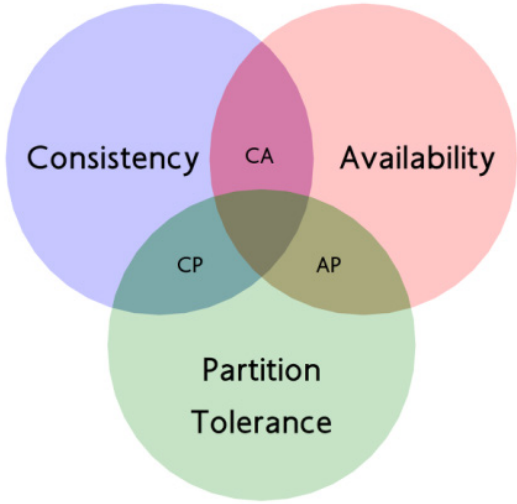


图 1-1

图 1-2

图 1-3

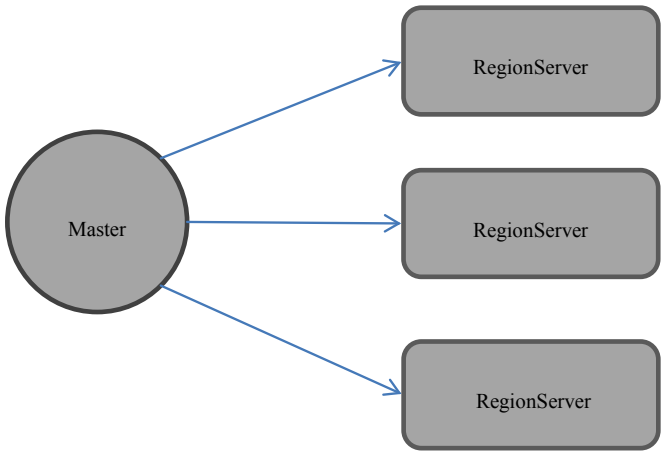


图 1-4

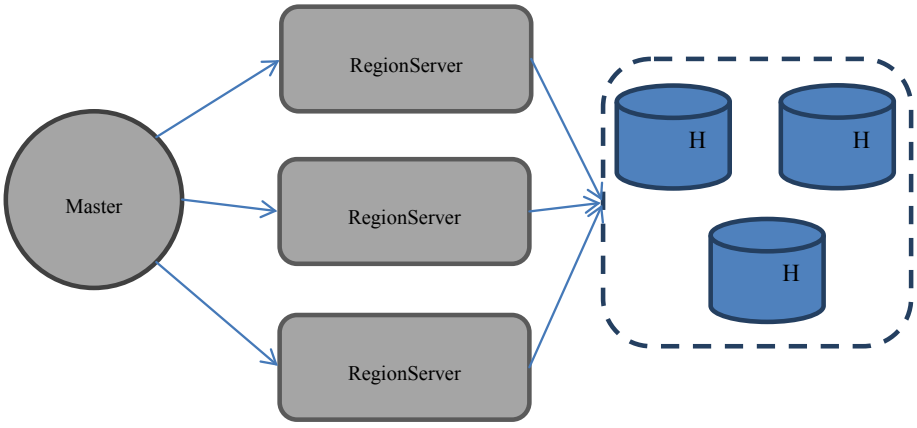

RegionServer
Master
RegionServer
RegionServer
H
H
H

图 1-5

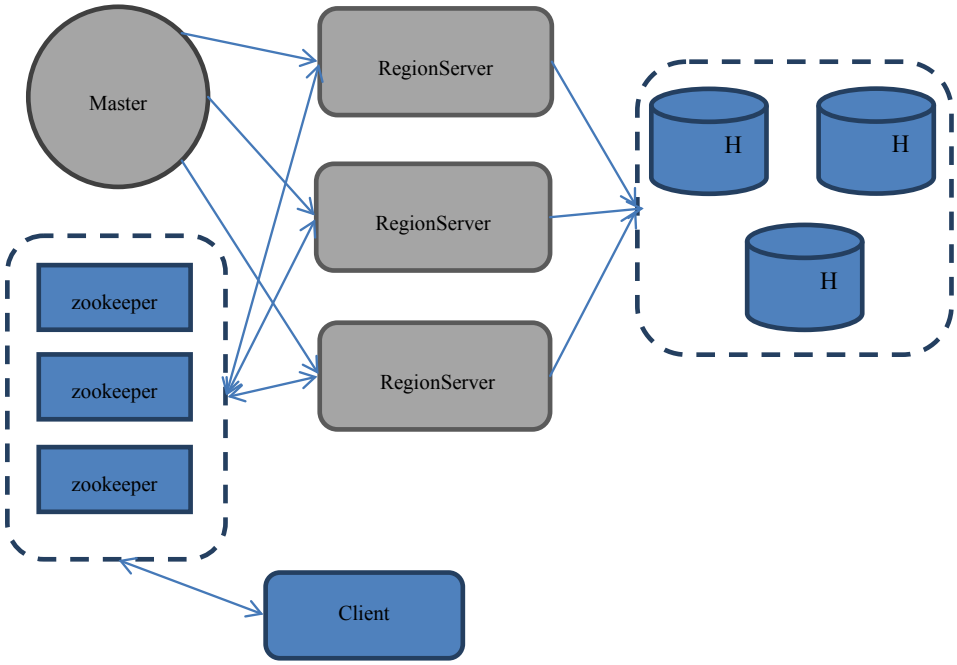

Master
RegionServer
RegionServer
RegionServer
H
H
H
zookeeper
zookeeper
zookeeper
Client

图 1-6

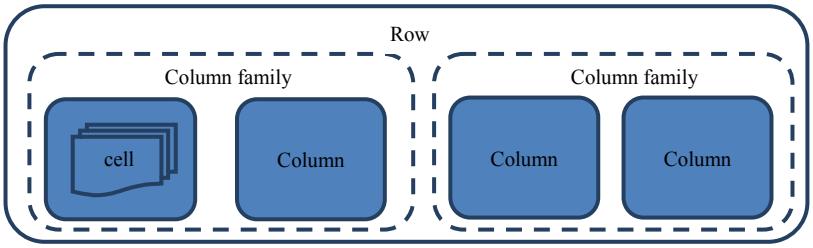

Row
Column family
Column family
cell
Column
Column
Column

图 1-7

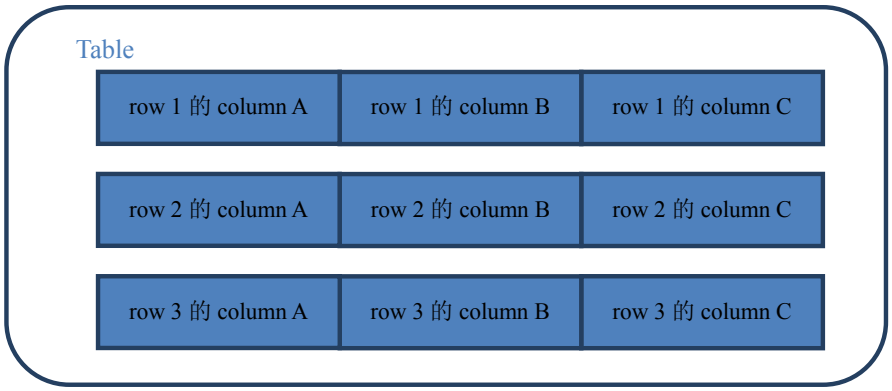


Table
row 1 的 column A
row 1 的 column B
row 1 的 column C
row 2 的 column A
row 2 的 column B
row 2 的 column C
row 3 的 column A
row 3 的 column B
row 3 的 column C

图 1-8

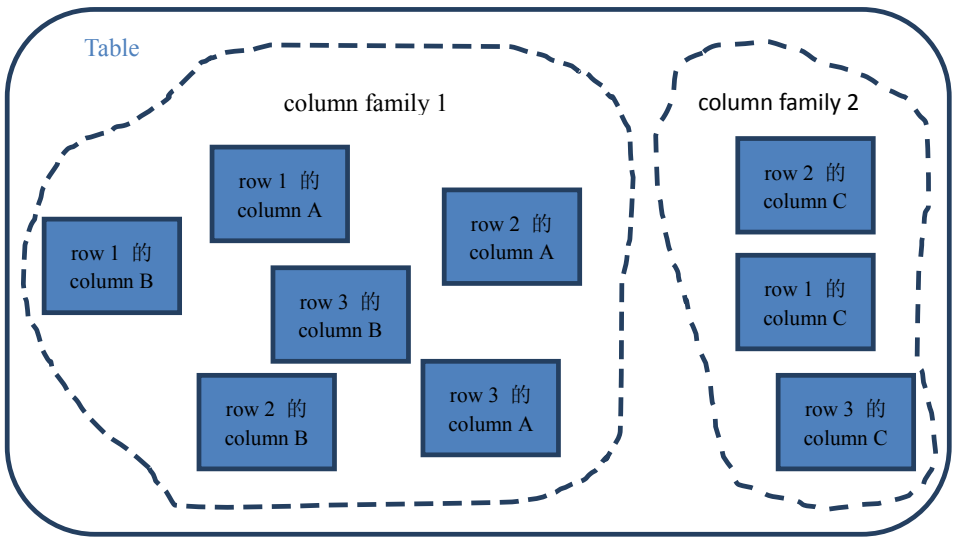
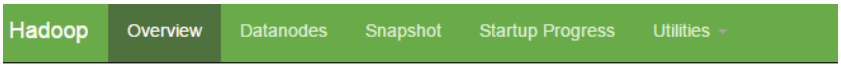


Table
column family 1
column family 2
row 1 的 column A
row 1 的 column B
row 2 的 column A
row 3 的 column B
row 2 的 column B
row 3 的 column A
row 2 的 column C
row 1 的 column C
row 3 的 column C

图 1-9

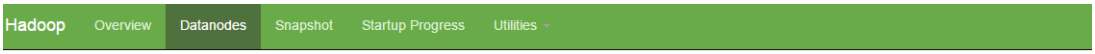
图 2-1



Overview 'nn01:8020' (active)

Started:	Wed Jul 06 08:56:10 CST 2016
Version:	2.6.4, r5082c73637530b0b7e115f9625ed7fac69f937e6
Compiled:	2016-02-12T09:45Z by jenkins from (detached from 5082c73)
Cluster ID:	CID-1fdf8209-3118-4001-bcf6-62f92980dfbc
Block Pool ID:	BP-179823345-0.0.0.50-1467766559324

图 2-3



Datanode Information

In operation

Node	Last contact	Admin State	Capacity	Used	Non DFS Used	Remaining	Blocks	Block pool used	Failed Volumes	Version
0.0.0.53 (1[redacted]78.53:50010)	0	In Service	48.39 GB	24 KB	3.03 GB	45.36 GB	0	24 KB (0%)	0	2.6.4
0.0.0.54 ([redacted]78.54:50010)	1	In Service	48.39 GB	24 KB	3.03 GB	45.36 GB	0	24 KB (0%)	0	2.6.4
0.0.0.52 ([redacted]78.52:50010)	0	In Service	48.39 GB	24 KB	3.19 GB	45.2 GB	0	24 KB (0%)	0	2.6.4

图 2-4

图 2-5

Overview 'nn02:8020' (standby)

图 2-7

图 2-8

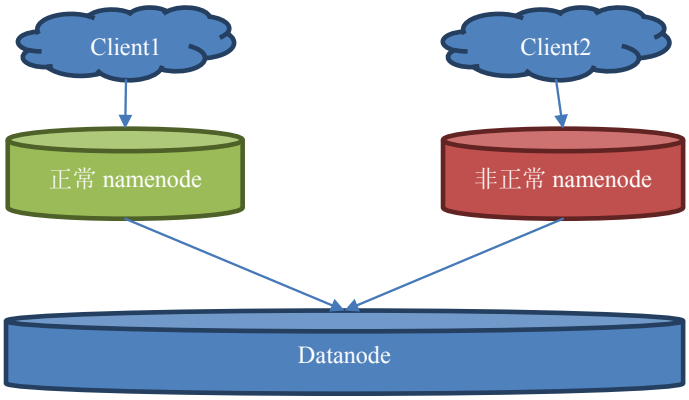


图 2-9

图 2-10

Datanode Information

In operation

Node	Last contact	Admin State	Capacity	Used	Non DFS Used
0.0.0.53 (78.53:50010)	1	In Service	48.39 GB	24 KB	3.04 GB
0.0.0.54 (.78.54:50010)	1	In Service	48.39 GB	24 KB	3.04 GB
0.0.0.52 (.78.52:50010)	1	In Service	48.39 GB	24 KB	3.2 GB

图 2-11

Name	Last modified	Size	Description
Parent Directory		-	
0.98.20/	14-Jun-2016 01:31	-	
1.1.5/	13-May-2016 20:32	-	
1.2.2/	12-Jul-2016 13:47	-	
hbase-0.94.27/	14-Oct-2015 11:10	-	
hbase-1.0.3/	08-May-2016 20:05	-	
stable/	12-Jul-2016 13:47	-	

图 2-12

```
Jan 29 15:43 bin
Jul  1 20:41 CHANGES.txt
Jan 29 15:33 conf
Jul  1 21:27 docs
Jul  1 20:49 hbase-webapps
Jul  1 21:32 LEGAL
Jul 25 02:38 lib
Jul  1 21:32 LICENSE.txt
Jul  1 21:32 NOTICE.txt
Dec 27  2015 README.txt
```

