



大数据处理

*Hadoop*生态

毛波

厦门大学

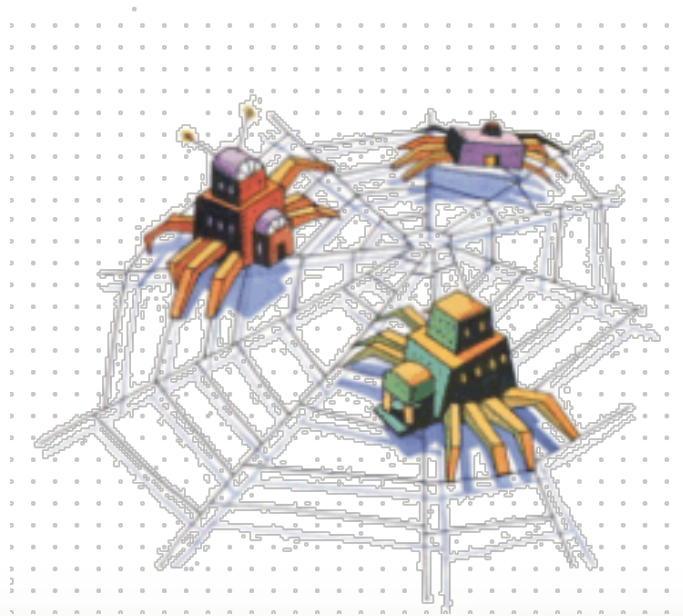
What is Hadoop?

- Apache top level project, open-source implementation of frameworks for reliable, scalable, distributed computing and data storage.
- It is a flexible and highly-available architecture for large scale computation and data processing on a network of commodity hardware.



Brief History of Hadoop

- Designed to answer the question: **“How to process big data with reasonable cost and time?”**



Search engines in 1990s



MetaCrawler Parallel Web Search Service

by [Erik Selberg](#) and [Oren Etzioni](#)

Try the new [MetaCrawler Beta!](#)
If you're searching for a person's home page, try [Abov!](#)

• [Examples](#) • [Beta Site](#) • [Add Site](#) • [About](#) •

Search for:
☐ as a Phrase ☒ All of these words ☐ Any of these words

For better results, please specify:
Search Region: Search Sites:

Performance parameters:
Max wait: minutes Match type:

[[About](#) | [Help](#) | [Problems](#) | [Add Site](#) | [Search](#)]

webmaster@metacrawler.com
© Copyright 1995, 1996 Erik Selberg and Oren Etzioni

[Serious Sports Fans Only \\$1,000,000 in Cash and Prizes!](#)
[For serious sports fans only! Play Fantasy Football!](#)



It's amazing where
Go Get It will get you.

Find:

[Enhance your search.](#)



[New Search](#) • [Top News](#) • [Sites by Subject](#) • [Top 5% Sites](#) • [City Guide](#) • [Pictures & Sounds](#)
[PeopleFind](#) • [Point Review](#) • [Road Maps](#) • [Software](#) • [About Lycos](#) • [Club Lycos](#) • [Help](#)

[Add Your Site to Lycos](#)

Copyright © 1996 Lycos[™], Inc. All Rights Reserved.
Lycos is a trademark of Carnegie Mellon University.
[Questions & Comments](#)

1996

The Excite search engine interface features a red header with the 'excite' logo. Navigation links include 'search', 'reviews', 'city.net', 'NEW live!', and 'reference?'. Below these are links for 'excite home', 'maps', 'news', and 'people finder'. The main search area has a 'What:' text input, a 'Where:' dropdown menu set to 'World Wide Web', and a 'search' button. To the left, there are links for 'Researching stocks?', 'Buying a car?', 'Planning a wedding?', 'Check out Excite Seeing Tours.', and 'Bill Mitchell: Satire that clicks!'. To the right, there's a section for 'Excite Reviews: site reviews by the web's best editorial team.' with a grid of category links: Arts, Business, Computing, Education, Entertainment, Health, Hobbies, Life & Style, Money, News & Reference, Personal Pages, Politics & Law, Regional, Science, Shopping, and Sports.

