和封神一起“深挖”Spark

转载

weixin_34167819 于 2016-08-24 16:36:40 发布 112 收藏
文章标签： 大数据 运维 java
原文链接：https://yq.aliyun.com/articles/59364

2016云栖大会·北京峰会于8月9号在国家会议中心拉开帷幕，在云栖社区开发者技术专场中，来自阿里云技术专家曹龙（封神）为在场的听众带来《Deep dive into Spark》精彩分享。

关于分享者

曹龙，花名封神，专注在大数据领域，6年分布式引擎研发经验。先后研发上万台Hadoop、ODPS集群。先后负责阿里YARN、Spark及自主研发内存计算引擎。目前为广大公共云用户提供专业的Hadoop服务，即：E-mapreduce产品。

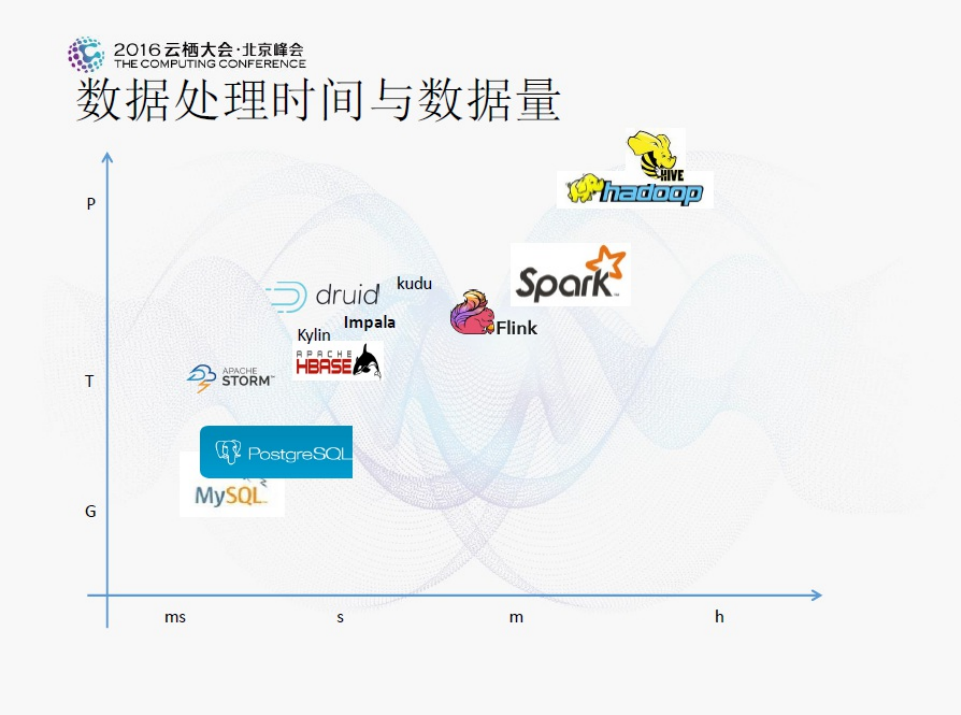
演讲内容架构

- 数据处理技术介绍
- Spark 介绍
- Spark Plus
- Spark 应用场景
- Spark 在云上
- Spark 常见的问题
- E-MapReduce大数据平台

