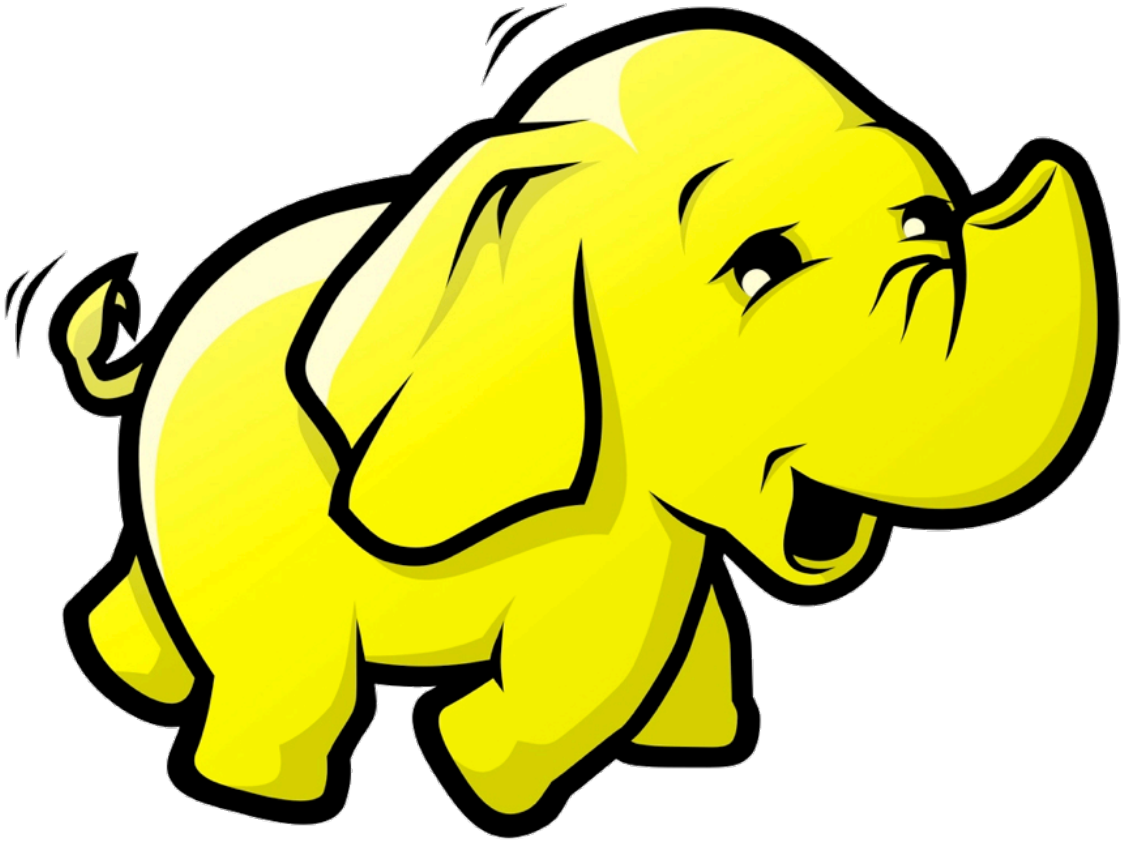


Hadoop

York Keyser

10 May 2012

Map Reduce

- Paradigm by Google
- Framework to hide the distributed work from the developer
- Most common languages are C++, Java, or Python
- Originally designed for x86 Architecture (desktop/pc)

