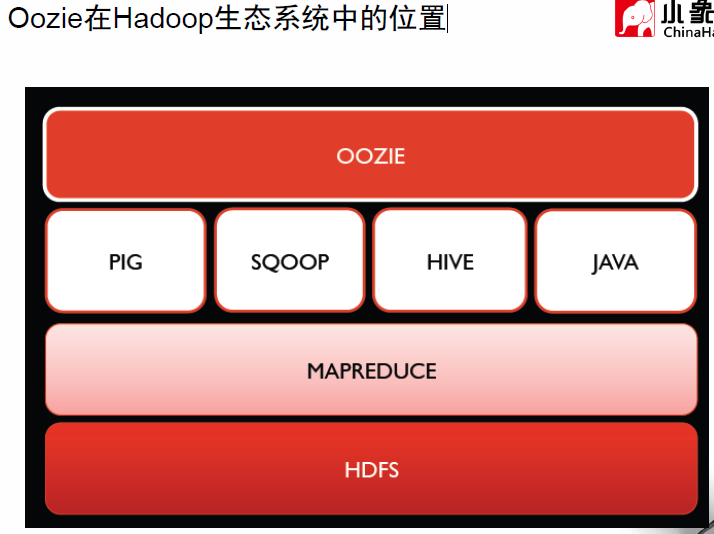
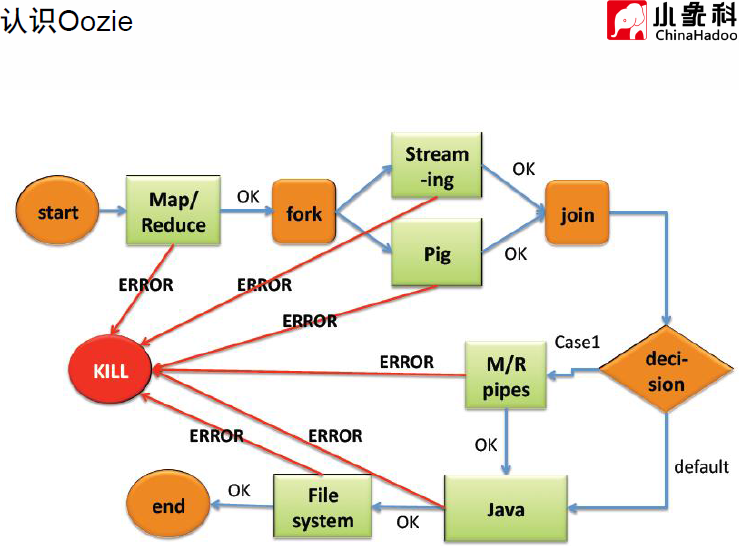
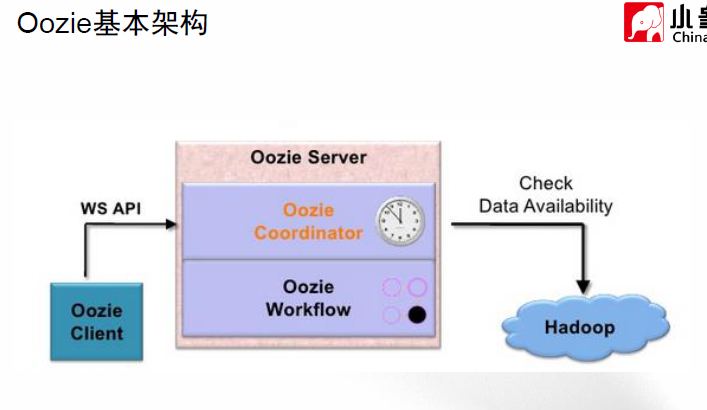
基本概念：

