HDFS

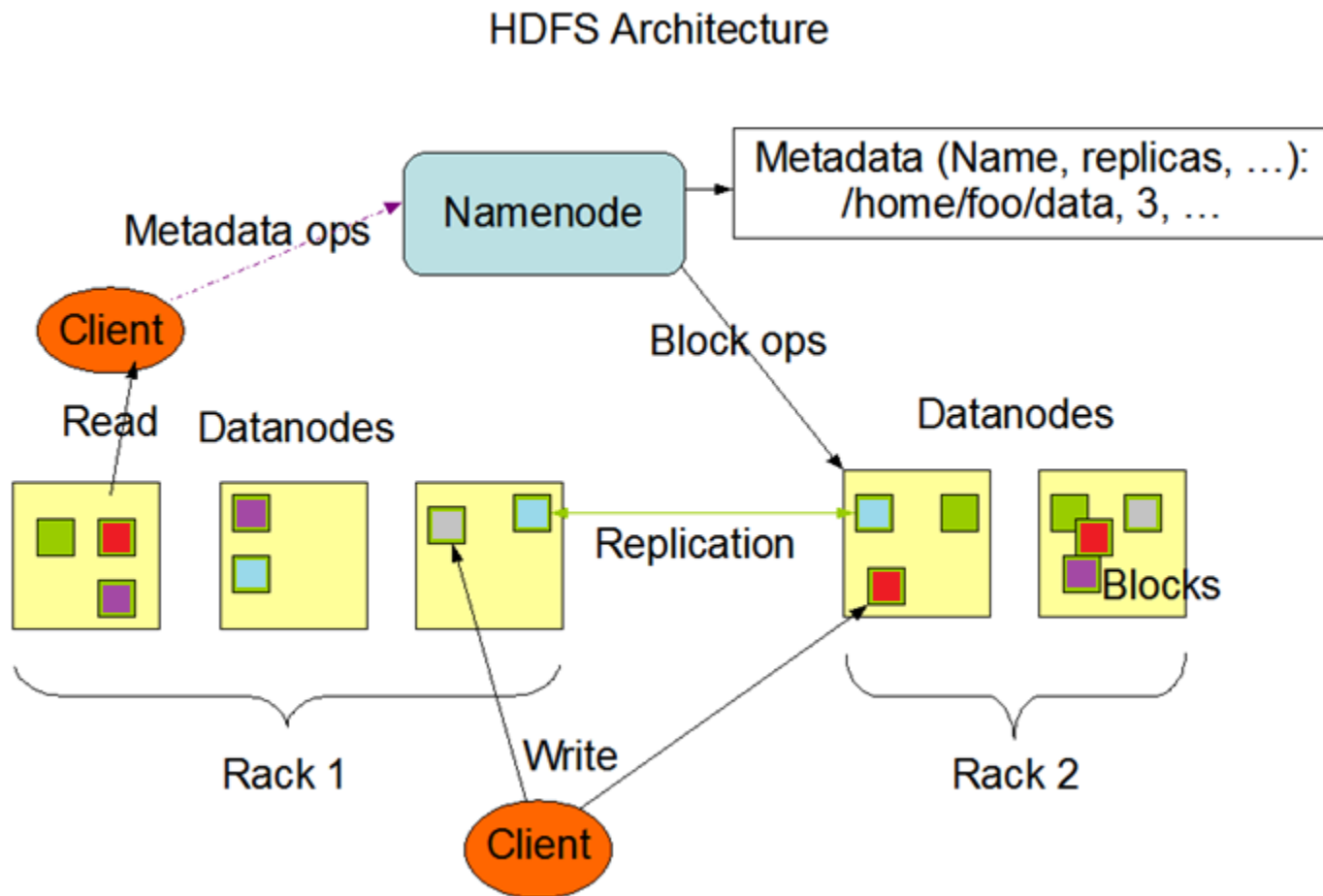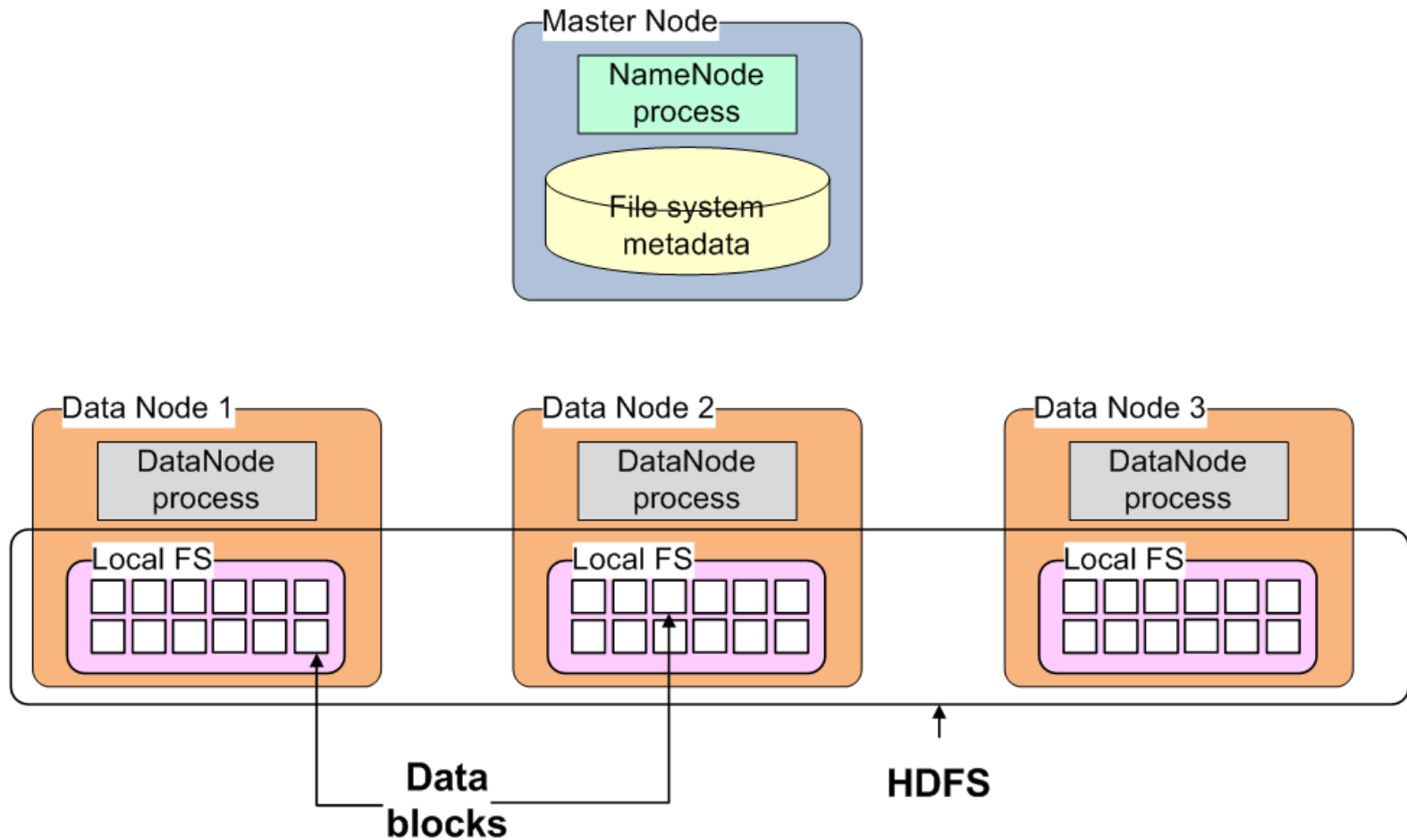## Hadoop Distributed File System

