164cc)
10. 离线数仓: 我看的是电商数仓这一套, 但是业务指标统计的太多了, 后面看不下去[https://www.bilibili.com/video/BV1AT411j7hu/?spm\\_id\\_from=333.999.0.0&vd\\_source=4c1702a0e944e2290a9681d1aeb164cc](https://www.bilibili.com/video/BV1AT411j7hu/?spm_id_from=333.999.0.0&vd_source=4c1702a0e944e2290a9681d1aeb164cc), 现在推荐这一套, 知识点和电商数仓那一套讲的都差不多, 区别就是统计指标没那么多, 所以时长也更短:



或者这一套:



## 额外内容 (我准备找实习前看这部分内容) :

hive sql练习: [https://www.bilibili.com/video/BV1og4y1w7W3/?spm\\_id\\_from=333.999.top\\_right\\_bar\\_window\\_history.content.click&vd\\_source=4c1702a0e944e2290a9681d1aeb164cc](https://www.bilibili.com/video/BV1og4y1w7W3/?spm_id_from=333.999.top_right_bar_window_history.content.click&vd_source=4c1702a0e944e2290a9681d1aeb164cc)

以上只是数仓的理论，不贴近企业实际，学完后可以再玩一下阿里云的DataWorks：<https://help.aliyun.com/zh/dataworks/user-guide/data-modeling-10/?spm=a2c4g.11186623.0.0.3e8c5420sHHpTX>

**阶段三：spark部分（有时间可以学一下spark和scala语言，并做一个spark的离线数仓项目，我准备不学spark，直接学flink的内容）**

## **阶段四：实时数仓部分**

flink教程：[https://www.bilibili.com/video/BV1eg4y1V7AN/?spm\\_id\\_from=333.999.0.0&vd\\_source=4c1702a0e944e2290a9681d1aeb164cc](https://www.bilibili.com/video/BV1eg4y1V7AN/?spm_id_from=333.999.0.0&vd_source=4c1702a0e944e2290a9681d1aeb164cc)

redis教程：[https://www.bilibili.com/video/BV13R4y1v7sP/?spm\\_id\\_from=333.337.search-card.all.click&vd\\_source=4c1702a0e944e2290a9681d1aeb164cc](https://www.bilibili.com/video/BV13R4y1v7sP/?spm_id_from=333.337.search-card.all.click&vd_source=4c1702a0e944e2290a9681d1aeb164cc)

实时数仓项目：选择跟离线数仓项目配套的实时数仓项目就行