1. 目前计算框架和作业类型繁多：
   1. MapReduce Java、Streaming、HQL、Pig
2. 如何对这些框架和作业进行统一管理和调度：
   1. 不同作业间存在依赖关系（DAG）
   2. 周期性作业
   3. 定时执行作业
   4. 作业执行状态监控与报警
3. 多种解决方案
   1. Crontab
   2. 自己设计调度系统
   3. 直接使用开源系统
4. Hadoop作业留调度系统基本功能：
   1. 对常见作业类型进行调度
      1. MapReduce（Java、Streaming等）作业
      2. Hive作业
      3. Pig作业
      4. Shell…
   2. 根据需要对作业进行调度
      1. 按周期调度
      2. 定时调度…
   3. 支持容错与报警
   4. 可视化作业流运行过程
5. 
6. 缺点：
   1. 编写复杂，不灵活
   2. 不易管理
   3. 难以与监控、报警相结合
7. 常见的开源作业流调度系统
   1. Oozie
      1. Yahoo！开源，基于XML表达作业依赖关系；
   2. Azkaban
      1. Linkedin开源，通过Java property配置作业依赖关系
   3. Zeus（宙斯）
      1. 阿里开源，通过界面配置作业依赖关系
   4. 其他开源系统
      1. Cascading(通过Java API编程实现作业依赖关系)
8. Oozie基本使用方法
   1. Oozie在Hadoop生态系统中的位置



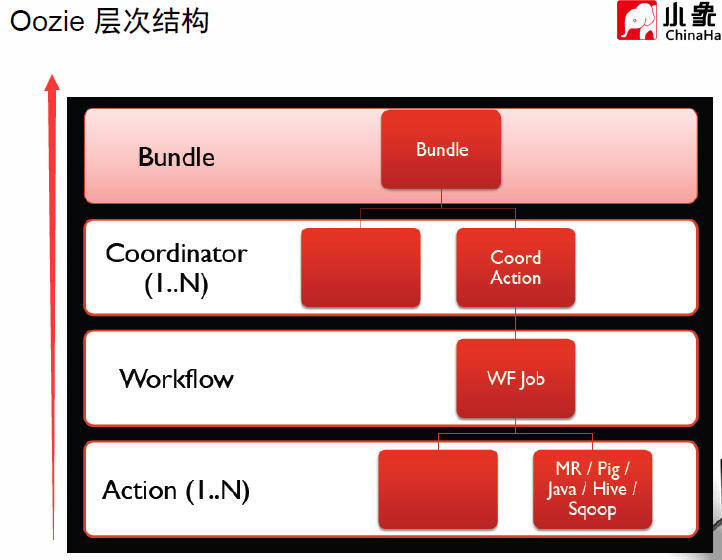




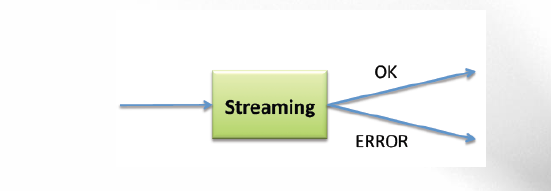
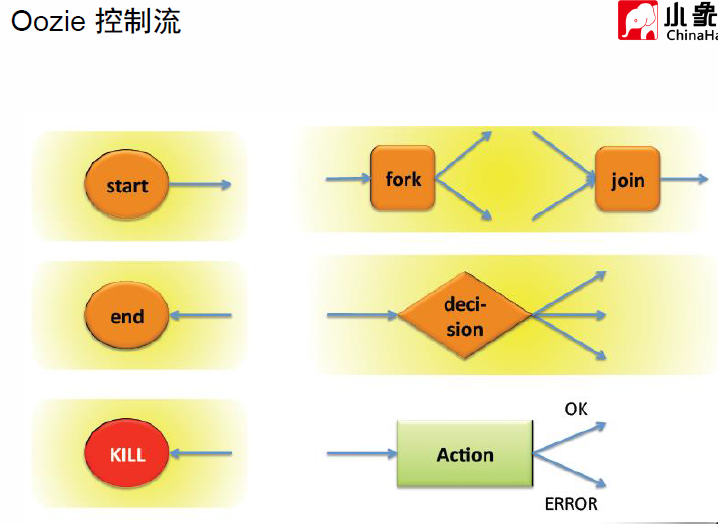