История

- 2002 – запуск проекта Nutch
- 2003 – публикация с описанием GFS
- 2004 – создание NDFS (Nutch Distributed File System)
- 2004 – публикация Google и MapReduce
- 2005 – реализация MR в Nutch
- 2006 – выделение подпроекта Hadoop
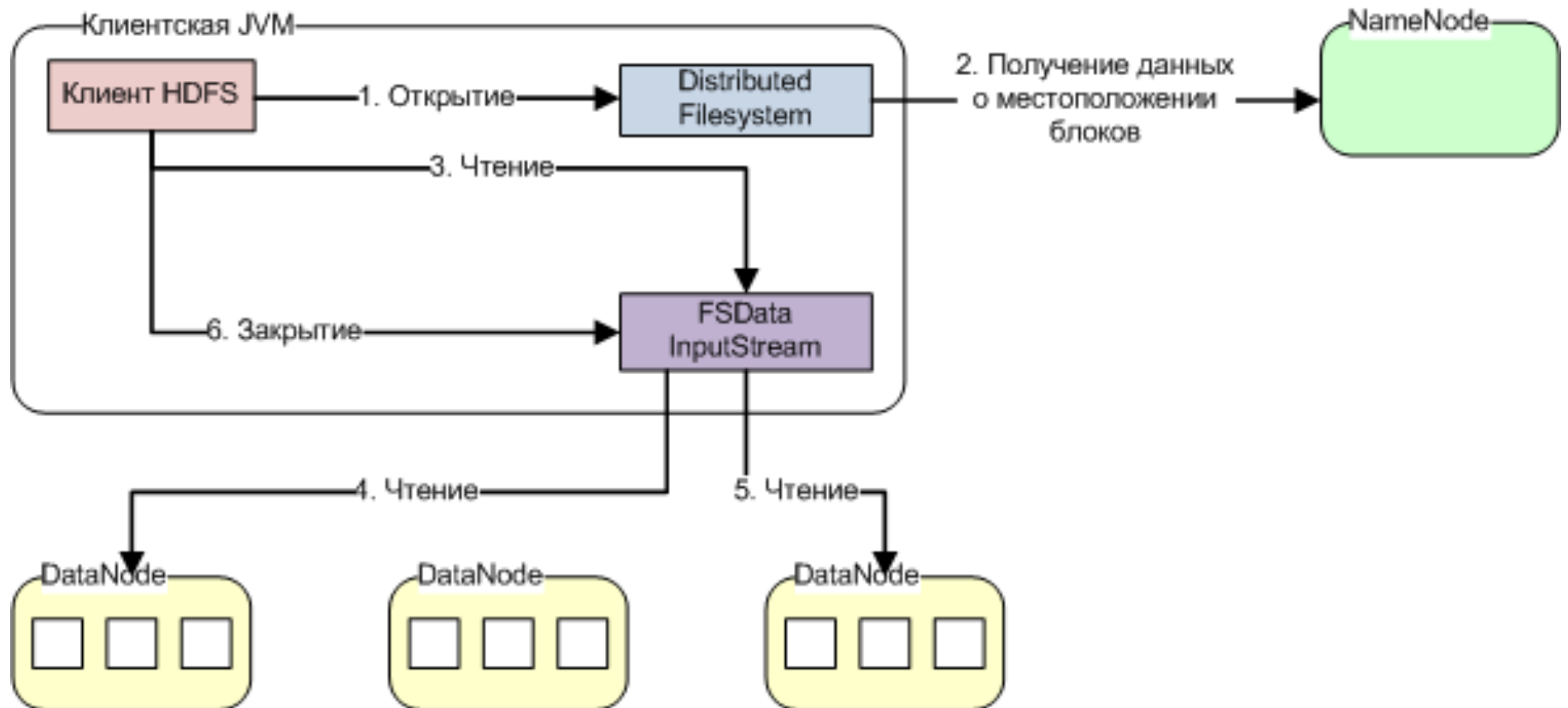- 2008 – выход Hadoop в лидеры ASF (Apache Software Foundation)

