2012 Hadoop 与大数据技术大会参会感言

By LijieXu@ISCAS
2012 年 12 月 1 日

【题外话】这个冬天依旧冷

【摘要】

从 5 开始是想继续全年这个时候写的《Hadoop in China 2011 参会感言》

5、 参会记录

6、 个人观点

7、 对比工业界，学术界表示我们来了

8、 以后会怎样

【参会记录】

2012 年 11 月 30 日和 2012 年 12 月 1 日在北京云南大厦举行了特别盛大的 Hadoop 与大数据技术”大会，参会人明显超过了 1000。会议讨论主题涉及 Hadoop 生态系统、大数据应用、大数据共享平台与应用、NoSQL、NewSQL、大数据的技术挑战与发展趋势的内容。参会人来自学术界、工业界、商界、媒体等。

与去年碰上了第一场大雪不同，2012 的雪来的比以往来的早一些，月初就下了。最近已经一个月都没雨水了，干冷干冷的。今年地点选的还好，市内，交通方便，旁边也有饭馆。与去年不同的是，今年没买餐票，中午只能在外面吃了，不同听说会议餐不怎么样。去年就买了 4 张票，今年实验室慷慨，加上我们这些老人，和积极的师弟师妹们一块组团去了，欢乐了许多。

与去年最大的不同，当然是心情了。去年啥都不懂，看什么都是新鲜的。今年提前看了下会议议题，期待还是有，只是少了很多。也可能是最近被各种事情纠结，心里乱糟糟，不能全力 focus 这些话题。

今年夏天米国举办了 Hadoop Summit，基本上将各种成熟、老练、新奇、有潜力的系统和技术介绍了遍。这次的会算是今年国内的发展状况概览。可能是我听的议题比较集中，感觉演讲嘉宾各个都像架构师，恨不得把系统的方方面面都讲清楚。

【第一天】

第一天都是全体报告，一千多人坐在一个超大报告厅，几千只眼睛注视着演讲者，不知嘉宾有没有鸭梨。

先是大会组织者计算所的一些大佬开场白、祝词、谈趋势、谈发展。然后是计算机学会大佬祝词，世界是你们的，也是我们的，但归根结底是你们的，好好干。比较喜欢清华的一个老教授，一开口就感觉是“慢羊羊”，听他说话挺高兴的。然后 Hortonworks（被 Yahoo 踢

出来新成立的，专门做 Big data 的公司，Hadoop 最早的版本出自他们）首席技术官讲了 Hadoop 的昨天、今天和明天。这次大会竟然还有同声传译，当然水平不得而知，反正感觉带个耳机听报告容易被当成怪叔叔的。之后，Gartner（一个号称自己是先知的公司）以商业分析师的视角分析了大数据的前景和钱景，鉴于师兄总是引用他们的白皮书，我还是觉得他讲的挺不错的，其实他主要讲了大数据对数据中心架构的挑战。然后一个 IBM 信息管理副总裁讲了“IThink Big, Turning Promise to Reality”，到底讲了什么？我也不记得了。最后是中移动常出场的支撑研究所所长，讲了中移动研究院为支持业务需求（也就是各种计费、分析话费）而建造的数据存储、处理与分析平台。作为这次会议的主要参与者，中移动派来了好多演讲嘉宾，看来自从股份制改革后，移动还是认识到要有自己的研发人员。

中午去吃了旁边的鸿毛饺子，不禁有点怀旧，因为曾经常常跟同届的童鞋一块去吃，他们现在已经毕业工作了，住的也远了。

下午的报告比较给力，哈工大的李建中教授给我们科普了大数据中存在一些计算、分析方法问题，较为理论。每个问题都像是一篇博士论文题目（估计就是），由于之前做的了很多无线传感器方面的数据处理与分析，李教授在数据库和无线网络的顶级会议上都有很多文章发表。但在互联网数据方面涉及不多，也没牵涉目前的一些主流大数据系统。

接下来也是常客的 Ohio State Univ.张晓东系主任总结了他们在大数据方面的具体工作，由于跟 Facebook 合作，他们做的工作基本都落到了实处.他们团队设计了新的 Hive 表文件存储方式 RCFile，更优秀的 SQL-to-MapReduce 引擎 Ysmart，以及理论模型上的创新——DOT（张教授将 DOT 提到了与 BSP 模型对等的地位）。张教授以前做高性能计算，现在 focus 到大数据上，做出了又能发 paper，又有实际价值的工作，厉害。

接下来是 Teradata（在数据仓库领域，市场占有率第一）的事业部总监，讲了他们做的 Aster，也是 SQL-to-MapReduce。不同的是，他们主要改进的功能方面。他们的工具与自己的数据仓库结合紧密，将很多常用的分析算法做成功能包，最后用户直接使用功能包提供的关键字或语句简单结合 SQL 就能完成类似 Mahout+Hive 的功能。Teradata 的 ppt 功底深厚，我表示他画的层次图真跟楼房的层次一样啊！演讲者也挺有意思，普通话有点磕绊，却很逗。

