



Grid Dynamics

Scalable eCommerce Platform Solutions



by Oleksiy Gayduk

Data Everywhere



The project Hadoop

Apache Hadoop is an open source Java software framework for running data-intensive applications on large clusters of commodity hardware.

Hadoop Components

Storage

HDFS

Self-healing high-
bandwidth clustered
storage

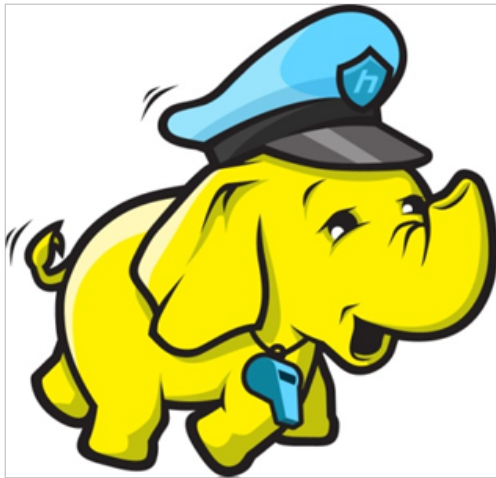
Processing

MapReduce

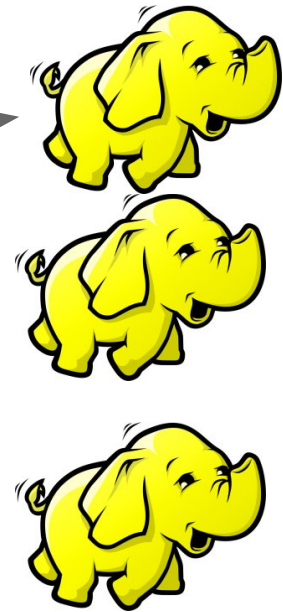
Fault-tolerant
distributed processing
HDFS MapReduce

2 Kinds of Nodes

Master Nodes



Slave Nodes



Master Nodes

Name Node

- Only 1 per cluster
- Meta server
- Secondary NameNode for fault tolerance.

