

尊敬的各位老师，专家下午好，今天我们项目组汇报的内容是**Hadoop大数据处理行为痕迹记录的创新应用**，我是今天的主讲人:杨秉学，今天我们汇报一下之前我们取得的成就以及未来的规划。

接下来给大家讲述的内容是四个方面,前3个方面是我们已经完成的,最后一个是我们项目组的未来计划。

一、在服务器上成功部署了完全分布式Hadoop平台

我们可以通过观察各节点的运行状态来监控hadoop平台运行程序以及平台稳定性。

二、根据实际需求优化部署

最初在3台物理服务器部署9个虚拟节点，在还没有运行程序的时候虚拟机的开销就占用内存80%，所以我们决定减少节点到7个虚拟节点，来减少额外开销。

三、记录分析Hadoop的计算和IO开销状况与文件输出时间。

第3方面，我们根据完成的程序进行单节点运行程序，处理24M用户行为数据，用时255s，经过分析之后，我们发现程序的开销比较大的是I/O开销

四、优化Hadoop计算框架算法

我们项目组总结之后,提出如下的问题,以及我们解决的问题.
最后在学校网络教学平台进行测试，以推广使用。