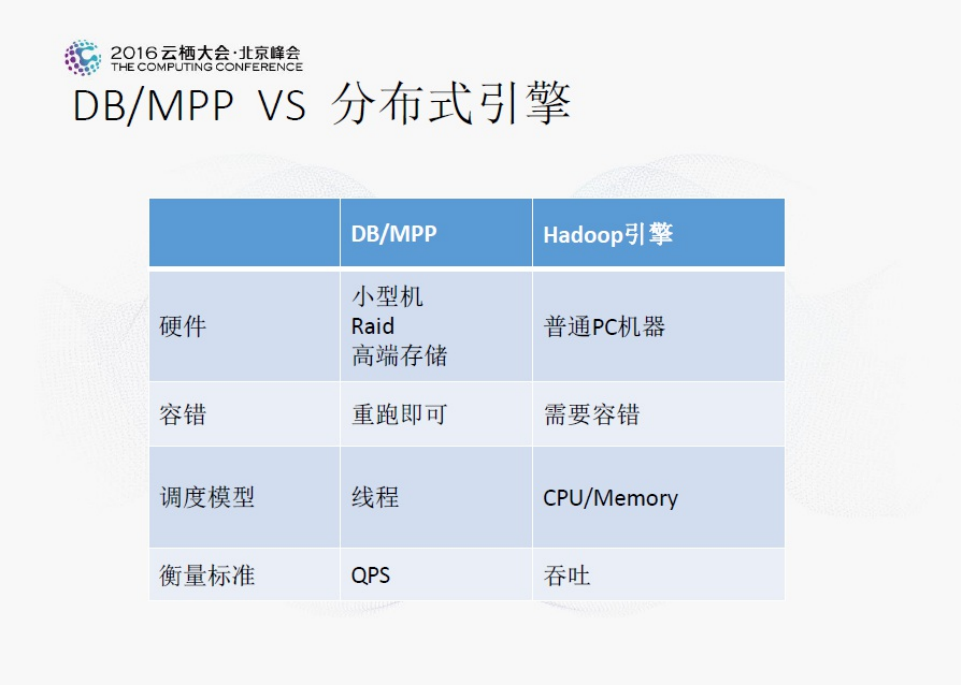
演讲主要内容

大数据通常自上而下分为大数据产品、数据治理/作业生命周期、作业管理/作业流、分布式计算、分布式存储、分布式调度、硬件/机房七层。本次演讲的重点在于分布式计算层。

在以时间、数据量的坐标抽上列出目前引擎大致擅长处理数据的坐标，应该还需要加上数据复杂度、成本等维度，才能更好的体现侧重点，这里不列出。没有哪个软件能解决所有的问题，能解决问题也是在一个范围内，即使是spark、flink等。目前存在有意思的事情是：greenplum类似的MPP引擎想处理大数据的需求，hadoop等被定位为大数据的引擎也想解决小数据的问题（列式存储、或者也加入一些索引）。图中右上角的想往左边靠，减少延迟，图中左下角的想往上面靠，增大能处理的数据量。