图 2-13

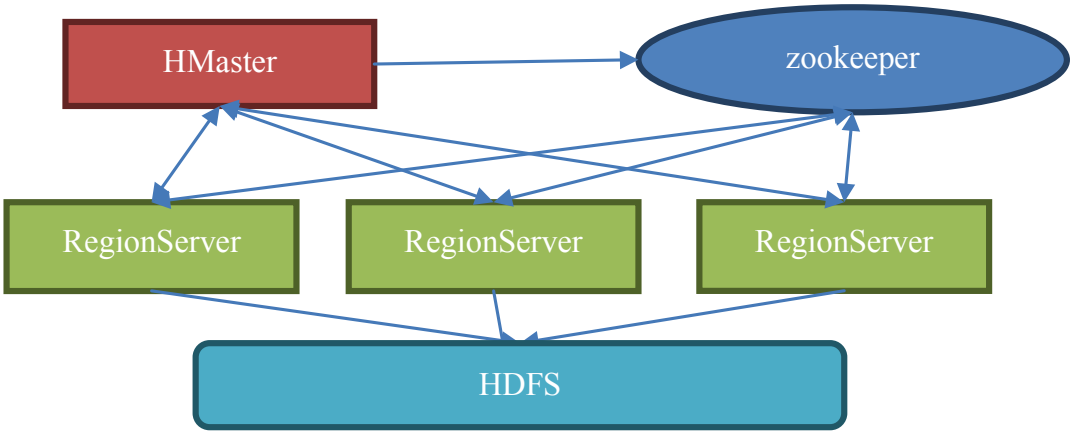


图 2-14

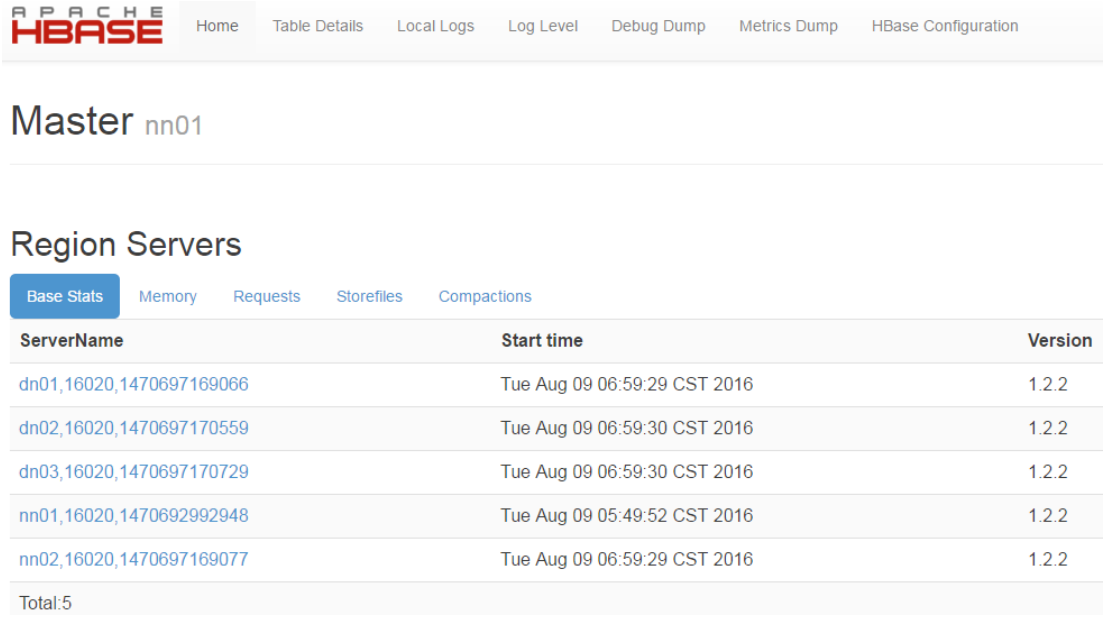


图 2-15

图 2-16

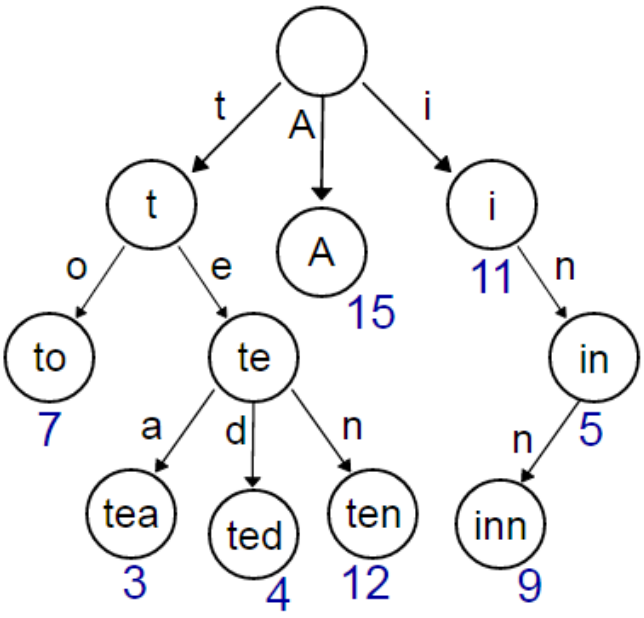

t
A
i
t
A
i
o
e
11
n
15
to
te
in
7
a
d
n
n
5
tea
ted
ten
inn
3
4
12
9

图 2-17

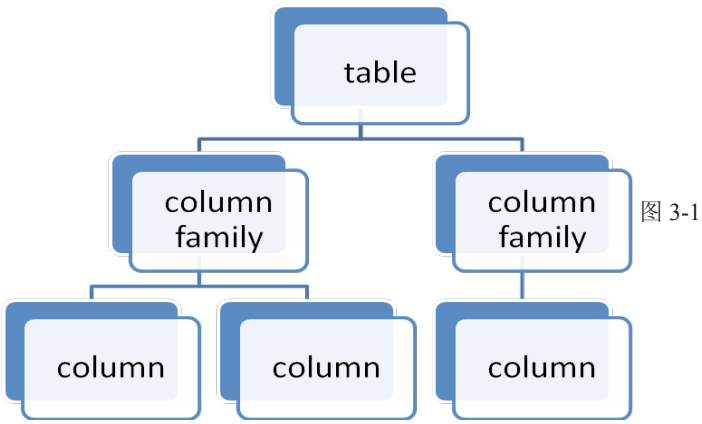


table
column family
column family
column
column
column

图 3-1

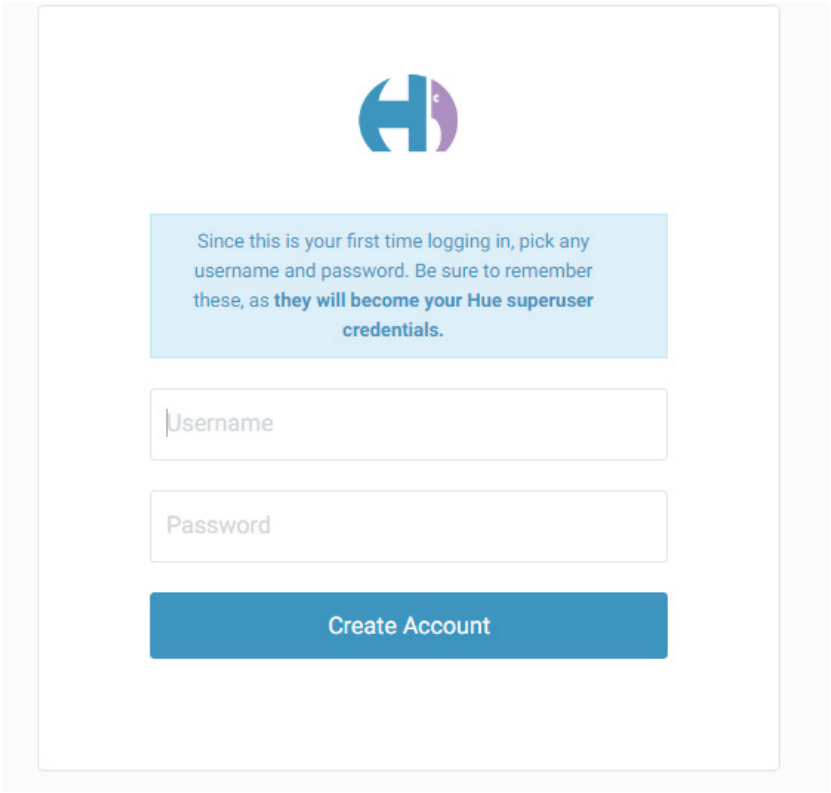

Since this is your first time logging in, pick any username and password. Be sure to remember these, as they will become your Hue superuser credentials.
Username
Password
Create Account

图 3-2

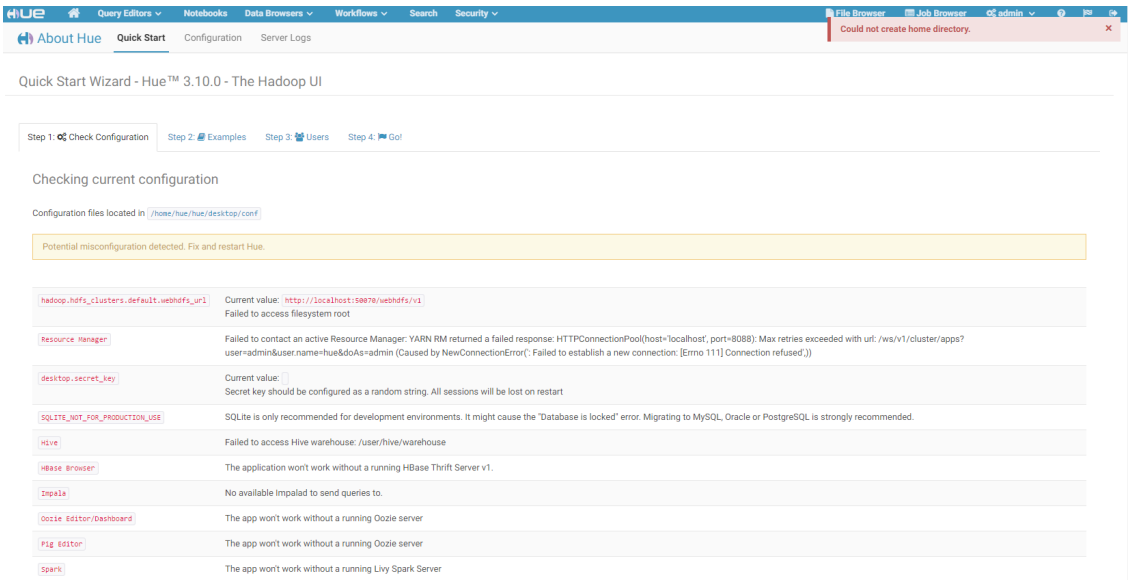

Query Editors
Notebooks
Data Browsers
Workflows
Search
Security
File Browser
Job Browser
admin
About Hue
Quick Start
Configuration
Server Logs
Could not create home directory.
Quick Start Wizard - Hue™ 3.10.0 - The Hadoop UI
Step 1: Check Configuration
Step 2: Examples
Step 3: Users
Step 4: Go!
Checking current configuration
Configuration files located in /home/hue/hue/desktop/conf
Potential misconfiguration detected. Fix and restart Hue.
hadoop.hdfs_clusters.default.webhdfs_url
Current value: http://localhost:50070/webhdfs/v1
Failed to access filesystem root
Resource Manager
Failed to contact an active Resource Manager: YARN RM returned a failed response: HTTPConnectionPool(host='localhost', port=8088): Max retries exceeded with url: /ws/v1/cluster/apps?user=admin&user.name=hue&doAs=admin (Caused by NewConnectionError(': Failed to establish a new connection: [Errno 111] Connection refused',))
desktop.secret_key
Current value:
Secret key should be configured as a random string. All sessions will be lost on restart
SQLITE_NOT_FOR_PRODUCTION_USE
SQLite is only recommended for development environments. It might cause the "Database is locked" error. Migrating to MySQL, Oracle or PostgreSQL is strongly recommended.
Hive
Failed to access Hive warehouse: /user/hive/warehouse
HBase Browser
The application won't work without a running HBase Thrift Server v1.
Impala
No available Impalad to send queries to.
Oozie Editor/Dashboard
The app won't work without a running Oozie server
Pig Editor
The app won't work without a running Oozie server
Spark
The app won't work without a running Livy Spark Server

图 3-3

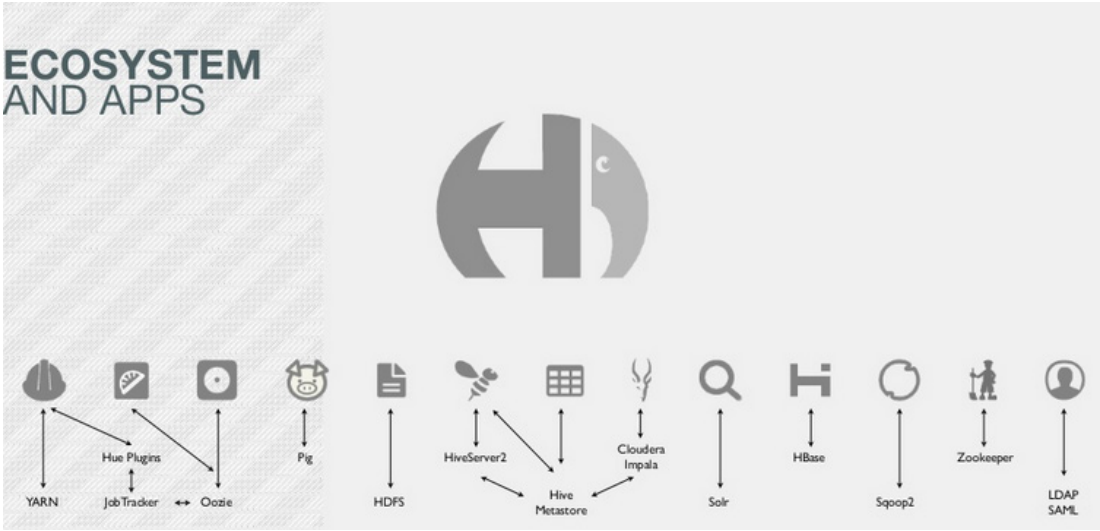

ECOSYSTEM
AND APPS
YARN
Hue Plugins
JobTracker
Oozie
Pig
HDFS
HiveServer2
Hive
Metastore
Cloudera
Impala
Solr
HBase
Sqoop2
Zookeeper
LDAP
SAML

图 3-4

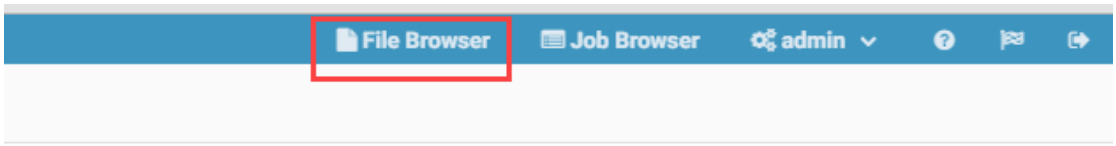

File Browser
Job Browser
admin

图 3-5

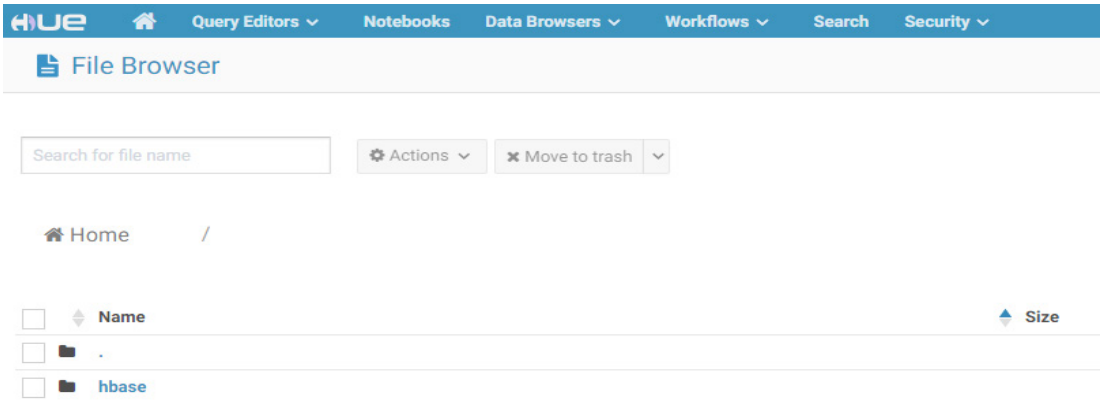

Query Editors
Notebooks
Data Browsers
Workflows
Search
Security
File Browser
Search for file name
Actions
Move to trash
Home
/
Name
Size
.
hbase

图 3-6

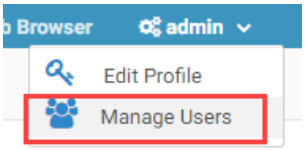

Browser
admin
Edit Profile
Manage Users

图 3-7

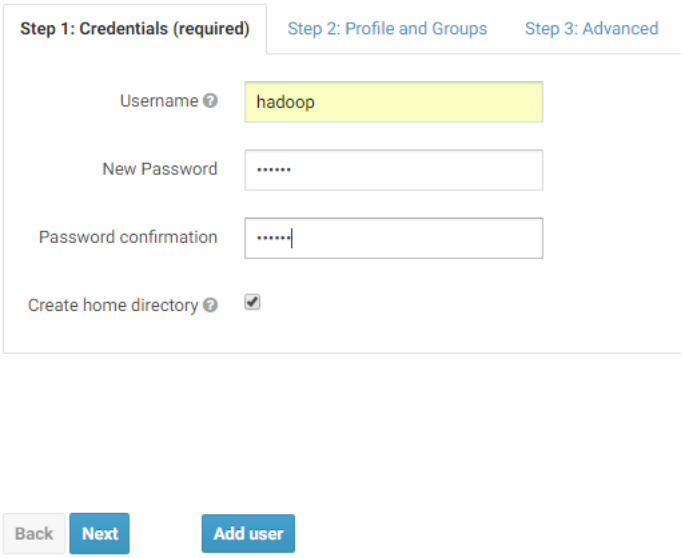

Step 1: Credentials (required)
Step 2: Profile and Groups
Step 3: Advanced
Username
hadoop
New Password
Password confirmation
Create home directory
Back
Next
Add user