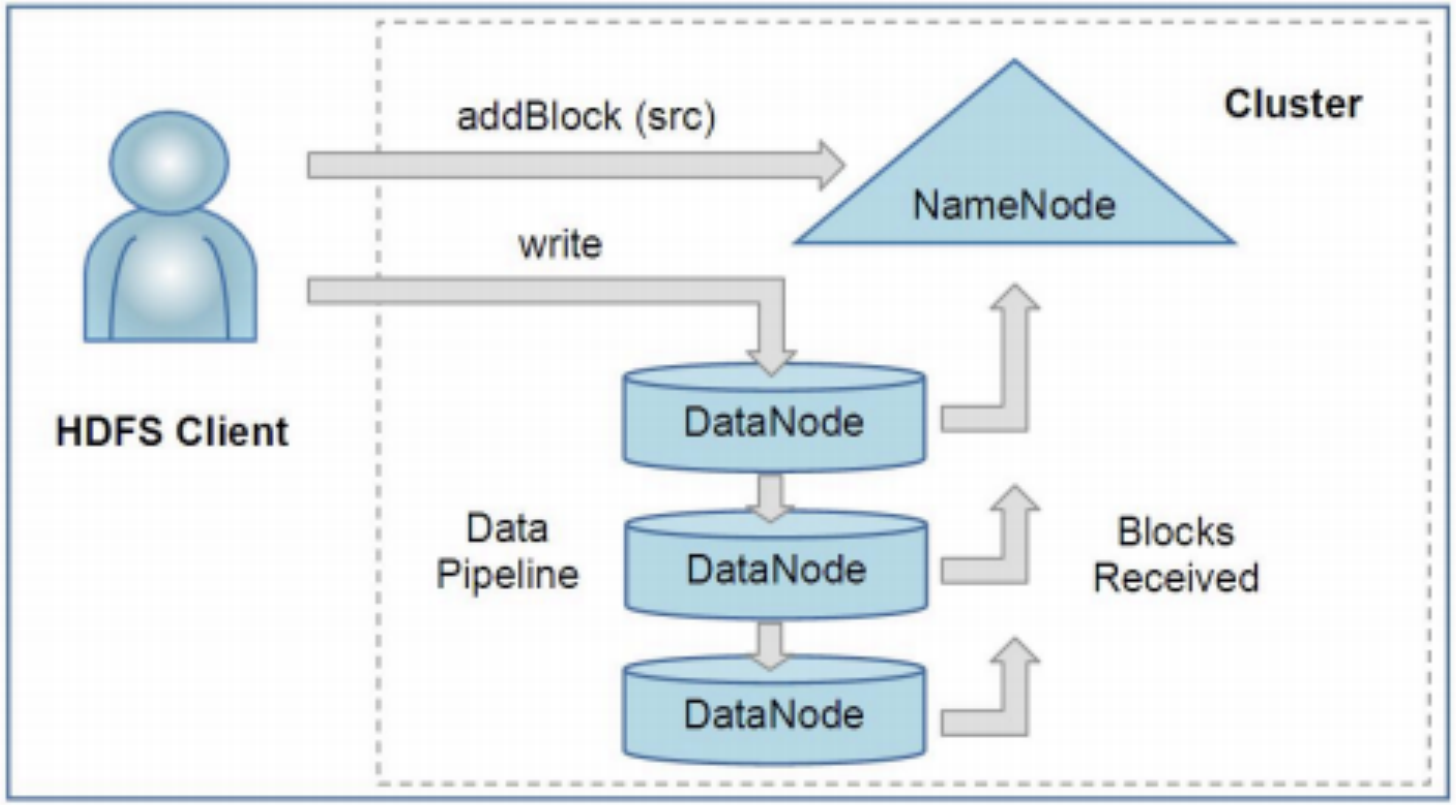
Hadoop

- Implementation of Map Reduce paradigm by Apache Software Foundation
- Language is Java
- Top Level Project since 2008
- Hadoop Distributed File System (HDFS)

HDFS

- NameNode (1 per cluster)
  - Metadata
    - permission, modification, namespace, ...
- DataNode (n per cluster)
  - Data block default 128MB
- CheckpointNode BackupNode
- Client (m per cluster)

HDFS


addBlock (src)
Cluster
NameNode
write
HDFS Client
DataNode
Data
Pipeline
DataNode
Blocks
Received
DataNode

Hadoop components

HDFS	Distributed file system Subject of this paper!
MapReduce	Distributed computation framework
HBase	Column-oriented table service
Pig	Dataflow language and parallel execution framework
Hive	Data warehouse infrastructure
ZooKeeper	Distributed coordination service
Chukwa	System for collecting management data
Avro	Data serialization system

Who use hadoop

- EBay
  - 532 nodes cluster (8 * 532 cores, 5.3PB).
  - Heavy usage of Java MapReduce, Pig, Hive, HBase.
- Facebook
  - Currently we have 2 major clusters:
    - A 1100-machine cluster with 8800 cores and about 12 PB raw storage.
    - A 300-machine cluster with 2400 cores and about 3 PB raw storage.
  - Usage of Hadoop HDFS and Hive

side note: 1 Petabyte (PB) = 10^{15} Byte

Install Hadoop Map Reduce

- Debian 2.6.36
- Java (1.6.0_30)
- Hadoop 1.0.1

Amazon

- Elastic Compute Cloud (EC2)
- Simple Storage Service (S3)
- Elastic MapReduce (EMR)