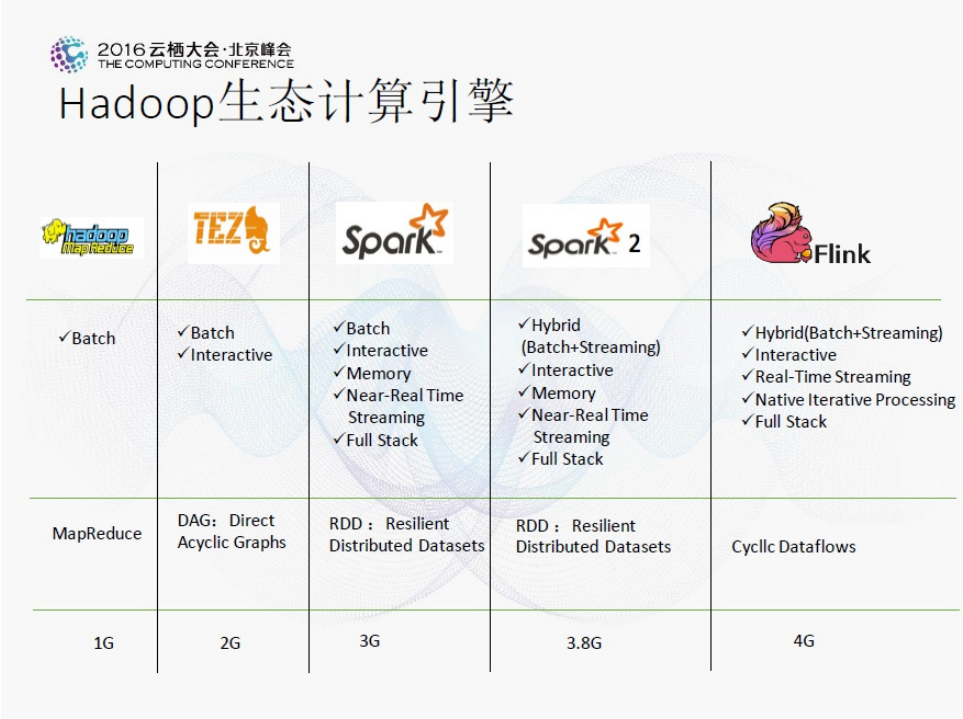


DB/MPP跟Hadoop引擎相对比，两者有很大的不同，具体差异参见下图。从硬件、容错、调度模型及衡量标准方面各自都侧重一方面，对于事务性、index等，Hadoop引擎当前是不支持的。另外MPP其实也在跟Hadoop在融合，比如MPP on HDFS，Spark on DB也在实现。



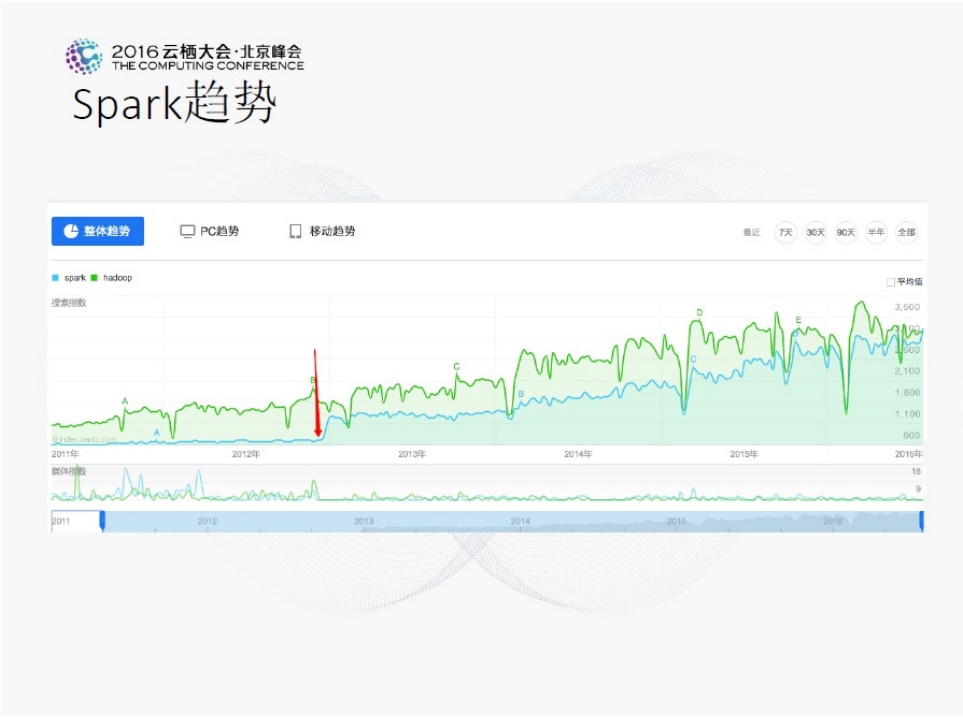
	DB/MPP	Hadoop引擎
硬件	小型机 Raid 高端存储	普通PC机器
容错	重跑即可	需要容错
调度模型	线程	CPU/Memory
衡量标准	QPS	吞吐

Hadoop生态计算引擎目前包括：Hadoop MapReduce、Spark/Spark 2.0、TEZ、Flink等，这里从计算模型，各自的特点分为了1G、2G、3G、3.8G、4G，分别代表其理论先进程度。Spark理论上并不是最先进的，但是目前来讲应该是最适合的。