Архитектура HDFS
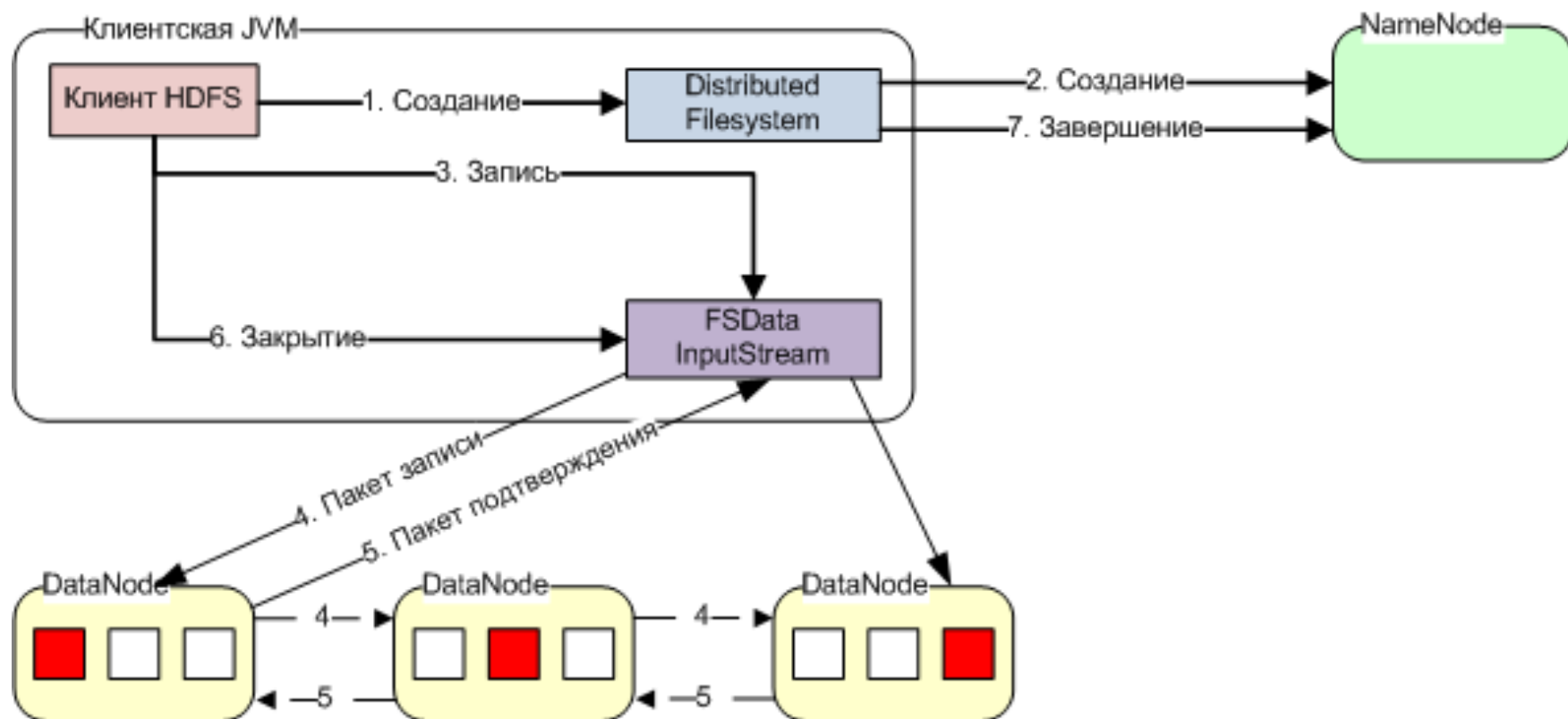

HDFS Architecture

Name Node

Чтение HDFS
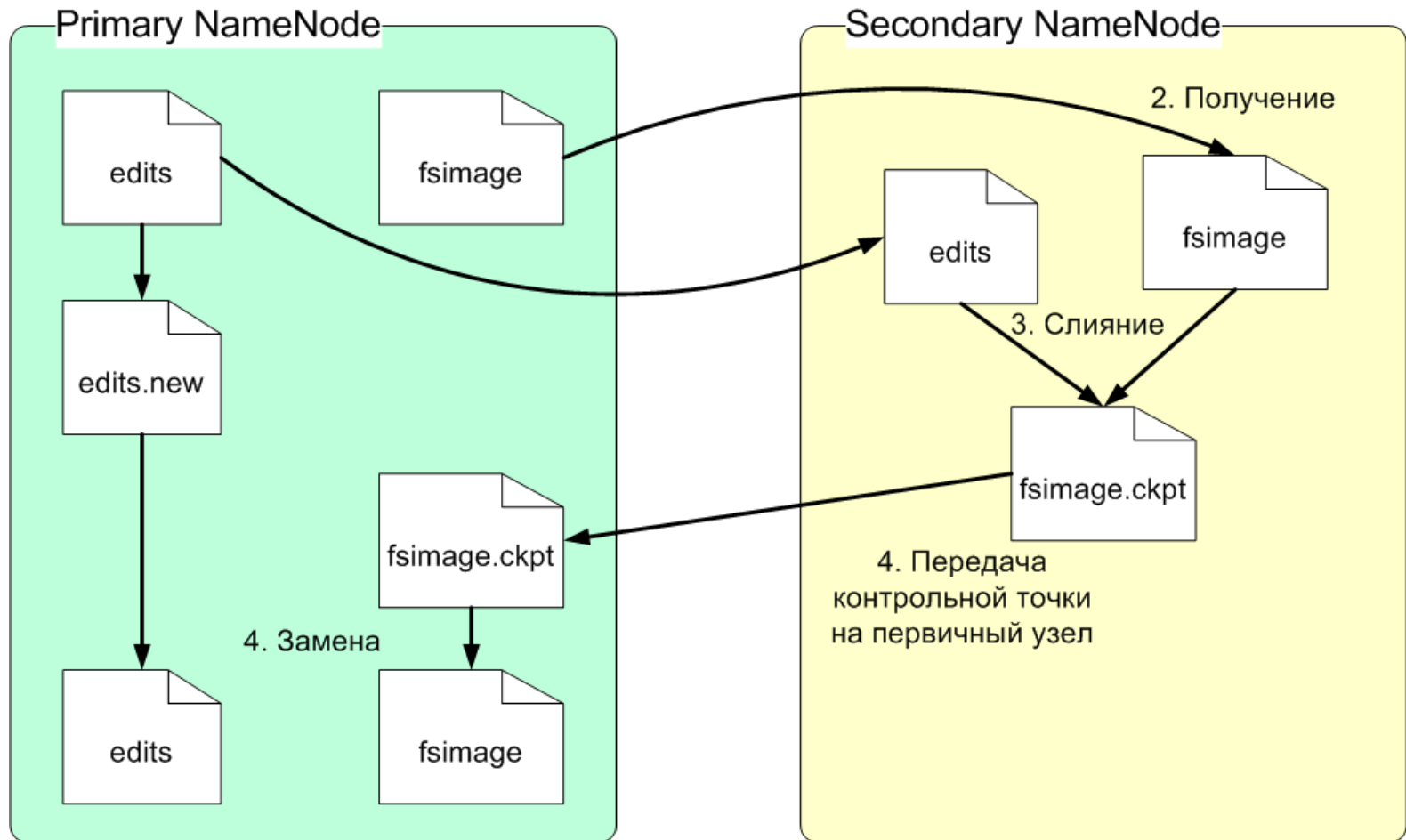
Запись HDFS

Name Node структура хранения

- ${dfs.name.dir}/
  - VERSION   (информация о версии HDFS)
  - edits   (журнал изменений)
  - fsimage   (контрольная точка метаданных)
  - fstime   (время создания контрольной точки)

Secondary Name Node

Сведения о блоках файла

[root@hdp-15 ~]# **hdfs fsck /user/hive/warehouse/big_cdr_parquet/000000_0 -files -blocks -locations**
Connecting to namenode via http://hdp-7:50070
FSCK started by root (auth:SIMPLE) from /192.168.91.141 for path /user/hive/warehouse/big_cdr_parquet/000000_0 at Mon May 18 14:00:22 MSK 2015
/user/hive/warehouse/big_cdr_parquet/000000_0 1133129924 bytes, 9 block(s): OK
0. BP-1972162810-192.168.91.133-1428693610895:blk_1073747244_6432 len=134217728 repl=3
[192.168.91.141:50010, 192.168.91.139:50010, 192.168.91.133:50010]
1. BP-1972162810-192.168.91.133-1428693610895:blk_1073747245_6433 len=134217728 repl=3
[192.168.91.136:50010, 192.168.91.139:50010, 192.168.91.141:50010]
