

Yandex

HDFS

Web UI, REST API

Datanode Information

In operation

Node	Last contact	Capacity		Blocks	Block pool used	Version
virtual-node1.atp-fivt.org (138.201.93.209:50010)	Sun Apr 09 10:18:09 + 0300 2017	560.1 GB	<div><div></div></div>	7173	124.62 GB (22.25%)	2.6.0- cdh5.9.0
virtual-node2.atp-fivt.org (138.201.94.59:50010)	Sun Apr 09 10:18:09 + 0300 2017	560.1 GB	<div><div></div></div>	7191	124.65 GB (22.26%)	2.6.0- cdh5.9.0
virtual-node3.atp-fivt.org (138.201.95.93:50010)	Sun Apr 09 10:18:09 + 0300 2017	560.1 GB	<div><div></div></div>	7201	124.62 GB (22.25%)	2.6.0- cdh5.9.0

<http://namenode-server:50070/dfshealth.html#tab-datanode>

Browse Directory

/data/wiki/en_articles_part

Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name
-rw-r--r--	hdfs	supergroup	73.3 MB	Sat Jan 07 22:59:55 +0300 2017	3	128 MB	articles-part

Browse Directory

/data/wiki/en_artic

Permission	Owner	Group
-rw-r--r--	hdfs	sup

File information - articles

Download

Block information -- Block 77

Block ID: 1073808554

Block Pool ID: BP-858288134-138.201.91.191-1481279621434

Generation Stamp: 67733

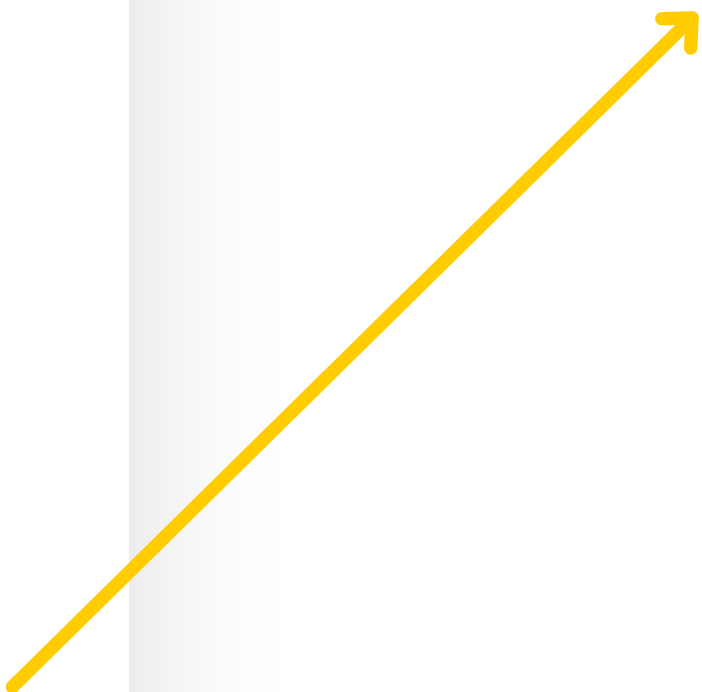
Size: 134217728

Availability:

- virtual-node1.atp-fivt.org
- virtual-node3.atp-fivt.org
- virtual-node2.atp-fivt.org

Close

Block Size	Name
128 MB	articles-part



Browse Directory

/data/wiki/en_artic

Permission	Owner	Group
-rw-r--r--	hdfs	sup

File information - articles

Download

Block information -- Block 77

Block ID: 1073808554

Block Pool ID: BP-858288134-138.201.91.191-1481279621434

Generation Stamp: 67733

Size: 134217728

Availability:

- virtual-node1.atp-fivt.org
- virtual-node3.atp-fivt.org
- virtual-node2.atp-fivt.org

Close

Block Size	Name
128 MB	articles-part

Browse Directory

/data/wiki/en_artic

Permission	Owner	Group
------------	-------	-------

-rw-r--r--	hdfs	sup
------------	------	-----



File information - articles

Download

Block information -- Block 77

Block ID: 1073808554

Block Pool ID: BP-858288134-138.201.91.191-1481279621434

Generation Stamp: 67733

Size: 134217728

Availability:

- virtual-node1.atp-fivt.org
- virtual-node3.atp-fivt.org
- virtual-node2.atp-fivt.org

Close

Block Size	Name
------------	------

128 MB	articles-part
--------	---------------

Browse Directory

/data/wiki/en_artic

Permission	Owner	Group
-rw-r--r--	hdfs	sup



File information - articles

Download

Block information -- Block 77

Block ID: 1073808554

Block Pool ID: BP-858288134-138.201.91.191-1481279621434

Generation Stamp: 67733

Size: 134217728

Availability:

