**phoenix**

Apache Phoenix 是运行在Hbase之上的高性能关系型数据库，通过Phoenix可以像使用jdbc访问关系型数据库一样访问hbase。Phoenix操作的表以及数据存储在hbase上。phoenix只需要和hbase进行表关联。然后在用工具进行一些读写操作。对于简单的低延迟查询，其量级为毫秒；对于百万级别的行数来说，其量级为秒。Phoenix并不是像HBase那样用于map-reduce job的，而是通过标准化的语言来访问HBase数据的。

Phoenix查询引擎会将SQL查询转换为一个或多个HBase scan，并编排执行以生成标准的JDBC结果集。直接使用HBase API、协同处理器与自定义过滤器，对于简单查询来说，其性能量级是毫秒，对于百万级别的行数来说，其性能量级是秒。

**Falcon鹰隼**

[Apache Falcon](http://falcon.apache.org/) 是一个面向Hadoop的、新的数据处理和管理平台，设计用于数据移动、数据管道协调、生命周期管理和数据发现。它使终端用户可以快速地将他们的数据及其相关的处理和管理任务“上载（onboard）”到Hadoop集群。

**Livy：基于Apache Spark的REST服务**

Apache Spark作为当前最为流行的开源大数据计算框架，广泛应用于数据处理和分析应用，它提供了两种方式来处理数据：一是交互式处理，比如用户使用spark-shell或是pyspark脚本启动Spark应用程序，伴随应用程序启动的同时Spark会在当前终端启动REPL(Read–Eval–Print Loop)来接收用户的代码输入，并将其编译成Spark作业提交到集群上去执行;二是批处理，批处理的程序逻辑由用户实现并编译打包成jar包，spark-submit脚本启动Spark应用程序来执行用户所编写的逻辑，与交互式处理不同的是批处理程序在执行过程中用户没有与Spark进行任何的交互。

两种处理交互方式虽然看起来完全不一样，但是都需要用户登录到Gateway节点上通过脚本启动Spark进程。这样的方式会有什么问题吗?

首先将资源的使用和故障发生的可能性集中到了这些Gateway节点。由于所有的Spark进程都是在Gateway节点上启动的，这势必会增加Gateway节点的资源使用负担和故障发生的可能性，同时Gateway节点的故障会带来单点问题，造成Spark程序的失败。

其次难以管理、审计以及与已有的权限管理工具的集成。由于Spark采用脚本的方式启动应用程序，因此相比于Web方式少了许多管理、审计的便利性，同时也难以与已有的工具结合，如Apache Knox。

同时也将Gateway节点上的部署细节以及配置不可避免地暴露给了登陆用户。

为了避免上述这些问题，同时提供原生Spark已有的处理交互方式，并且为Spark带来其所缺乏的企业级管理、部署和审计功能，本文将介绍一个新的基于Spark的REST服务：Livy。

**Hcatalog是apache开源的对于表和底层数据管理统一服务平台**

HCatalog将每份结构化的hdfs数据定义schema和访问信息（db、table、partition），然后读和写的时候使用db、table、partition(对于无partition这个可以为空)这三部分信息来访问相应的表数据，屏蔽掉表底层InputFormat、OutFormat以及path信息

由于hcatalog中所有的底层数据信息都是保存在hive metastore里，所以hive版本升级后schema变动或者api变动会对hacatalog产生影响，因此在hive 0.11中已经集成了了hcatalog，以后也会成为hive的一部分，而不是独立的项目。

该WebHCat服务器提供HCatalog一个类似REST的网络API。应用程序发出HTTP请求从应用程序内运行猪，蜂房，并HCatalog DDL。

**SmartSense™**

提供了一组工具和服务以主动预防常见的群集问题。利用 SmartSense，客户可以快速捕获解决问题所需要的诊断信息。此信息还用于生成主动建议，从而改善群集性能、安全性和运营。利用 Hortonworks 工程和支持团队的行业最佳实践和专业领域知识，SmartSense 可以扩展客户运营团队的影响力。