2. BP-1972162810-192.168.91.133-1428693610895:blk_1073747246_6434 len=134217728 repl=3
[192.168.91.142:50010, 192.168.91.136:50010, 192.168.91.141:50010]
3. BP-1972162810-192.168.91.133-1428693610895:blk_1073747247_6435 len=134217728 repl=3
[192.168.91.134:50010, 192.168.91.142:50010, 192.168.91.137:50010]
4. BP-1972162810-192.168.91.133-1428693610895:blk_1073747248_6436 len=134217728 repl=3
[192.168.91.135:50010, 192.168.91.133:50010, 192.168.91.137:50010]
5. BP-1972162810-192.168.91.133-1428693610895:blk_1073747249_6437 len=134217728 repl=3
[192.168.91.140:50010, 192.168.91.137:50010, 192.168.91.142:50010]
6. BP-1972162810-192.168.91.133-1428693610895:blk_1073747250_6438 len=134217728 repl=3
[192.168.91.142:50010, 192.168.91.139:50010, 192.168.91.141:50010]
7. BP-1972162810-192.168.91.133-1428693610895:blk_1073747251_6439 len=134217728 repl=3
[192.168.91.139:50010, 192.168.91.140:50010, 192.168.91.135:50010]
8. BP-1972162810-192.168.91.133-1428693610895:blk_1073747252_6440 len=59388100 repl=3 [192.168.91.141:50010, 192.168.91.137:50010, 192.168.91.135:50010]