图 3-8

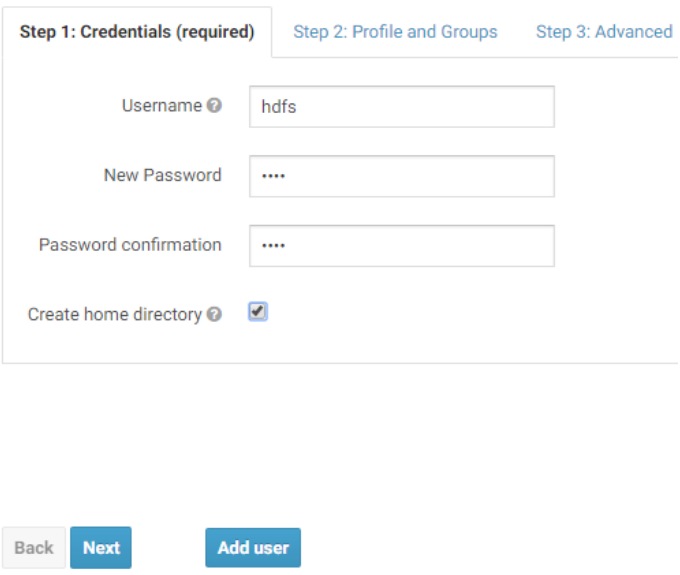

Step 1: Credentials (required)
Step 2: Profile and Groups
Step 3: Advanced
Username
hdfs
New Password
Password confirmation
Create home directory
Back
Next
Add user

图 3-9

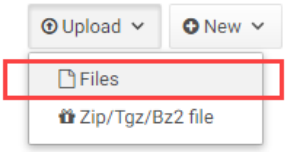


图 3-10

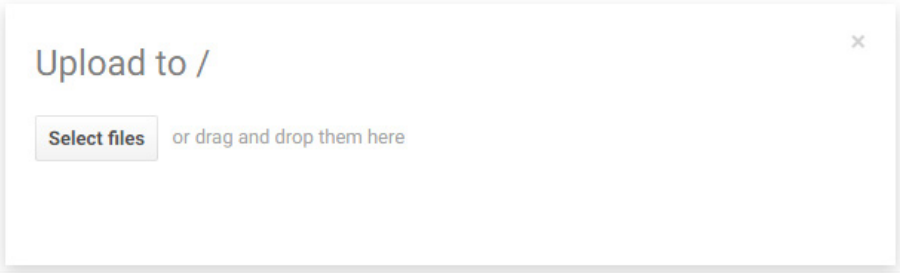


图 3-11

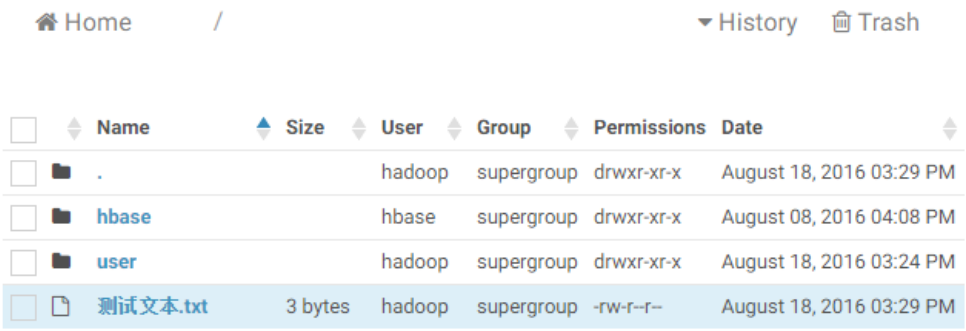


图 3-12

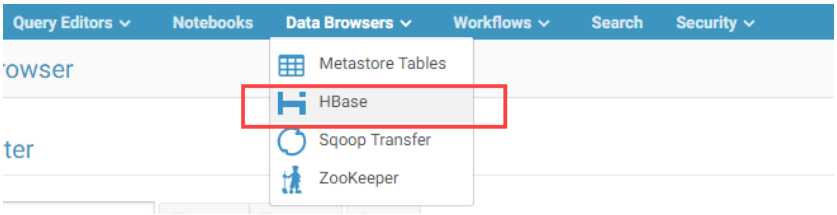


图 3-13

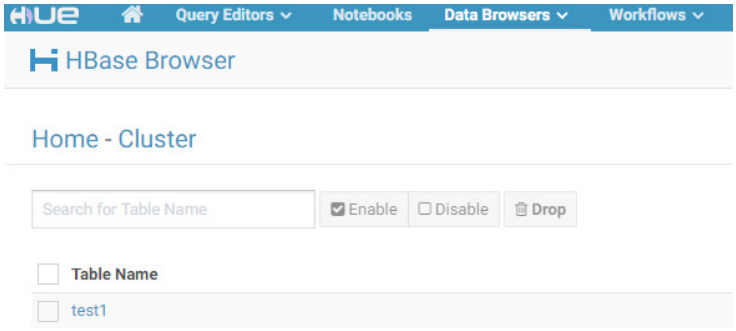

HUE
Query Editors
Notebooks
Data Browsers
Workflows
HBase Browser
Home - Cluster
Search for Table Name
Enable
Disable
Drop
Table Name
test1

图 3-14

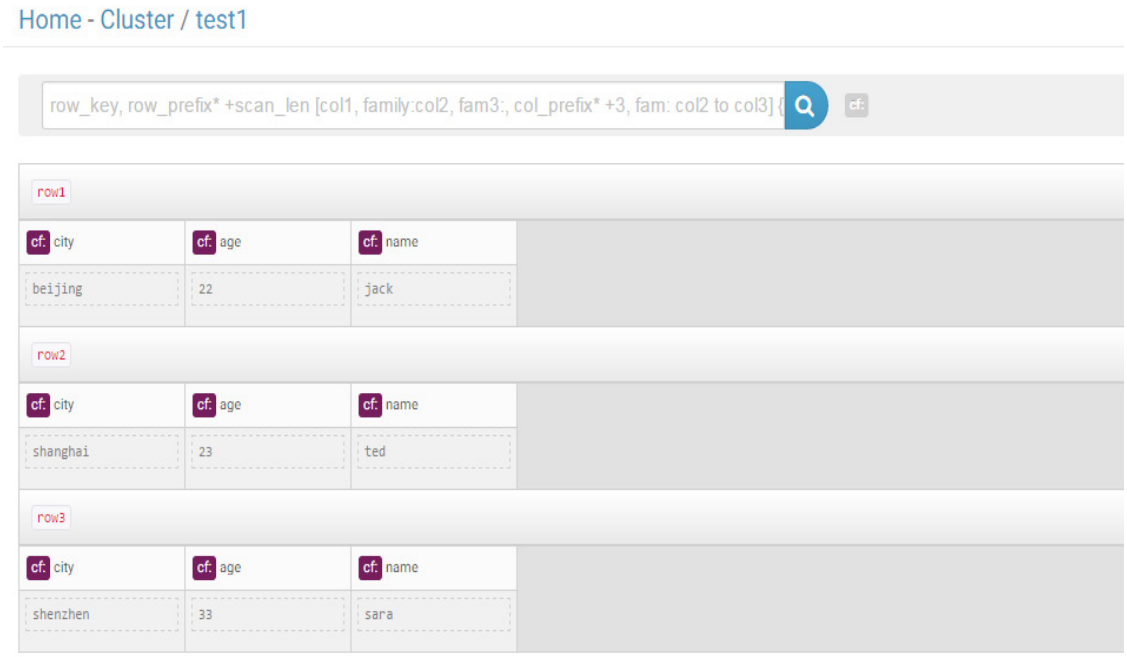

Home - Cluster / test1
row_key, row_prefix* +scan_len [col1, family:col2, fam3:, col_prefix* +3, fam: col2 to col3]
cf:
row1
cf: city
cf: age
cf: name
beijing
22
jack
row2
cf: city
cf: age
cf: name
shanghai
23
ted
row3
cf: city
cf: age
cf: name
shenzhen
33
sara

图 3-15

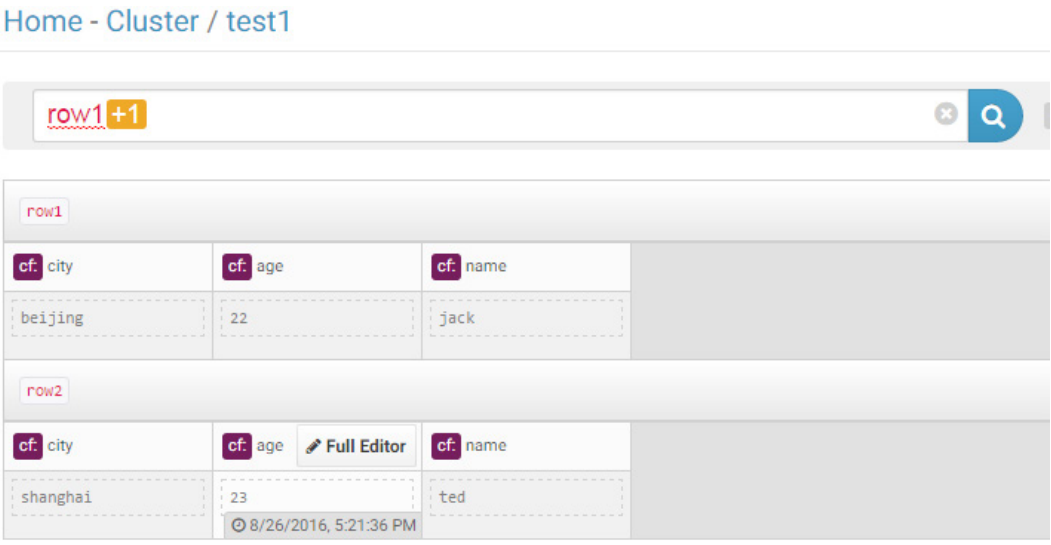

Home - Cluster / test1
row1 +1
row1
cf: city
cf: age
cf: name
beijing
22
jack
row2
cf: city
cf: age
Full Editor
cf: name
shanghai
23
ted
8/26/2016, 5:21:36 PM

图 3-16

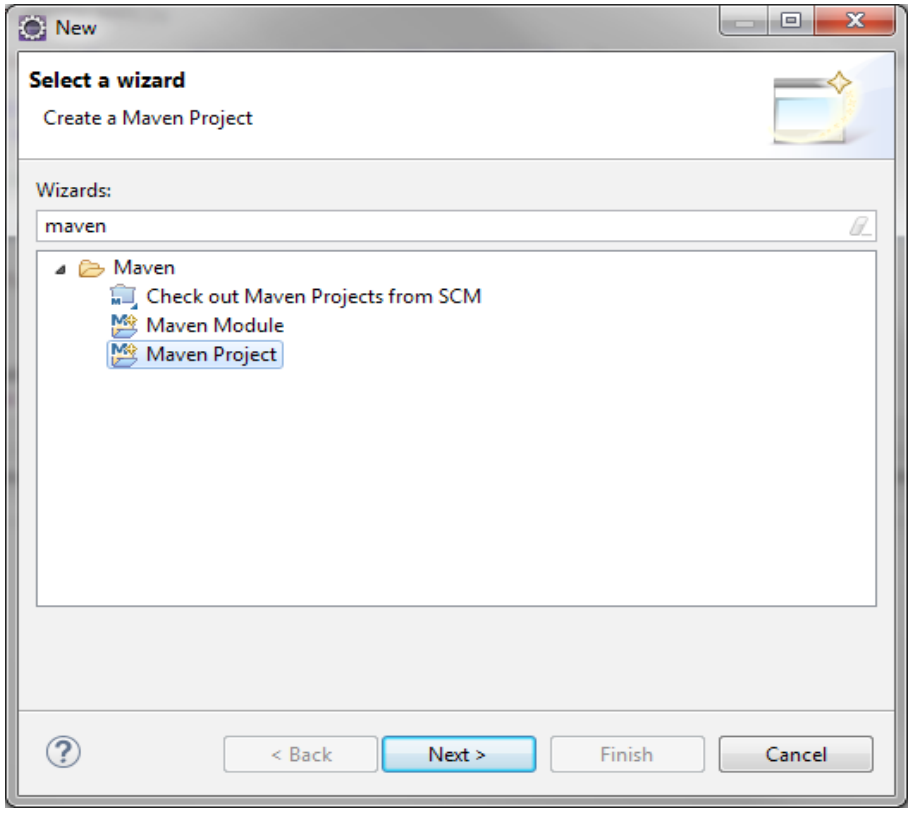

New
Select a wizard
Create a Maven Project
Wizards:
maven
Maven
Check out Maven Projects from SCM
Maven Module
Maven Project
< Back
Next >
Finish
Cancel

图 4-1

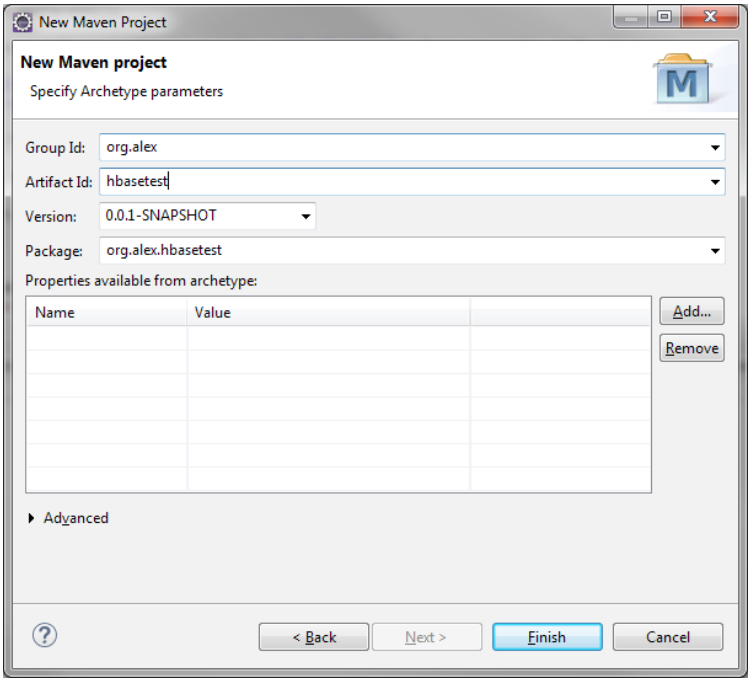

New Maven Project
New Maven project
Specify Archetype parameters
Group Id: org.alex
Artifact Id: hbasetest
Version: 0.0.1-SNAPSHOT
Package: org.alex.hbasetest
Properties available from archetype:
Name
Value
Add...
Remove
Advanced
< Back
Next >
Finish
Cancel

图 4-2

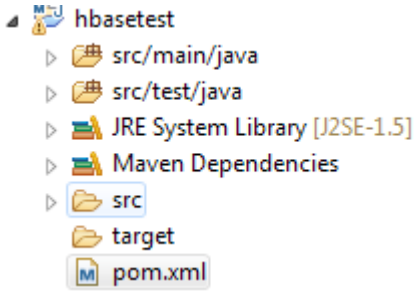

hbasetest
src/main/java
src/test/java
JRE System Library [J2SE-1.5]
Maven Dependencies
src
target
pom.xml