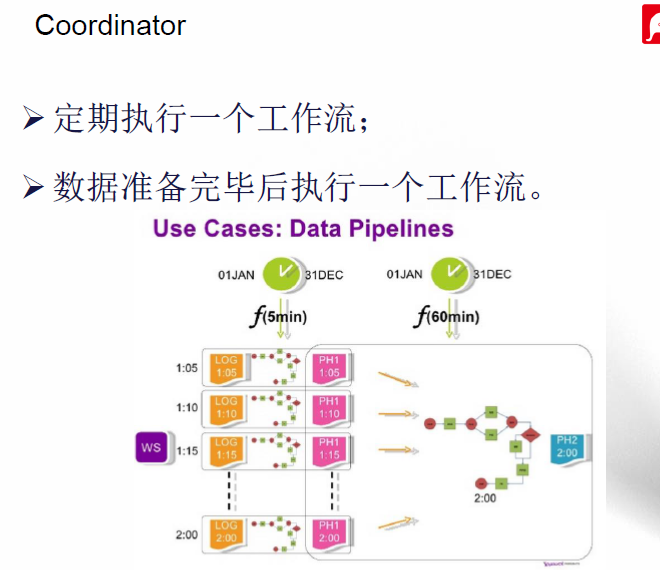
Oozie基本概念：

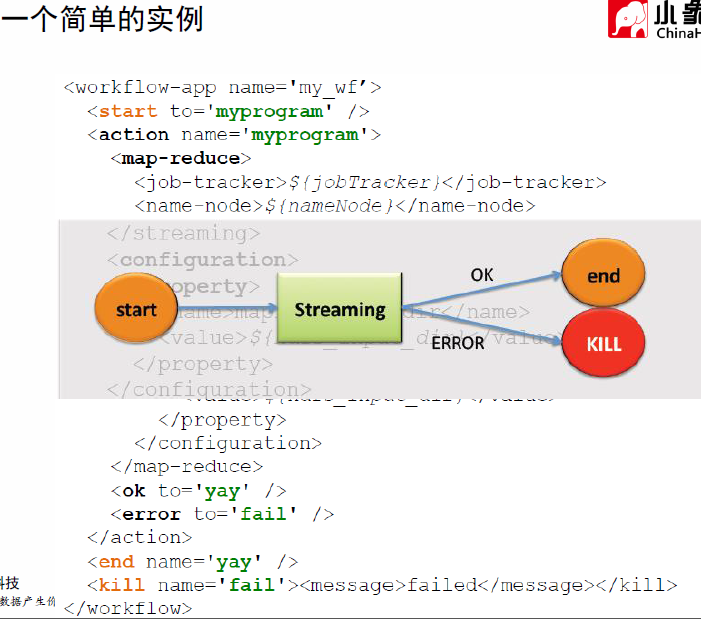
1. 作业流是由一组节点（Hadoop MR作业、Hive作业等）组成，且节点通过控制流节点相连；
   1. 控制流节点定义了工作流的起始于结束，并控制着工作流的执行路径。
   2. 行为节点包含一个计算/处理任务，包括Hadoop MapReduce，HDFS，PIG，SSH，HTTP。Email等。
2. Oozie工作流是通过hPDL语言（一种XML过程定义语言）编写的。

