

A Review on Data Lake

Shravan R Poojary¹, Pradeep Nayak², Shrivitha³, Shreya Rai⁴, Shrujan Kumar⁵, Chandana N M⁶

Department of Information Science and Engineering
Alvas Institute of Engineering and Technology, Mijar, Karnataka, India

Abstract: *One of the contentious ideas to emerge in the big data era. The original concept for Data Lake originated in the business sphere rather than the academic one. Due to its revolutionary features and recent development, Data Lake faces a variety of adoption problems. However, Because data lakes have the ability to alter the data environment, it is worthwhile to conduct research on them.*

Keywords: Data Lake

REFERENCES

- [1] Toon Calders, Oscar Romero, Alberto Abell'o, and Ayman Alserafi, Towards Information Profiling: Data Lake Content: 16th IEEE International Conference on Data Mining Workshops on Metadata Management.
- [2] According to Brian Stein and Alan Morrison's article from 2014, "The Enterprise Data Lake: Better Integration and Deeper Analytics, Technology Forecast: Rethinking Integration," Issue 1, www.pwc.com/us/en/technology-forecast/2014/cloudcomputing/assets/pdf/pwc-technology-forecast-data-lakes.pdf, retrieved on August 25, 2017.
- [3] Chris Campbell, "Top 5 Data Lake vs. Data Warehouse Differences," 2015-01-26, Blue Granite Accessed August 25, 2017. <https://www.bluegranite.com/blog/bid/402596/top-five-differences-between-data-lakes-and-datawarehouses>.
- [4] 5 Keys to Creating a Killer Data Lake by Chuck Yarbrough 2017-07-21, retrieved. <http://www.pentaho.com/blog/5-keys-creating-killer-data-lake>.
- [5] According to Dan Wood, "Big data requires a big new architecture," Forbes, accessed August 8, 2017. <https://www.forbes.com/sites/ciocentral/2011/07/21/big-data-requires-a-big-new-architecture/#66609cb61157>.
- [6] Gartner.Inc., Gartner Says Beware of the Data Lake Fallacy, STAMFORD, Conn., July 28, 2014, retrieved on August 29, 2017. <http://www.gartner.com/newsroom/id/2809117>.
- [7] Hassan Alrehamy Personal Data Lake With Data Gravity Pull by Coral Walker, 5th IEEE International Conference on Big Data and Cloud Computing, Dalian, China, 26-28 August 2015.
- [8] The 5th Annual IEEE International Conference on Cyber Technology in Automation, Control and Intelligent Systems, June 8–12, 2015, Shenyang, China, Huang Fang, Managing Data Lakes in the Big Data Era: What's a Data Lake and Why Has It Become Popular in the Data Management Ecosystem
- [9] Pentaho, Hadoop, and Data Lakes, James Dixon, retrieved 10 August 2017. <https://jamesdixon.wordpress.com/2010/10/14/pentaho-hadoop-and-data-lakes/>
- [10] MapReduce: Simplified Processing on Large Cluster, Jeffrey Dean and Sanjay Ghemawat, Communication of the ACM, Vol. 51, No. 1, Jan 2008.
- [11] Uur Etintemel, Michael Stonebraker, Proceedings: One Size Fits All: An Idea Whose Time Has Come and Gone. April 2005, 21st International Conference on Data Engineering. 2005 ICDE in Tokyo, Japan.
- [12] Application of Big Data, Fast Data, and Data Lake Concepts to Information Security Issues, Natalia Miloslavskaya and Alexander Tolstoy, 2016 4th International Conference on Future Internet of Things and Cloud Workshops.
- [13] Top 10 Big Data Trends for 2017, according to Tableau, accessed 30 August 2017:
- [14] Tamara Dull, Data Lake Vs. Data Warehouse: Key Differences, <http://www.kdnuggets.com/2015/09/data-lake-vs.-data-warehouse-key-differences.html>, retrieved Sep. 26, 2017.
- [15] Timothy King in Best Practices, What's the Difference Between a Data Lake and a Data Warehouse? <https://solutionsreview.com/datamanagement/data-warehouse-vs-data-lake-whats-the-difference/> (accessed September 10, 2017)
- [16] Timothy King, "The Emergence of Data Lake: Pros and Cons," March 3, 2016, accessed September 15, 2017 at <https://solutionsreview.com/data-integration/the-emergence-of-data-lake-pros-and-cons/>.

[17] Hadoop: The Definitive Guide, 4th edition, Storage and Analysis at Internet Scale, Tom White, O'Reilly Media, Inc., 1005 Gravenstein Highway North, Sebastopol, USA, 2015.