

Hadoop 气数已尽？

Hadoop

我先从一个悲观的观点说起：Hadoop 正在迅速失去市场，我们可以从 Google 趋势走向看出这个现象：



如果想及时了解Spark、Hadoop或者HBase相关的文章，欢迎关注微信公众号：iteblog_hadoop

下面的炒作生命周期表也上面的趋势很类似：



如果想及时了解Spark、Hadoop或者HBase相关的文章，欢迎关注微信公众号：iteblog_hadoop

看起来 Hadoop 正处于炒作图的下坡轨道上，正在走向灭亡。我们都知道前段时间 Cloudera 已经收购了 Hortonworks，这意味着市场上最大的两个 Hadoop 厂商现在只有一个。尽管收购成功进行了，但是 Cloudera 远未在股市上取得成功，特别是6月6日 Cloudera 的股价几乎腰斩：

Cloudera NYSE: CLDR

5.81 USD **+0.080 (1.40%)** ↑

6月21日 GMT-4 上午10:29 · 免责声明

1天 5天 1个月 **6个月** YTD 1年 5年 最大



开盘	5.96	股息收益率	-
最高	6.01	昨日收盘价	5.73
最低	5.78	52 周高点	20.18
市值	15.93亿	52 周低点	4.89
市盈率	-		

如果想及时了解Spark、Hadoop或者HBase相关的文章，欢迎关注微信公众号：iteblog_hadoop

从本质上讲，市场上只剩下一个主要的 Hadoop 服务商 Cloudera。如果我告诉你 Cloudera 很久以前就不是主要搞 Hadoop 的呢？这是一个有趣的事情：根据[互联网档案](#)，过去几年中 Cloudera 首页 (<http://cloudera.com>) 上出现“Hadoop”这个词的次数如下：

- 2008年 - 4次
- 2009年 - 11次
- 2010年 - 29次
- 2011年 - 37次
- 2012年 - 23次
- 2013年 - 9次
- 2014年 - 4次
- 2015年 - 8次
- 2016年 - 6次
- 2017年 - 1次
- 2018年 - 1次
- 2019年 - 2次

如今，Cloudera 在其网站首页以粗体字写着：“我们为任何数据提供企业数据云，从 Edge 到 AI”（We deliver an Enterprise Data Cloud for any data, anywhere, from the Edge to AI）。我们可以清楚地看到焦点的转变 - 不再是 Hadoop 和 CDH，不再是大数据。现在他们做企业云和人工智能，只能在“Quickstart VMs”连接进入的页面上找到对 CDH 相关的东西。

但是 Hadoop 真的很糟糕吗？一点也不！事实上，这并不是 Hadoop 在走向终结，而是“大数据”的炒作。在介绍这个之前，让我们先来看看 Apache Spark。

Apache Spark

Apache Spark 是搭上“大数据”最后一班车的成员，下面是 Apache Hadoop 和 Apache Spark 的 Google 全球搜索趋势：



如果想及时了解Spark、Hadoop或者HBase相关的文章，欢迎关注微信公众号：iteblog_hadoop

从图中可以看到，Spark 最近几年迅速崛起，与 Hadoop 并列成为大数据最火的框架。基于此图表，我们可以清楚地看到它已经达到了“大数据”市场的上限。这意味着没有更多的水平增长空间，唯一的前进方法是垂直增长。这就是为什么 2019 年我们不再有 Spark Summit，而出现一个闪亮的新 Spark + AI 峰会。

Big Data

大数据是处理大量数据的问题。但这个词被大肆炒作，现在它有明显的消极成份。在炒作的高峰时期，任何东西都可能被贴上“大数据”的标签来提升销量。然而，很明显“大数据”本身并不是一件事，并且本身没有任何价值。

“大数据”是 2000 - 2005 年几个大型互联网公司所面临的问题。在那个时间点，这是一个非常具

有挑战性的问题。没有关于如何处理它的知识，当然也没有这样做的开源解决方案。许多大型互联网公司已成为该行业的领导者，并向我们赠送了我们现在称之为“大数据”的东西：谷歌的 GFS，MapReduce 和 BigTable，雅虎的 Hadoop，Facebook 的 Cassandra 和 Hive，Twitter 的 Storm，LinkedIn 的 Kafka。大型互联网公司通过发明新方法和工具来利用他们必须处理的大量数据来推动革命。它们中的许多公司都开源了它们的软件，使得这些软件可供全世界使用。这是一个关键时刻，因为它诞生了一系列创业公司，其使命是将所有这些解决方案出售给传统企业。其中包括 Cloudera，Hortonworks，MapR 和其他许多公司。

围绕“大数据”的炒作很大程度上是由于上述创业公司对其营销的巨额投资以及传统企业中 IT 人员的短视而造成的。市场营销已经利用了大型互联网公司生产的“大数据”技术与该公司的成功之间的联系。他们的营销材料并没有直接说明这一点，但它的字面意思是“使用 Cassandra 并且像 Facebook 一样成功”，“使用 Kafka 并达到 LinkedIn 的规模”，“使用 Hadoop 并变得像 Google 一样富有”。总体而言，“大数据”并不是在销售技术，而是将大型 IT 巨头的成功卖给传统公司。

不出所料，许多企业正在购买这些技术，并在其堆栈中实施这些技术。由于这一实施，他们通常大胆宣布他们正在利用“大数据”的力量，他们的企业在这个问题上取得了进步。然而，通常实现本身更像是一个实验 -

除了主要的数据处理管道之外，一些小而孤立的案例，甚至可能无法交付给生产并保留在 PoC 或 MVP 级别。

然而，许多小型企业正在购买大型企业的这一信息及其成功案例，并将其资金和努力投入到“大数据”中。通过这种方式，大肆宣传成为一种大雪球，越来越多的资深人士直言不讳或不能说出完整的真相，营销人员利用他们的话语（有时会删除重要的背景）来进一步推广他们的解决方案。

一个时代的结束

所以，我并不是说一些新的突破性技术已经取代了“大数据”，我也不是说 Hadoop 不再是一种可行的技术，不再值得投资。我说的是“大数据”时代即将结束，从炒作的高峰下降到最低点。新的趋势 AI 和 ML，已经取代它们，生命的循环再次开始，新的技术在炒作图上攀升，营销人员推销新软件，以科技巨头的成功为代表，以及传统企业购买这种软件，消灭了下一个科技泡沫。

Hadoop 时代真的结束了吗？

并没有！Hadoop 是一项伟大的技术，但它本质上是一个很好的解决方案，但是只有少数企业真正需要它。作为一项技术，它与提供替代大规模存储解决方案的主要云厂商竞争：AWS 包含 S3，GCP 包含云存储，Microsoft 包含 Azure 存储等。云计算一点一点地吞噬了自建部署市场，云计算提供商及其分布式存储解决方案在我看来是 Hadoop 的主要竞争对手，Hadoop 未来将面临更多的挑战。

本博客文章除特别声明，全部都是原创！
原创文章版权归过往记忆大数据（[过往记忆](#)）所有，未经许可不得转载。
本文链接: [【】（）](#)