Сведения о блоках файла

```
Status: HEALTHY
 Total size:    1133129924 B
 Total dirs:    0
 Total files:   1
 Total symlinks:            0
 Total blocks (validated):     9 (avg. block size 125903324 B)
 Minimally replicated blocks:   9 (100.0 %)
 Over-replicated blocks:       0 (0.0 %)
 Under-replicated blocks:      0 (0.0 %)
 Mis-replicated blocks:       0 (0.0 %)
 Default replication factor:    3
 Average block replication:    3.0
 Corrupt blocks:              0
 Missing replicas:           0 (0.0 %)
 Number of data-nodes:         10
 Number of racks:             1
FSCK ended at Mon May 18 14:00:22 MSK 2015 in 1 milliseconds
```
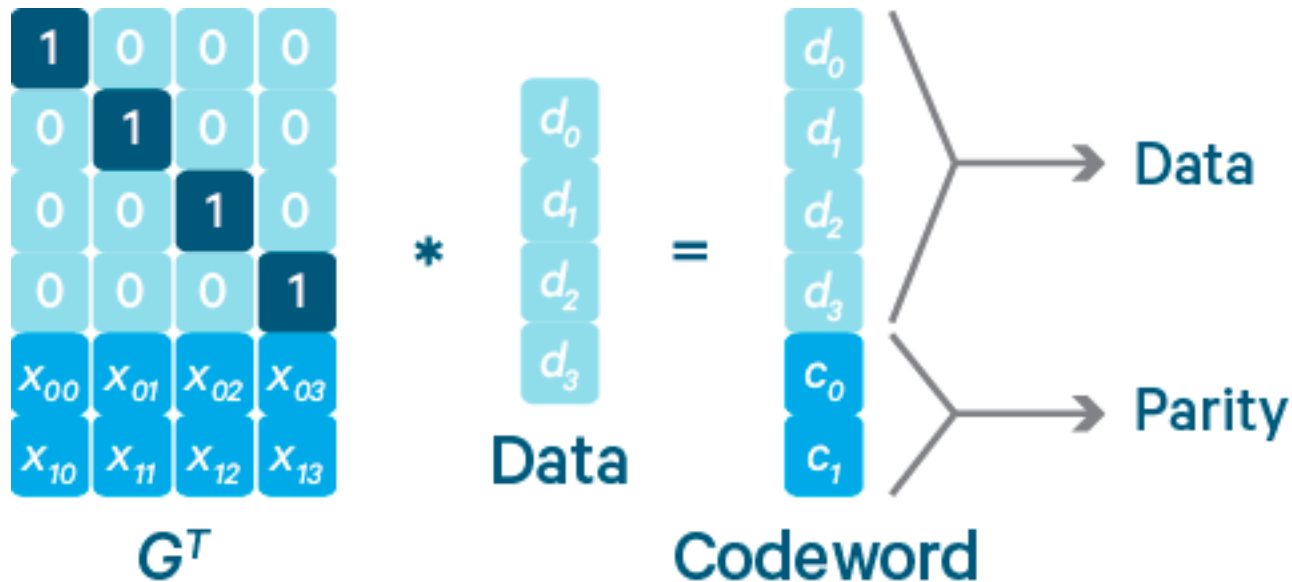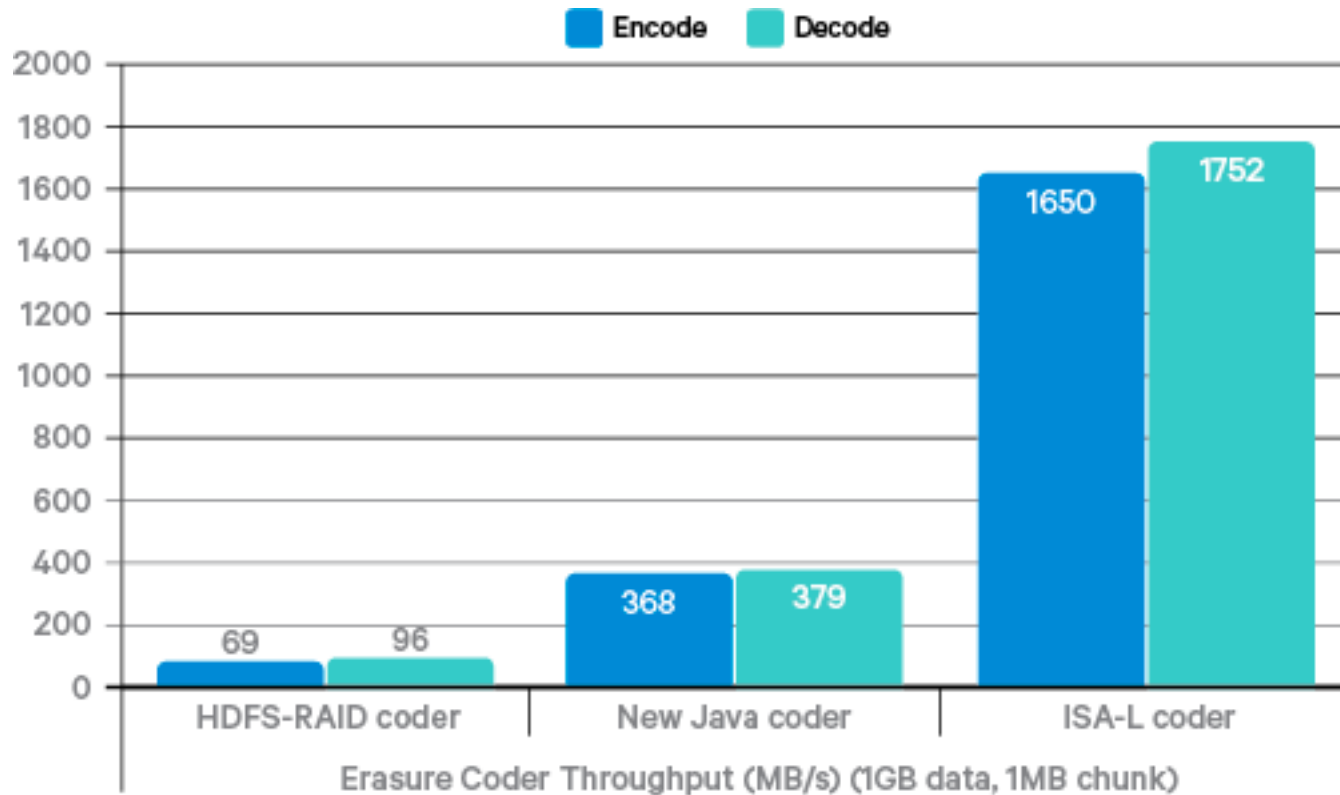
HDFS Erasure Coding

XOR

Input		Output
A	B	
0	0	0
0	1	1
1	0	1
1	1	0

HDFS Erasure Coding

	Data Durability	Storage Efficiency
Single replica	0	100%
Three-way replication	2	33%
XOR with six data cells	1	86%
RS(6,3)	3	67%
RS(10,4)	4	71%

$$\frac{k}{k+m}$$

HDFS Erasure Coding

Erasure Coder Throughput (MB/s) (1GB data, 1MB chunk)

Legend: Encode, Decode

HDFS-RAID coder: Encode 69, Decode 96
New Java coder: Encode 368, Decode 379
ISA-L coder: Encode 1650, Decode 1752

HDFS Erasure Coding

HDFS Write Throughput (MB/s)

