

Spark Summit China 2014 参会报告

报告人：曾丹

2014-4-19

1. The Spark Project Today and What's Next

Spark 创始团队成员 Andy Konwinski 主要介绍了 spark 的发展趋势以及未来的发展目标。报告中使用了相关数据清晰地展示了 spark 的发展速度以及强大的生命力。Andy 从 Standard libraries, Deployment, Ease of use 等方面介绍了目前 spark 的特性。spark 在宏观上的发展目标主要在于 spark sql, sparkR, java 8 支持以及企业级应用等。在实现细节上, Andy 提出 spark 现在面临的一个问题是参数调优对易用性的影响。不同的参数配置会影响 spark 程序的性能, 这要求使用者对参数配置有一定的了解, 在一定程度上降低了易用性。因此 spark 的长期发展目标之一是 remove need to tune beyond defaults.

感想: spark 有着很大的发展潜力, 随着 spark 企业级应用的开发, 其使用范围会更广。之前测试过一些程序, 对参数配置这方面感触比较深。虽然是一些小的测试, 但是参数配置不好直接会导致运行出错。

2. Spark 内核探索与性能优化

皮皮网数据平台经理陈超主要介绍了 spark 程序的执行流程, 以及在使用时需要注意的细节(比如 persist 和 unpersist 的区别、groupbykey 与 reducebykey 的区别、MEMORY_AND_DISK 存储等级的含义等)。他指出了 spark 官方文档中不正确的地方, 看来是对 spark 源码有一定程度的了解。

另外, 对于如何提升 spark 的性能也提供了一些建议。主要包括使用广播变量解决 task 序列化后太大的问题, 空任务/小任务的处理, 倾斜现象, reducer 数目的设定, collect 返回结果太大, 序列化方法, jvm 中 cache 所占比例等。

感想: 要透彻理解 spark, 还是要看源码。至于后来对提升 spark 性能的建议, 主要依赖于使用经验, 企业在这方面有着很好的优势, 可以通过真实的应用场景观察 spark 的运行情况。这方面的经验在开发/运行具体的程序时会比较有用。

3. Spark 在闪存中的优化: Broaden Use Cases and Reduce TCO

SanDisk 副总裁 John Busch 介绍了利用闪存来提升 spark 性能。虽然现在内存很大, 但是还是不能满足大数据的要求, 因此 SanDisk 从存储介质入手, 增加了 SSD 这一层, 扩充了数据容量, 并对 SSD 进行优化以达到很好的性能。

感想: 之前了解的性能方面的优化都是基于软件的, SanDisk 这种通过改变硬件来提升系统性能的方法对我来说比较新颖, 感觉到对系统的优化可以从多个角度进行。对于具体实现细节, 由于对 SSD 不了解以及个人英语听力差的原因, 没太听懂。从报告展示这方面来看, John 的 ppt 太多字了, 看得很晕, 不如 Andy 的简洁明了, 我们自己在做 presentation 的时候也需要注意这一点, 要多从听众的角度考虑。

4. Transwarp Inceptor: 基于 Spark 引擎的高速内存分析和挖掘工具

星环科技联合创始人孙元浩介绍了一种基于 spark 的分析和挖掘工具。

从整体上来看, 商业产品在包装上做得很好, Inceptor 提供了可视化的界面, 用户只需要简单的操作就可以获得复杂的查询结果, 相对于 sql 语句来说, 进一步提高了易用性。

从实现上看, Inceptor 跟我们想要做的工作比较相关, 它在 HDFS 和 spark 执行引擎之间构建了一个存储层, 采用列式存储和辅助索引, 来加速查询处理。Inceptor 提出了 Off-heap 的概念, 由于 java 的垃圾回收机制受 headp 的影响, 所以将任务从 JVM heap 中搬出以降低这种影响。为解决内存容量不足的问题, Inceptor 使用 SSD 作为缓存层。此外, Inceptor 支持标准 SQL 查询, 与 Hive 不同, Inceptor 可以将多条 SQL 语句一起编译执行, SQL 语句之间的关系用控制流来表示, 孙元浩提出 spark 的 rdd 模型的代表能力还存在一定限制, 在他们的应用场景下, 控制流模型会更好地表示应用程序逻辑。

感想: 听完 Inceptor 的介绍之后, 我不禁在想我们未来的工作与之相比, 创新点在哪里? Inceptor 作为一个商业产品, 在用户体验以及后台的系统实现方面做得已经比较好了(从展示来看)。