Elastic Compute Cloud

- Different plans based on
  - power
  - time
- Regions
- Scaleable
- Balancing

EC2 Pricing

Region: EU (Irland)

	Linux/UNIX-Nutzung	Windows-Nutzung
Standard On-Demand Instances		
Small (Standard)	$0,090 pro Stunde	$0,115 pro Stunde
Medium	$0,180 pro Stunde	$0,230 pro Stunde
Large	$0,360 pro Stunde	$0,460 pro Stunde
Extra Large	$0,720 pro Stunde	$0,920 pro Stunde
Micro On-Demand Instances		
Micro	$0,025 pro Stunde	$0,035 pro Stunde
Hi-Memory On-Demand Instances		
Extra Large	$0,506 pro Stunde	$0,570 pro Stunde
Double Extra Large	$1,012 pro Stunde	$1,140 pro Stunde
Quadruple Extra Large	$2,024 pro Stunde	$2,280 pro Stunde
Hi-CPU On-Demand Instances		
Medium	$0,186 pro Stunde	$0,285 pro Stunde
Extra Large	$0,744 pro Stunde	$1,140 pro Stunde
Cluster Compute Instances		
Quadruple Extra Large	N/A*	N/A*
Cluster GPU Instances		
Quadruple Extra Large	N/A*	N/A*

* Cluster Compute und Cluster GPU Instances sind gegenwärtig nur in der Region USA Ost (Virginia) verfügbar.

Instance Type	RAM (GB)	Compute Units	Disk Drive (GB)	Platform (bits)	I/O Performance	Name
Small (default)	1.7	1	160	32	Moderate	m1.small
Large	7.5	4	850	64	High	m1.large
Extra Large	15	8	1690	64	High	m1.xlarge
High-CPU Medium	1.7	5	350	32	Moderate	c1.medium
High-CPU Extra Large	7	20	1690	64	High	c1.xlarge
High-Memory Extra Large	17.1	6.5	420	64	Moderate	m2.xlarge
High-Memory Double Extra Large	34.2	13	850	64	Moderate	m2.2xlarge
High-Memory Quadruple Extra Large	68.4	26	1690	64	High	m2.4xlarge
Cluster Compute Quadruple Extra Large Instance*	23	33.5	1690	64	Very High (10 Gigabit Ethernet)	cc1.4xlarge
Cluster Compute Eight Extra Large*	60.5	88	3370	64	Very High (10 Gigabit Ethernet)	cc2.8xlarge
Cluster GPU Instance*	23**	33.5	1690	64	Very High (10 Gigabit Ethernet)	cg1.4xlarge