JobTracker

- Only 1 per cluster
- Job scheduler

Slave Nodes

DataNodes

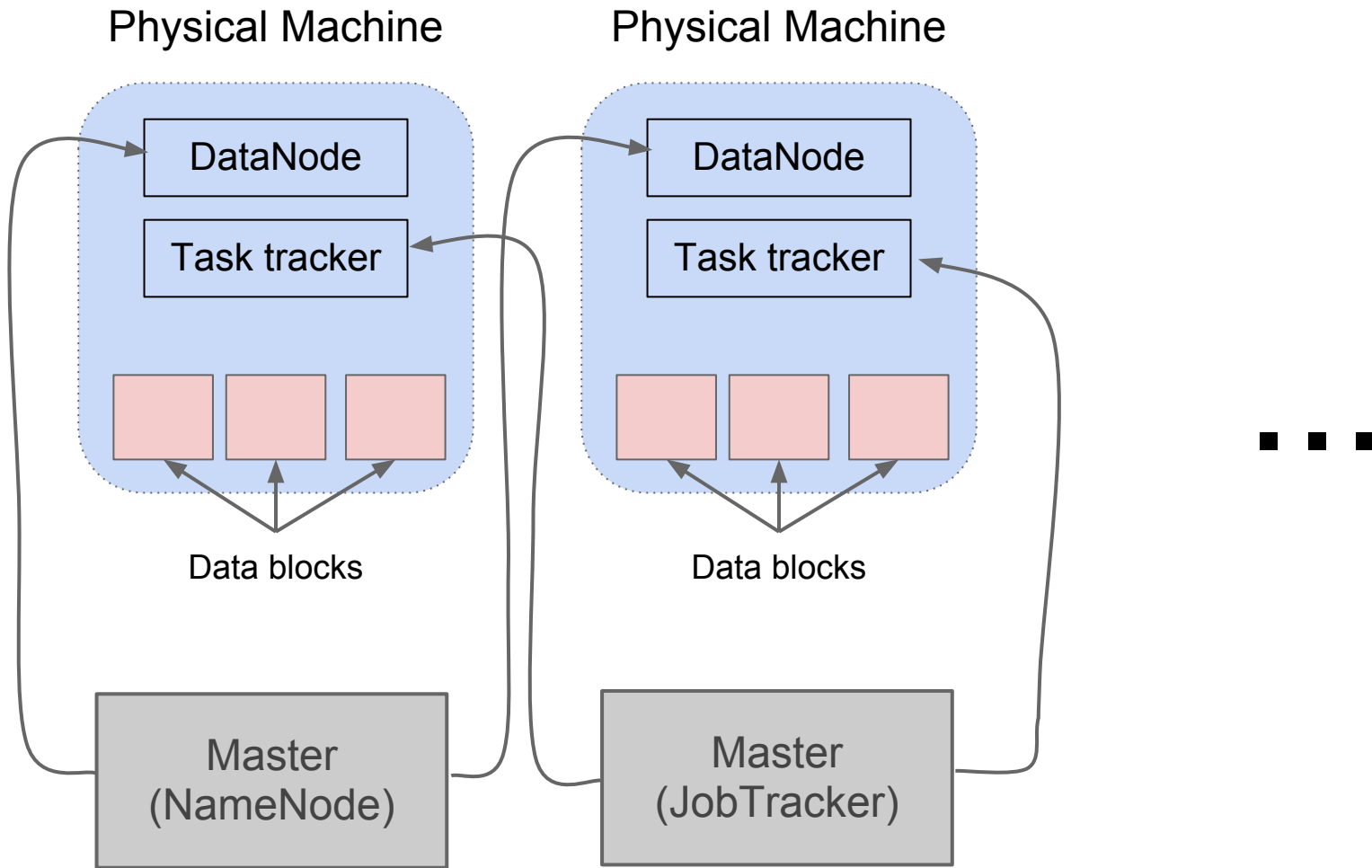
- 1-4000 per cluster

Block data storage

TaskTrackers

- 1-4000 per cluster
- Task execution

High Level Architecture



HDFS Basics

- HDFS is file system written in Java
- Sits on top of native file system
- Provides redundant storage for massive amounts of data
- Use cheap, unreliable computers

HDFS Data

- Data is split into blocks and stored on multiple nodes in the cluster.
Each block is usually 64 MB or 128 MB
- Each block is replicated multiple times.
Replicas stored on different data nodes
- Large files, 100 MB+

Name Node

- A single NameNode stores all metadata
- Filenames, locations on DataNodes of each block, owner, group, etc.
- All information maintained in RAM for fast lookup
- Filesystem metadata size is limited to the amount of available RAM on the NameNode