图 4-3

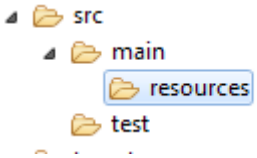

src
main
resources
test

图 4-4

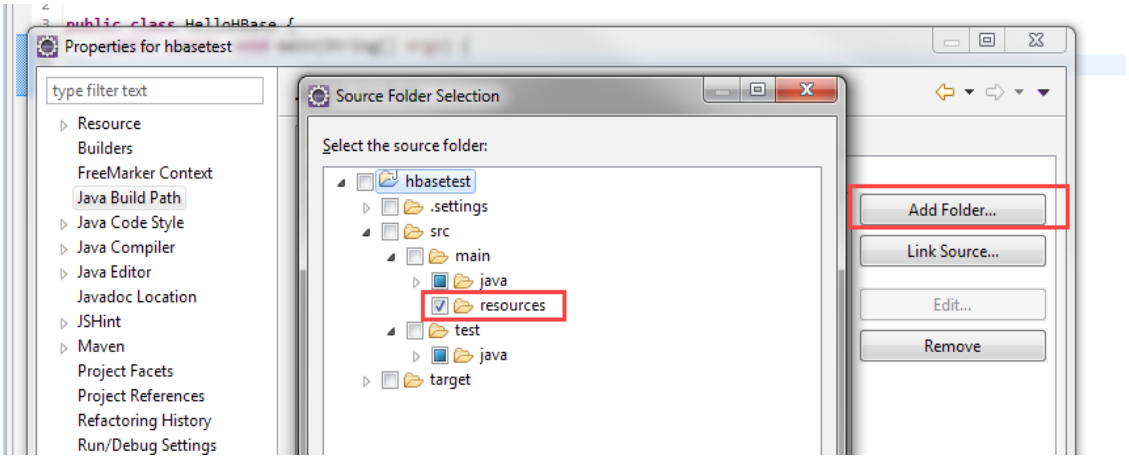

Properties for hbasetest
type filter text
Resource
Builders
FreeMarker Context
Java Build Path
Java Code Style
Java Compiler
Java Editor
Javadoc Location
JSHint
Maven
Project Facets
Project References
Refactoring History
Run/Debug Settings
Source Folder Selection
Select the source folder:
hbasetest
.settings
src
main
java
resources
test
java
target
Add Folder...
Link Source...
Edit...
Remove

图 4-5

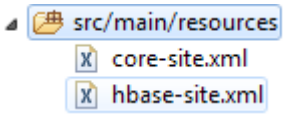

src/main/resources
core-site.xml
hbase-site.xml

图 4-6

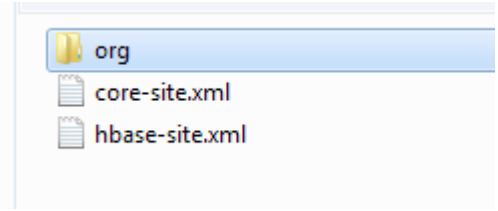

org
core-site.xml
hbase-site.xml

图 4-7

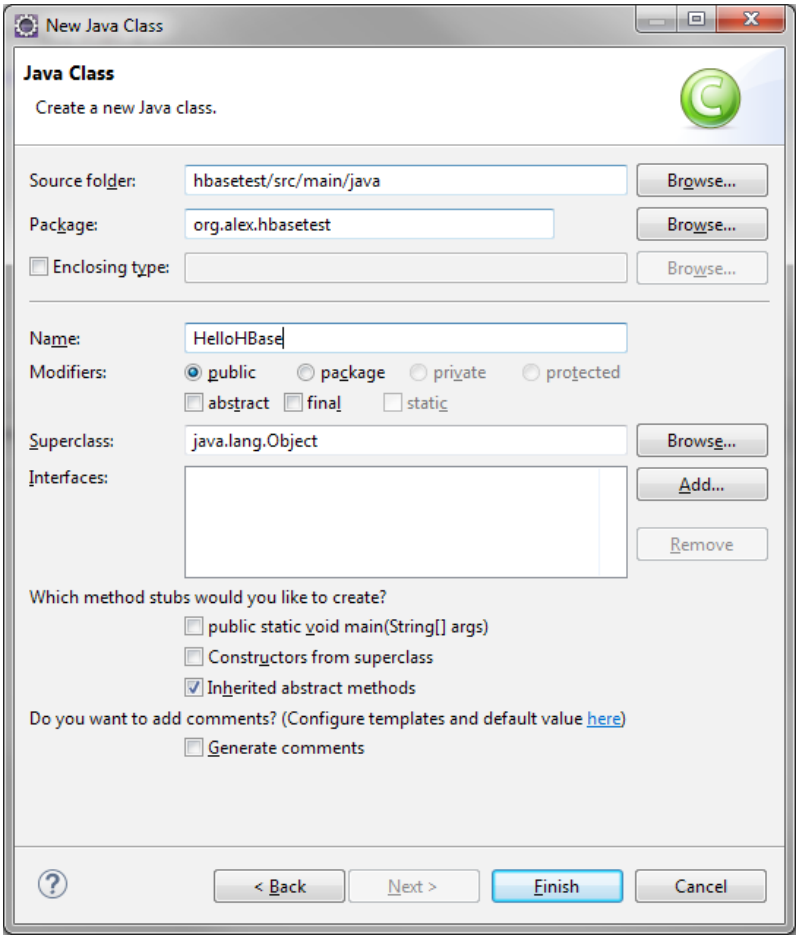

New Java Class
Java Class
Create a new Java class.
Source folder:
hbasetest/src/main/java
Browse...
Package:
org.alex.hbasetest
Browse...
Enclosing type:
Browse...
Name:
HelloHBase
Modifiers:
public
package
private
protected
abstract
final
static
Superclass:
java.lang.Object
Browse...
Interfaces:
Add...
Remove
Which method stubs would you like to create?
public static void main(String[] args)
Constructors from superclass
Inherited abstract methods
Do you want to add comments? (Configure templates and default value here)
Generate comments
< Back
Next >
Finish
Cancel

图 4-8

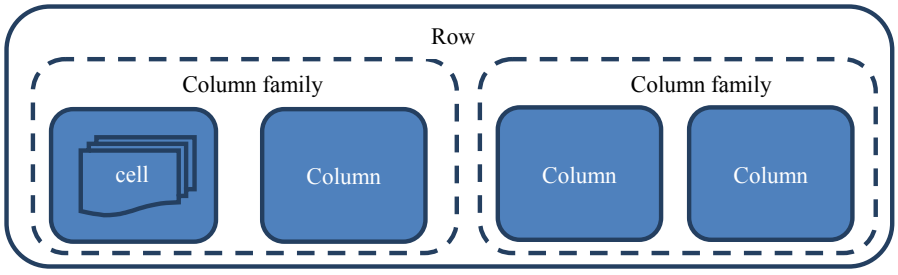

Row
Column family
cell
Column
Column family
Column
Column

图 5-1

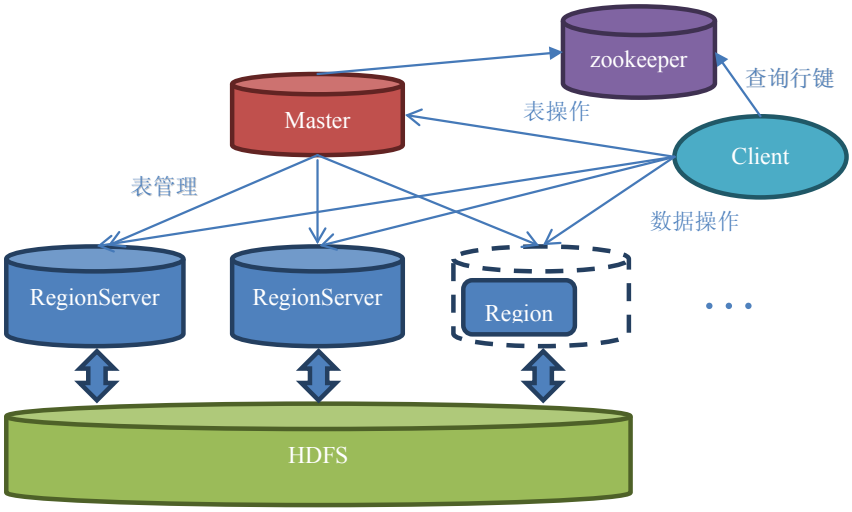

zookeeper
查询行键
Master
表操作
Client
表管理
数据操作
RegionServer
RegionServer
Region
HDFS

图 5-2

RegionServer
Region

图 5-3

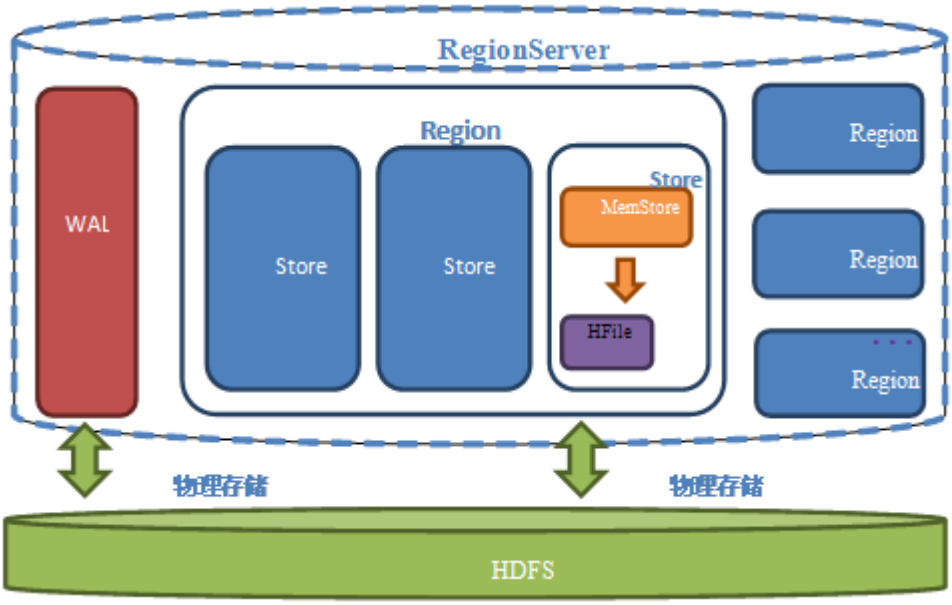

RegionServer
WAL
Region
Store
Store
Store
MemStore
HFile
Region
Region
Region
物理存储
物理存储
HDFS

