**每周工作报告**

2018年9月20日 王宇鹏

# **工作计划**

研究云平台存储架构、计算方式

**工作进展**

阅读资料：java虚拟机、Hadoop实战、Zookeeper

数据采集（3d扫描仪）—通讯协议（http、tcp、udp）\*1—》消息中间件Kafka，在分布式环境下Kafka的吞吐性能非常优秀，并且其持久化和订阅/发布的功能与物联网的场景非常匹配——》混合数据库（HBase、RDBMS Mysql、Redis）分为结构化和非结构化数据\*2

HBase适合于存储半结构化的数据，可以很好的支持海量物联网终端的历史数据的查询

RDBMS适合于存储结构化的数据，通常根据具体的数据库采用不同的高可用部署方案，在系统中主要用来存储终端基础数据、字典数据和数据分析的结果等。

Redis是基于内存的KV数据库，在系统中通常用来缓存需要频繁更新和访问的数据。

——》数据处理 spark即可批处理也可流处理 （实时计算spark streaming/storm 离线计算MapReduce和Hive）

——》DataxDataX 是一个异构数据源离线同步工具，致力于实现包括关系型数据库(MySQL、Oracle等)、HDFS、Hive、MaxCompute(原ODPS)、HBase、FTP等各种异构数据源之间稳定高效的数据同步功能。——》后台

# **遇到问题**

\*1扫描仪上的传输协议及方式

\*2数据格式结构是怎么样的

\*3需要服务器平台去学习部署数据库

**下周计划**

* 1. 确定前端和后台技术方案
  2. 学习分布式数据库部署