• virtual-node1.atp-fivt.org

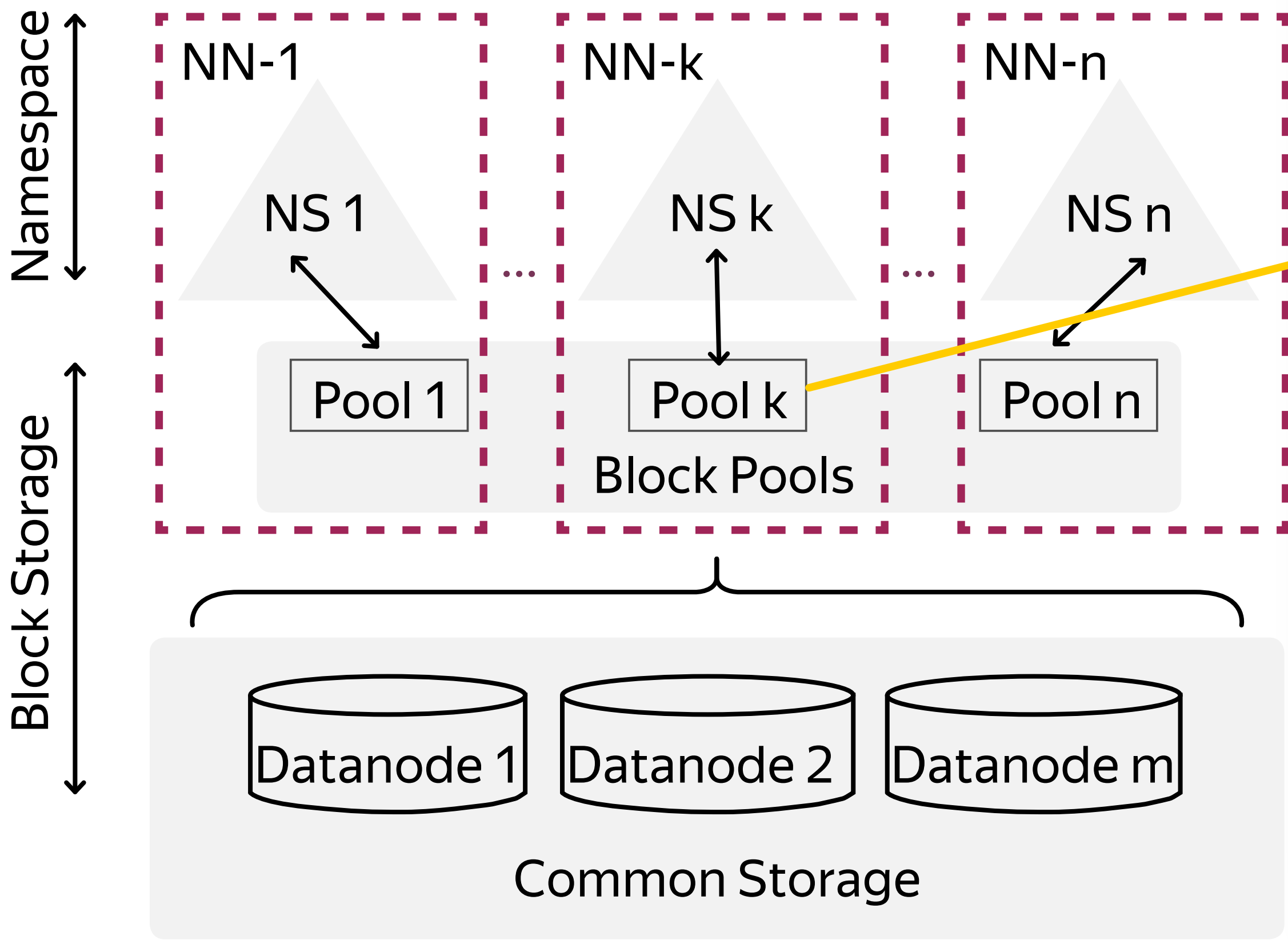
• virtual-node3.atp-fivt.org

• virtual-node2.atp-fivt.org

Close

Block Size	Name
128 MB	articles-part

HDFS Federationv



File information - articles

Download

Block information -- Block 77

Block ID: 1073808554

Block Pool ID: BP-858288134-138.201.91.191-1481279621434

Generation Stamp: 67733

Size: 134217728

Availability:

