**数据仓库与大数据工程-第16小组第一次会议纪要**

时 间：2021年5月17日14:30

地 点：微信视频会议

主 持 人：徐文凯

记 录 人：吴亚伦

参会人员：陈奇、徐文凯、唐国洁、吴亚伦

内容概要：会议中确定了小组成员的分工，并对下一步的工作制定了大致的计划，具体会议内容如下：

1. **小组分工**

会议决定，由陈奇同学负责担任小组组长，负责组织协调内部工作以及课堂PPT展示汇报，由徐文凯、唐国洁、吴亚伦同学负责资料的整理和收集。

1. **大致工作流程**

经小组成员讨论，本次工作分为如下几部分：

1. 确定业务场景。从某个业务场景入手，研究相关的数仓架构。
2. 搜集资料。组内成员从自己公司内部或者网上查找公开的数据仓库架构设计资料（暂不限于某家公司），将资料整理发到小组群内进行讨论。
3. 确定课题。根据组员收集的数仓架构资料，确定最终的研究方向和展示内容。
4. 展示汇报。整理资料、制作PPT，准备课堂汇报。

**数据仓库与大数据工程-第16小组第二次会议纪要**

时 间：2021年5月19日11:30

地 点：腾讯会议

主 持 人：CQ

记 录 人：WYL

参会人员：CQ、XWK、TGJ、WYL

内容概要：会议中根据组员收集的资料，分析了课堂展示的要求，确定了关于本小组的选题方向，制定了下一步工作安排，具体会议内容如下：

1. **关于选题方向的讨论**

根据小组成员近期收集的资料，如：《京东大数据专家认证培训—数据仓库业务篇》，《360政企安全之基于Flink的PB级数据即席查询实践》等进行了讨论。同时，考虑到组员所在的公司内部数据安全和架构设计方案的保密性，决定选择以360公司的基于Flink的PB级数据即席查询实践架构作为本次调研的对象。

1. **关于演示内容的讨论**

经会上讨论，确定演示内容大致分为以下三部分：

第一部分，介绍“Threat Hunting平台的架构与设计”。从介绍“平台的演进”开始，依次介绍“架构设计”和“索引结构”。

第二部分，介绍“以降低IO为目标的优化与探索”。介绍“裁剪率的理论上限及Hilbert空间填充曲线”，“字典索引上Like的优化”，“基于索引对Join的优化”以及“Alluxio 作为对象存储的缓存”。

第三部分，对未来的相关工作，如“BI预算”、“容器化与JVM预热”、“提升多用户的并发查询能力”做一个展望。

1. **下一步工作安排**

陈奇同学负责PPT制作和演讲稿的准备，唐国洁同学协助处理PPT的相关事宜（包括但不限于：确定PPT的主题风格、控制演示时间等），吴亚伦和徐文凯同学负责会后会议纪要的整理和课程群的日常事务。