图 5-4

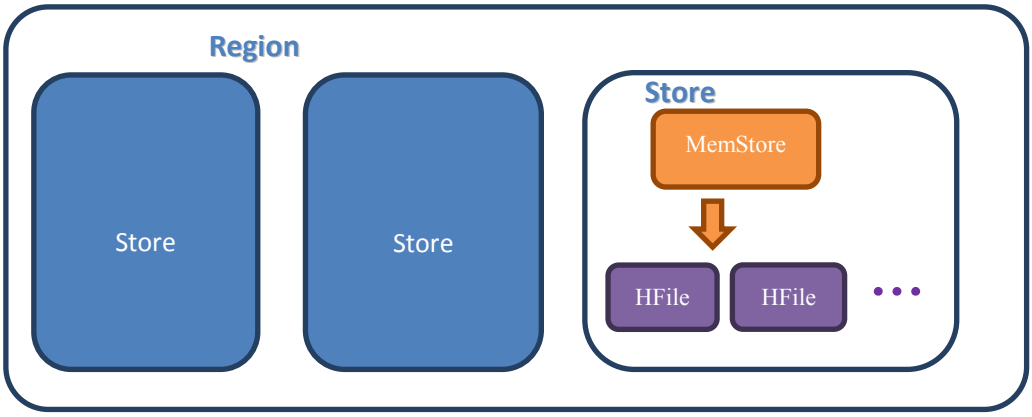


图 5-5

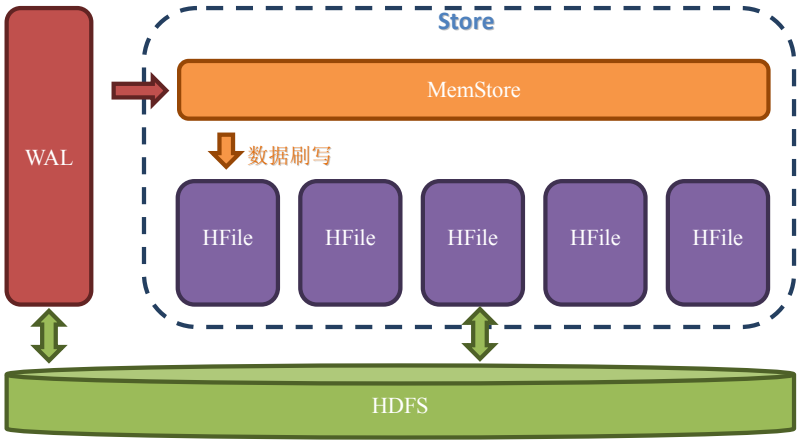


图 5-6

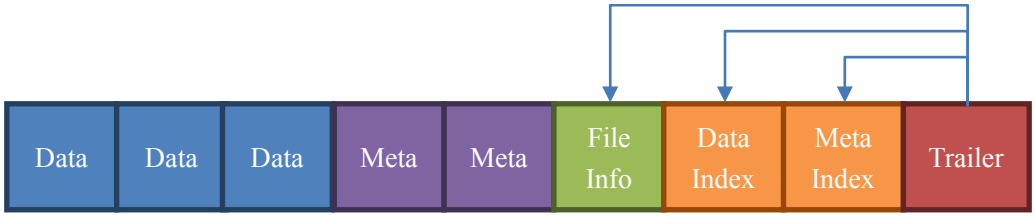


图 5-7

图 5-8

图 5-9

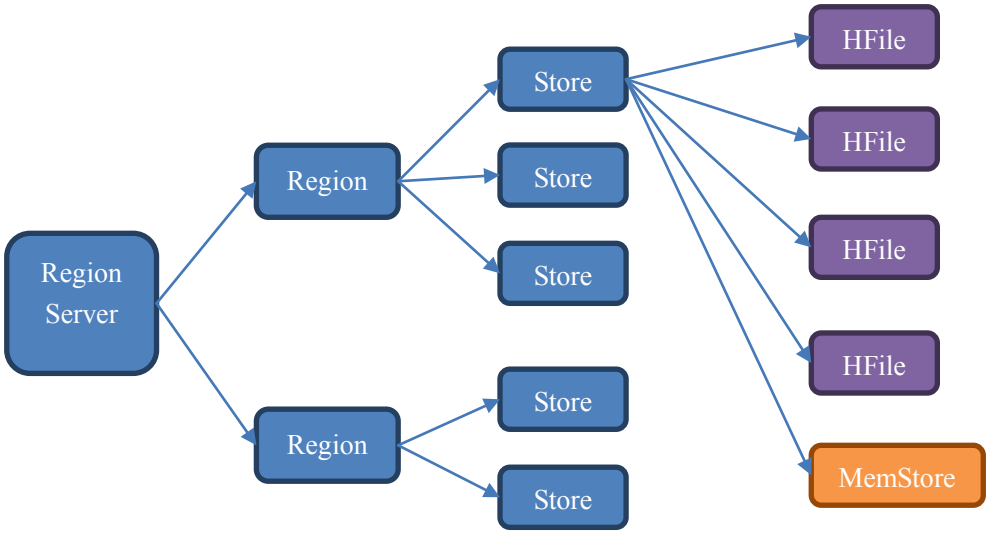


图 5-10

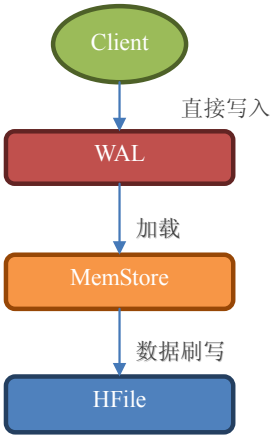


图 5-11

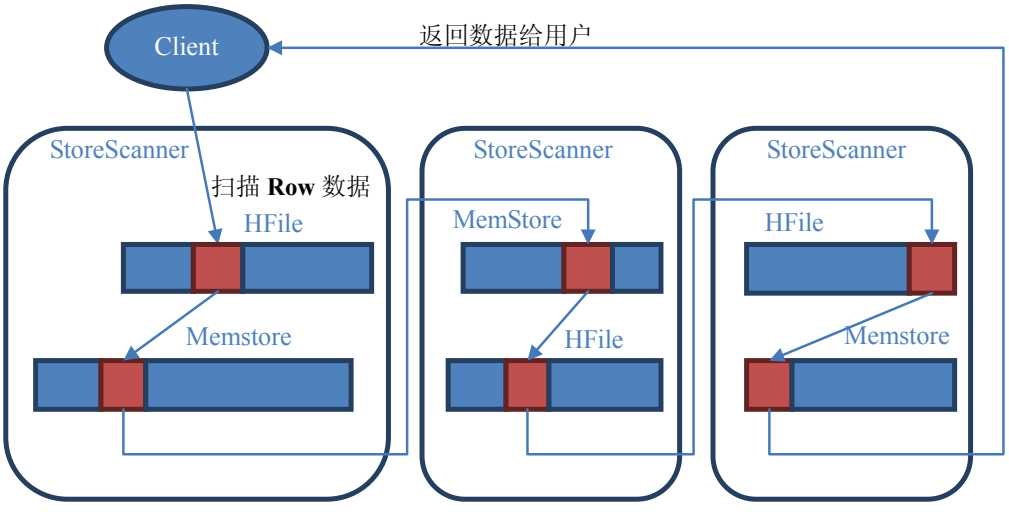


图 5-12

图 5-13

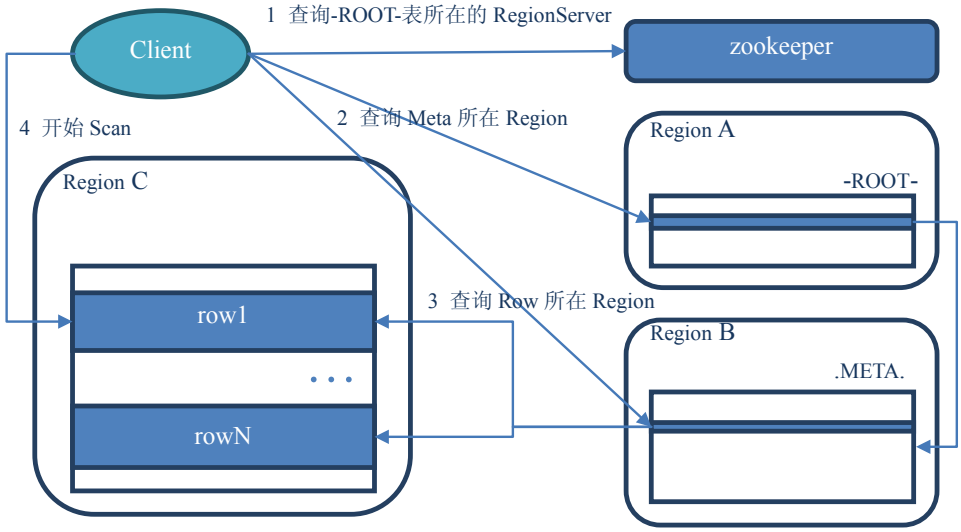


图 5-14

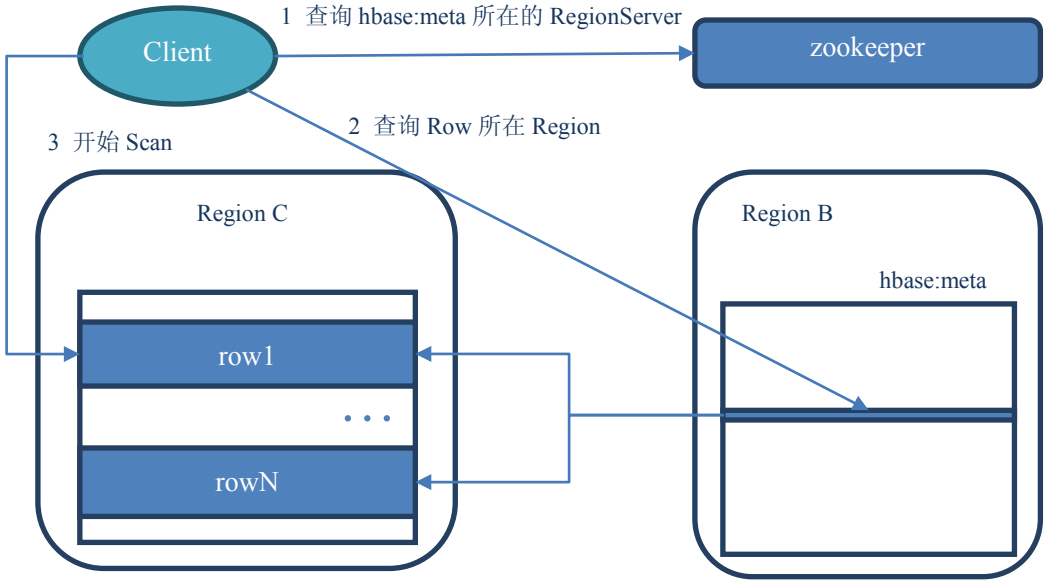


图 5-15

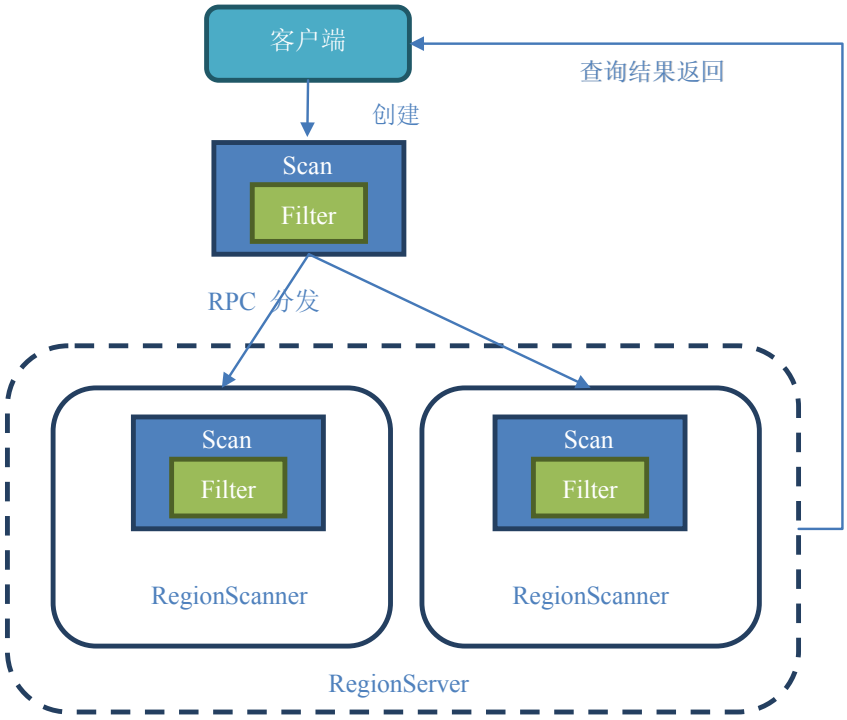

客户端
查询结果返回
创建
Scan
Filter
RPC 分发
Scan
Filter
Scan
Filter
RegionScanner
RegionScanner
RegionServer

图 6-1

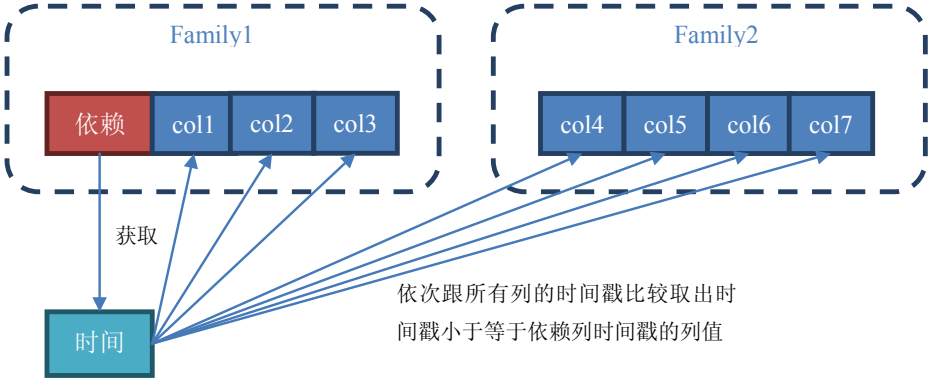

Family1
Family2
依赖
col1
col2
col3
col4
col5
col6
col7
获取
依次跟所有列的时间戳比较取出时
间戳小于等于依赖列时间戳的列值
时间

图 6-2

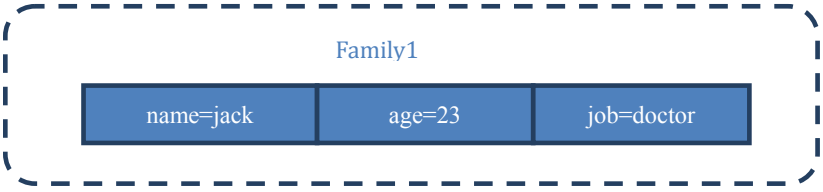

Family1
name=jack
age=23
job=doctor

图 6-3

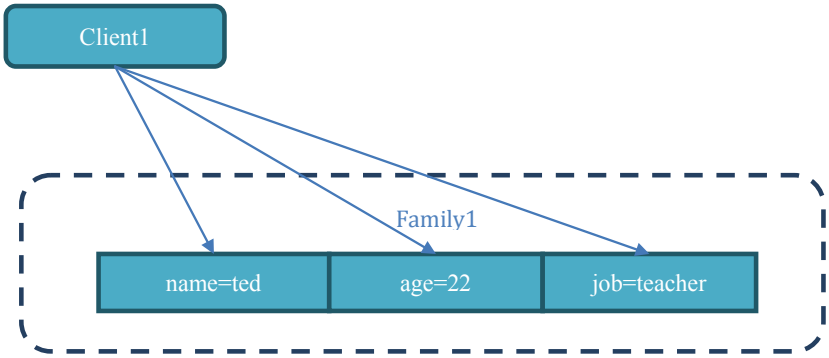


图 6-4

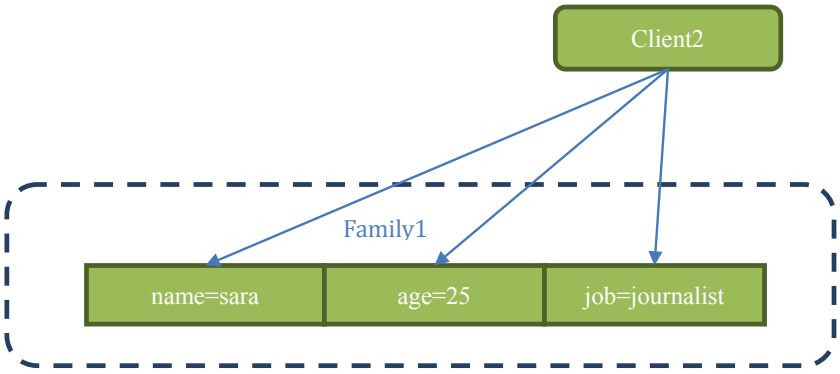


图 6-5

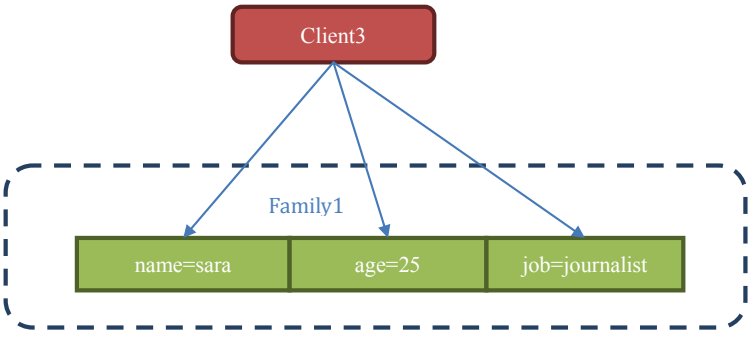


图 6-6

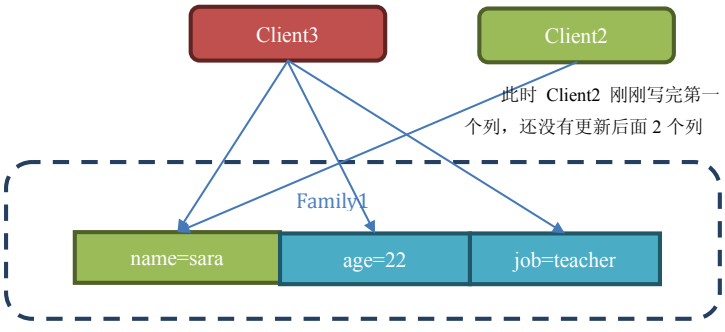


图 6-7

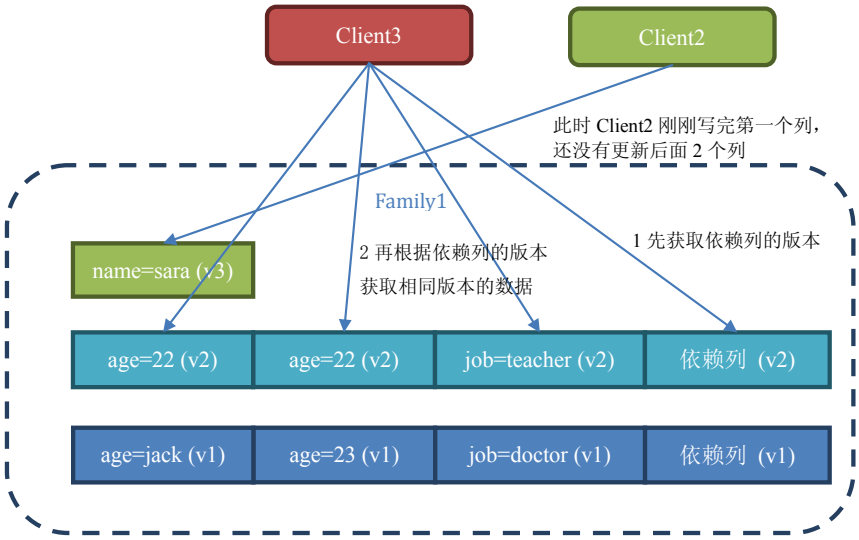

Client3
Client2
此时 Client2 刚刚写完第一个列，
还没有更新后面 2 个列
Family1
name=sara (v3)
2 再根据依赖列的版本
获取相同版本的数据
1 先获取依赖列的版本
age=22 (v2)
age=22 (v2)
job=teacher (v2)
依赖列 (v2)
age=jack (v1)
age=23 (v1)
job=doctor (v1)
依赖列 (v1)

图 6-8

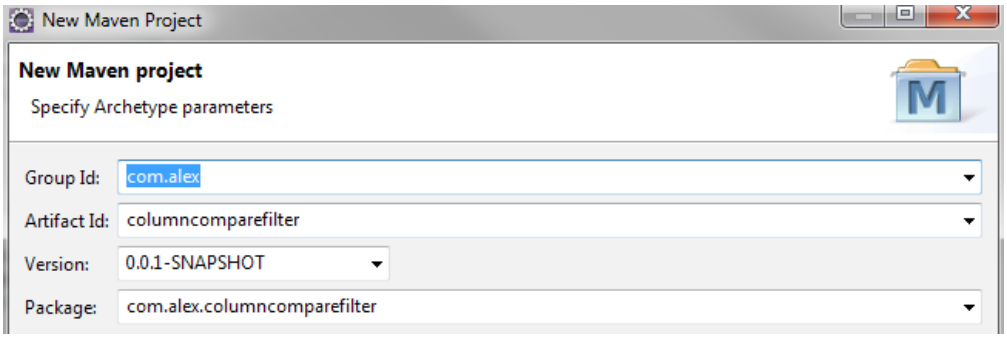

New Maven Project
New Maven project
Specify Archetype parameters
Group Id: com.alex
Artifact Id: columncomparefilter
Version: 0.0.1-SNAPSHOT
Package: com.alex.columncomparefilter

图 6-9

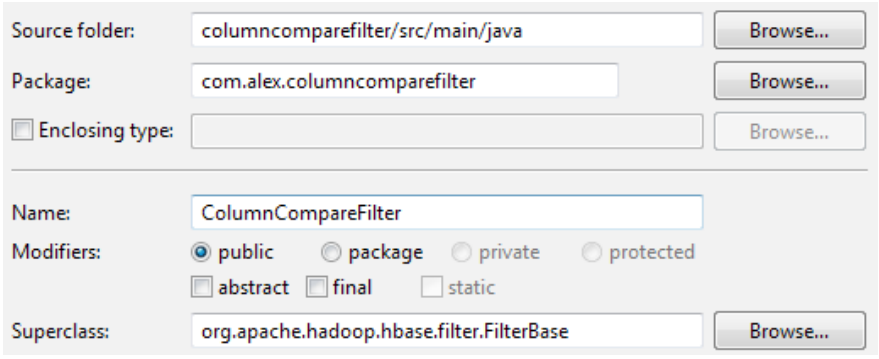

Source folder: columncomparefilter/src/main/java
Browse...
Package: com.alex.columncomparefilter
Browse...
Enclosing type:
Browse...
Name: ColumnCompareFilter
Modifiers: public package private protected
abstract final static
Superclass: org.apache.hadoop.hbase.filter.FilterBase
Browse...