- virtual-node1.atp-fivt.org
- virtual-node3.atp-fivt.org
- virtual-node2.atp-fivt.org

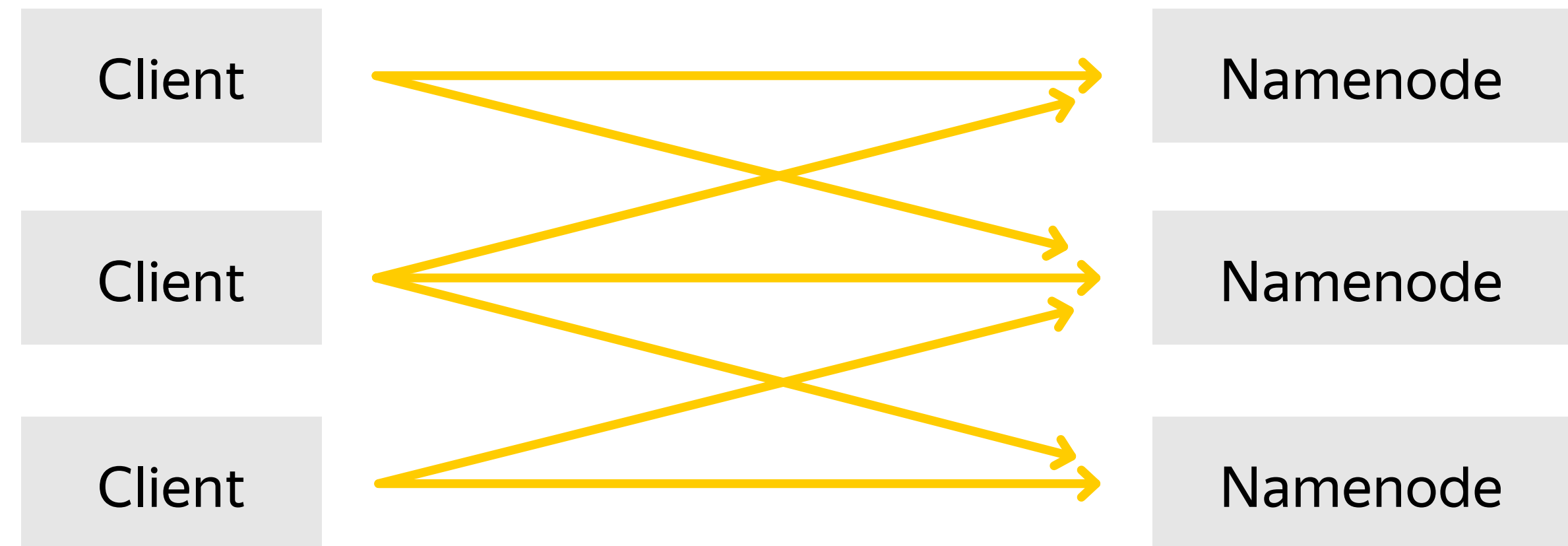
Close

WebHDFS

————→ HTTP request → RPC request ----→ block request