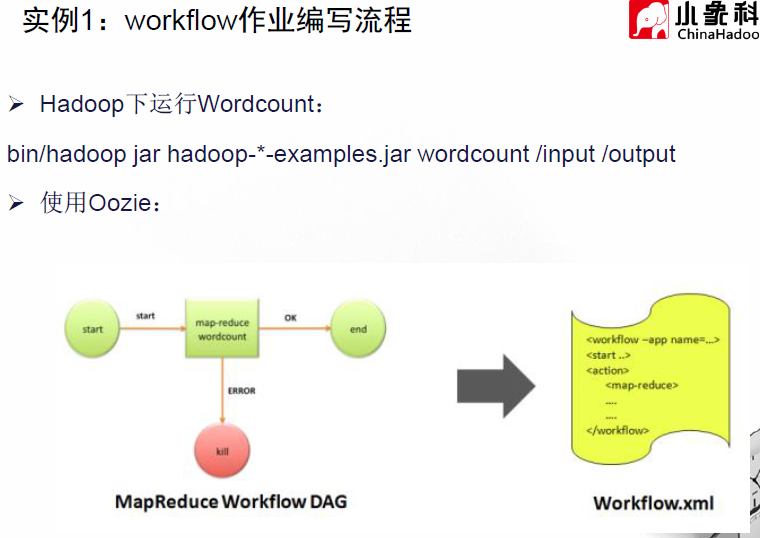


Oozie Action：

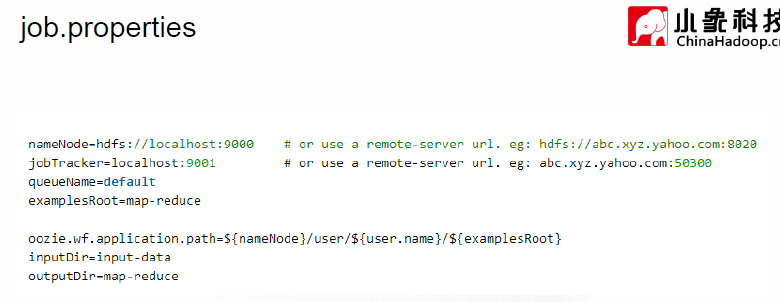
1. 支持Hadoop MR、HDFS、PIG、SSH、HTTP、Email等；
2. 可设置重试次数。
3. 
4. 

