

Ceph RGW-Prefetching

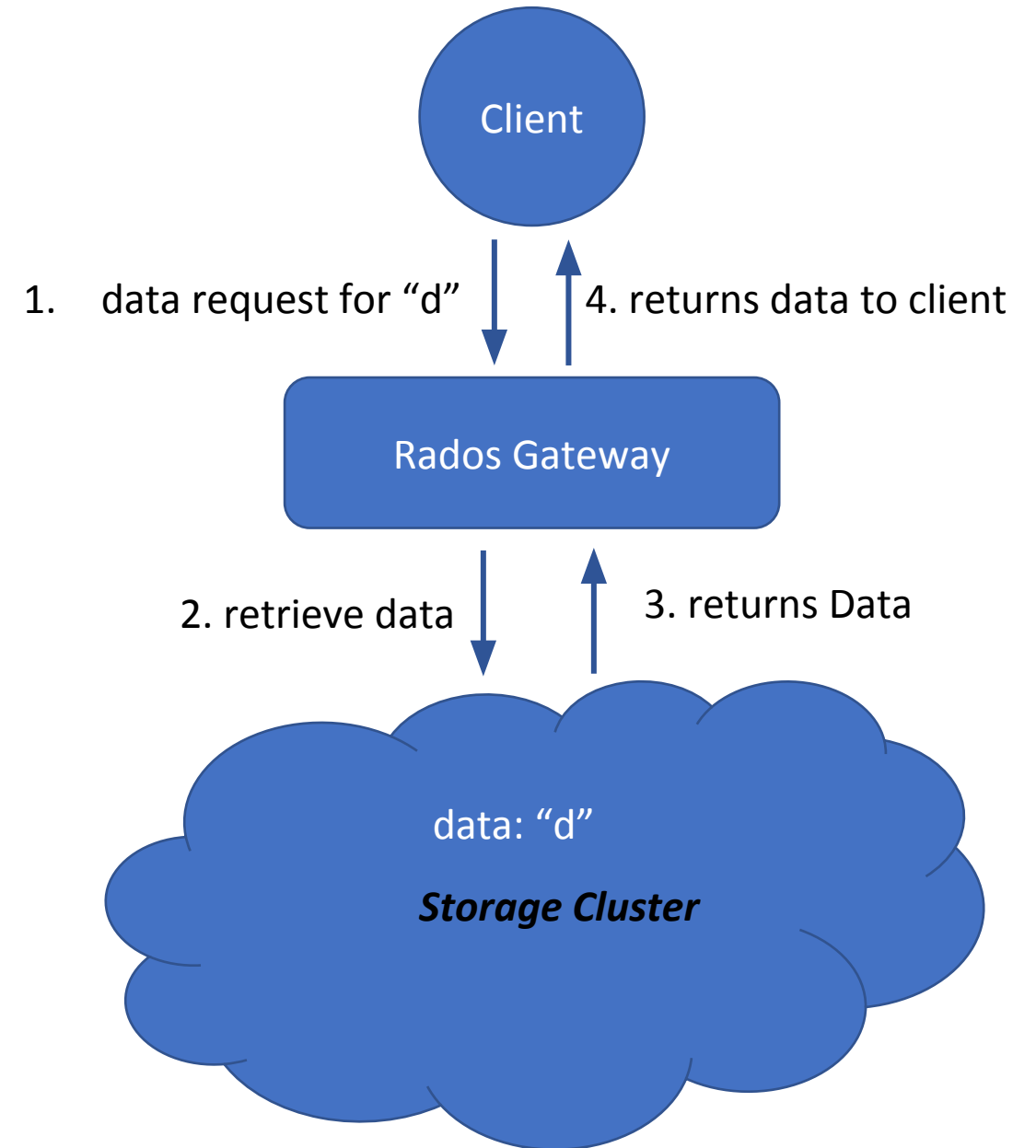
Ali R., Amin M., Bissenbay D.

What is CEPH?

- ❑ Distributed Storage System

- ❑ **Key Components** [Related to Project]

