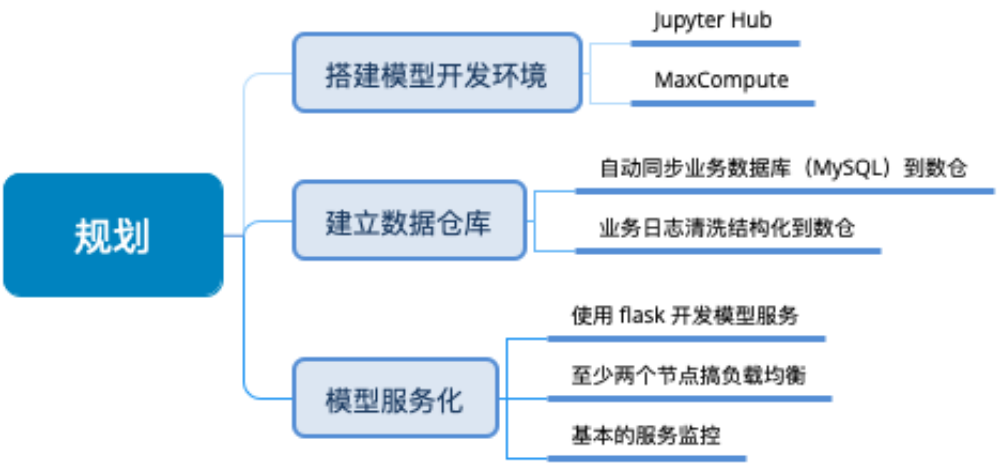

搭建模型开发环境

Jupyter Notebook + Python 是比较常规的模型开发开发套件，Jupyter Hub 是 Notebook 的多用户部署方案，包含了用户管理、资源隔离等特性。模型训练需要大量的计算，通常需要用到一些并行计算框架，比如 Spark、Hadoop 等，不过对于刚起步的公司不建议自建（维护成本高、资源浪费）。像 Aliyun、AWS 等云服务商都提供了类似的服务（MaxCompute、EMR 等），我们用 Aliyun 平台直接选用 MaxCompute 就好，MaxCompute 可以按使用量收费，很合适。

建立数据仓库

数据仓库主要应用是OLAP（Online Analytical Processing），用于处理复杂的数据规模大的查询；数据库主要应用是事务处理。两种技术底层实现方案也不一样，简单来说数据仓库主要是读优化，数据库主要是读写优化。同样 Aliyun、AWS 等云服务商都提供了类似的服务（Redshift、ADB），我们用 ADB 相对合适一些。

另外一部分是数据统一收集的工作，可以使用数据集成服务将业务数据库 MySQL 自动同步到 ADB，日志数据可以使用 MaxCompute 写一些数据清洗的 Job。数据仓库搭建完成之后，模型开发的时候就能很方便、快速的拉取数据。

模型服务化

这个主要是从正确性和灵活性两方面考虑的。考虑到模型是用 Python 开发的，服务如果也用 Python 框架兼容性会好一些，另一方面就是独立部署更容易做模型的动态更新和迭代。To B 的服务稳定性很重要，所以这里还需要做一些基本的服务监控。