之后的英特尔亚太研发总经理一上来就表示，我们已经与时俱进，不仅仅是一家芯片厂，而是做前沿软件的了。英特尔多年来默默开发、维护并最后发布了自己的独家大数据处理平台，实际上是改良过的 Hadoop+HBase+Mahout+etc.。其实我对他们的印象还停留在他们在 USENIX 上发的 HiBench 论文上，论文主要关注了 MapReduce 的 Benchmark，从论文看，分析的挺到位的，以后搞 MapReduce 测试的时候可以深入研究下。

接下来的，云计算、虚拟化方面领头羊 VMware 介绍了他们开发的容易在虚拟机上部署 Hadoop 的工具 Serengeti（搜了一下，意思是坦桑尼亚的塞伦盖蒂国家公园），号称 10 分钟就可以配一个虚拟机上的 Hadoop 集群。还讲了让用户最担心的性能问题，意思就是只要你愿意花钱买我们的企业版，不用担心虚拟机带来的性能损耗。其实 VMware 做的东西可以直接拿来发 paper。

然后出场的是从雅虎北京研发中心的总监，由于中文没英文流利，就用英文讲了。当时已经困了，加上 ppt 字太小，不记得讲啥了。最后出场的是 MemSQL 的创始人，虽然从华盛顿大学毕业并在微软呆了 8 年，却还是年轻的样子。装束一看就是 Geek，大冬天黑色短袖，后面印着霸气侧漏的“Our database is 30X faster than yours”，演讲前还在 Mac 上 Linux shell（Sorry，是 Unix Shell）。MemSQL 是今年推出的主存数据库，一个字“快”，强调数据

的实时分析，有段时间在微博上很火。我在他 slides 上只看到一行震惊的话“Efficient SQL-to-C++ conversion”。

第一天完了，回去给师弟师妹们一块吃了饭，感觉年轻了不少。

【第二天】

场面有条不紊，有文科生参与组织的活动确实和只有理工科的不一样。

五个分会场讨论五个相关议题，议题前面已经提到过。我主要去听了“Hadoop 生态系统”和“大数据应用”。

先是腾讯的“Hive 实战”，想起这个项目应该也有毕业的易师兄的参与贡献，索性认真听了一下，发现腾讯走了 Facebook 一样的路子，将 Hive 作为全公司的主要数据仓库。腾讯将自己的 Hive 叫做 TDW。该系统揉合了 Oracle，PostgreSQL，Hive，Python 的特性，支持权限管理、过程语言、窗口函数、多维分析、不完美的单行数据插入等功能，另外也有配套的集成开发环境等等。哎，不知道多少该系统的开发人员加了多少班来满足各种狗血需求，做平台的人伤不起。从这个嘉宾的演讲中捕获了一点点对我目前工作有用的信息，也算有所收获了。

接下来换了场地去听 VMware 的“Hadoop 内核在虚拟化平台上的优化与扩展”，由于之前在 Xen，KVM 上都实验过 Hadoop 的性能，还萌发过做一些这方面的研究工作，所以比较好奇 VMware 自己做了什么有价值的工作。除了第一天听到的 Serengeti，他们还做了数据存储方面的改进，我表示这个改进很 cool。当然，由于存储方案改进，调度器要进行相应的修改。总之，VMware 强调高度的伸缩性，无论是 scale-up 还是 scale-out，表示都能妥妥地 hold 住。

然后会折返到原来的会场，听了计算所新开的公司的技术总监做 HBase 的用例分析，演讲者算是跟我同学并且是一级的，气质不错。哎，看看人家都混到 CTO 了，要加油啊。该嘉宾帮助淘宝用 HBase 重新实现了数据魔方业务，从原来只能查询到 7 天的数据，扩展到 30 天。本质问题是 HBase 的多维查询，他利用了很多数据存储 trick 来提升性能。另外还介绍了两个具体应用，尝试过很多解决方案，Redis 啊等等，分析了各种解决方案的利弊，强调了带宽的瓶颈。实战经验丰富，善于通过改动提升性能，满足需求。

上午最后一个是 Facebook 的软件工程师，主要讲 HDFS 和 HBase，由于主讲人主要做 HDFS，我也就主要听了 HDFS。Facebook 前一段公布了自己改造过的 Hadoop 和调度器，主要看点是 HDFS 的 HA（高可用性），包括 HDFS Federation，HDFS 不停机升级技术，AvatarNode，通过 Reed Solomon 校验技术来降低数据存储空间，都是很 cool 的技术。HBase 没多听。