HDFS Read Throughput (MB/s) with 2 datanodes killed

Legend: 3-Replica, HDFS-RAID coder, New Java coder, ISA-L coder
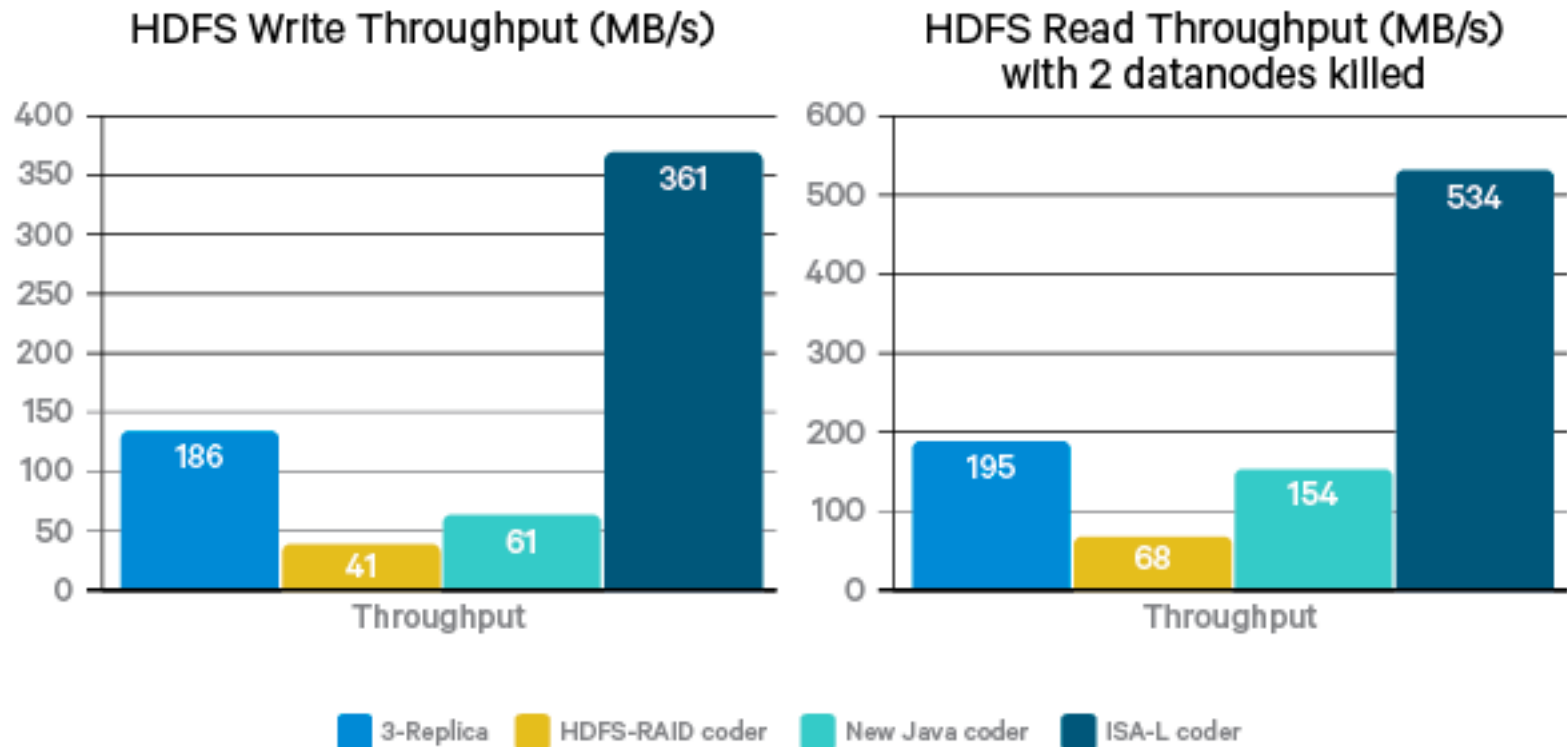
11 nodes (1 NameNode, 9 DataNodes, 1 client node) with 10 GigE network.
- client writing a 12GB file to HDFS;
- client reading a 12GB file from HDFS

Работа с HDFS (CLI)

Команда	Пример
appendToFile	hdfs dfs -appendToFile localfile /user/hadoop/hadoopfile
cat	hdfs dfs -cat hdfs://nn1.example.com/file1
copyFromLocal **copyToLocal**	hdfs dfs -copyFromLocal localfile /user/hadoop/data/ hdfs dfs -copyToLocal localfile /tmp/data/ localfile
cp	hdfs dfs -cp [-f] [-p | -p[topax]] URI [URI …] <dest>
du	hdfs dfs -du -s /tmp/test.data
expunge	hdfs dfs -expunge
get **getmerge**	hdfs dfs -get /user/hadoop/file localfile hdfs dfs -getmerge <src> <localdst> [addnl]
ls	hdfs dfs -ls /user/hadoop/file1
mkdir	hdfs dfs -mkdir /user/hadoop/dir1 /user/hadoop/dir2
mv	hdfs dfs -mv /user/hadoop/file1 /user/hadoop/file2

Работа с HDFS (CLI)

Команда	Пример
put	hdfs dfs -put localfile /user/hadoop/hadoopfile
rm	hdfs dfs -rm [-f] [-r|-R] [-skipTrash] URI [URI ...]
tail	hdfs dfs -tail pathname
setrep	hdfs dfs -setrep [-R] [-w] <numReplicas> <path>

Права доступа

Команда	Пример
chmod	hdfs dfs chmod [-R] mode file
chgrp	hdfs dfs chgrp [-R] group file ...
chown	hdfs dfs chown [-R] [owner][:[group]] file ...
getfacl	hdfs dfs -getfacl /file hdfs dfs -getfacl -R /dir
setfacl	hdfs dfs -setfacl -m user:hadoop:rw- /file

```
[root@hdp-15 ~]# hdfs dfs -ls  /user/hive/000000_0
-rwxr-xr-x   3 hive hive 1133129924 2015-04-23 11:48 /user/hive/000000_0
```

Квоты

Команда	Пример
setQuota	hdsf dfsadmin -setQuota <N> <directory>...<directory>
clrQuota	hdfs dfsadmin -clrQuota <directory>...<directory>
setSpaceQuota	hdfs dfsadmin -setSpaceQuota <N> <directory>...<directory>
clrSpaceQuota	hdfs dfsadmin -clrSpaceQuota <directory>...<director>
count	hdfs dfs -count -q <directory>...<directory>