图 6-10

HBase Releases

Please make sure you're downloading from a nearby mirror site, not from v

We suggest downloading the current stable release.

The 1.2.x series is the current stable release line, it supercedes 1.1.x, 1.0.x, EOL'd September 1st, 2014.

For older versions, check the apache archive.

Name	Last modified	Size	Description
Parent Directory		-	
0.98.12.1/	2015-05-21 17:42	-	
0.98.13/	2015-08-19 00:03	-	
0.98.14/	2015-08-31 19:30	-	
0.98.15/	2015-10-14 16:12	-	
0.98.16.1/	2015-11-27 05:47	-	
0.98.16/	2015-11-17 21:04	-	
0.98.17/	2016-01-25 21:26	-	
0.98.18/	2016-03-26 17:08	-	
0.98.19/	2016-05-09 01:05	-	
0.98.20/	2016-06-14 06:32	-	
0.98.21/	2016-08-16 05:11	-	
0.98.22/	2016-09-13 02:44	-	

图 6-11

Index of /dist/hbase/1.2.2

Name	Last modified	Size	Description
Parent Directory		-	
hbase-1.2.2-bin.tar.gz	2016-07-12 18:46	103M	
hbase-1.2.2-bin.tar.gz.asc	2016-07-12 18:46	819	
hbase-1.2.2-bin.tar.gz.md5	2016-07-12 18:46	73	
hbase-1.2.2-bin.tar.gz.mds	2016-07-12 18:46	1.1K	
hbase-1.2.2-bin.tar.gz.sha	2016-07-12 18:46	216	
hbase-1.2.2-src.tar.gz	2016-07-12 18:46	15M	
hbase-1.2.2-src.tar.gz.asc	2016-07-12 18:46	819	
hbase-1.2.2-src.tar.gz.md5	2016-07-12 18:46	73	
hbase-1.2.2-src.tar.gz.mds	2016-07-12 18:46	1.1K	
hbase-1.2.2-src.tar.gz.sha	2016-07-12 18:46	216	

图 6-12

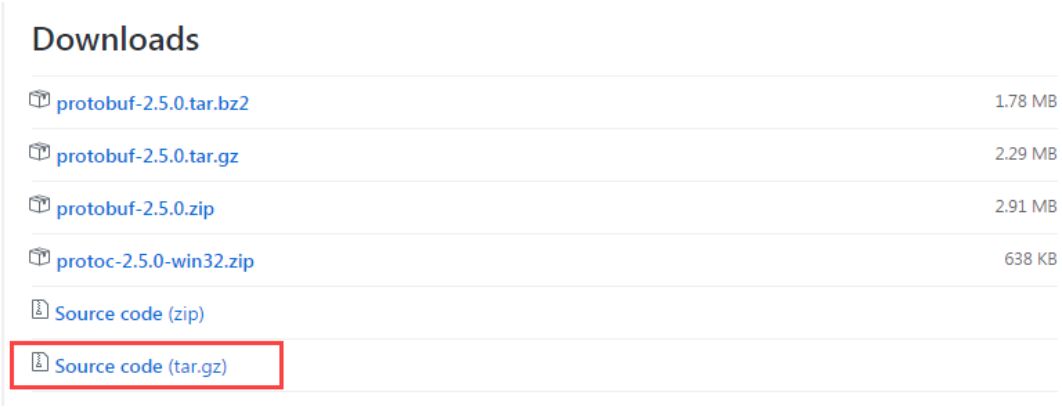

Downloads
protobuf-2.5.0.tar.bz2
1.78 MB
protobuf-2.5.0.tar.gz
2.29 MB
protobuf-2.5.0.zip
2.91 MB
protoc-2.5.0-win32.zip
638 KB
Source code (zip)
Source code (tar.gz)

图 6-13

on Apr 18, 2011
release-1.6.0 ...
800f542 zip tar.gz
on Apr 15, 2010
release-1.5.0 ...
c9d6f5e zip tar.gz
on Oct 2, 2009
release-1.4.0 ...
b4d333f zip tar.gz

图 6-14

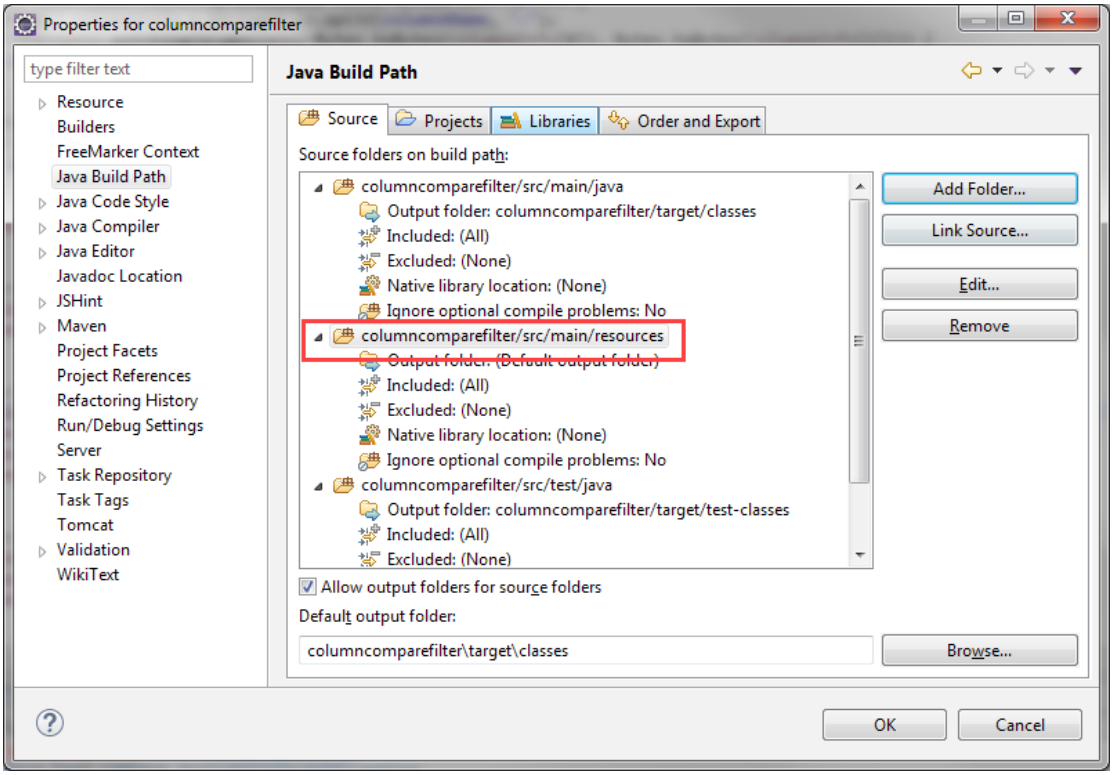


图 6-15

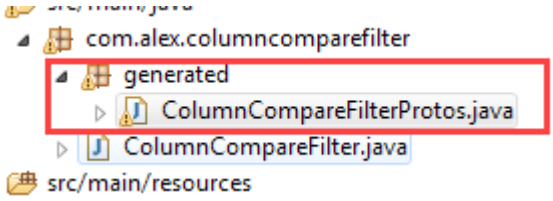


图 6-16

图 6-17

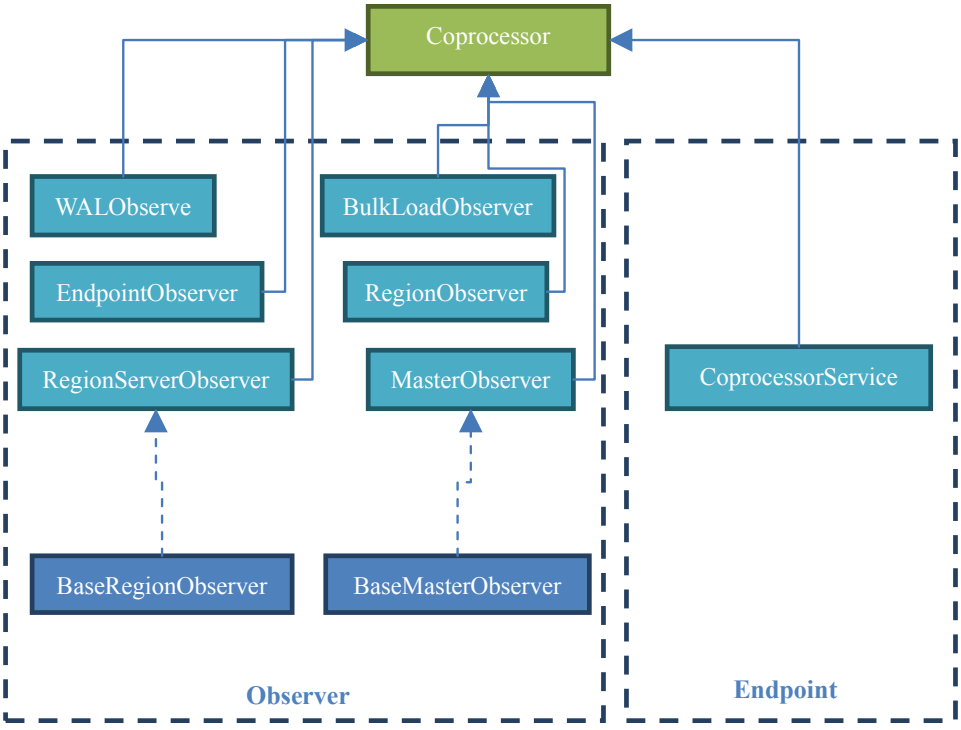

Coprocessor
WALObserve
BulkLoadObserver
EndpointObserver
RegionObserver
RegionServerObserver
MasterObserver
CoprocessorService
BaseRegionObserver
BaseMasterObserver
Observer
Endpoint

图 6-18

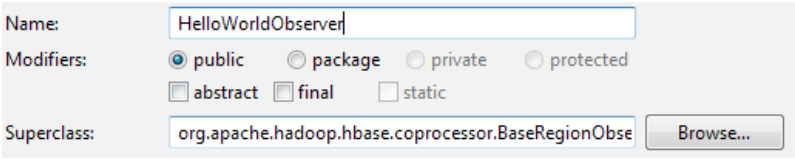

Name:
HelloWorldObserver
Modifiers:
public
package
private
protected
abstract
final
static
Superclass:
org.apache.hadoop.hbase.coprocessor.BaseRegionObse
Browse...

图 6-19

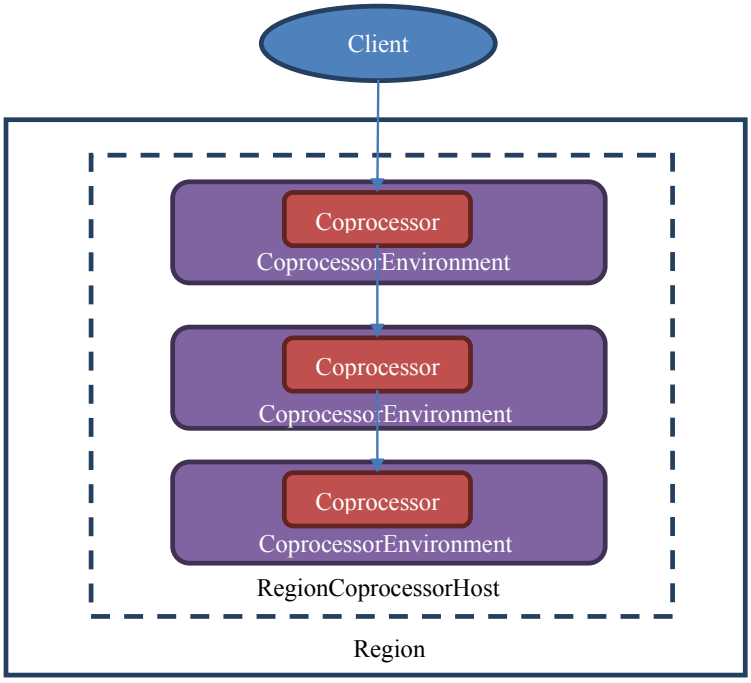

Client
Coprocessor
CoprocessorEnvironment
Coprocessor
CoprocessorEnvironment
Coprocessor
CoprocessorEnvironment
RegionCoprocessorHost
Region

图 6-20

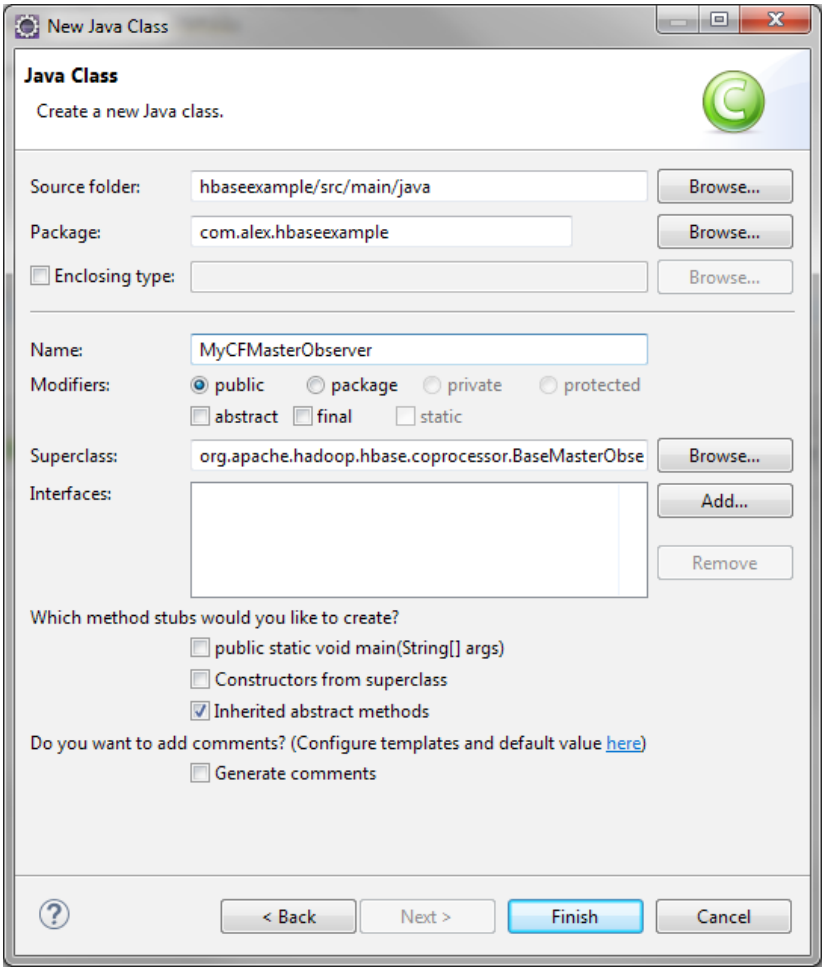

New Java Class
Java Class
Create a new Java class.
Source folder:
hbaseexample/src/main/java
Browse...
Package:
com.alex.hbaseexample
Browse...
Enclosing type:
Browse...
Name:
MyCFMasterObserver
Modifiers:
public
package
private
protected
abstract
final
static
Superclass:
org.apache.hadoop.hbase.coprocessor.BaseMasterObse
Browse...
Interfaces:
Add...
Remove
Which method stubs would you like to create?
public static void main(String[] args)
Constructors from superclass
Inherited abstract methods
Do you want to add comments? (Configure templates and default value here)
Generate comments
< Back
Next >
Finish
Cancel