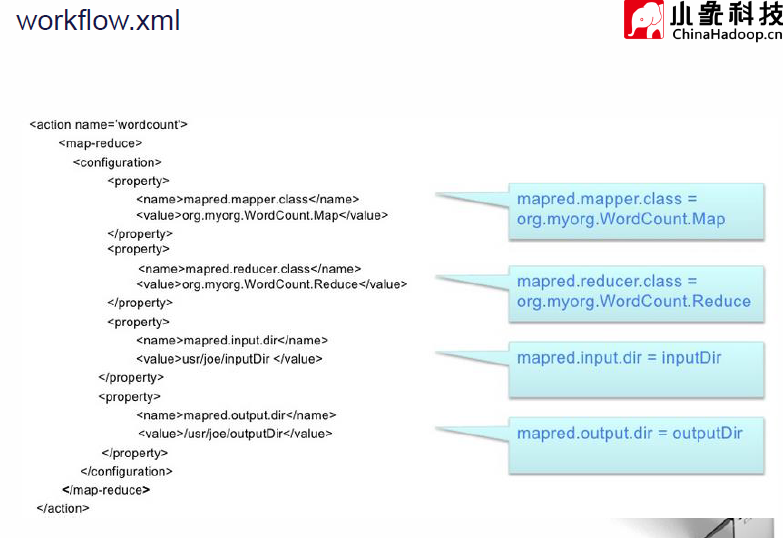


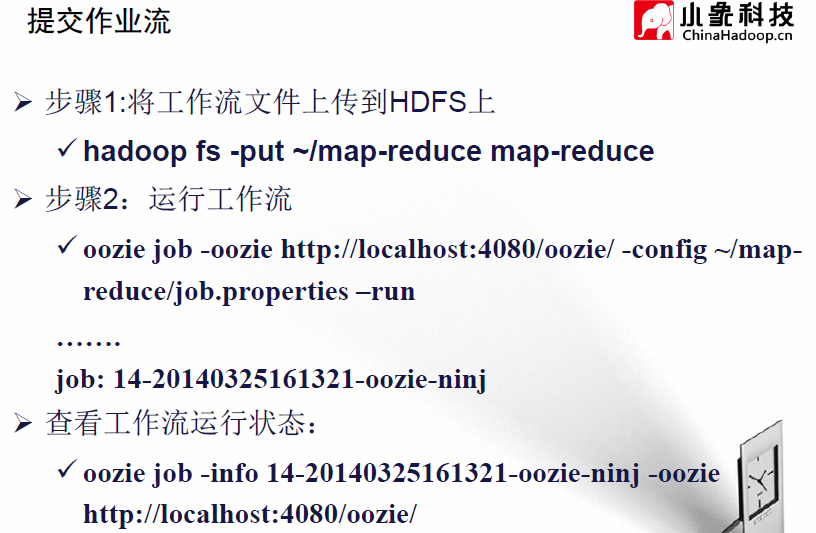








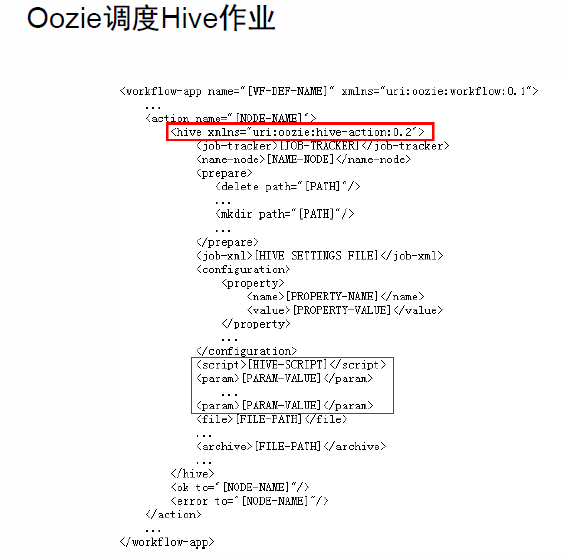


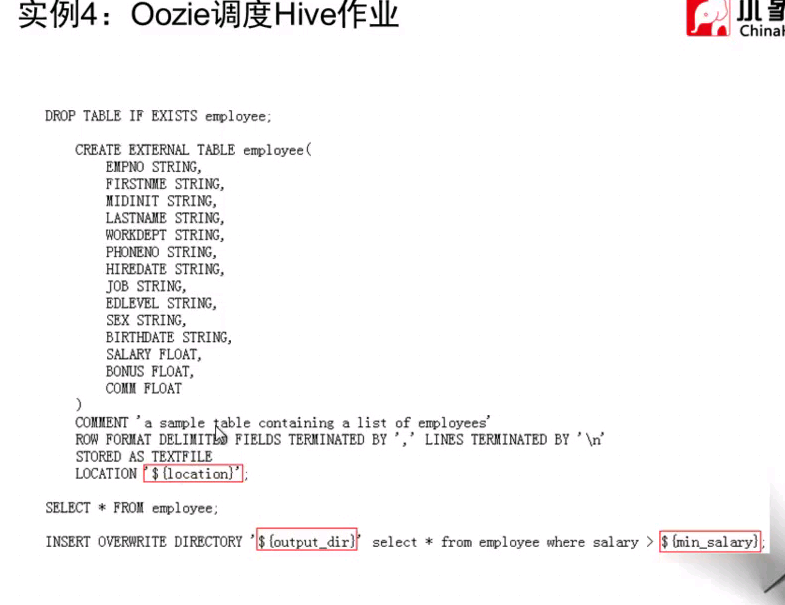