Квоты

[hdfs@hdp-15 ~]$ **hdfs dfsadmin -setQuota 2 /tmp/tests**
[hdfs@hdp-15 ~]$ **hdfs dfs -count -q /tmp/tests**

| 2 | 1 | none | inf | 1 | 0 | 0 /tmp/tests |

[hdfs@hdp-15 ~]$ hdfs dfs -count -q /tmp/tests/

| 2 | 1 | none | inf | 1 | 0 | 0 /tmp/tests |

[hdfs@hdp-15 ~]$ hdfs dfs -touchz /tmp/tests/test1
[hdfs@hdp-15 ~]$ hdfs dfs -count -q /tmp/tests/

| 2 | 0 | none | inf | 1 | 1 | 0 /tmp/tests |

[hdfs@hdp-15 ~]$ hdfs dfs -touchz /tmp/tests/test2
touchz: The NameSpace quota (directories and files) of directory /tmp/tests is exceeded: quota=2 file count=3

-count -q:
QUOTA   REMAINING_QUOTA SPACE_QUOTA   REMAINING_SPACE_QUOTA
DIR_COUNT   FILE_COUNT   CONTENT_SIZE   FILE_NAME

Стандартное API

- Основная библиотека: hadoop-client
- Пример:

```
private static void ImportDirToHiveTable(Path sourceDir, Path
targetDir) throws IOException {
    Configuration conf = new Configuration();
    FileSystem fs = FileSystem.get(conf);
    for (FileStatus fsts : fs.listStatus(sourceDir)) {
        FileUtil.copy(fs, fsts.getPath(), fs, targetDir, false, conf);
        logger.info(String.format("File %s copied to %s",
                fsts.getPath().getName(),
                        targetDir.getName()));
    }
}
```

Web API

Команда	
CREATE	Создание и запись данных в файл
APPEND	Дописываение файла
CONCAT	Объединение файлов
OPEN	Открытие и чтение файла
MKDIRS	Создание каталога
RENAME	Переименование файла/каталога
DELETE	Удаление файла/каталога
GETFILESTATUS	Получение информации о файле/каталоге
LISTSTATUS	Просмотр информации о каталоге
…	

Web API (пример)

```
[root@hdp-15 ~]# curl -i "http://192.168.91.139:14000/webhdfs/v1/user/hive?
op=LISTSTATUS&user.name=hdfs"
{"FileStatuses":
          {"FileStatus":
          [{
                    "pathSuffix":"user.csv",
                    "type":"FILE",
                    "length":22628,
                    "owner":"hive",
                    "group":"hive",
                    "permission":"644",
                    "accessTime":1429262046873,
                    "modificationTime":1429262048992,
                    "blockSize":134217728,
                    "replication":3
          }]
          }
}
```

Монтирование HDFS

- FUSE (Filesystem in Userspace)
  - sudo yum install hadoop-hdfs-fuse
  - mkdir -p <mount_point>
  - hadoop-fuse-dfs dfs://
    <NN_hostname>:<NN_port> <mount_point>
- export LIBHDFS_OPTS="-Xmx128m"
- /etc/fuse.conf – файл конфигурации FUSE

SQOOP

- Импорт данных из RDBMS
- Экспорт данных в RDBMS
- Вся БД / таблица / запрос
- Многопоточность
- Поддержка Hbase, Accumulo

SQOOP
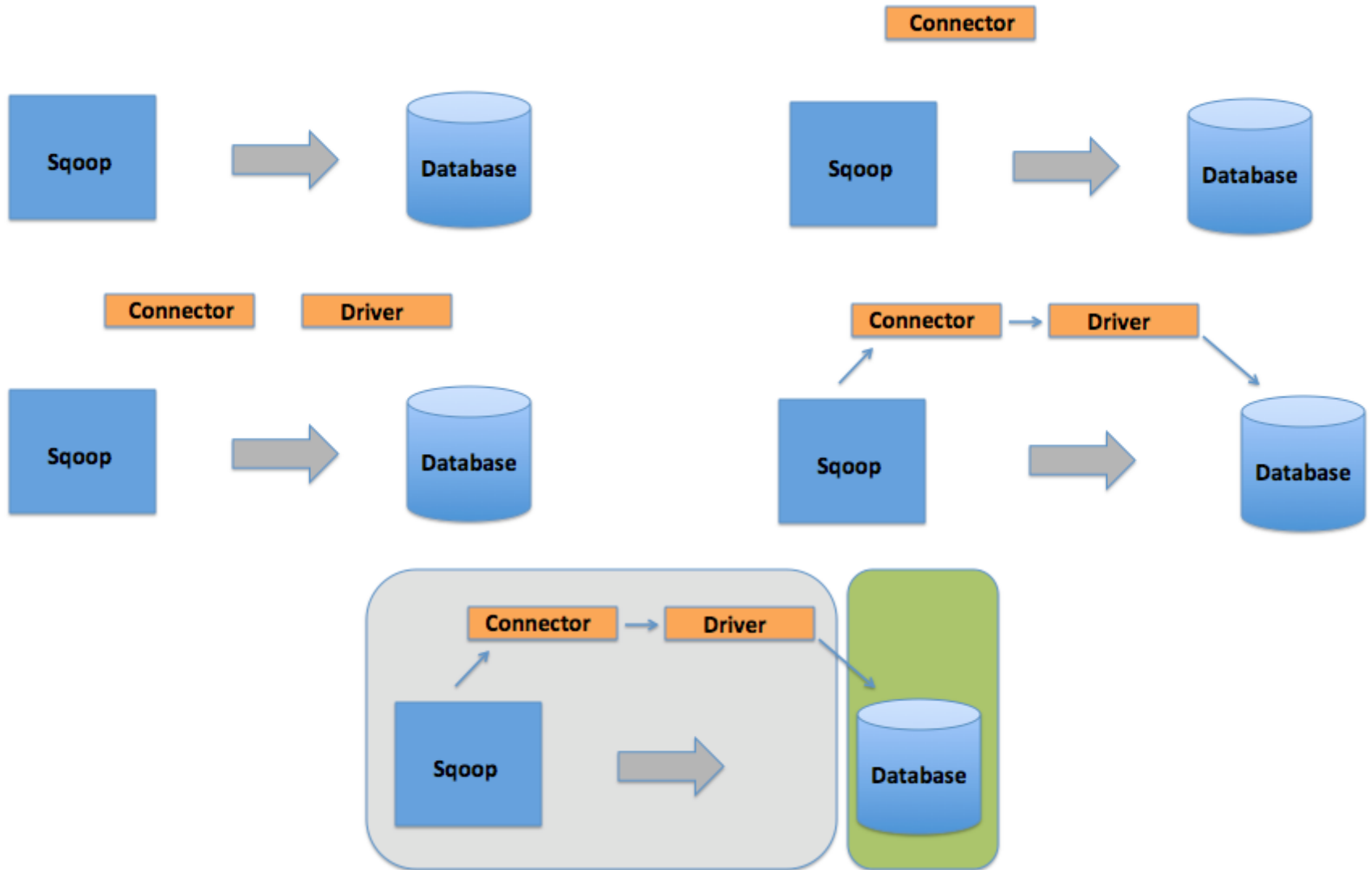
SnakeBite (spotify.com)