图 6-21

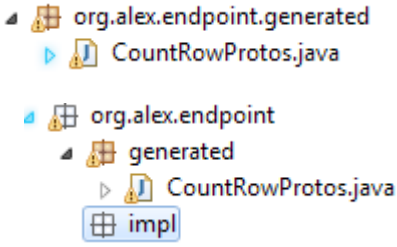

org.alex.endpoint.generated
CountRowProtos.java
org.alex.endpoint
generated
CountRowProtos.java
impl

图 6-22

。

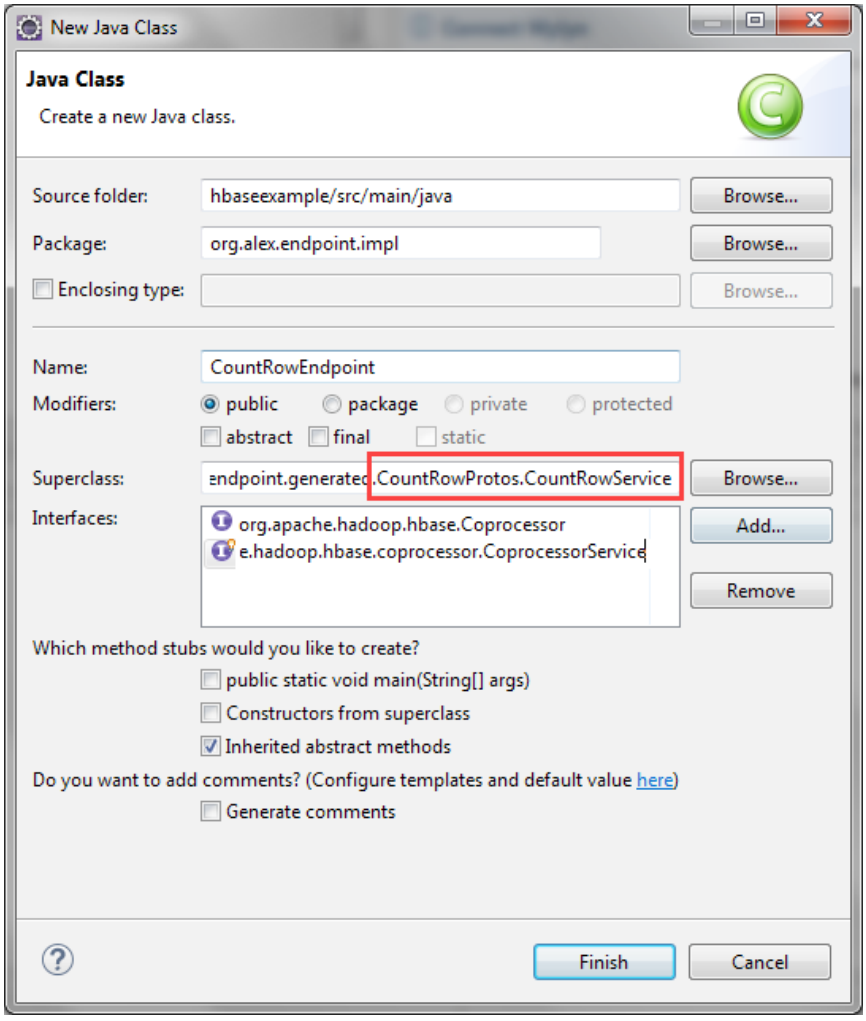


图 6-23

```
Util.cloneRow(cell);
            count++;
```

Region Servers

Base Stats | Memory | Requests | Storefiles | Compactions

ServerName	Start time
dn01,16020,1491937622761	Wed Apr 12 03:07:02 CST 2017
dn02,16020,1491937622195	Wed Apr 12 03:07:02 CST 2017
dn03,16020,1491937622779	Wed Apr 12 03:07:02 CST 2017
nn01,16020,1491937622739	Wed Apr 12 03:07:02 CST 2017
nn02,16020,1491937622190	Wed Apr 12 03:07:02 CST 2017

Total:5

图 6-24

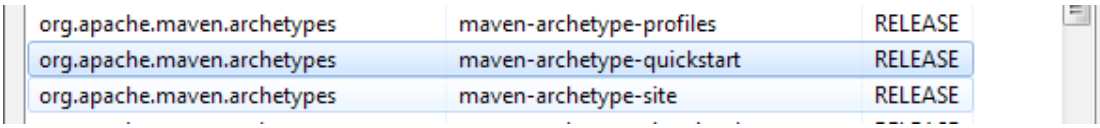


图 6-25

项目名叫 callendpoint。如图 6-26 所示。

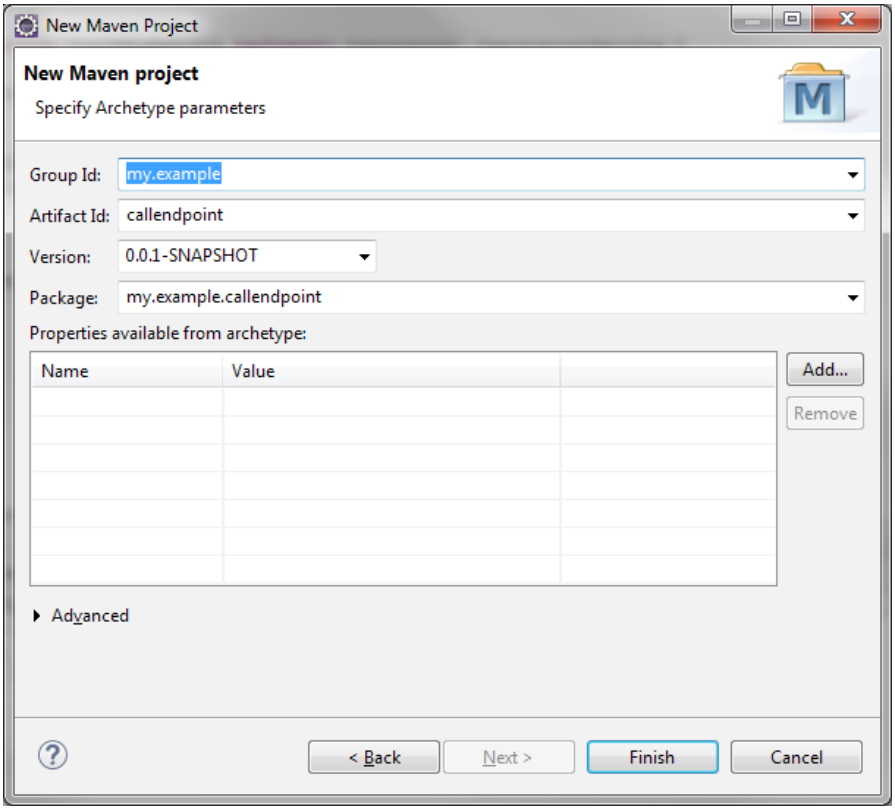


图 6-26

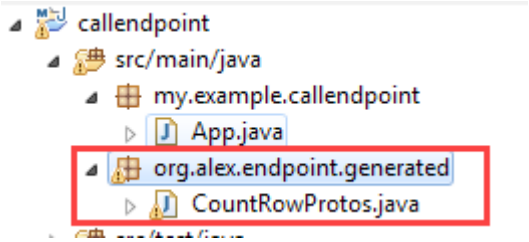


图 6-27

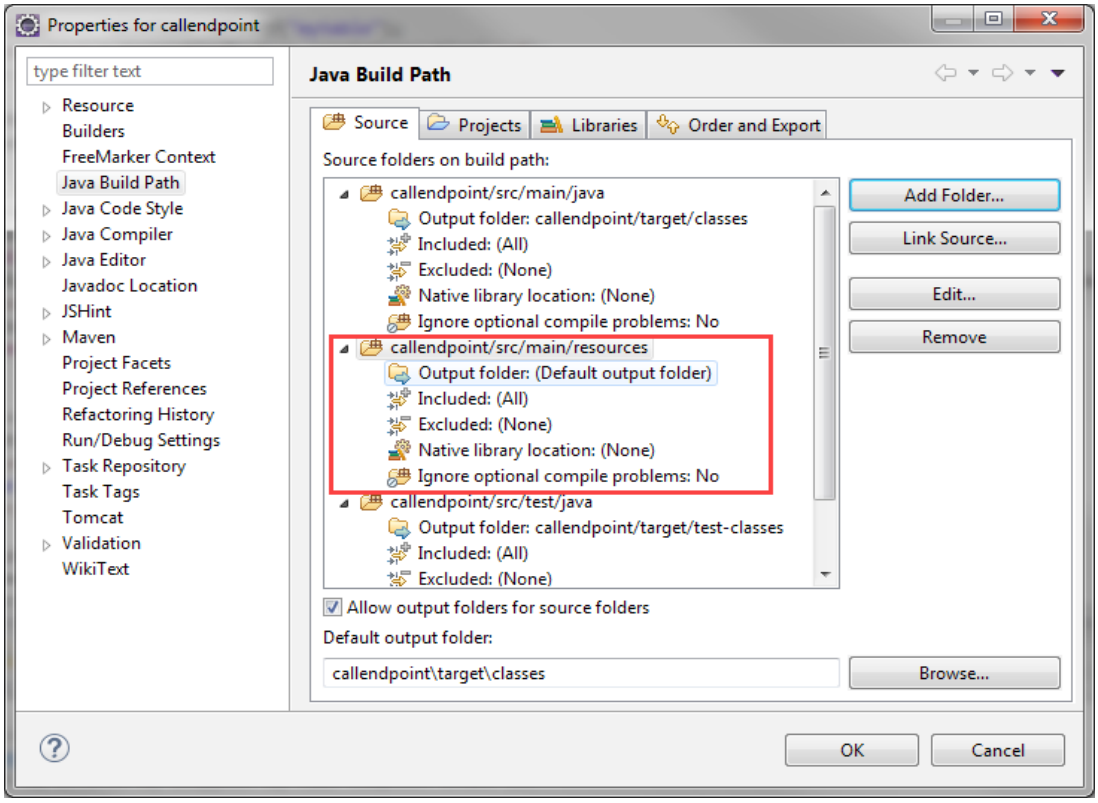

Properties for callendpoint
type filter text
Resource
Builders
FreeMarker Context
Java Build Path
Java Code Style
Java Compiler
Java Editor
Javadoc Location
JSHint
Maven
Project Facets
Project References
Refactoring History
Run/Debug Settings
Task Repository
Task Tags
Tomcat
Validation
WikiText
Java Build Path
Source
Projects
Libraries
Order and Export
Source folders on build path:
callendpoint/src/main/java
Output folder: callendpoint/target/classes
Included: (All)
Excluded: (None)
Native library location: (None)
Ignore optional compile problems: No
callendpoint/src/main/resources
Output folder: (Default output folder)
Included: (All)
Excluded: (None)
Native library location: (None)
Ignore optional compile problems: No
callendpoint/src/test/java
Output folder: callendpoint/target/test-classes
Included: (All)
Excluded: (None)
Add Folder...
Link Source...
Edit...
Remove
Allow output folders for source folders
Default output folder:
callendpoint\target\classes
Browse...
OK
Cancel