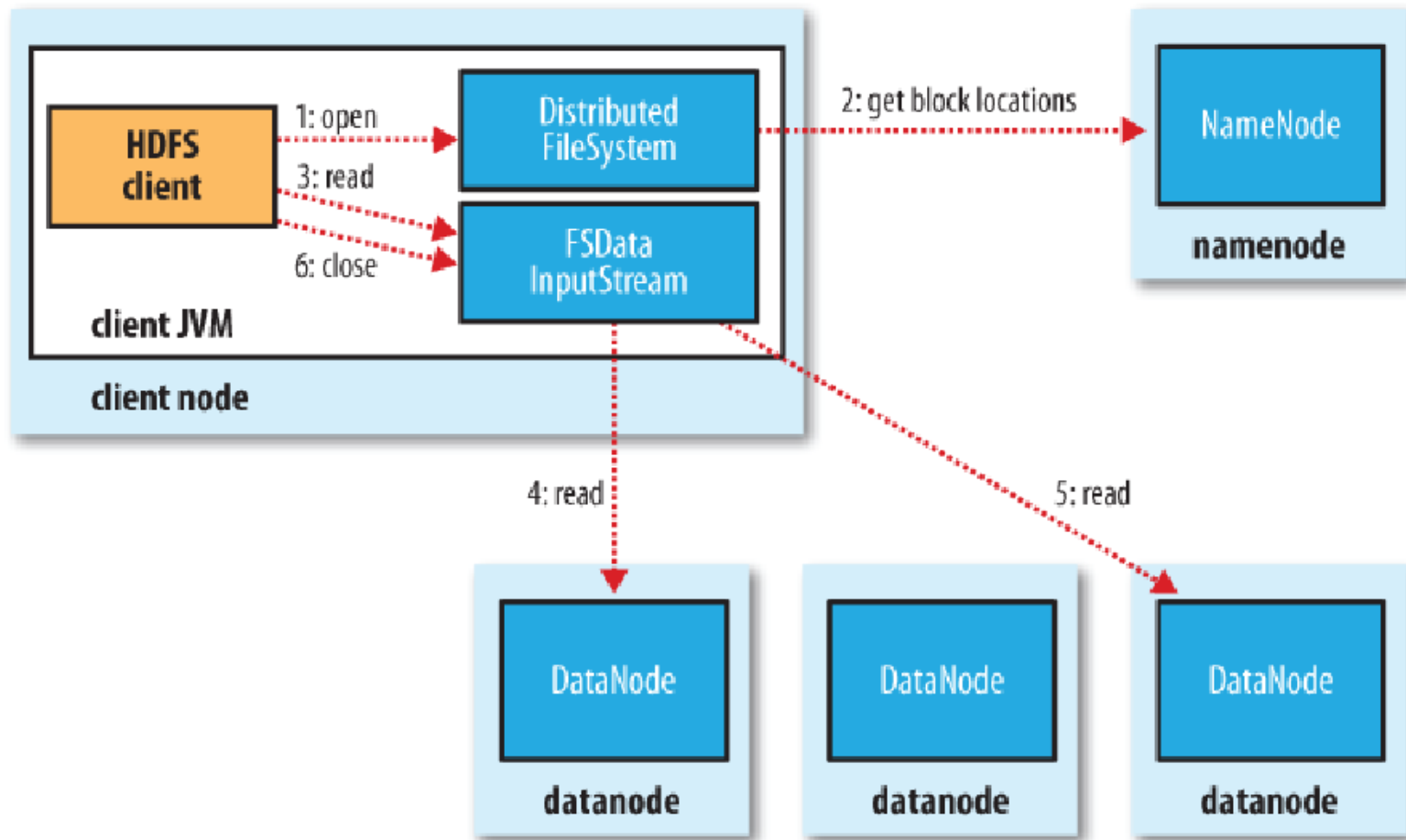
Data Node

- DataNodes store file contents
- Stored as data 'blocks' on the underlying filesystem
- Different blocks of the same file will be stored on different DataNodes
- Same block is stored on three (or more) DataNodes for redundancy

