Big Data

MapReduce vs. RDBMS

Hadoop

❖ MapReduce vs. RDBMS

- RDBMS (Relational Database Management System) Characteristics
  - RDBMS is good for updating a small proportion of a big database
  - RDBMS uses a traditional B-Tree, which is highly dependent in the time required to perform seek operations

Hadoop

❖ MapReduce vs. RDBMS

- MapReduce Characteristics
  - MapReduce is good for updating all (or a majority) of a big database
  - MapReduce uses Sort and Merge to rebuild the database, which depends more on transfer operations

Hadoop

❖ MapReduce vs. RDBMS

- RDBMS is good for applications that require the datasets of the database to be very frequently updated (e.g., point queries or small dataset updates)
- MapReduce is better for WORM (Write Once and Read Many times) based data applications
- MapReduce is a complementary system to RDBMS

Hadoop

❖ MapReduce vs. RDBMS

	RDBMS	MapReduce
Data Size	Gigabytes (10^9)	Petabytes (10^{12})
Access	Interactive & Batch	Batch
Updates	Read & Write Many Times	WORM (Write Once, Read Many Times)
Data Structure	Static Schema	Dynamic Schema
Integrity	High	Low
Scalability	Nonlinear	Linear

Hadoop

❖ MapReduce vs. RDBMS: Data Types

- *Structured* Data: Data that has a formal defined structure (e.g., XML documents or database tables)
- *Semi-Structured* Data: Data that has a looser format where the data structure is used as a guide and may be ignored
- *Unstructured* Data: Data that does not have any formal structure (e.g., plain text or image data)

Hadoop

❖ MapReduce vs. RDBMS: Data Types

- MapReduce is very effective on unstructured and semi-structured data
- Why?
  - MapReduce interprets data during the data processing sessions
  - MapReduce does not use intrinsic properties of the data as input keys or input values
  - Parameters used are selected by the person analyzing the data

Hadoop

❖ MapReduce vs. RDBMS: Data Types

- MapReduce has a programming model that is linearly scalable
- MapReduce Functions: 2 types
  - Map function
  - Reduce function
- Both of these functions define a Key-Value pair mapping relation (e.g., Key-Value pair 1 ➔ Key-Value pair 2)

Big Data

References

References

- V. Mayer-Schönberger, and K. Cukier, Big data: A revolution that will transform how we live, work, and think. Houghton Mifflin Harcourt, 2013.
- T. White, Hadoop: The Definitive Guide. O'Reilly Media, 2012.
- J. Venner, Pro Hadoop. Apress, 2009.
- S. LaValle, E. Lesser, R. Shockley, M. S. Hopkins, and N. Kruschwitz, "Big Data, Analytics and the Path From Insights to Value," MIT Sloan Management Review, vol. 52, no. 2, Winter 2011.
- B. Randal, R. H. Katz, and E. D. Lazowska, "Big-data Computing: Creating revolutionary breakthroughs in commerce, science and society," Computing Community Consortium, pp. 1-15, Dec. 2008.
- G. Linden, B. Smith, and J. York. "Amazon.com Recommendations: Item-to-Item Collaborative Filtering," IEEE Internet Computing, vol. 7, no. 1, pp. 76-80, Jan/Feb. 2003.

References

- J. R. GalbRaith, "Organizational Design Challenges Resulting From Big Data," Journal of Organization Design, vol. 3, no. 1, pp. 2-13, Apr. 2014.
- S. Sagiroglu and D. Sinanc, "Big data: A review," Proc. IEEE International Conference on Collaboration Technologies and Systems, pp. 42-47, May 2013.
- M. Chen, S. Mao, and Y. Liu, "Big Data: A Survey," Mobile Networks and Applications, vol. 19, no. 2, pp. 171-209, Jan. 2014.
- X. Wu, X. Zhu, G. Q. Wu, and W. Ding, ''Data Mining with Big Data,'' IEEE Transactions on Knowledge and Data Engineering, vol. 26, no. 1, pp. 97–107, Jan. 2014.
- Z. Zheng, J. Zhu, and M. R. Lyu, ''Service-Generated Big Data and Big Data-as-a-Service: An Overview,'' Proc. IEEE International Congress on Big Data, pp. 403–410, Jun/Jul. 2013.

References

- I. Palit and C.K. Reddy, "Scalable and Parallel Boosting with MapReduce," IEEE Transactions on Knowledge and Data Engineering, vol. 24, no. 10, pp. 1904-1916, 2012.
- M.-Y Choi, E.-A. Cho, D.-H. Park, C.-J Moon, and D.-K. Baik, "A Database Synchronization Algorithm for Mobile Devices," IEEE Transactions on Consumer Electronics, vol. 56, no. 2, pp. 392-398, May 2010.
- IBM, What is big data?, http://www.ibm.com/software/data/bigdata/what-is-big-data.html [Accessed June 1, 2015]
- Hadoop Apache, http://hadoop.apache.org
- Wikipedia, http://www.wikipedia.org

Image sources

- Walmart Logo, By Walmart [Public domain], via Wikimedia Commons
- Amazon Logo, By Balajimuthazhagan (Own work) [CC BY-SA 3.0 (http://creativecommons.org/licenses/by-sa/3.0)], via Wikimedia Commons