- ❑ Client (Applications)
- ❑ Rados Gateway
- ❑ Storage Cluster



Prefetching in Ceph

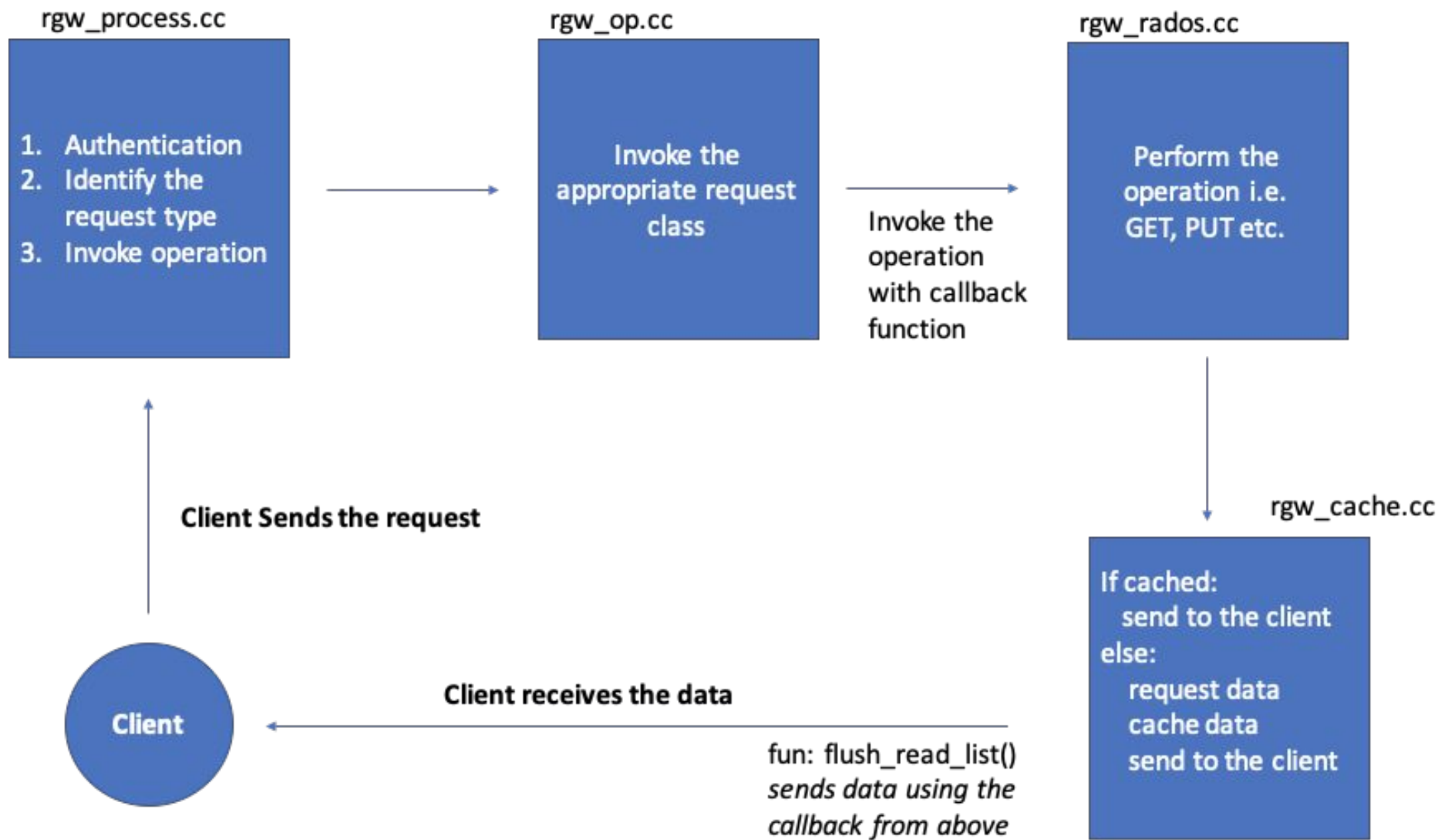
- ❑ Current caching scheme is re-active
- ❑ Goal: Improve performance in first access
 - ❑ Lower latency
 - ❑ Higher throughput
- ❑ Two type of prefetching
 - ❑ on-demand prefetching
 - ❑ pro-active prefetching in rgw

Sprint 5

- ❑ Done:
 - ❑ Prefetching based on user request
 - ❑ Pro-active prefetching on block level
 - ❑ Full code testing of the above prefetching code

- ❑ In progress:
 - ❑ Evaluation (in terms of performance)
 - ❑ Improving the design
 - ❑ Stretch goals

Prefetching Based on User Request



Prefetching based on user request

- ❑ Overload the GET request (add header)
- ❑ Catch the header in `rgw_op.cc`
 - ❑ Reply the user with normal response headers (200 OK etc. and Content-length: 0)
 - ❑ Update the `handle_data_cb` so that instead of sending back the data, place it in the cache