5. YAFIM: 基于 Spark 的并行化频繁项集挖掘算法

南京大学计算机系博士生顾荣介绍了传统的频繁项集挖掘算法 Apriori 在 spark 上的实现, 他们所做的工作已经发了一篇论文。

感想: 个人感觉创新性不大, 在算法实现上没有什么改进, 只是利用 spark 实现了 Apriori 算法, 应该算是 spark 的一个新的应用。

6. Spark 实际应用模式的经验共享

英特尔架构师易岚分享了使用 spark 的经验。从实际应用来看, 得出了三个结论:

(1) spark streaming 使用不多。由于 spark streaming 限制太多, 实际应用中, 主要是使用 mini-batch 的方式执行在线处理。

(2) shark 在处理复杂查询上的优势更明显。hive 在处理复杂查询时, 性能显著降低。另外需要注意 spark/shark 与 hadoop/hive 的配置不同, 比如核数的设置等, 不能把 hadoop 中的那一套直接搬过来用。

(3) combined application 比较少。spark 是一个统一的平台, 可以处理多种应用, 如图计算、机器学习、交互式查询、流处理等。但在实际应用中, 同时设计到多个方面的应用比较少。

感想: 在实际应用中会了解 spark 的核心优势, 虽然 spark 可以应用不同的场景, 但是对一个企业来说这一优势可能体现得并不明显, 其最大的竞争力还是在于速度上。

7. Spark 在中国移动省公司试点应用

广州浩微数据服务公司 CEO 邢钢介绍了 spark 在移动公司的应用。

感想: spark 在通信行业的应用范围也很广, 但是对于移动公司这样的大型企业来说, 对 spark 的应用还是比较慎重的。由于企业内部没有研发及维护 spark 的相关人员, 使用 spark 的成本还是比较高的, 因此目前只是处于初步尝试的阶段。

8. Datacenter Management With Mesos

Andy 除了是 Spark 创始团队人员之外, 还是 Mesos 作者。本次报告主要介绍了 Mesos 的开发背景以及对集群资源的宏观管理。

感想: 演讲思路比较清晰, 让我对 Mesos 有了一个大致的了解, Mesos 的核心思想是 OS for the "Datacenter Computer", 主要使用 Linux Control Groups 进行资源的隔离与复用。

9. 基于 Spark Graphx 的大规模用户图计算和应用

淘宝技术部明风介绍了淘宝公司内部 spark 在图计算方面的应用, 主要涉及 Graphx 函数库。

感想: spark 丰富的函数库使得部署在其上的应用可以简单地进行, 这也是 spark 能够

吸引众多用户的一个很重要的方面。另外，明风也给出了一些参数配置的建议，如 `spark.storage.memoryFraction` 的值，并行度的设置等。可以看出，参数配置对于实际应用的执行还是很重要的。

10. Catalyst: Spark a Chain Reaction

Spark 贡献者连城介绍了 Spark 内部的 sql 支持 Spark SQL 以及查询优化组件 Catalyst。

首先，他提出了 Spark SQL 的研究背景。他指出 Shark 在处理 sql 查询时一个很大的问题在于对 Hive 的过度依赖。SQL 语句的执行分为三大步骤：查询解析、逻辑计划生成和物理计划生成。Shark 在前两个阶段直接复用了 Hive，只是在最后一个阶段将原来 hive 的 MapReduce 作业转化为 spark 中的作业。复用 hive 主要存在两个方面的不足：代码维护开销较大，Shark 内部的线程安全得不到保障。Spark 内部运行机制是基于多线程的，而 Hadoop 运行机制是多进程，不存在线程安全的问题，因此在 spark 内部添加对 sql 的支持显得十分重要。

其次，连城介绍了查询优化组件 Catalyst。Catalyst 主要工作是查询优化，对查询计划应用一些规则，Catalyst 的核心亦即这些规则。相比于 hive，Catalyst 的优化规则使用起来更简单，更全面。

Spark SQL 将会在 spark 1.0 发布 alpha 版本，目前包含了对 Hive 的兼容，对 Parquet 存储格式的支持以及查询计划的自动优化（如 filter push down 等）。Spark SQL 的优势主要体现在以下四个方面：Stability, Compatibility, Performance 和 Integration With Spark ecosystem。