The Wired Search Center interface has a purple header with 'HELP' and navigation links for 'WIRED NEWS', 'HOTWIRED', 'WIRED MAGAZINE', and 'SUCK.COM'. The main area is green and features a 'The WIRED Search Center' section with a 'look for' dropdown set to 'all the words' and a 'SEARCH' button. Below this are filters for 'Date' (set to 'in the last week') and 'Content' (set to 'North America (.com)'). There's also an 'Include media type' section with checkboxes for 'Image', 'Audio', 'Video', and 'Shockwave'. A 'Return Results' section shows '10' results and 'full descriptions'. On the left, there's a sidebar with links: 'Search: The Web', 'Usenet', 'Top News Sites', 'Classifieds', 'Domain Names', 'Stocks', 'Discussion Groups', 'ShareWare', 'Find:', 'Businesses', 'People', and 'Email Addresses'. On the right, there's a 'Sandbox Entertainment' section with a 'Shop WIRED Holiday Gift Guide' and a 'SOMETHING HAS SURVIVED.' advertisement. At the bottom right, there's a 'Cybernet Outpost' and 'Microsoft Expedia Travel' section.

1997



1996

Google search engines

Google!
BETA

1998

Search the web using Google!

Special Searches
[Stanford Search](#)
[Linux Search](#)

[Help!](#)
[About Google!](#)
[Company Info](#)
[Google! Logos](#)

Get Google!
updates monthly:

 [Archive](#)

Copyright ©1998 Google Inc.

2013

Google

Google Search

I'm Feeling Lucky



Hadoop's Developers



Doug Cutting



2005: Doug Cutting and Michael J. Cafarella developed Hadoop to support distribution for the [Nutch](#) search engine project.

The project was funded by Yahoo.

2006: Yahoo gave the project to Apache Software Foundation.



Google Origins

2003

The Google File System

Sanjay Ghemawat, Howard Gobioff, and Shun-Tak Leung
Google*



2004

MapReduce: Simplified Data Processing on Large Clusters

Jeffrey Dean and Sanjay Ghemawat

jeff@google.com, sanjay@google.com

Google, Inc.



2006

Bigtable: A Distributed Storage System for Structured Data

Fuy Chang, Jeffrey Dean, Sanjay Ghemawat, Wilson C. Hsieh, Deborah A. Wallach
Mike Burrows, Tushar Chandra, Andrew Fikes, Robert E. Gruber
{fuy.jeff.sanjay.wilson.chandra.willach@cs.stanford.edu, tushar.fikes.gruber}@google.com

Google, Inc.

Abstract

Bigtable is a distributed storage system for managing structured data that is designed to scale to a very large number of nodes across thousands of commodity servers. Many projects at Google store data in Bigtable, including web indexing, Google Earth, and Google Fx. These applications place very different demands on Bigtable, both in terms of data size (from URLs to

achieved scalability and high performance, but Bigtable provides a different interface than such systems. Bigtable does not support a full relational data model; instead, it provides clients with a simple data model that supports dynamic control over data layout and format, and allows clients to reason about the locality properties of data represented in the underlying storage. Data is indexed using row and column names that can be arbitrary strings. Bigtable also treats data as uninterpreted str



Some Hadoop Milestones

- **2008 - Hadoop Wins Terabyte Sort Benchmark** (sorted 1 terabyte of data in 209 seconds, compared to previous record of 297 seconds)
- 2009 - Avro and Chukwa became new members of Hadoop Framework family
- 2010 - Hadoop's Hbase, Hive and Pig subprojects completed, adding more computational power to Hadoop framework
- **2011 - ZooKeeper Completed**
- **2013 - Hadoop 1.1.2 and Hadoop 2.0.3 alpha.**
 - Ambari, Cassandra, Mahout have been added