WebHDFS

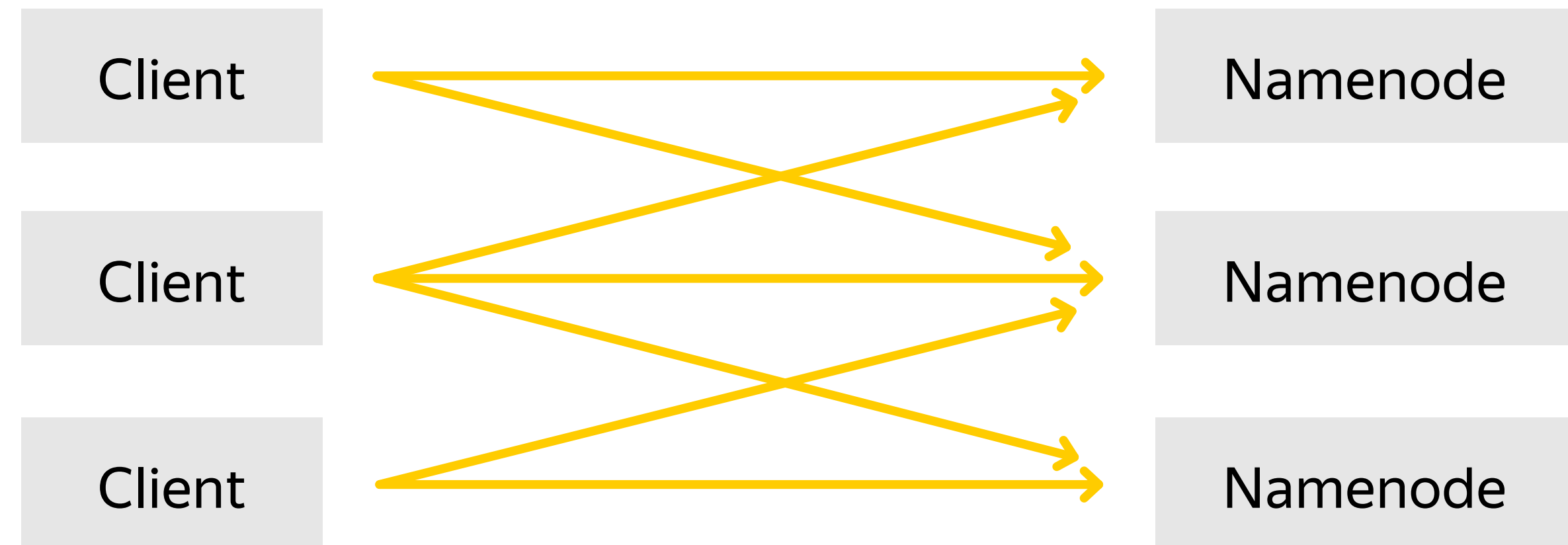
› Direct access



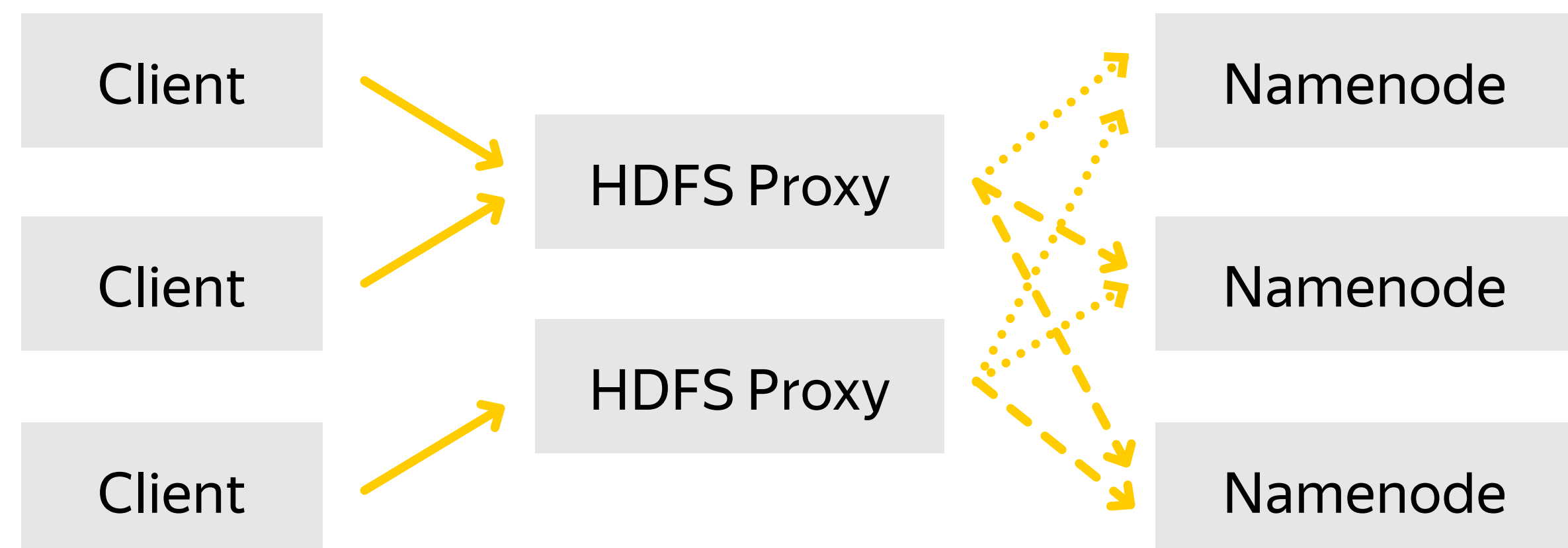
————→ HTTP request → RPC request ----→ block request