Spark 介绍

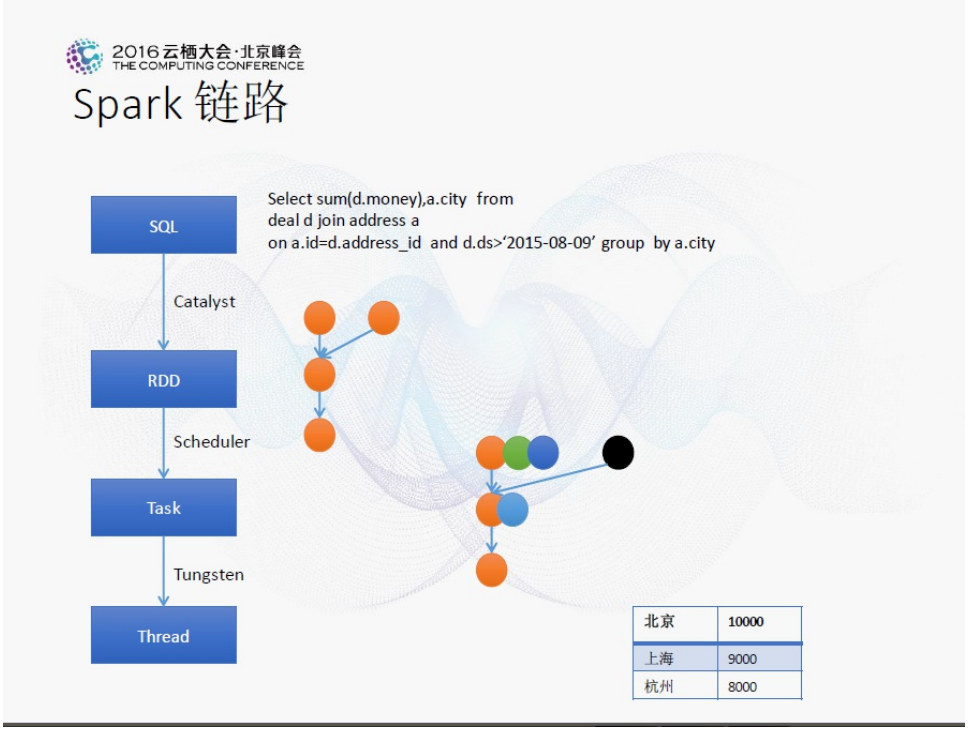
下图展示的是Spark的趋势，可以清楚地看到，在2012年至2013年间，Spark有了一个很大的转折，在那时候，阿里也在逐步使用Spark，到今天，Spark和Hadoop逐渐持平发展。



Spark 提供 SQL、机器学习库 MLlib、流计算 Streaming 和图计算 Graphx，同时也支持 Scala、Java、Python 和 R 语言开发的基于 API 的应用程序。下图显示的是Spark 1.0的基础架构。

下图是Spark 2.0的基础架构，对比于1.0，Spark 2.0主要聚焦于两个方面：（1）对标准的SQL支持。（2）统一的DataFrame和Dataset（逻辑执行计划）API。特别的以后一些的API都是基于Catalyst的。

完整的Spark链路如下图所示，主要包括SQL、RDD、Task、Thread。



Spark Plus

常见的Spark puls有：Spark部署模式、Spark弹性伸缩、Spark+aliuxio（加速）、与业务系统融合（解耦，业务系统与大数据系统）、Spark+数据库服务、Spark+存储格式。

