Apache HBase

Functional Area: Database:NoSQL:Column

Overview:

Apache HBase is an open-source software package modeled after Google Bigtable database [1], a part of the Apache Big Data Stack that serves as a scalable database capable of managing and organizing petabyte size data sets [2]. These data sets form an enormous system of tables, billions of rows by millions of columns, which provide storage space and meet head-on the challenges associated with Big Data.

Features presented in the Big Table paper which have been implemented in HBASE include in-memory operation and the application of Bloom filters to columns. HBase can be accessed through a number of APIs, including Java, REST, Avro or Thrift.

HBase originated with a company called Powerset, which was developing natural language search engines. As an Apache project, it began as a subproject of Hadoop and became a top-level project in 2010.

References:

[1] Chang F, Dean J, Ghemawat S, Hsieh WC, Wallach DA, Burrows M, Chandra T, Fikes A, E GR: “**Bigtable: A distributed storage system for structured data.”**In Seventh Symposium on Operating System Design and Implementation. Seattle, WA: Usenix Association; 2006. [OpenURL](http://www.biomedcentral.com/sfx_links?ui=1471-2105-11-S12-S1&bibl=B32)

[2] Ronald Taylor, “An overview of the Hadoop/MapReduce/HBase framework and its current applications in bioinformatics,” in Proceedings of the 11th Annual Bioinformatics Open Source Conference, 2010 © BioMed Central Ltd. doi: 10.1186/1471-2105-11-S12-S1