Prefetching based on user request

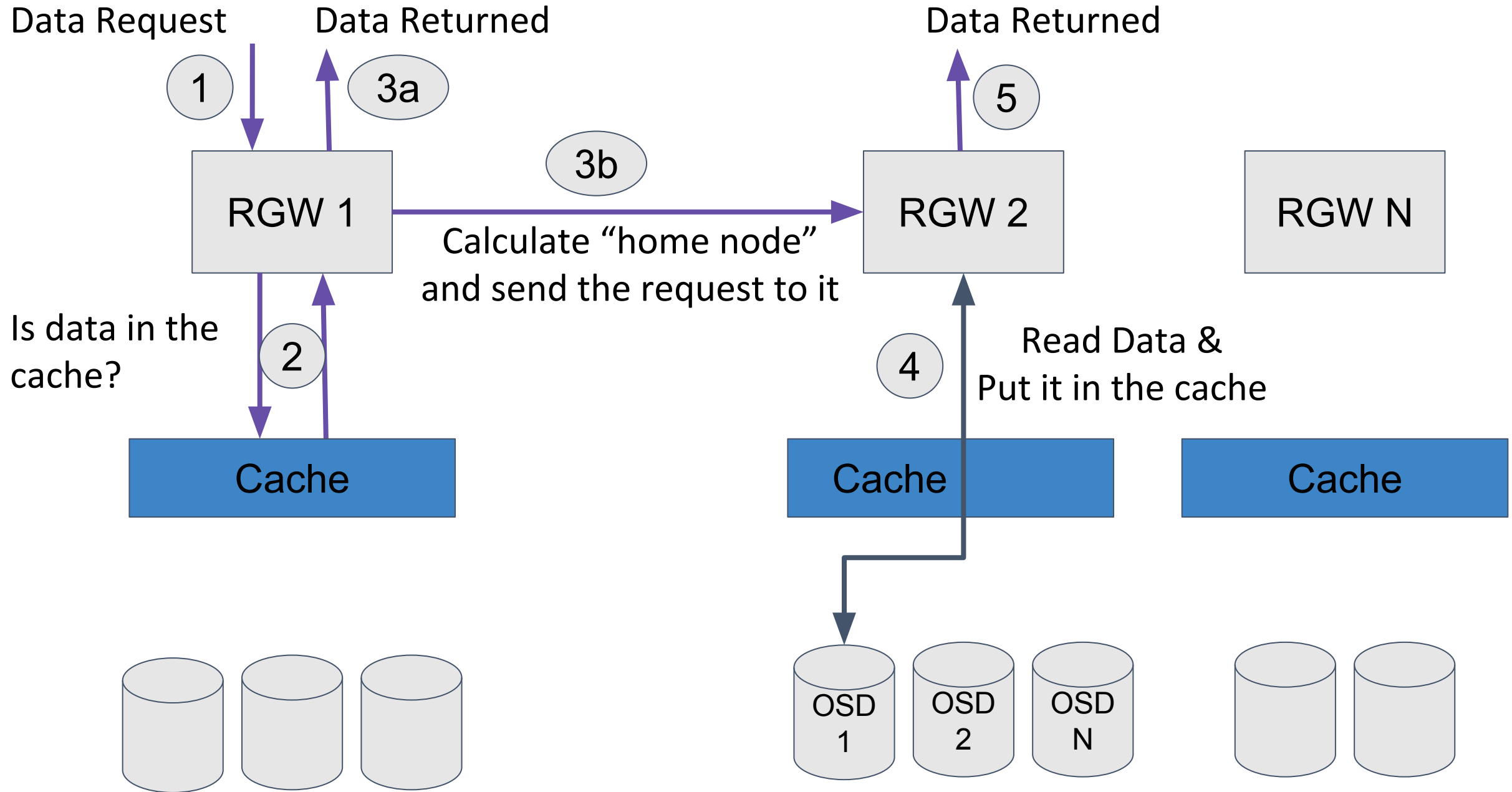
- ❑ Overload the GET request (add header)
- ❑ Catch the header in `rgw_op.cc`
 - ❑ Reply the user with normal response headers (200 OK etc. and Content-length: 0)
 - ❑ Update the `handle_data_cb` so that instead of sending back the data, place it in the cache