连城在演讲中也提到了 scala 语言可能影响性能的地方，如 Option 的开销。这方面主要涉及具体开发过程中的一些考虑，在未来的工作中可能会有借鉴意义。另外，他也讲到 GC 和序列化对性能的影响。

感想：这一部分跟我们的工作比较相关，目前 spark 生态系统中主要是使用 Shark 处理 sql，我们的调研重点也放在 Shark 上。Spark SQL 的发布无疑会影响我们后续的工作，我想接下来在这部分要深入了解。

11. 面向实时计算和 ETL 处理的 Shark 应用

网易大数据研究员王健宗介绍了 Shark 的实际应用，体现了 Shark 相比于 Hive 的优势。Shark 的优势主要体现在底层的快速执行引擎 spark，与 Hive 兼容，Shark 中的数据可以用于机器学习。但是，通过具体数据，可以看到 Shark 比 Impala 的性能差，其中一个原因在于 Shark 复用了 Hive 的 metastore，维护 metastore 会带来开销。

在技术实现上，王健宗介绍了 Shark 与 Hive 不同的主要技术特点，如 Map-side Join, Shuffle Improvement 等。另外，他指出 Shark 使用中需要注意的地方：

- (1) 内存消耗明显
- (2) 对硬件/配置的调优要求较高，版本更新换代速度快
- (3) 权限控制方面
- (4) hive 支持度方面：不支持桶，文件输入输出格式限制
- (5) 原始数据集较大时 Out of memory 问题。可通过设置并行度解决。

感想：本次报告也提到了配置调优的问题，看来 spark 在自动调优方面还需进一步地改进。Shark 相比于 Impala 性能差，那么 Spark SQL 在这方面的表现又如何？比较期待。

12. 当机器学习遇到大数据系统：从 Spark MLib 看两者的混合发展

Intel 工程师尹绪森介绍了 Spark 在机器学习方面的应用。

感想：跟我们的工作不太相关，不过这是 spark 相比于 MapReduce 的一个很大的优势，

即对机器学习的支持。机器学习的应用非常广泛，这在一定程度上增加了 spark 的用户群。

13. 快速模型更新在精准广告推荐中的应用

腾讯广告推荐负责人薛伟介绍了利用 spark 实现快速模型更新的例子。

感想：快速模型更新主要是利用机器学习算法快速地训练出模型，在这方面，spark 能够得到很好的应用。其一，spark 对机器学习算法的支持使得用户只需要调用库函数就可以执行复杂的机器学习算法。其二，spark 的速度也很好地符合快速模型更新的要求。又一次感受到了 spark 的强大。

参会总结

1. 应用场景。

Spark 主要专注于两方面的应用：迭代式计算和交互式查询。从这次 spark 峰会上看，其应用范围也没有超出这两个方面。

2. 参数调优。

在进行 Shark 及 Hive 测试时碰到过参数配置不当导致出错的情况。Andy 指出 spark 未来的发展目标是不需要用户进行参数配置。从用户来看，比较 nice，省去了很多麻烦。但是从开发者来看，感觉实现起来比较困难，因为各个集群的运行环境不同，如果不进行参数配置，也只能获得一个相对较好的性能。

对于如何进行最佳的参数设置，我也咨询了相关讲师，他们的意见主要是靠经验，多思考。

3. 优化方向。

会上有几个报告都提出了使用 SSD 解决内存空间不足的问题，典型代表是 SanDisk 公司的工作。这是我以前没有想到的，之前关注的优化方向主要在于执行引擎的加速以及数据的存储格式，没有想过存储介质方面也可以提升系统的性能。这次会让我认识到一个好的系统必须要充分考虑硬件和软件，软硬结合才能达到更好的效果，因此了解一些硬件的知识也很有必要。同时，我也感觉到自己的思维存在着很大的局限性，以后在这方面要多加锻炼。

4. 程序开发方面。

会上多次提到 GC 和序列化对性能的影响，连城的报告中也指出了 scala 语言的某些特性会导致性能下降。以前写程序就只满足于功能方面，没有考虑过对性能的影响，这次会让我看到了开发者对底层实现的透彻理解，这是产生高质量代码的前提。

5. 与自身工作的联系。

Spark 峰会上跟我们工作比较相关的是 Inceptor 和 Spark SQL。听了报告之后，我觉得需要更进一步地了解这些工作，以调整我们的工作计划。