Hadoop Distributed File System

Sinziana Gafitanu

What is HDFS?

HDFS is a Java-based file system that provides scalable and reliable data storage, and it was designed to span large clusters of commodity servers.

One of the main components of Hadoop.

Source: https://hortonworks.com/apache/hdfs/#section_1

Goals/Assumptions of HDFS

- Fault Tolerance (Hardware failure is the norm not the exception and as such detection of faults and quick automatic recovery is a core goal).
- Streaming Data Access (HDFS is designed for batch processing and because the requirements are different than a multi purpose FS it does break POSIX requirements).
- Large Data Sets (tens of millions of files in a single instance)
- Write Once/Read Many (Files should not change)
- Move computation to where the data is
- Portability (on different hardware and software platforms)

Source: https://hadoop.apache.org/docs/r1.2.1/hdfs_design.html

Main Components

NameNode

- Master server
- Manages the file system namespace
- Regulates Access to the files
- Repository for all HDFS metadata

DataNode

- Manages the storage attached to the node

Source: https://hadoop.apache.org/docs/r1.2.1/hdfs_design.html

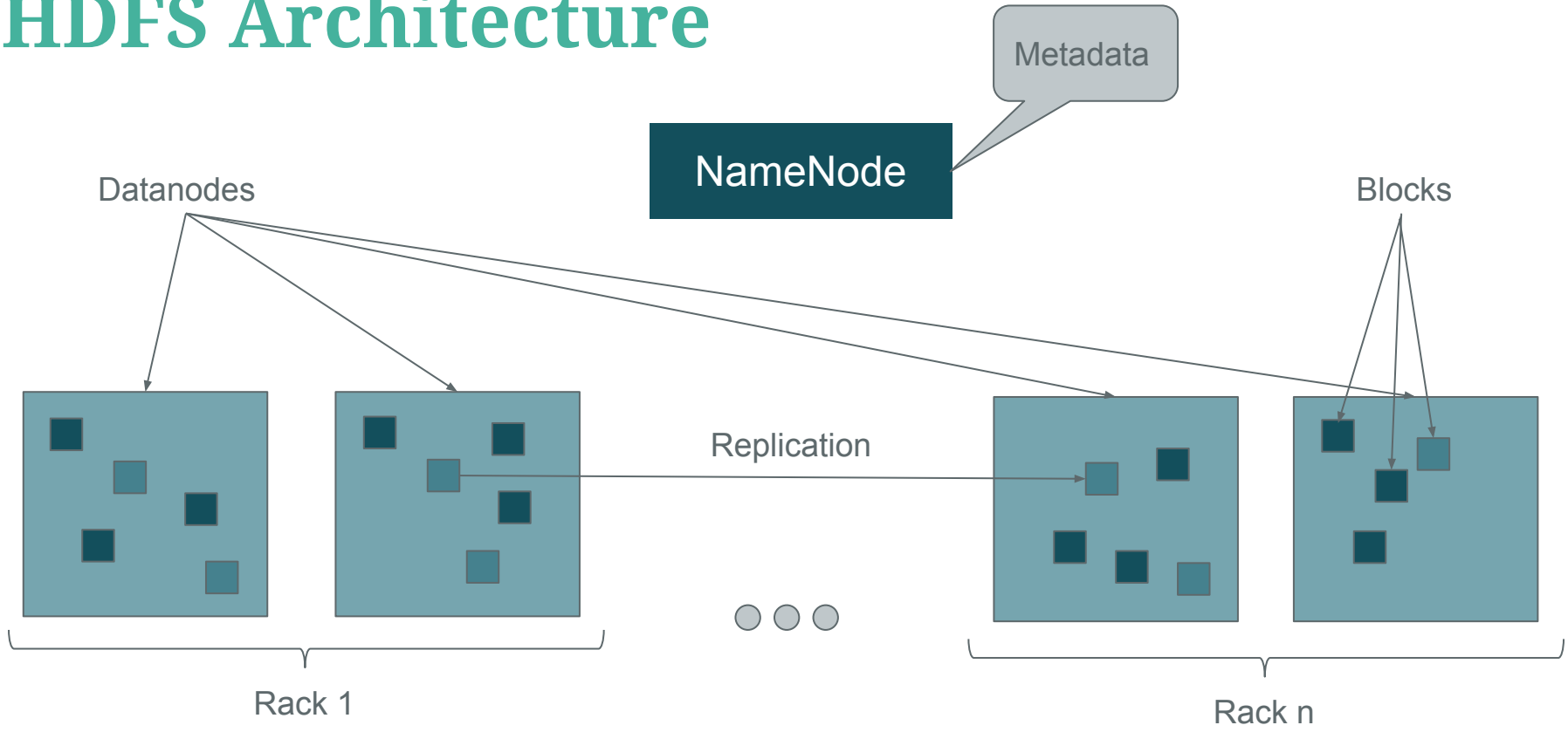
Main Components (II)

Blocks

- Each file is stored as a sequence of blocks
- All blocks (except the last one) are the same size (configurable)
- Replication is configurable and can be assigned at creation and changed later
- NameNode controls replication
- All blocks have associated checksums stored in the same namespace as the blocks.