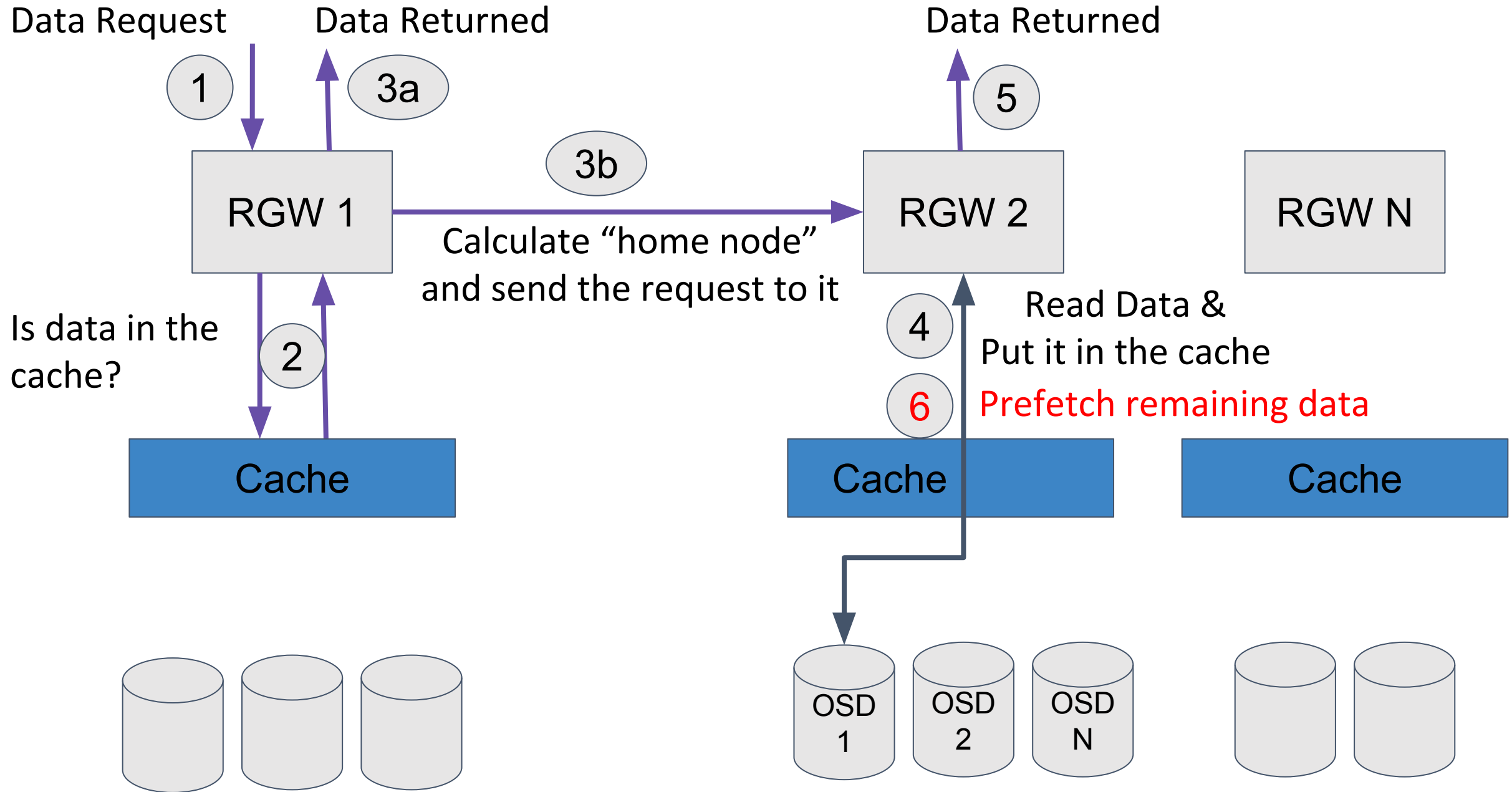
Demo

Challenges

- ❑ Fairness
- ❑ Another request for the “data” that is being prefetched

Pro-active Prefetching





Pro-active prefetching

- ❑ A prefetching C++ class to talk with cache
 - ❑ sends a HTTP prefetching request to the RGW
 - ❑ RGW forwards it to home node RGW
 - ❑ Home node returns data to RGW
- ❑ A thread pool for prefetching
- ❑ Adding prefetching task to the thread pool

Evaluation & Challenges

- ❑ Original plan was to evaluate CEPH's performance using COSbench
- ❑ Challenges with COSBench

So, we came up with simple performance test that we will present next

but if time allows we will benchmark CEPH + Prefetching performance using COSbench