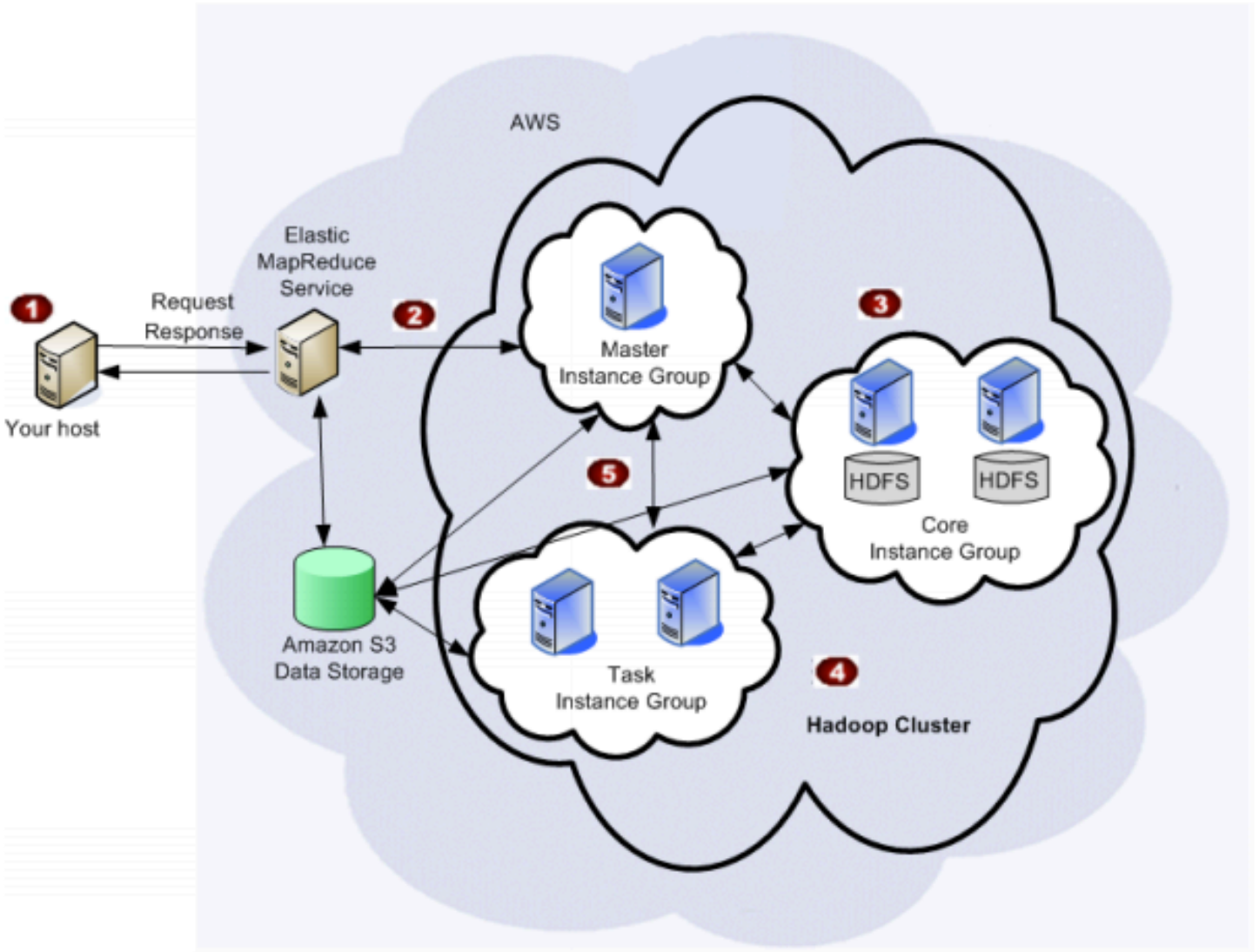
Simple Storage Service

- Storage Service of Amazon
- Possibility to encrypt your data
- Possibility to share data through different accounts
- File limit 1 byte up to 5 Terabyte
- No specified upload limit

Elastic MapReduce

- Hadoop (Apache)
- HDFS (Apache)
- max 19 nodes

Elastic MapReduce



Test environment

- Text file filled with Lorem Ipsum
- Word count **384.426.368**
- File size **2.2GB**
- Assignment **count all words which start with the letter ‘e’**

Test result

- Single node **ca. 12min**
- Pseudo distributed node **ca.18min**
- Amazon 19 nodes **ca. 7min**
- Amazon single node **ca. 8min**

Benchmark by yahoo

- approximately **3800 nodes** (in such a large cluster, some nodes are always down)
- 2 quad core Xeons @ 2.5ghz per node
- 4 SATA disks per node
- 8G RAM per node (upgraded to 16GB before the petabyte sort)
- 1 gigabit ethernet on each node
- 40 nodes per rack
- 8 gigabit ethernet uplinks from each rack to the core
- Red Hat Enterprise Linux Server Release 5.1 (kernel 2.6.18)
- Sun Java JDK (1.6.0 05-b13 and 1.6.0 13-b03) (32 and 64 bit)

Results

- 62 sec to sort 1Terabyte
- 16.25 h to sort 1 Petabyte

Bytes	Nodes	Maps	Reduces	Replication	Time
500,000,000,000	1406	8000	2600	1	59 seconds
1,000,000,000,000	1460	8000	2700	1	62 seconds
100,000,000,000,000	3452	190,000	10,000	2	173 minutes
1,000,000,000,000,000	3658	80,000	20,000	2	975 minutes

Thank you

Papers

- MapReduce: Simplified Data Processing on Large Clusters
- Hadoop at Home: Large-Scale Computing at a Small College
- MapReduce: Simplified Data Processing on Large Clusters
- Towards Quantitative Analysis of Data Intensive Computing: A Case Study of Hadoop
- Apache Hadoop Goes Realtime at Facebook
- The Hadoop Distributed File System