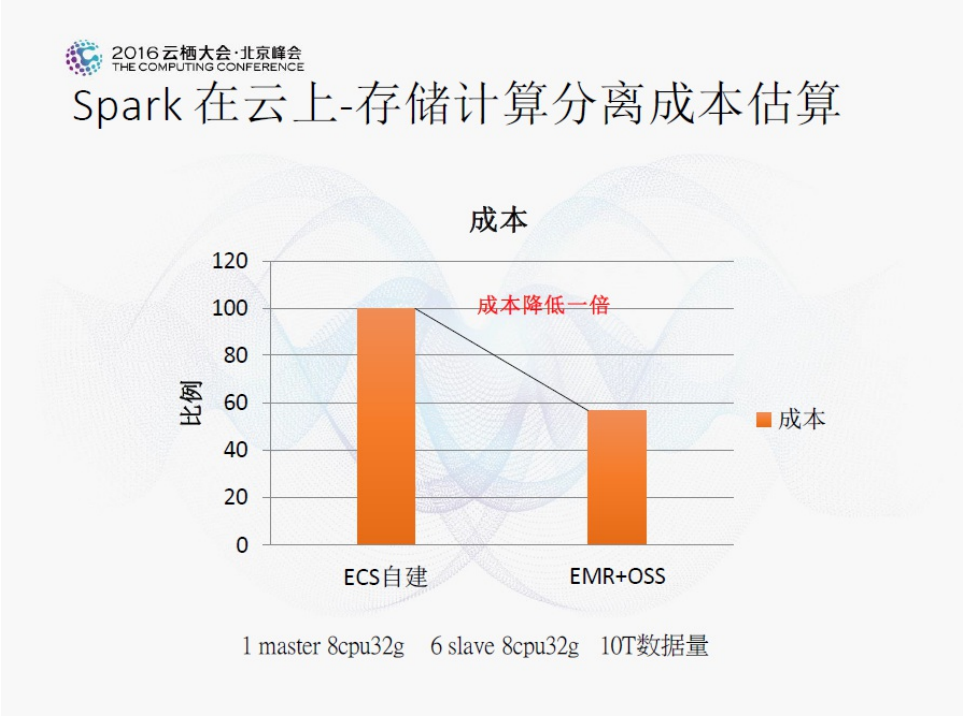
其中弹性伸缩让Spark上大集群成为了可能；在Spark+存储格式中：1 TB数据的存储相对比文本节省了将近 75%；性能按照不同的query提高从几倍到数十倍不等。

常见的Spark应用场景包括：ETL、机器学习、流式计算、即时查询。

其中，在ETL场景中，通过Spark SQL 、Spark API、Dataset实现图片、语音、视频等信息的在线/离线数据抽取、转化为结构化数据，便于后续分析处理。

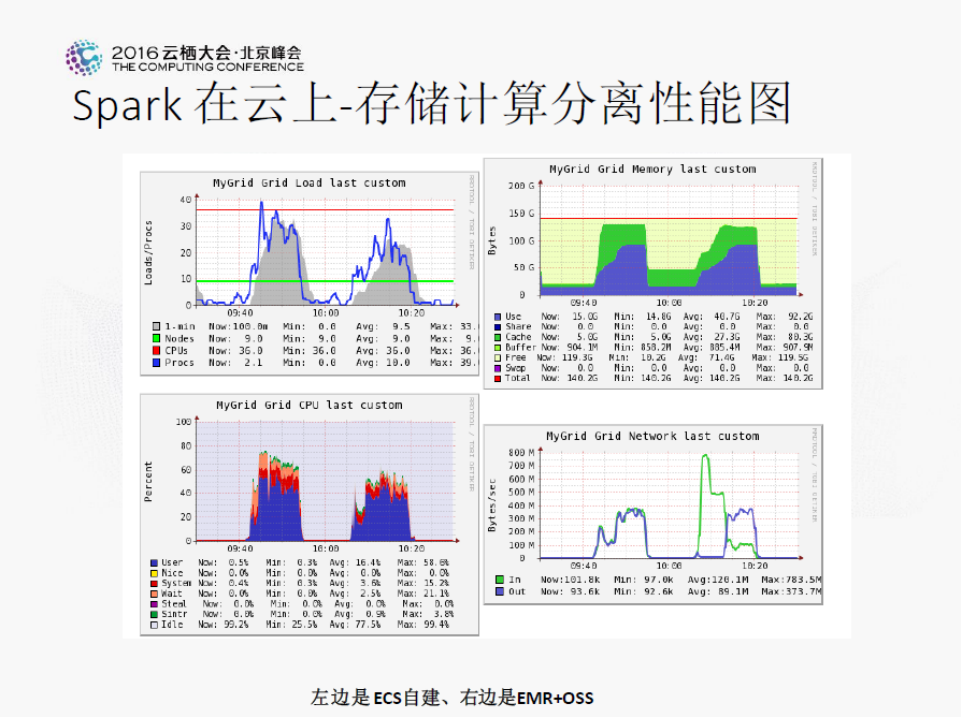
Spark 在云上

Spark在云上的最佳实践是将存储与计算分离，下图展现了自建ECS和EMP+OSS的存储计算分离成本估算对比情况。



下图展示的是自建ECS和EMP+OSS的terasort时间对比，这里自建ECS配置参数是1 master 4cpu 16g和8 Slave 4cpu 16g；EMR+OSS的参数是1 master 4cpu 16g和8 Slave 4cpu 16g。