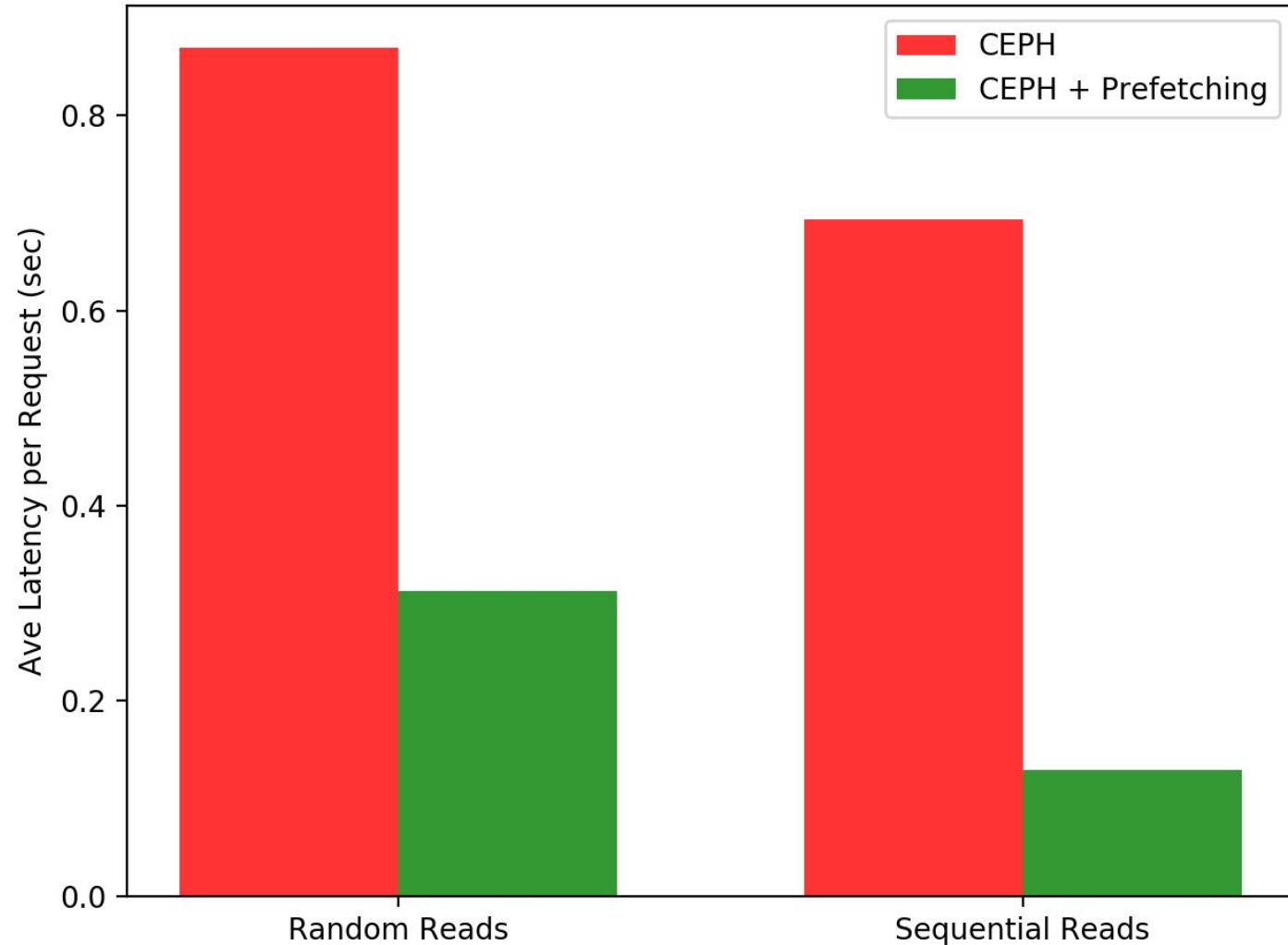
Initial Results

Experimental Setup

- ❑ Stored two files in CEPH each with file size (500MB)
- ❑ For each file created 10 request for the data chunks (0-50, 50-100, 100-150, ..., 450-500)

Results:

- Random Reads
- Sequential Reads



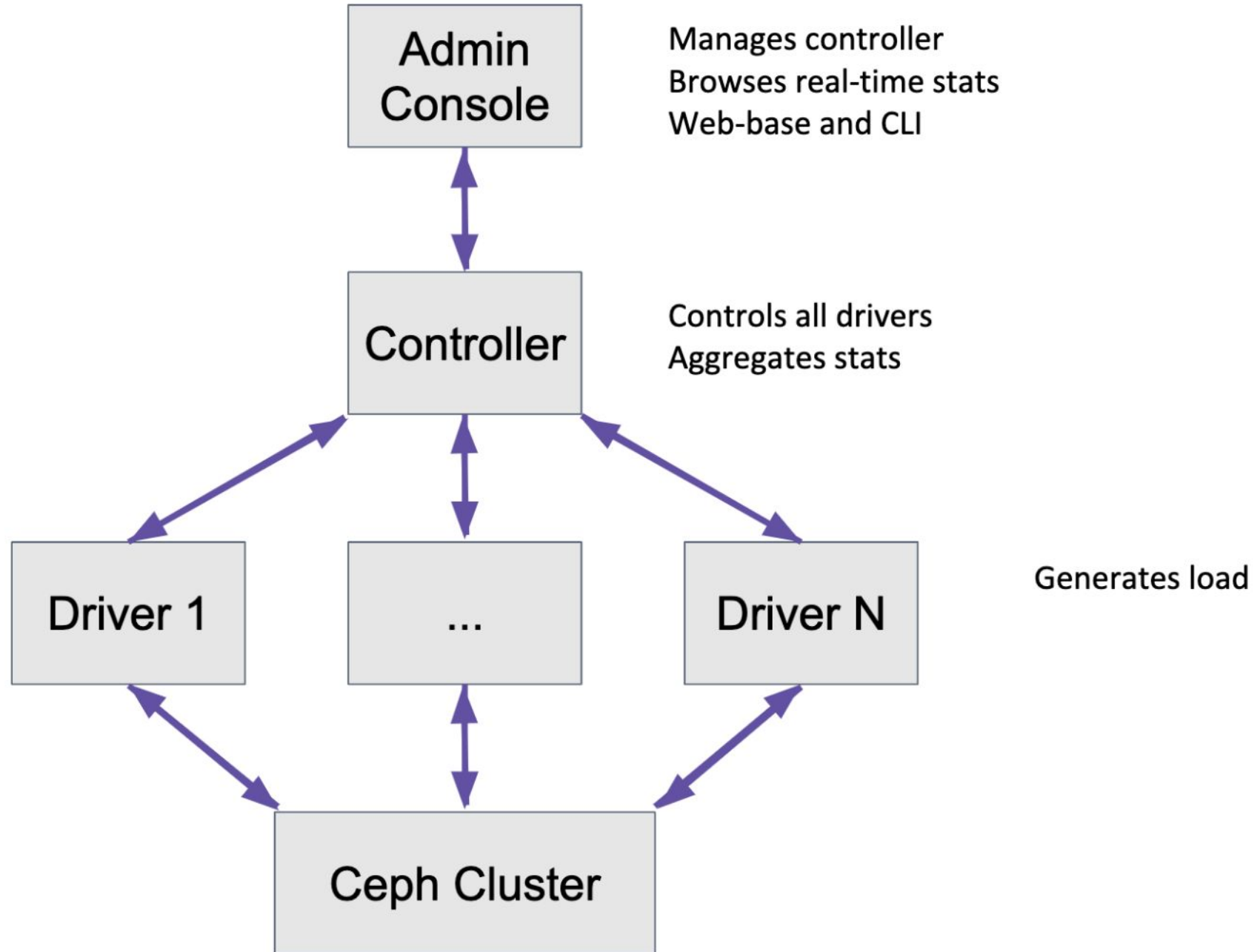
Improvements

- ❑ *More Performance*
 - ❑ Home node does not need to send an HTTP request to itself
 - ❑ Prefetch data in background or after serving the client
- ❑ *Concurrent requests for the same data can cause multiple reads/prefetches of the same data*
 - ❑ Create a queue of the requests for same data
 - ❑ Need to record on-flight prefetching requests for fairness purpose

COSBench Challenges

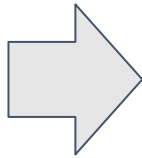
- 1. Running Workload**
- 2. Setting Up DEV Environment**

COSBench



COSBench Challenges - Run Workload

1. Drivers & Controllers cannot be launched! Solution v0.4.2.c4 instead v0.4.2

[illegible]

```
Listening on port 0.0.0.0/0.0.0.0:18089 ...
Persistence bundle starting...
Persistence bundle started.
-----
!!! Service will listen on web port: 18088 !!!
-----
~

Listening on port 0.0.0.0/0.0.0.0:19089 ...
Persistence bundle starting...
Persistence bundle started.
-----
!!! Service will listen on web port: 19088 !!!
-----
~
```

COSBench is not under active development for the last 2/3 years!!!

COSBench Challenges - Run Workload

```
<?xml version="1.0" encoding="UTF-8" ?>
<workload name="s3-b1" description="b1 sample benchmark for s3">

  <storage type="s3" config="accesskey=0555b35654ad1656d804;secretkey=h7GhxuBLTrlhVUyxSPUKUV8r/2EI4ngqJxD7iBdBYLhwluN30JaT3Q==;endpoint=http://localhost:8000" />

  <workflow>

    <workstage name="init">
      <work type="init" workers="1" config="cprefix=s3b1;containers=r(1,2)" />
    </workstage>

    <workstage name="prepare">
      <work type="prepare" workers="1" config="cprefix=s3b1;containers=r(1,2);objects=r(1,10);sizes=c(64)KB;" />
    </workstage>

    <workstage name="main">
      <work name="main" workers="1" runtime="30">
        <operation type="read" ratio="100" config="cprefix=s3b1;containers=u(1,2);objects=u(1,10)" />
      </work>
    </workstage>

    <workstage name="cleanup">
      <work type="cleanup" workers="1" config="cprefix=s3b1;containers=r(1,2);objects=r(1,10)" />
    </workstage>

    <workstage name="dispose">
      <work type="dispose" workers="1" config="cprefix=s3b1;containers=r(1,2)" />
    </workstage>

  </workflow>

</workload>
```

Contributors is not responding for emails!!!

COSBench Challenges - Run Workload

2. SignatureDoesNotMatch

```
2019-04-17 16:28:38,140 [INFO] [Log4jLogManager] - will append log to file /home/0.4.2.c4/log/mission/MA2C20F2005.log
2019-04-17 16:28:39,057 [INFO] [NoneStorage] - performing PUT at /s3body1
2019-04-17 16:28:39,348 [INFO] [NoneStorage] - performing PUT at /s3body2
```

```
2019-04-17 16:28:44,010 [INFO] [Log4jLogManager] - will append log to file /home/0.4.2.c4/log/mission/M62C210B0C6.log
2019-04-17 16:28:44,241 [INFO] [NoneStorage] - performing PUT at /s3body1/myobjects1
2019-04-17 16:28:46,402 [ERROR] [AbstractOperator] - worker 1 fail to perform operation s3body1/myobjects1
com.intel.cosbench.api.storage.StorageException: com.amazonaws.services.s3.model.AmazonS3Exception: null (Service: Amazon S3; Status Code:
403; Error Code: SignatureDoesNotMatch; Request ID: tx0000000000000000000006-005cb7543e-102b-default), S3 Extended Request ID: 102b-default-
default
    at com.intel.cosbench.api.S3Stor.S3Storage.createObject(S3Storage.java:170)
    at com.intel.cosbench.driver.operator.Writer.doWrite(Writer.java:98)
    at com.intel.cosbench.driver.operator.Preparer.operate(Preparer.java:97)
    at com.intel.cosbench.driver.operator.AbstractOperator.operate(AbstractOperator.java:76)
    at com.intel.cosbench.driver.agent.WorkAgent.performOperation(WorkAgent.java:197)
    at com.intel.cosbench.driver.agent.WorkAgent.doWork(WorkAgent.java:177)
    at com.intel.cosbench.driver.agent.WorkAgent.execute(WorkAgent.java:134)
    at com.intel.cosbench.driver.agent.AbstractAgent.call(AbstractAgent.java:44)
    at com.intel.cosbench.driver.agent.AbstractAgent.call(AbstractAgent.java:1)
    at java.util.concurrent.FutureTask.run(FutureTask.java:266)
    at java.util.concurrent.ThreadPoolExecutor.runWorker(ThreadPoolExecutor.java:1149)
    at java.util.concurrent.ThreadPoolExecutor$Worker.run(ThreadPoolExecutor.java:624)
    at java.lang.Thread.run(Thread.java:748)
Caused by: com.amazonaws.services.s3.model.AmazonS3Exception: null (Service: Amazon S3; Status Code: 403; Error Code:
SignatureDoesNotMatch; Request ID: tx0000000000000000000006-005cb7543e-102b-default), S3 Extended Request ID: 102b-default-default
```