中午去吃了名叫“阿拉侬”的上海口味快餐，面做的还可以，小笼包也不错，大拌菜好吃，只是排队有点长。这让我想起来 5 月底在上海呆的一周，感觉很美好。

由于时间安排的紧，下午报告开始才进场，第一个听了阿里巴巴 Hadoop 集群的架构和服务体系，这哥们讲他们在管理维护 Hadoop 集群，对外提供服务方面的各种考虑和辛酸史。我听了眼泪都快流下来了，你说的我感同身受啊（这个成语貌似用错了）。具体内容讲的很

细，只能说是各种小改进，大部分内容都有体会。他提到一个我最近比较感兴趣的问题，以后要向他多多请教。

接下来又是 VMware 的，我不想再听第三遍，去旁边会场听了会 IBM 的大数据战略，实在感觉讲的让我想睡觉，就跑去楼下瞻仰了上海的周傲英教授，周教授很有激情地分析了传统数据库和现在的大数据平台。

回到上面，去听了这次大会我认为最 awesome 的报告“海量数据分布式数据库的探索：Wasp”，阿里巴巴的海量数据架构师代志远讲的。之所以这么关注，是在开会前一天看到好像是 CSDN 对他的采访，他提到了一个我最近关注的问题，并且说了一些令我震惊并强烈赞同的观点。当然，本次他讲的这个研究项目是看起来颇具技术含量的——山寨 MegaStore（Google 08 年发表的第二代数据库，虽然现在 Google 已经发展到了 Spanner，F1），是构建在 BigTable，支持 BigQuery 的分布式数据库。与 OLAP 的数据仓库有着质的区别。之所以说这个山寨难做，是因为 Transaction，而且是分布式事务。这里来个双关语，transactions 有多难发，这个项目就有多难做。一句话表示就是“分布式系统没有简单问题”。他们目前的参考只有这篇论文，其他就要凭自己的理论和实践经验去探索、去实现、去改进。他们目前还在原型阶段，基于 HBase，并结合了的消息队列，目前原型支持一些基本的事务，正在完善中。他的目标是做到 Apache 的顶级项目中去，要知道目前 Apache 中没有一个顶级项目来自中国。其实之前淘宝开源的 OceanBase 也是海量数据库，支持事务，不过该系统不能算是一个真正意义上的分布式数据库，因为方案看起来不完美。

后面听了同样来自 Hortonworks 的工程师，讲了 Pig 的性能优化，Pig 在 Yahoo 大量使用，而且发展很快，现在已经是 0.11 版本了，支持了很多很 cool 的功能。相比 Hive 的类 SQL 式查询方式，我更喜欢 Pig 的脚本式的查询方式，因为作为码农，更习惯偏向过程式的开发语言。以后关注 Pig 时，可以考虑回顾下他的 slides。

【个人观点】

去年写到 MapReduce 已经沦为白菜技术，今年的看法有所改变。其实对用户来说，需要一个分布式文件系统、一个分布式 OLTP 数据库，一个分布式 OLAP 数据仓库，一个支持挖掘和机器学习的分布式计算框架。至于 MapReduce 不 MapReduce，用户不 care，care 的是去实现这些系统的架构师和苦逼开发、测试人员。一切苦难的来源于分布式，刚才强调过“分布式没有简单问题”。单机上的上述系统做的太好，用户都被惯坏了，任何对原有功能的不支持或者支持不好都会受到用户的批评，因此，压力大。

【对比工业界，学术界表示我们来了】

去年写到“对比工业界，学术界表示压力很大”，是因为之前学术界没有一个可用、好用的大数据软件平台。所谓“三十年河东，三十年河西”。以 University of California, Berkeley 的 AMPLab 为首的学术研究人员开始慢慢改变这种格局。已经 0.6 版本的 Spark 和 0.2 版本的 Shark 正在颠覆 Hadoop 的核心。Spark 使用了同样来自学术界（瑞士的洛桑联邦理工学院 EPFL）的 Scala 兼容函数式面向对象的静态类型编程语言，揉合了 MapReduce/Dryad/Pig 的特性，以内存计算为目标，极大提高了大数据的处理性能。更 amazing 的是，稍加在 Spark

上改动，就可以顺便实现 Pregel 等系统的功能。Shark 是 Spark 上的 Hive，解决 OLAP 的问题，性能是最大卖点，八成是明年 Sigmod 或 VLDB 的 best paper。

在大规模图处理方面，CMU 的 GraphLab 今年以 2 篇 OSDI 和实际可用、好用的系统来确立自己的大规模图处理方面的地位。

在以后的开源分布式 OLTP 数据库实现中，要求更高了，需要开发人员熟悉分布式事务理论，需要多看 paper，光凭经验是不够的。

【以后会怎样】

如果真有 2012，这个不用多操心了。

如果没有，谁知道咋样，走着说着吧，focus 该 focus 的事情，写 paper 要是有些总结的效率就好了。Anyway，return to normal。