下图展现了自建ECS和EMP+OSS的存储计算分离性能对照图，左边是ECS自建，右边是EMP+OSS。



左边是ECS自建、右边是EMR+OSS

Spark常见的问题包括卡住、内存溢出、GC频繁。

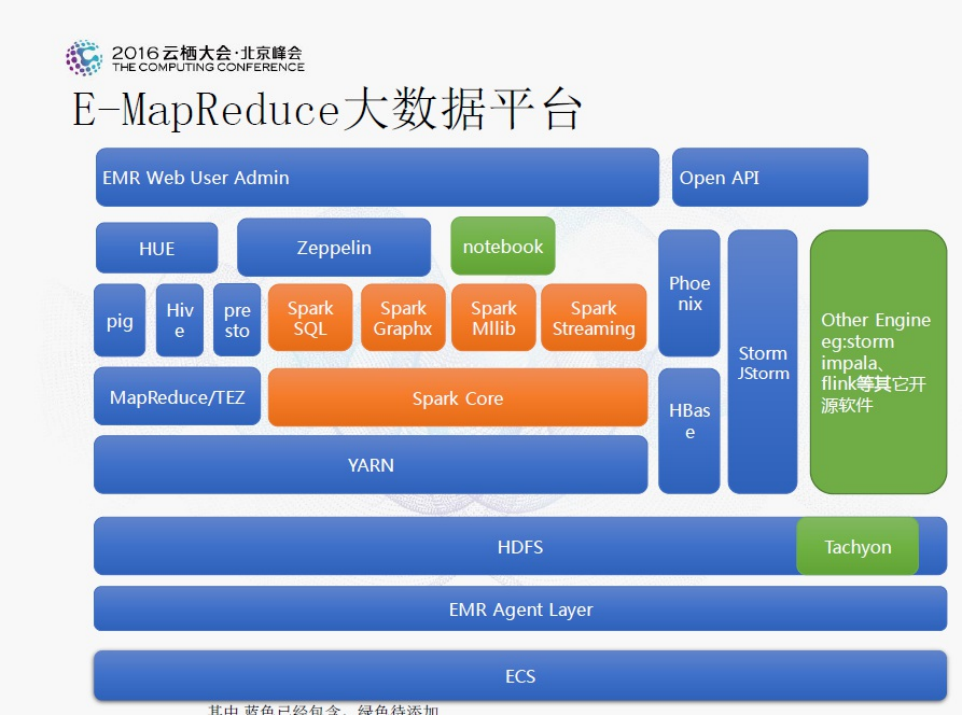
随着Spark 2.0的发布，Spark逐渐趋于成熟，未来Spark的发展方向：

- 支持ANSI SQL
- 性能接近MPP数据仓库
- 一切基于优化（Catalyst）
- 新硬件的支持，比如：大内存、GPU
- 更加友好的支持云

E-MapReduce大数据平台

E-MapReduce 是运行在阿里云平台上的开源大数据处理系统解决方案。它能够让用户将Apache Hadoop和Apache Spark等开源引擎运行在阿里云的云平台上，提供给用户在云上的分析和处理大数据的平台。我们提供管控系统、运维系统及后续的专家系统帮用户解决自动化的问题，并提供专家服务帮助客户解决疑难杂症。

E-MapReduce产品的架构如下图所示:



从上图可以看出，Spark生态是E-MapReduce引擎的一部分，我们还有支持了其它非常多的引擎，如在离线处理、在线流式、在线存储及交互式查询等各个方面。基于我们过去许多年在阿里内部的沉淀，在易用性、成本、性能、运维等各方面具有阿里开源大数据的技术能力，欢迎大家使用。