COSBench Challenges - Run Workload

2. Solution adding env var “-Dcom.amazonaws.services.s3.disableGetObjectMD5Validation=true” & “path_style_access=true” to config attribute

cosbench-start.sh

```
rm -f $BOOT_LOG
mkdir -p log

echo "Launching osgi framework ..."

/usr/bin/nohup java -Dcosbench.tomcat.config=$TOMCAT_CONFIG -Dcom.amazonaws.services.s3.disableGetObjectMD5Validation=true -server -cp main/* org.eclipse.equinox.launcher.Main -configuration $OSGI_CONFIG -console $OSGI_CONSOLE_PORT 1> $BOOT_LOG 2>&1 &

if [ $? -ne 0 ];
then
    echo "Error in launching osgi framework!"
    cat $BOOT_LOG
    exit 1
fi
```

.xml configuration file

```
<?xml version="1.0" encoding="UTF-8" ?>
<workload name="s3-b1" description="b1 sample benchmark for s3">

  <storage type="s3" config="accesskey=0555b35654ad1656d804;secretkey=h7GhxuBLTrlhVUyxSPUKUV8r/2EI4ngqJxD7iBdBYLhwluN30JaT3Q==;endpoint=http://localhost:8000;path_style_access=true" />

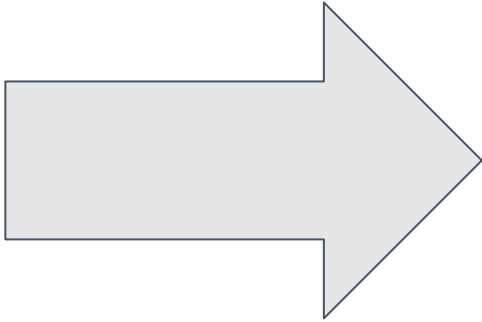
  <workflow>

    <workstage name="init">
      <work type="init" workers="1" config="cprefix=s3b1;containers=r(1,2)" />
    
```

Just 1 hour ago resolved with Red Hat Team

COSBench Challenges - DEV Env

- ▶ cosbench-ampli [cosbench master]
- ▶ cosbench-api [cosbench master]
- ▶ cosbench-castor [cosbench master]
- ▶ cosbench-cdmi-base [cosbench master]
- ▶ cosbench-cdmi-swift [cosbench master]
- ▶ cosbench-cdmi-util [cosbench master]
- ▶ cosbench-config [cosbench master]
- ▶ cosbench-controller [cosbench master]
- ▶ cosbench-core [cosbench master]
- ▶ cosbench-core-web [cosbench master]
- ▶ cosbench-driver [cosbench master]
- ▶ cosbench-driver-web [cosbench master]
- ▶ cosbench-ecs [cosbench master]
- ▶ cosbench-gcs [cosbench master]
- ▶ cosbench-http [cosbench master]
- ▶ cosbench-httpauth [cosbench master]
- ▶ cosbench-keystone [cosbench master]
- ▶ cosbench-librados [cosbench master]
- ▶ cosbench-log [cosbench master]
- ▶ cosbench-log4j [cosbench master]
- ▶ cosbench-mock [cosbench master]
- ▶ cosbench-openio [cosbench master]
- ▶ cosbench-oss [cosbench master]
- ▶ cosbench-s3 [cosbench master]
- ▶ cosbench-scality [cosbench master]
- ▶ cosbench-swauth [cosbench master]
- ▶ cosbench-swift [cosbench master]
- ▶ cosbench-tomcat [cosbench master]
- ▶ cosbench-web-controller [cosbench master]