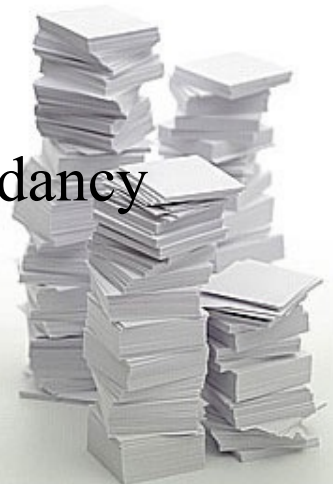
What is Hadoop?

- **Hadoop:**

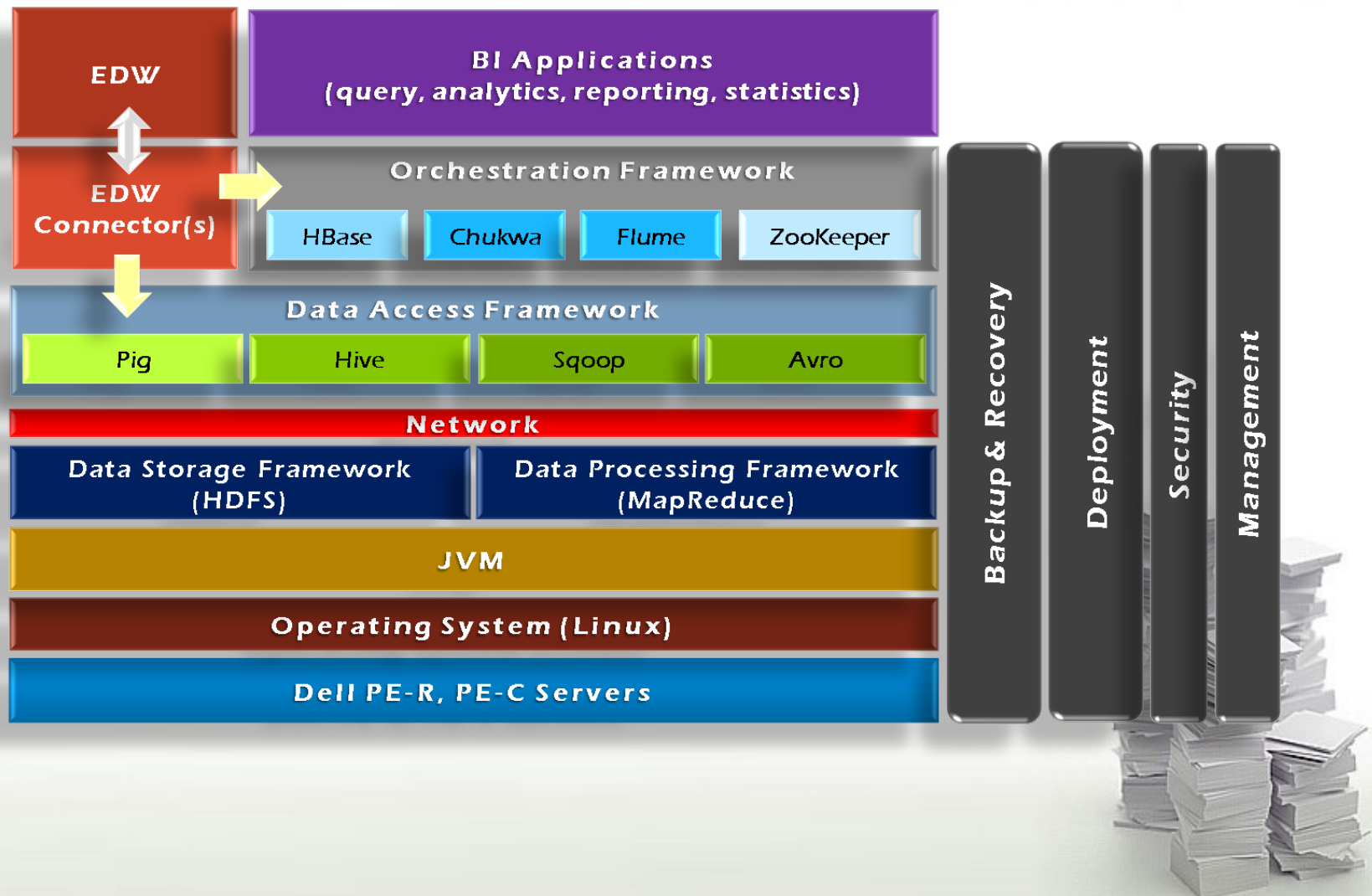
- an open-source software framework that supports data-intensive distributed applications, licensed under the Apache v2 license.