WebHDFS

› Direct access



› HDFS proxies



—————> HTTP request > RPC request - - - -> block request

```
$ curl -i http://virtual-master.atp-fvt.org:50070/webhdfs/v1/data/access_logs/big_log/
access.log.2015-12-10\?op=OPEN
HTTP/1.1 307 TEMPORARY_REDIRECT
Cache-Control: no-cache
Expires: Sun, 09 Apr 2017 08:32:02 GMT
Date: Sun, 09 Apr 2017 08:32:02 GMT
Pragma: no-cache
Expires: Sun, 09 Apr 2017 08:32:02 GMT
Date: Sun, 09 Apr 2017 08:32:02 GMT
Pragma: no-cache
Content-Type: application/octet-stream
X-FRAME-OPTIONS: SAMEORIGIN
Location: http://virtual-node1.atp-fvt.org:50075/webhdfs/v1/data/access_logs/
big_log/access.log.2015-12-10?op=OPEN&namenoderpcaddress=virtual-master.atp-fvt.
org:8020&offset=0
Content-Length: 0
Server: Jetty(6.1.26.cloudera.4)
```

```
$ curl -i http://virtual-node1.atp-fvt.org:50075/webhdfs/v1/data/access_logs/big_log/  
access.log.2015-12-10\?op\=OPEN\&namenoderpcaddress\=virtual-master.atp-fvt.org:8020\&of  
fset\=0\&length\=100
```

HTTP/1.1 200 OK

Access-Control-Allow-Methods: GET

Access-Control-Allow-Origin: *

Content-Type: application/octet-stream

Connection: close

Content-Length: 100

```
41.190.60.121 - - [10/Dec/2015:00:00:00 +0400] "GET /cgi-bin/commandit.cgi HTTP/1.1"  
404 0 "-" "Mozilla/5.0 (Windows NT 6.0; WOW64; rv:31.0) Gecko/20100101 Firefox/31.0"
```

actual data

```
$ curl -i -L http://virtual-master.atp-fvt.org:50070/webhdfs/v1/data/access_logs/big_log/
access.log.2015-12-10\?op=OPEN\&length\=100
```

HTTP/1.1 307 TEMPORARY_REDIRECT

...

Content-Type: application/octet-stream

X-FRAME-OPTIONS: SAMEORIGIN

Location: http://virtual-node2.atp-fvt.org:50075/webhdfs/v1/data/access_logs/big_log/access.log.2015-12-10?op=OPEN&namenoderpcaddress=virtual-master.atp-fvt.org:8020&length=100&offset=0

Content-Length: 0

Server: Jetty(6.1.26.cloudera.4)

HTTP/1.1 200 OK

Access-Control-Allow-Methods: GET

Access-Control-Allow-Origin: *

Content-Type: application/octet-stream

Connection: close

Content-Length: 100

41.190.60.121 - - [10/Dec/2015:00:00:00 +0400] "GET /cgi-bin/commandit.cgi HTTP/1.1"

404 0 '-' '-' 'Mozi%

```
$ curl -i -L http://virtual-master.atp-fvt.org:50070/webhdfs/v1/data/access_logs/big_log/
access.log.2015-12-10\?op=OPEN\&length\=100
```

HTTP/1.1 307 TEMPORARY_REDIRECT

...

Content-Type: application/octet-stream

X-FRAME-OPTIONS: SAMEORIGIN

Location: http://virtual-node2.atp-fvt.org:50075/webhdfs/v1/data/access_logs/big_log/access.log.2015-12-10?op=OPEN&namenoderpcaddress=virtual-master.atp-fvt.org:

8020&length=100&offset=0 ←

Content-Length: 0

Server: Jetty(6.1.26.cloudera.4)

HTTP/1.1 200 OK

Access-Control-Allow-Methods: GET

Access-Control-Allow-Origin: *

Content-Type: application/octet-stream

Connection: close

Content-Length: 100

41.190.60.121 - - [10/Dec/2015:00:00:00 +0400] "GET /cgi-bin/commandit.cgi HTTP/1.1"

404 0 '-' '-' Mozi%

WebHDFS Operations (examples)

HTTP GET

- › OPEN — read data
- › GETFILESTATUS — get file meta information
- › LISTSTATUS — list directory

HTTP PUT

- › CREATE
- › MKDIRS
- › RENAME
- › SETREPLICATION

HTTP POST

- › APPEND

HTTP DELETE

- › DELETE

WebHDFS Operations (examples)

HTTP GET

- › OPEN — read data
- › GETFILESTATUS — get file meta information
- › LISTSTATUS — list directory

HTTP PUT

- › CREATE
- › MKDIRS
- › RENAME
- › SETREPLICATION

HTTP POST

- › APPEND

HTTP DELETE

- › DELETE

Summary

Summary

- › you can **use** HDFS Web UI to list folders and find detailed information about file blocks

Summary

- › you can **use** HDFS Web UI to list folders and find detailed information about file blocks
- › you can **use** curl to move files through WebHDFS REST API, read / write files, ...

BigDATAteam