Big Data Hyperlinks

**2017**

**November**

**Can we Trust Hadoop Benchmarks?**

How to approach online benchmark tests by vendors. [HIGHLY Recommended]

[**Double-Digit Growth Forecast for the Worldwide Big Data and Business Analytics Market Through 2020 Led by Banking and Manufacturing Investments, According to IDC**](https://www.idc.com/getdoc.jsp?containerId=prUS41826116) **.**

[**Big Data. Why the boom is already over**](http://www.zdnet.com/article/big-data-why-the-boom-is-already-over/). Summary of Gartner report. *The rumors of my death ARE BEING greatly exaggerated.*

[**Gartner Survey Reveals Investment in Big Data Is Up but Fewer Organizations Plan to Invest**](http://www.gartner.com/newsroom/id/3466117). Those who plan to invest in big data within the next two years fell from 31 to 25 percent.

[**Apache Spark Overview**](https://jaceklaskowski.gitbooks.io/mastering-apache-spark/spark-overview.html). A review moving progressively into technical and big data speak. Also a little biased towards forgiving Spark’s short comings.

**December**

[**Hadoop, we hardly knew ye**](http://www.infoworld.com/article/3141314/analytics/hadoop-we-hardly-knew-ye.html). “Hadoop was synonymous with big data. Now, Spark and cloud services are riding high as Hadoop heads off into the sunset.” [Recommended]

**2018**

**March**

[**Top 10 TED Talks for Data Scientists and Machine Learning Engineers**](https://www.kdnuggets.com/2018/01/top-10-ted-talks-data-scientists-machine-learning.html)**.** My favorite is [*How computers are learning to be creative*](https://www.ted.com/talks/blaise_aguera_y_arcas_how_computers_are_learning_to_be_creative).

[**2018 Gartner Magic Quadrant for Data Management Solutions for Analytics**](http://data.teradata.com/LP=218). In case you missed it, the MQ is free on the Teradata site with email ID login. [Recommended]

[**Driving Forces for Taking Analytics into the Cloud**](https://www.teradata.com/Resources/White-Papers/Driving-Forces-for-Taking-Analytics-into-the-Cloud)**.** White paper by Stephen Brobst on planning and configuring clouds. [Recommended]

**April**

[**Critical Capabilities for Data Management Solutions for Analytics**](https://www.gartner.com/doc/reprints?id=1-4P8TVEJ&ct=180119&st=sb)**.** URL only. Free copy of Garrtner report showing Teradata winning 3 of 4 use cases, earning three 5.0 scores (gold medals). <https://gtnr.it/2JghFzn> Pass URL to clients, prospects, & Mom.

Publishing rights purchased by Teradata [Recommended]

**May**

[**Big Data 2018 Cloud storage becomes the de facto data lake**](https://www.zdnet.com/article/big-data-2018-cloud-storage-becomes-the-de-facto-data-lake/)**. “**Or as Mike Olson prophetically stated back in 2014, [Hadoop would disappear](https://www.onstrategies.com/blog/2014/10/19/strata-2014-part-1-hadoop-bright-lights-big-city/)..” and “ Cloudera and Hortonworks decouple their data governance services from HDFS. So the data lake is wherever you store data.”

**June**

**Use Design Patterns to Increase the Value of Your Data Lake**. A comprehensive review of data lake designs, myths, macro and micro architectures. A good handbook to keep around. [Recommended]

**July**

[**What’s the Difference Between Data Integration and Data Engineering**](https://www.kdnuggets.com/2018/06/difference-between-data-integration-data-engineering.html). Bill Schmarzo, EMC’s self proclaimed Dean of Big Data, tackles definitions including data scientist.

[**Great Power, Great Responsibility. The 2018 Big Data & AI Landscape**](http://mattturck.com/bigdata2018/)**. URL only**. A VC at FirstMark Capital summarizes the major trends of the last year such as AI, Hadoop, cloud, and other topics. Bonus is the 3mb [Big Data and AI Landscape vendors](http://mattturck.com/wp-content/uploads/2018/06/Matt-Turck-FirstMark-Big-Data-Landscape-2018.png) poster map. Find Teradata logo in 2 places. Courtesy of Karthik Guruswamy.

[**Apache Spark. Scala vs. Java v. Python vs. R vs. SQL**](https://mindfulmachines.io/blog/2018/6/apache-spark-scala-vs-java-v-python-vs-r-vs-sql26). While SQL outperforms the other languages in the graphs, the conclusion is avoid R UDFs and use Scala or Java. [Recommended]

[**The EDW Lives On. The Beating Heart of the Data Lake**](https://hortonworks.com/info/data-lake-vs-data-warehouse/). Hortonworks gets Devlin (father of data warehousing) to include Hortonworks EDW Optimization Solution in his white paper. Proves IDW & lakes work together. [Click here](https://hortonworks.com/solutions/edw-optimization/) for HDP EDW optimization web page. [Recommended]