- **Goals / Requirements:**

- Abstract and facilitate the storage and processing of large and/or rapidly growing data sets
 - Structured and non-structured data
 - Simple programming models
- High scalability and availability
- Use commodity (cheap!) hardware with little redundancy
- Fault-tolerance
- Move computation rather than data

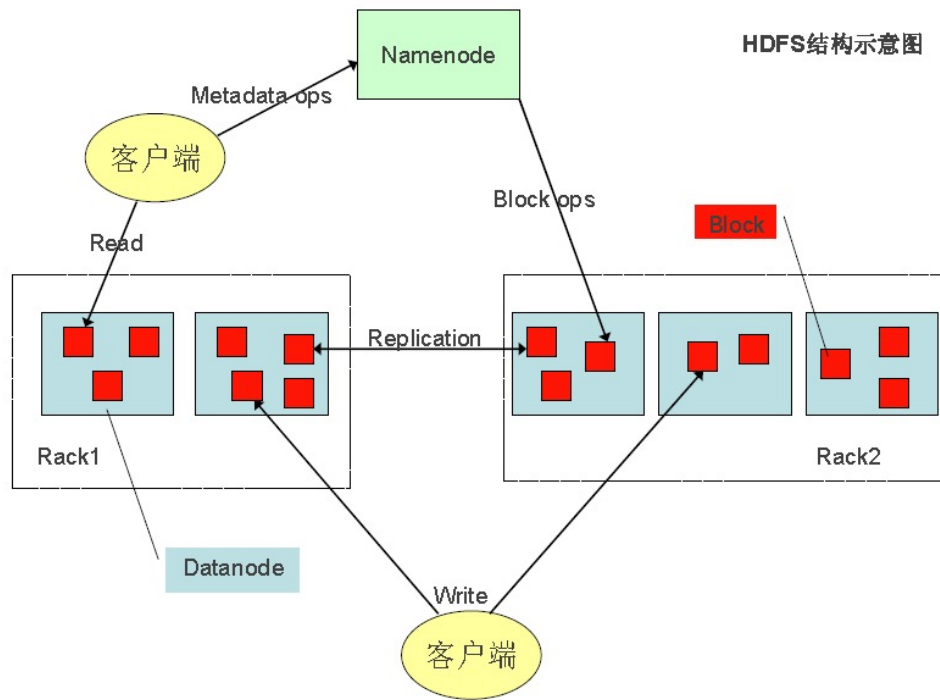


Hadoop Framework



HDFS体系结构

- NameNode → Master
- DataNode → Chunkserver



HDFS关键运行机制

--保障可靠性的措施

- 一个名字节点和多个数据节点
- 数据复制（冗余机制）

--存放的位置（机架感知策略）

- 故障检测

--数据节点

心跳包（检测是否宕机）

块报告（安全模式下检测）

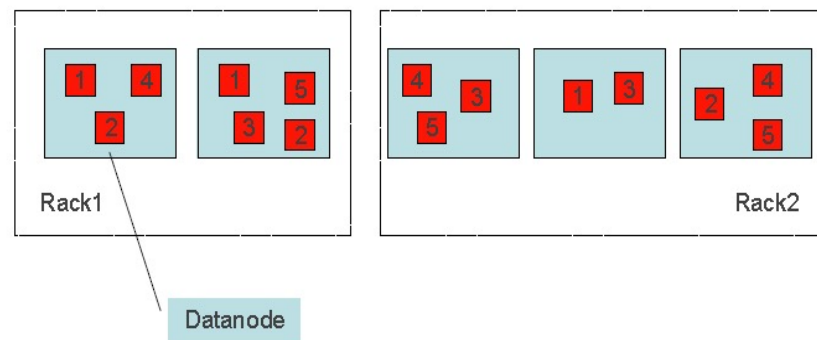
数据完整性检测（校验和比较）

--名字节点（日志文件，镜像文件）

- 空间回收机制

Block Replication

```
Namenode(Filename, nameReplicas, block-ids, ...)  
/user/grid/data/part-0, r:3, {1,2}, ...  
/user/grid/data/part-1, r:3, {3,4,5}, ...
```



HDFS关键运行机制

--写文件流程

- 客户端缓存
- 流水线复制
- 并发写控制
- 流程：
 1. 客户端把数据缓存到本地临时文件夹
 2. 临时文件夹数据超过64M，客户端联系NameNode，NameNode分配DataNode，DataNode依照客户端的位置被排列成一个有着最近物理距离和最小的序列
 3. 与序列的第一个数据服务器建立Socket连接，发送请求头，然后等待回应，依次下传，客户端得到回包，流水线建立成功，
 4. 正式发送数据



HDFS关键运行机制

