Hadoop

From Wikipedia, the free encyclopedia

Apache Hadoop is a free Java software framework that supports data intensive distributed applications.[1] It enables applications to work with thousands of nodes and petabytes of data. Hadoop was inspired by Google's MapReduce and Google File System (GFS) papers.

Hadoop is a top level Apache project, being built and used by a community of contributors from all over the world[2]. Yahoo! has been the largest contributor[3] to the project and uses Hadoop extensively in its Web Search and Advertising businesses.[4] IBM and Google have announced a major initiative to use Hadoop to support University courses in Distributed Computer Programming. [5]

Apache Hadoop

Developed by	Apache Software Foundation
Latest release	0.19.0 / 2008-11-21
Written in	Java
OS	Cross-platform
Type	Distributed File System
License	Apache License 2.0
Website	http://hadoop.apache.org/

Hadoop was created by Doug Cutting (now a Yahoo! employee), who named it after his child's stuffed elephant. It was originally developed to support distribution for the Nutch search engine project.[6]

Contents

Architecture

Hadoop consists of the *Hadoop Core*, which provides access to the filesystems that Hadoop supports. As of June 2008, the list of supported filesystems includes:

- HDFS: Hadoop's own filesystem. This is designed to scale to petabytes of storage, and run on top of the fileystems of the underlying operating systems.
- Amazon S3 filesystem. This is targeted at clusters hosted on the Amazon Elastic Compute Cloud server-on-demand infrastructure. There is no rack-awareness in this filesystem, as it is all remote.
- Kosmos Distributed File System -like HDFS, this is rack-aware.

- FTP Filesystem: all the data are stored on remotely accessible FTP servers.
- Read-only HTTP and HTTPS file systems.

Hadoop Distributed File System

The HDFS filesystem is a pure-Java filesystem, which stores large files (an ideal file size is 64 MB [7]), across multiple machines. It achieves reliability by replicating the data across multiple hosts, and hence does not require RAID storage on hosts. With the default replication value, 3, data is stored on three nodes: two on the same rack, and one on a different rack.

The filesystem is built from a cluster of *data nodes*, each of which serves up blocks of data over the network using a block protocol specific to HDFS. They also serve the data over HTTP, allowing access to all content from a web browser or other client. Data nodes can talk to each other to rebalance data, to move copies around, and to keep the replication of data high.

A filesystem requires one unique server, the *name node*. This is a single point of failure for an HDFS installation. If the name node goes down, the filesystem is offline. To reduce the impact of such an event, some sites use a secondary name node for failover. Many sites stick to a single name node, relying on the name node to replay all outstanding operations when it comes back up. This replay process can take over half an hour for a big cluster. [8]

Another limitation of HDFS is that it can not be directly mounted by an existing operating system. Getting data into and out of the HDFS file system is an action that often needs to be performed before and after executing a job, so this can be inconvenient. A Filesystem in Userspace has been developed to address this problem, at least for Linux and some other Unix systems.

Job Tracker and Task Tracker: the map/reduce engine

Above the file systems come the map/reduce engine, which consists of one *Job Tracker*, to which client applications submit map/reduce jobs. The Job Tracker pushes work out to available *Task Tracker* nodes in the cluster, striving to keep the work as close to the data as possible. With a rack-aware filesystem, the Job Tracker knows which node the data live on, and which other machines are nearby. If the work cannot be hosted on the actual node where the data live, priority is given to nodes in the same rack. This reduces network traffic on the main backbone network. If a Task Tracker fails or times out, that part of the job is rescheduled. If the Job Tracker fails, the entire job is lost and must be resubmitted.

Known limitations of this approach are:

- The Job Tracker is a Single Point of Failure for submitted work.
- There is (currently) no checkpointing or recovery within a single map/reduce job.
- The allocation of work to task trackers is very simple. Every task tracker has a number of available *slots* (such as "4 slots'). Every active map or reduce task takes up one slot. The Job Tracker allocates work to the tracker nearest to the data with an available slot. There is no consideration of the current active load of the allocated machine, and hence its actual availability.
- If one task tracker is very slow, it can delay the entire operation.

Other applications

The HDFS filesystem is not restricted to map/reduce jobs. It can be used for other applications, many of which are under way at Apache. The list includes the HBase database, the Apache Mahout

machine learning system, and matrix operations. Hadoop can in theory be used for any sort of work that is batch-oriented rather than real-time, very data-intensive, and able to work on pieces of the data in parallel.

Prominent users

Hadoop at Yahoo!

On February 19, 2008, Yahoo! launched what it claimed was the world's largest Hadoop production application. The Yahoo! Search Webmap is a Hadoop application that runs on a more than 10,000 core Linux cluster and produces data that is now used in every Yahoo! Web search query.[9]

There are multiple Hadoop clusters at Yahoo!, each occupying a single datacenter (or fraction thereof). No HDFS filesystems or Map/Reduce jobs are split across multiple datacenters; instead each datacenter has a separate filesystem and workload. The cluster servers run Linux, and are configured on boot using Kickstart. Every machine bootstraps the Linux image, including the Hadoop distribution. Cluster configuration is also aided through a program called *Zookeeper*. Work that the clusters perform is known to include the index calculations for the Yahoo! search engine.