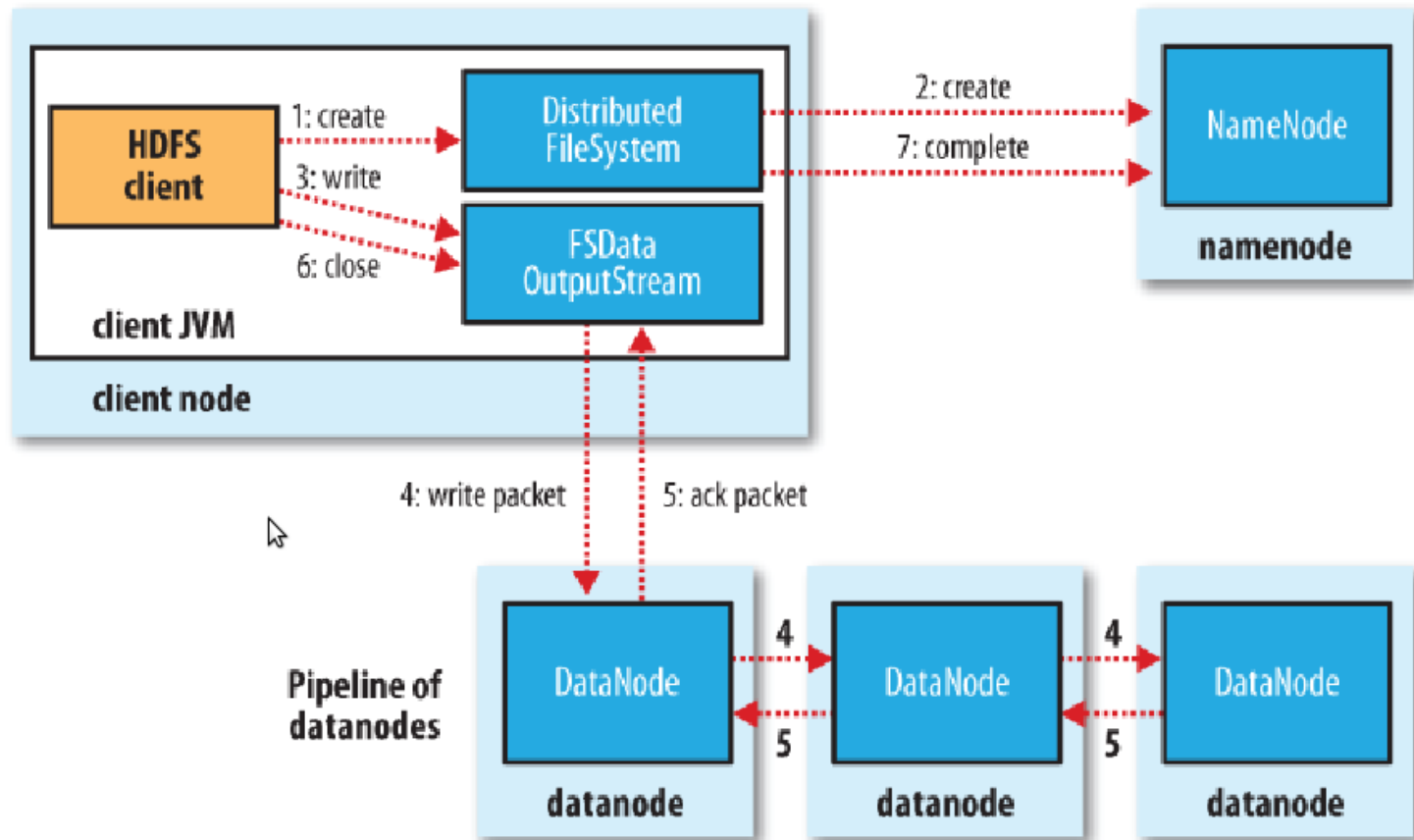
Self-healing

- DataNodes send heartbeats to the NameNode
- After a period without any heartbeats, a DataNode is assumed to be lost
- NameNode determines which blocks were on the lost node
- NameNode finds other DataNodes with copies of these blocks
- These DataNodes are instructed to copy the blocks to other nodes

Reading from HDFS



Writing to HDFS



What is MapReduce?

- MapReduce is a method for distributing a task across multiple nodes
- Automatic parallelization and distribution
- Each node processes data stored on that node

Features of MapReduce

- Fault-tolerance
- Status and monitoring tools
- A clean abstraction for programmers

JobTracker

- MapReduce jobs are controlled by a software daemon known as the JobTracker
- The JobTracker resides on a master node
- Assigns Map and Reduce tasks to other nodes on the cluster
- These nodes each run a software daemon known as the TaskTracker
- The TaskTracker is responsible for actually instantiating the Map or Reduce task, and reporting progress back to the JobTracker

Two Parts

Developer specifies two functions:

map()

reduce()

- The framework does the rest

map()