--读文件流程

- 客户端联系NameNode, 得到所有数据块信息, 以及数据块对应的所有数据服务器的位置信息
- 尝试从某个数据块对应的一组数据服务器中选出一个, 进行连接 (选取算法未加入相对位置的考虑)
- 数据被一个包一个包发送回客户端, 等到整个数据块的数据都被读取完了, 就会断开此链接, 尝试连接下一个数据块对应的数据服务器, 整个流程, 依次如此反复, 直到所有想读的都读取完了为止



Hadoop VS. Google

- 技术架构的比较
 - 数据结构化组件: Hbase → BigTable
 - 并行计算模型: MapReduce → MapReduce
 - 分布式文件系统: HDFS → GFS
 - 锁管理: ZooKeeper → Chubby

Hadoop 云计算应用

ZooKeeper

HBase

MapReduce

HDFS

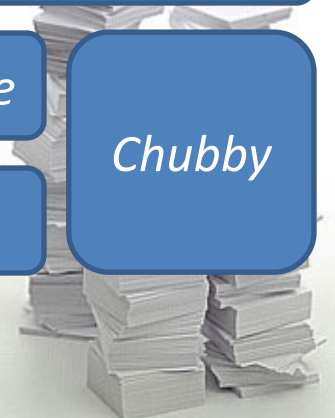
Google 云计算应用

BigTable

MapReduce

GFS

Chubby



Hadoop VS. Google

- HDFS与GFS比较

- 子服务器管理模式差异

- **GFS**: Chunk Server在Chubby中获取独占锁表示其生存状态，Master通过轮询这些独占锁获知Chunk Server的生存状态
 - **HDFS**: DataNode通过心跳的方式告知NameNode其生存状态
 - **GFS**中，Master损坏时，替补服务器可以快速获知Chunk Server的状态
 - **HDFS**中，NameNode损坏后，NameNode恢复时需要花费一段时间获知DataNode的状态



Hadoop VS. Google

- HDFS与GFS比较
 - HDFS具备空间回收机制
 - 文件删除时，仅删除目录结构
 - 实际数据的删除在等待一段时间后实施
 - 优点：便于恢复文件



Hadoop Related Subprojects

- Pig
 - High-level language for data analysis
- HBase
 - Table storage for semi-structured data
- Hive
 - SQL-like Query language and Metastore
- Mahout
 - Machine learning
-