- ▶ cosbench-ampli
- ▶ cosbench-api
- ▶ cosbench-castor
- ▶ cosbench-cdmi-base
- ▶ cosbench-cdmi-swift
- ▶ cosbench-cdmi-util
- ▶ cosbench-config
- ▶ cosbench-controller
- ▶ cosbench-core
- ▶ cosbench-core-web
- ▶ cosbench-driver
- ▶ cosbench-driver-web
- ▶ cosbench-gcs
- ▶ cosbench-http
- ▶ cosbench-httpauth
- ▶ cosbench-keystone
- ▶ cosbench-librados
- ▶ cosbench-log
- ▶ cosbench-log4j
- ▶ cosbench-mock
- ▶ cosbench-openio
- ▶ cosbench-s3
- ▶ cosbench-scality
- ▶ cosbench-swauth
- ▶ cosbench-swift
- ▶ cosbench-tomcat
- ▶ cosbench-web-controller

BaseConfiguration cannot be resolved to a type	KVConfigPars...	/cosbench-config/s...	line 37	Java Pro
BaseConfiguration cannot be resolved to a type	KVConfigPars...	/cosbench-config/s...	line 37	Java Pro
Catalina cannot be resolved to a type	TomcatLoader...	/cosbench-tomcat/s...	line 37	Java Pro
Catalina cannot be resolved to a type	TomcatLoader...	/cosbench-tomcat/s...	line 88	Java Pro
configFile cannot be resolved to a variable	TomcatLoader...	/cosbench-tomcat/s...	line 53	Java Pro
Configuration cannot be resolved to a type	COSBConfigA...	/cosbench-config/s...	line 32	Java Pro
Configuration cannot be resolved to a type	COSBConfigA...	/cosbench-config/s...	line 34	Java Pro
Configuration cannot be resolved to a type	COSBConfigA...	/cosbench-config/s...	line 35	Java Pro
Configuration cannot be resolved to a type	COSBConfigA...	/cosbench-config/s...	line 41	Java Pro
Configuration cannot be resolved to a type	COSBConfigA...	/cosbench-config/s...	line 46	Java Pro

0 errors, 270 warnings, 13 others (filter matched 200 of 283 items)	
Description	
▶ JRE Compiler Compliance Problem (22 items)	
▶ Java Build Path Problems (27 items)	
▶ Java Problems (100 of 186 items)	
▶ Java Task (8 items)	
▶ Plug-in Problems (48 items)	

Only one guy responded & told how painful it was to set up DEV Env & wished good luck!!!

COSBench Possible Changes

COSBench Possible Changes



```
<workload name="s3-sample" description="sample benchmark for s3">

  <storage type="s3" config="accesskey=0555b35654ad1656d804;secretkey=h7GhXuBLTrlhVUyxSPUKUV8r/2EI4ngqJxD7iBdBYLhwluN30JaT3Q==;endpoint=" />

  <workflow>
    <workstage name="init">
      <work type="init" workers="1" config="cprefix=s3testqwer;containers=r(1,2)" />
      <operation type="prefetch" ratio="100" config="cprefix=s3testqwer;containers=u(1,2);objects=u(1,10)" />
    </work>
  </workstage>

  <workstage name="prepare">
    <work type="prepare" workers="1" config="cprefix=s3testqwer;containers=r(1,2);objects=r(1,10);sizes=c(64)KB" />
  </workstage>

  <workstage name="main">
    <work name="main" workers="8" runtime="30">
      <operation type="read" ratio="80" config="cprefix=s3testqwer;containers=u(1,2);objects=u(1,10)" />
    </work>
  </workstage>

  <workstage name="cleanup">
    <work type="cleanup" workers="1" config="cprefix=s3testqwer;containers=r(1,2);objects=r(1,20)" />
  </workstage>

  <workstage name="dispose">
    <work type="dispose" workers="1" config="cprefix=s3testqwer;containers=r(1,2)" />
  </workstage>
</workflow>
</workload>
```

COSBench Possible Changes



S3Storage.java

```
@Override
public InputStream getObject(String container, String object, Config config) {
    super.getObject(container, object, config);
    InputStream stream = null;
    try {

        //    S3object s3obj = client.getObject(container, object);
//        BISSENBAY CODE
        GetObjectRequest getObjectRequest = new GetObjectRequest(container, object);
        // getObjectRequest.setRange(0, 4000);
        getObjectRequest.putCustomRequestHeader("prefetch", null);
        S3object s3obj = client.getObject(getObjectRequest);
//        BISSENBAY CODE

        stream = s3obj.getObjectContent();

    } catch (AmazonServiceException ase) {
        if(ase.getStatusCode() != HttpStatus.SC_NOT_FOUND) {
            throw new StorageException(ase);
        }
    } catch (AmazonClientException ace) { // recreate the AmazonS3 client connection if it is broken.
        logger.warn("below exception encountered when retrieving object " + object + " at bucket " + container + ": " + ace.getMessage());
        ace.printStackTrace();
        initClient();
    }

    return stream;
}
```

Taiga Chart