Other users

Besides Yahoo!, many other organizations are using Hadoop to run large distributed computations. Some of them include:[10]

- A9.com
- ADSDAQ by Contextweb
- EHarmony
- Facebook
- Fox Interactive Media
- IBM
- ImageShack
- ISI
- Joost
- Last.fm
- Powerset
- The New York Times
- Rackspace
- Veoh
- Metaweb

Hadoop on Amazon EC2/S3 services

It's possible to run Hadoop on Amazon Elastic Compute Cloud (EC2) and Amazon Simple Storage Service (S3)[11]. As an example The New York Times used 100 Amazon EC2 instances and a Hadoop application to process 4TB of raw image TIFF data (stored in S3) into 1.1 million finished PDFs in the space of 24 hours at a computation cost of about $240 (not including bandwidth).[12]

There is support for the S3 filesystem in Hadoop distributions, and the Hadoop team generates EC2 machine images after every release. From a pure performance perspective, Hadoop on S3/EC2 is inefficient, as the S3 filesystem is remote and delays returning from every write operation until the data are guaranteed to not be lost. This removes the locality advantages of Hadoop, which schedules

work near data to save on network load. However, as Hadoop-on-EC2 is the primary mass-market way to run Hadoop without one's own private cluster, the performance detail is clearly felt to be acceptable to the users.

Hadoop with Sun Grid Engine

Hadoop can also be used in compute farms and high-performance computing environments. Integration with Sun Grid Engine was released, and running Hadoop on Sun Grid (Sun's on-demand utility computing service) is possible. [13] Note that, as with EC2/S3, the CPU-time scheduler appears to be unaware of the locality of the data. A key feature of the Hadoop Runtime, "do the work in the same server or rack as the data" is therefore lost.

Sun also has the *Hadoop Live CD* OpenSolaris project, which allows running a fully functional Hadoop cluster using a live CD.[14] Sun plans to enhance the Grid Engine/Hadoop integration in the near future.[15]

References

1. ^ "Hadoop is a framework for running applications on large clusters of commodity hardware. The Hadoop framework transparently provides applications both reliability and data motion. Hadoop implements a computational paradigm named map/reduce, where the application is divided into many small fragments of work, each of which may be executed or reexecuted on any node in the cluster. In addition, it provides a distributed file system that stores data on the compute nodes, providing very high aggregate bandwidth across the cluster. Both map/reduce and the distributed file system are designed so that node failures are automatically handled by the framework." Hadoop Overview
2. ^ Hadoop Users List
3. ^ Hadoop Credits Page
4. ^ Yahoo! Launches World's Largest Hadoop Production Application
5. ^ Google Press Center: Google and IBM Announce University Initiative to Address Internet-Scale Computing Challenges
6. ^ "Hadoop contains the distributed computing platform that was formerly a part of Nutch. This includes the Hadoop Distributed Filesystem (HDFS) and an implementation of map/reduce." About Hadoop
7. ^ The Hadoop Distributed File System: Architecture and Design
8. ^ Improve Namenode startup performance. "Default scenario for 20 million files with the max Java heap size set to 14GB : 40 minutes. Tuning various Java options such as young size, parallel garbage collection, initial Java heap size : 14 minutes"
9. ^ Yahoo! Launches World's Largest Hadoop Production Application (Hadoop and Distributed Computing at Yahoo!)
10. ^ PoweredBy
11. ^ http://aws.typepad.com/aws/2008/02/taking-massive.html Running Hadoop on Amazon EC2/S3
12. ^ Self-service, Prorated Super Computing Fun! - Open - Code - New York Times Blog
13. ^ "Creating Hadoop pe under SGE". Sun Microsystems. 2008-01-16. http://blogs.sun.com/ravee/entry/creating_hadoop_pe_under_sge.
14. ^ "OpenSolaris Project: Hadoop Live CD". Sun Microsystems. 2008-08-29. http://opensolaris.org/os/project/livehadoop/.
15. ^ "OpenSolaris Live Hadoop with HPC Stack". Sun Microsystems. 2008-09-03. http://mail.opensolaris.org/pipermail/hpcdev-discuss/2008-September/000179.html.

See also

- HBase - BigTable-model database. Sub-project of Hadoop.
- Cloud computing

External links

- Video and podcast (49:44) - Yahoo's Parand Darugar explains Hadoop
- Hadoop website
- Hadoop wiki
- Yahoo's bet on Hadoop, an article about Yahoo's investment in Hadoop from Tim O'Reilly
- Yahoo's Doug Cutting on MapReduce and the Future of Hadoop
- An article on MapReduce Programming with Apache Hadoop
- A NYT blog mentioning that Hadoop and EC2 were used to reprocess all of The New York Times archive content
- Mention of Nutch and Hadoop in an article about Google
- IBM MapReduce Tools for Eclipse
- Problem Solving on Large Scale Clusters using Hadoop
- Pig, a high-level language over the Hadoop platform.
- Hive, a data warehousing framework for Hadoop that includes a SQL-like declarative query language.
- Cloudera, a company started by Hadoop and open source veterans from Google, Yahoo, Facebook and Oracle to provide commercial support and training for Hadoop.
- CloudBase, a data warehousing system built on top of Hadoop that uses ANSI-SQL as its query language. CloudBase creates a database system directly on flat files and converts input ANSI SQL expressions into map-reduce programs. CloudBase comes with a JDBC driver, so one can use any JDBC Database Manager application (e.g. Squirrel) as a client to connect to CloudBase.

Retrieved from "http://en.wikipedia.org/wiki/Hadoop"
Categories: Free system software | Distributed file systems
Hidden categories: Wikipedia external links cleanup