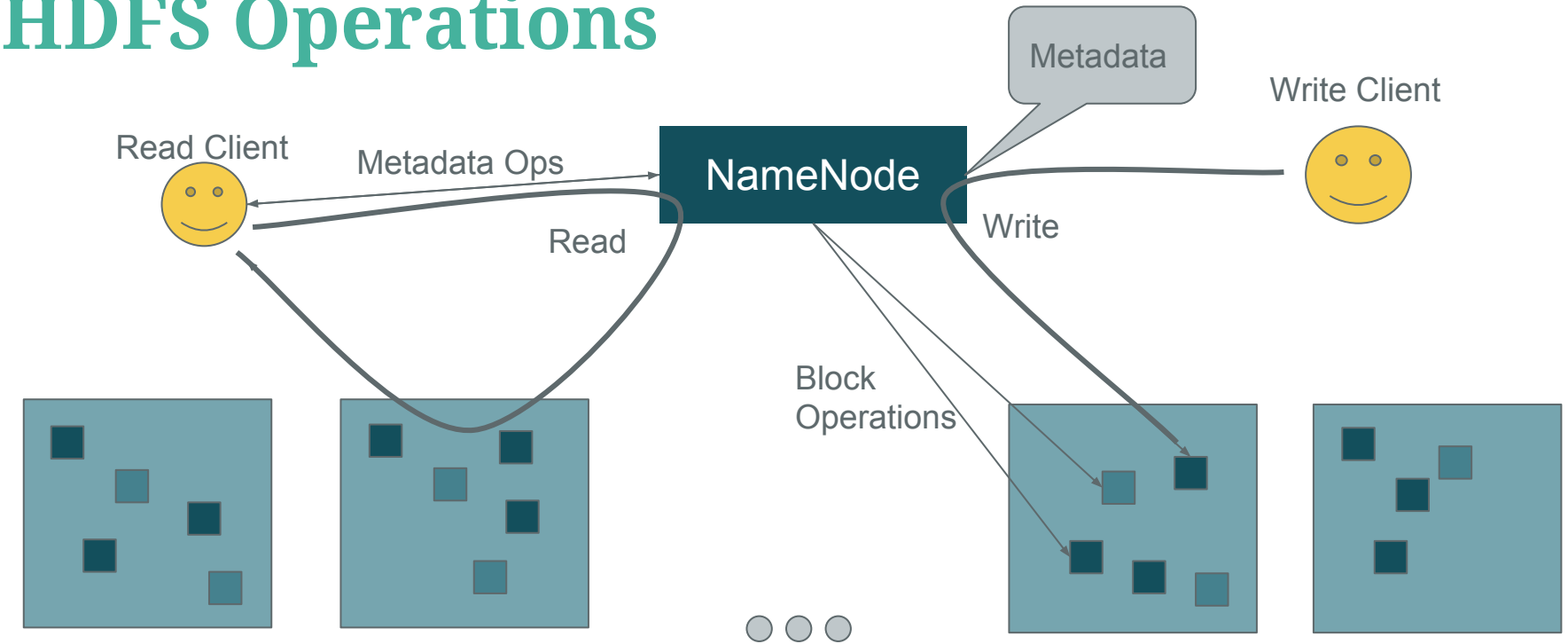
Source: https://hadoop.apache.org/docs/r1.2.1/hdfs_design.html

HDFS Architecture



Source: https://hadoop.apache.org/docs/r1.2.1/hdfs_design.html

HDFS Operations



Source: https://hadoop.apache.org/docs/r1.2.1/hdfs_design.html

Communication Protocols

All based on TCP/IP

Client Protocol

- Communication with the NameNode machine through configurable port
- Abstracted using a Remote Procedure Call layer

DataNode Protocol

- Communication from the DataNodes to the NameNode.

Specialized Messages

There are 2 specialized messages that the DataNode sends to the NameNode.

HeartBeat

- If the Heartbeat is absent the NameNode will mark the DataNode as dead and it will not forward any new IO messages to it.
- The death of a data node will make the NameNode trigger a replication for the files that no longer comply to their replication factor.

BlockReport

- Contains a list of data blocks contained in a DataNode

File System Namespace

- Similar to most other filesystems organized in directories hierarchical
- Create, remove, move, rename
- No support for hard or soft links
- Any change is managed by the Name Node.

Source: https://hadoop.apache.org/docs/r1.2.1/hdfs_design.html

Accessing HDFS

- Java API
- Wrapper over the Java API for C
- Http Browser for accessing an instance
- FS Shell
- DFSAdmin (HDFS administration)

FS Shell

Command line interface that works with commands that are very similar to other shells. Provides a simple way of interating with the cluster.

Example:

```
$> bin/hadoop dfs -mkdir /foodir

$> bin/hadoop dfs -rmr /foodir

$> bin/hadoop dfs -cat /foodir/myfile.txt
```

HDFS v2

- More than one NameNode (using HDFS federation)
- Horizontal scaling
- Performance improvement
- Multiple Namespaces

Demo