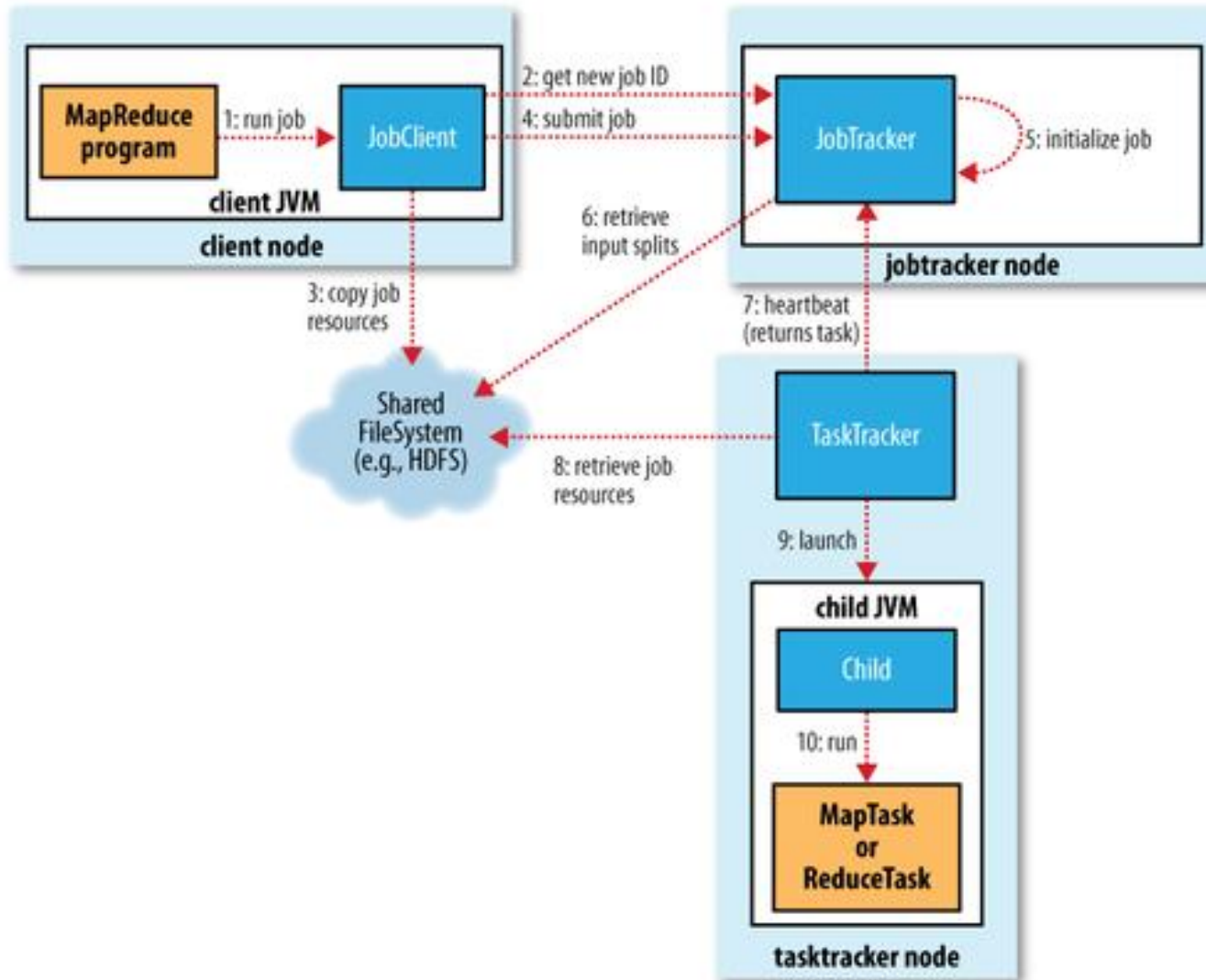
- The Mapper reads data in the form of key/value pairs
- It outputs zero or more key/value pairs

```
map(key_in, value_in) ->  
    (key_out, value_out)
```

reduce()

- After the Map phase all the intermediate values for a given intermediate key are combined together into a list
- This list is given to one or more Reducers
- The Reducer outputs zero or more final key/value pairs
- These are written to HDFS or another data storage

Architecture



Examples

- How we have built our infrastructure using Hadoop?
- Example of the application for building "Inverted Index"

References

- Hadoop: The Definitive Guide. <http://www.amazon.com/Hadoop-Definitive-Guide-Tom-White/dp/1449389732>

- Hadoop in Practice.

<http://www.manning.com/holmes/>

Hadoop in Action.

<http://manning.com/lam/>