- Написан на python
- Использует ProtoBuf (google.com)

```
wouter@foo:~$ time for i in {1..10}; do hadoop fs -ls / > /dev/n

real     0m14.464s
user     0m21.761s
sys      0m1.148s


wouter@foo:~$ time for i in {1..10}; do snakebite ls / > /dev/nu

real     0m1.639s
user     0m1.072s
sys      0m0.160s
```

Интеграция с Amazon s3

```
$HADOOP_HOME/conf/hdfs-site.xml:
<property>
    <name>fs.s3n.awsAccessKeyId</name>
    <value>AWS-ID</value>
</property>
<property>
    <name>fs.s3n.awsSecretAccessKey</name>
    <value>AWS-SECRET-KEY</value>
</property>
```
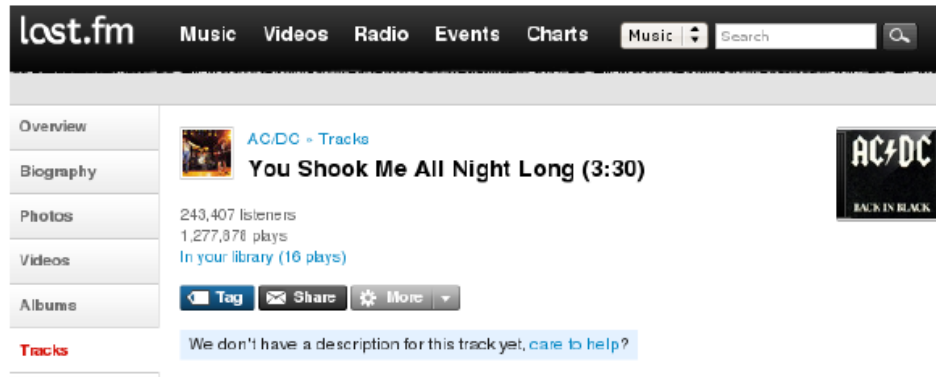
Интеграция с Amazon s3

$ hadoop jar hadoop-*-examples.jar wordcount **s3n://BUCKET-NAME/** s3n://BUCKET-NAME/DIRECTORY-NAME

или

$ hadoop jar hadoop-*-examples.jar wordcount **s3n://AWS-ID: AWS-SECRET-KEY@BUCKET-NAME/** s3n:// AWS-ID: AWS-SECRET-KEY@BUCKET-NAME/DIRECTORY-NAME

Hadoop в last.fm

Hadoop в last.fm

UserId	TrackId	Scrobble	Radio	Skip
111115	222	0	1	0
111113	225	1	0	0
111119	223	0	1	1
111115	225	1	0	0

Hadoop в last.fm
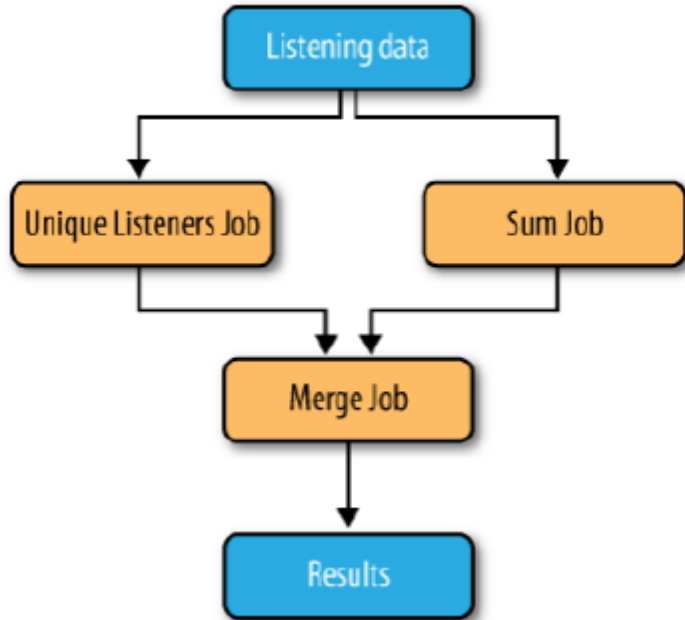
TrackId	#listeners
IntWritable	IntWritable
222	1
223	1
225	2

TrackId	#listeners	#plays	#scrobbles	#radio	#skips
IntWritable	IntWritable	IntWritable	IntWritable	IntWritable	IntWritable
222	1	1	0	1	0
223	1	1	0	1	1
225	1	2	2	0	0

Дополнительные материалы

- http://hadoop.apache.org/docs/stable/
- Hadoop. Подробное руководство

- http://blog.cloudera.com/