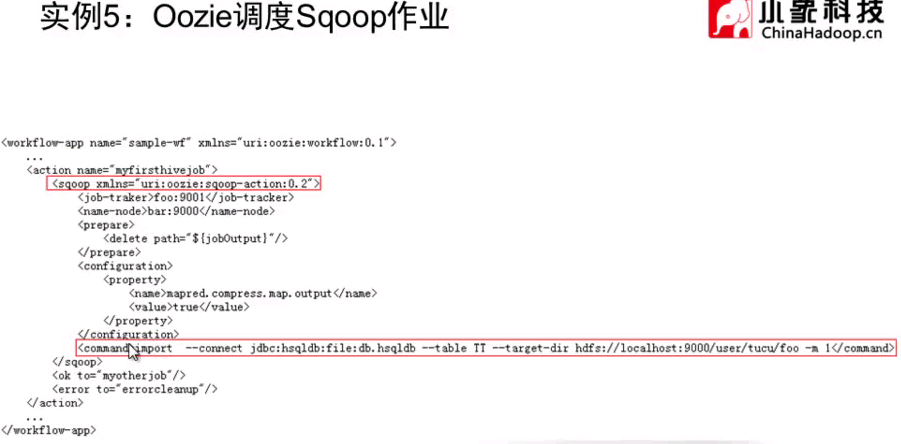




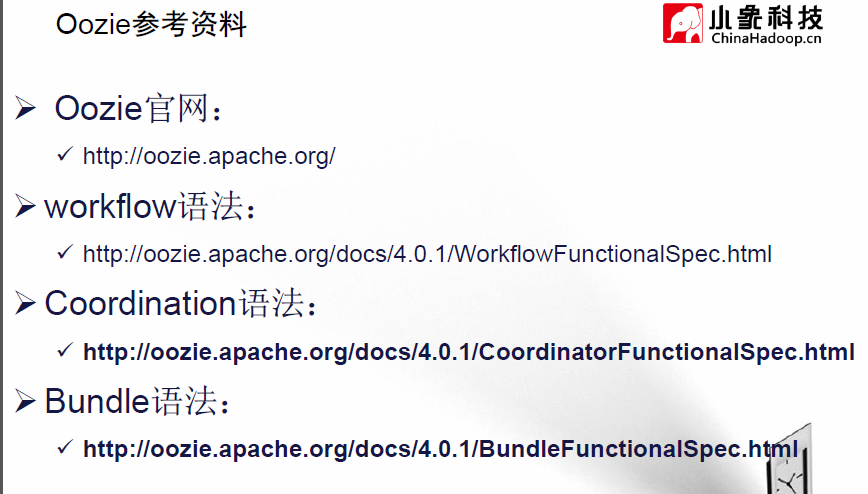




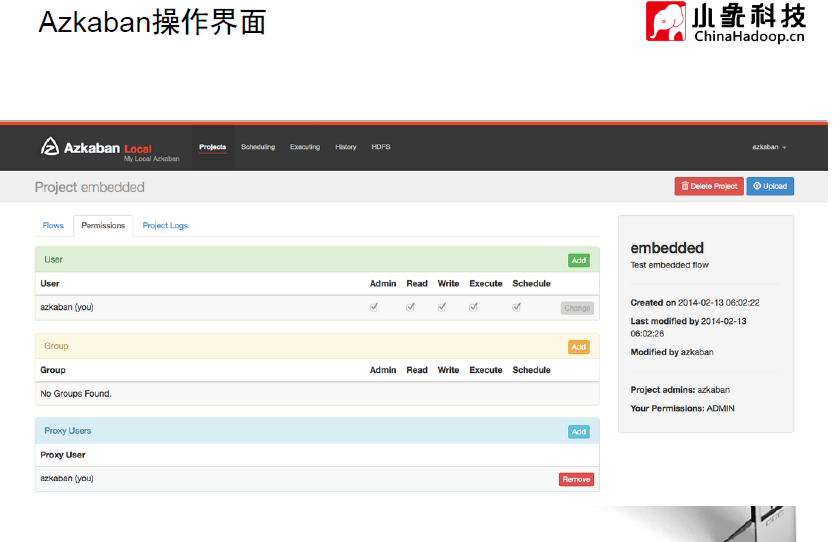








Azkaban基本架构：

1. 
2. 
3. 
4. 
5. 
6. 
7. 