图 6-28

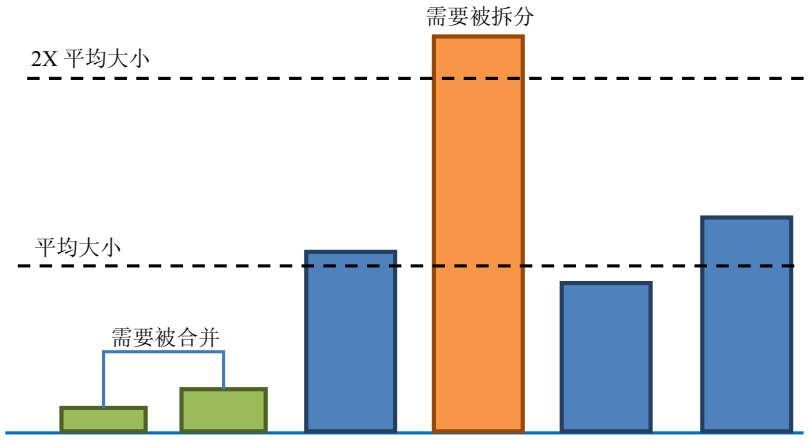

需要被拆分
2X 平均大小
平均大小
需要被合并

图 7-3

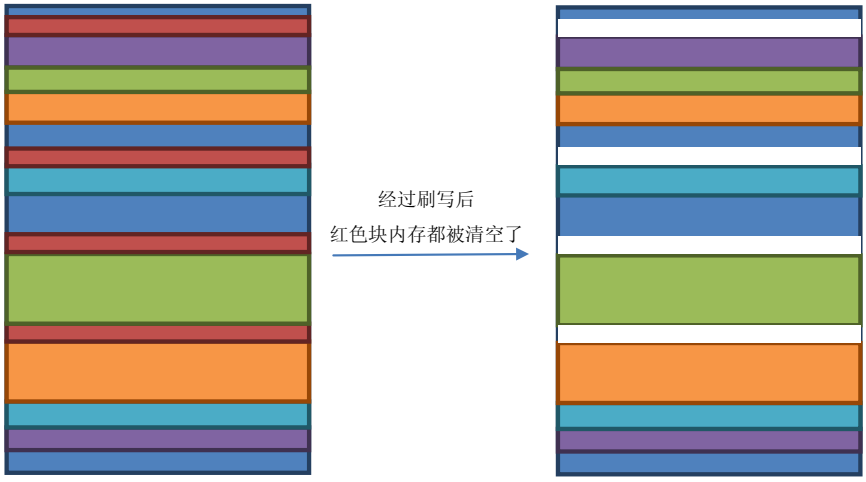


图 8-1

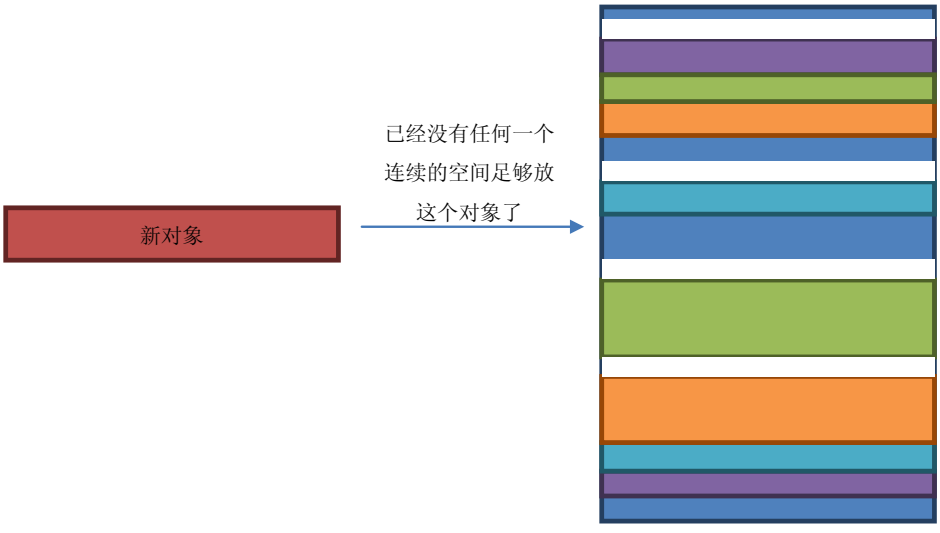


图 8-2

图 8-3

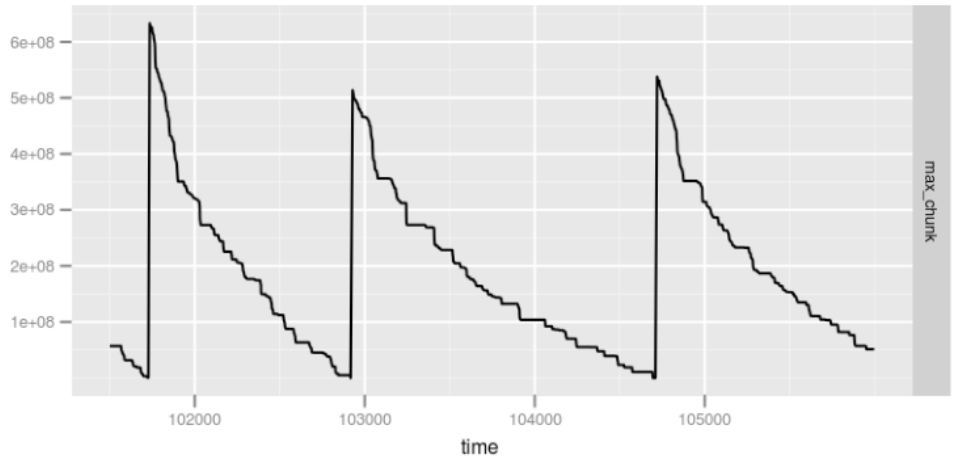

6e+08
5e+08
4e+08
3e+08
2e+08
1e+08
102000
103000
104000
105000
time
max_chunk

图 8-4

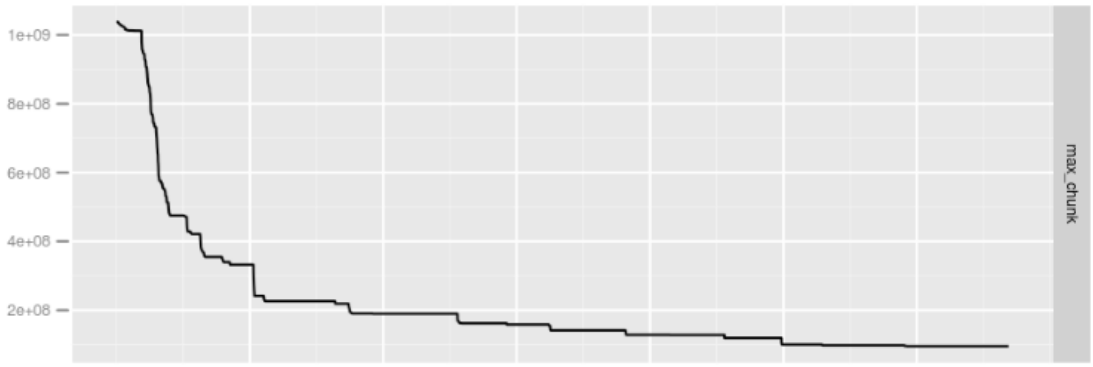

1e+09
8e+08
6e+08
4e+08
2e+08
max_chunk

图 8-5

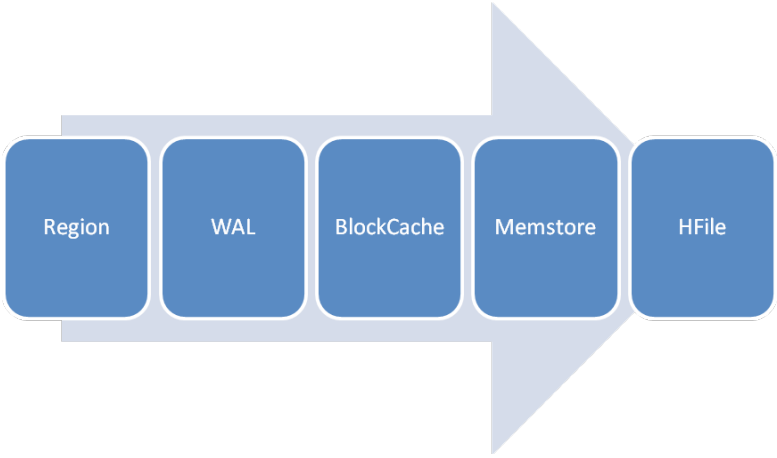

Region
WAL
BlockCache
Memstore
HFile

图 8-6

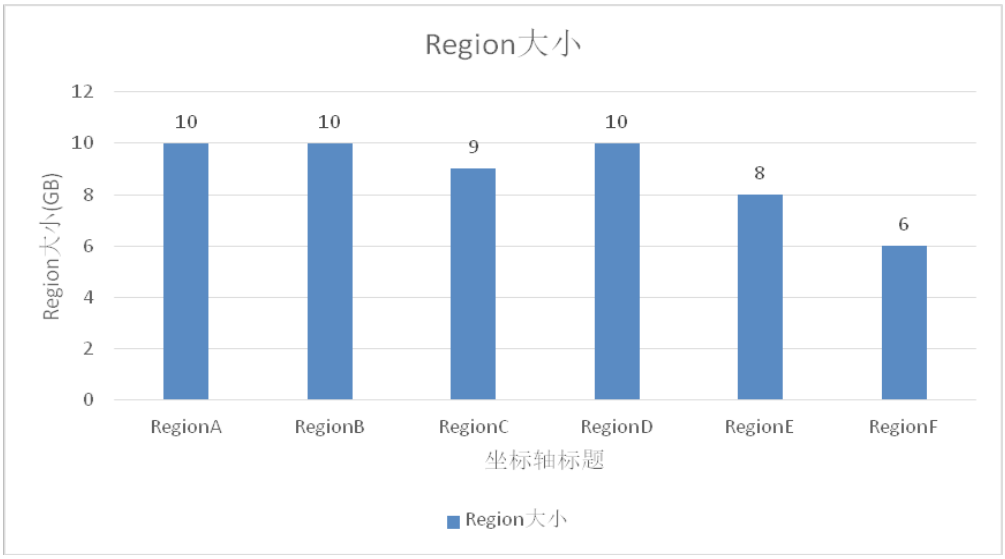

Region大小
12
10
8
6
4
2
0
Region大小(GB)
10
10
9
10
8
6
RegionA
RegionB
RegionC
RegionD
RegionE
RegionF
坐标轴标题
Region大小

图 8-7

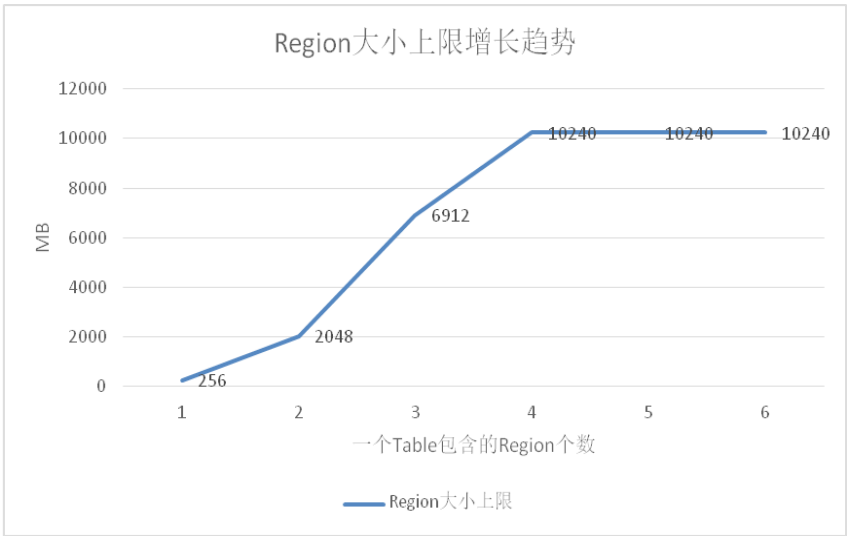

Region大小上限增长趋势
12000
10000
8000
6000
4000
2000
0
MB
256
2048
6912
10240
10240
10240
1
2
3
4
5
6
一个Table包含的Region个数
Region大小上限

图 8-8

用默认策略拆分的结果

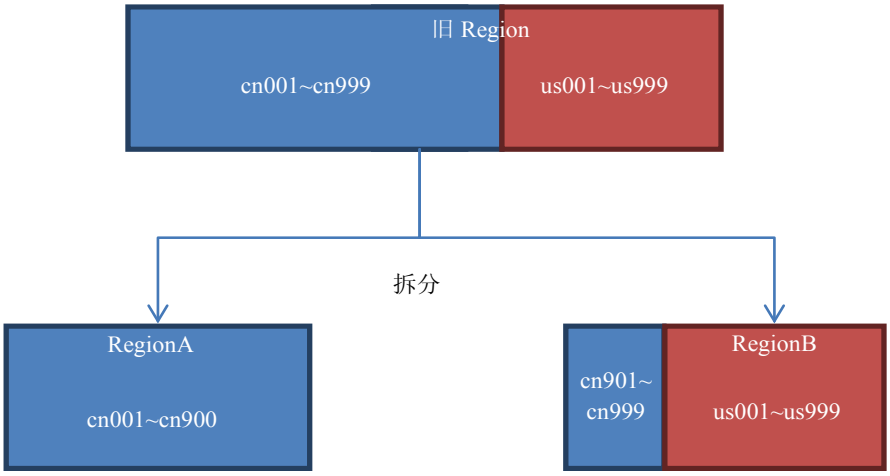


图 8-9

用默认策略拆分的情况

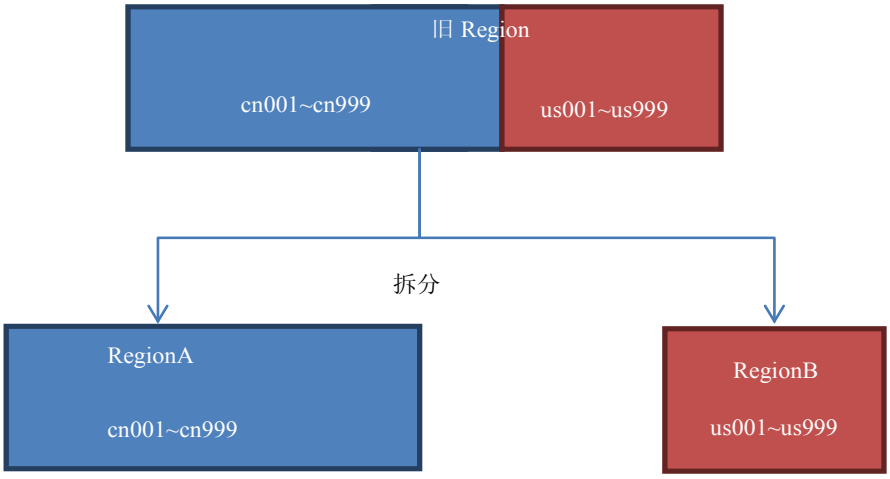


图 8-10

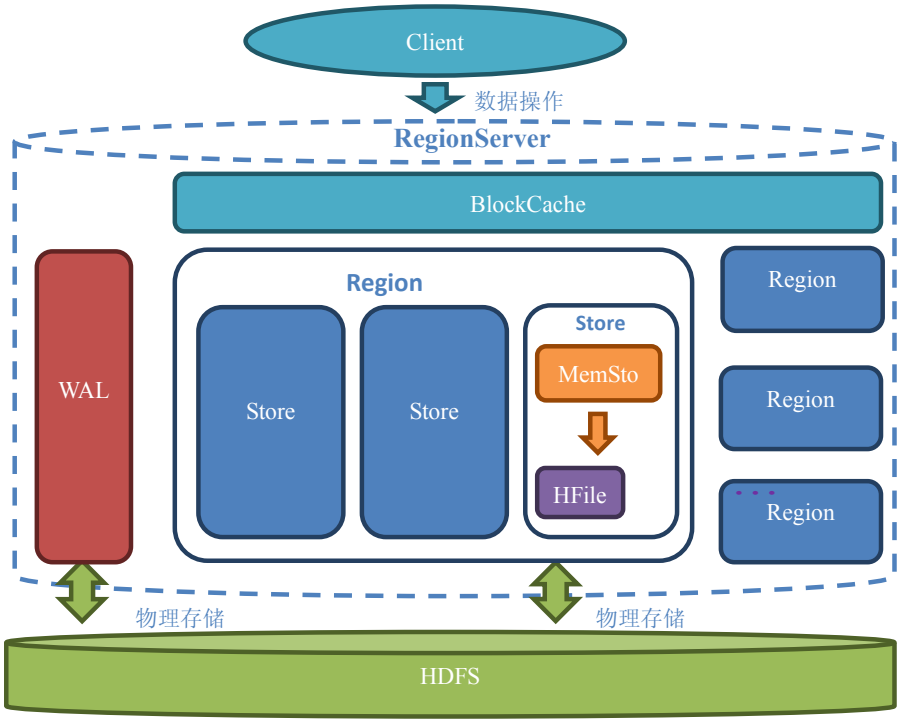

Client
数据操作
RegionServer
BlockCache
WAL
Region
Store
Store
Store
MemSto
HFile
Region
Region
Region
物理存储
物理存储
HDFS

图 8-11

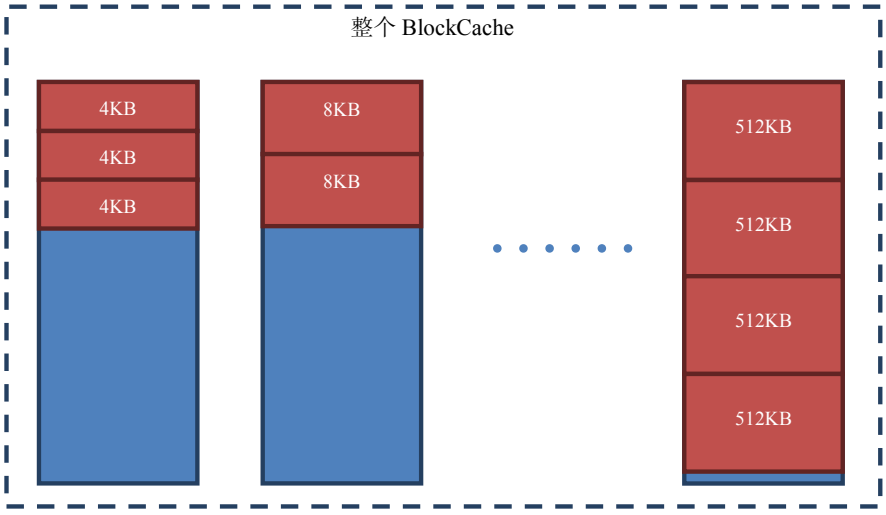

整个 BlockCache
4KB
4KB
4KB
8KB
8KB
512KB
512KB
512KB
512KB

图 8-12

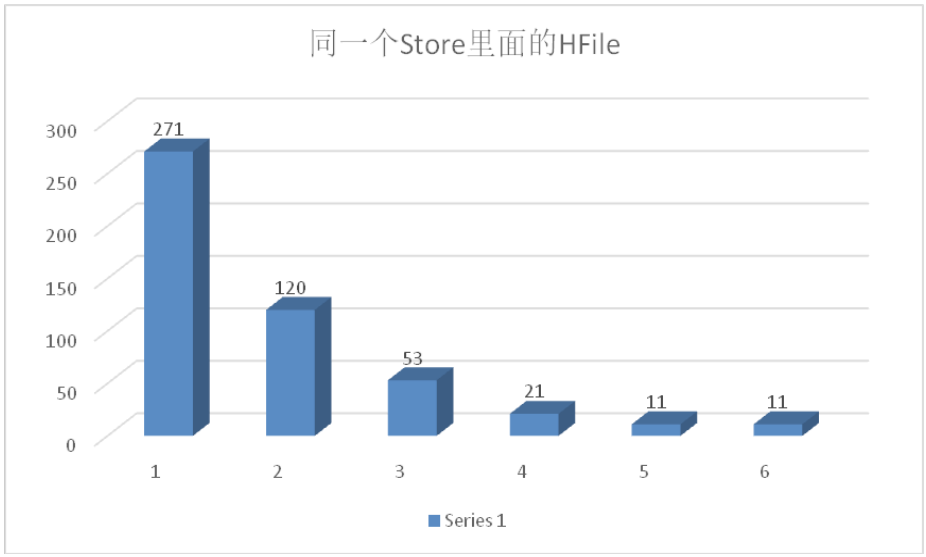

同一个Store里面的HFile
300
250
200
150
100
50
0
271
120
53
21
11
11
1
2
3
4
5
6
Series 1

图 8-13

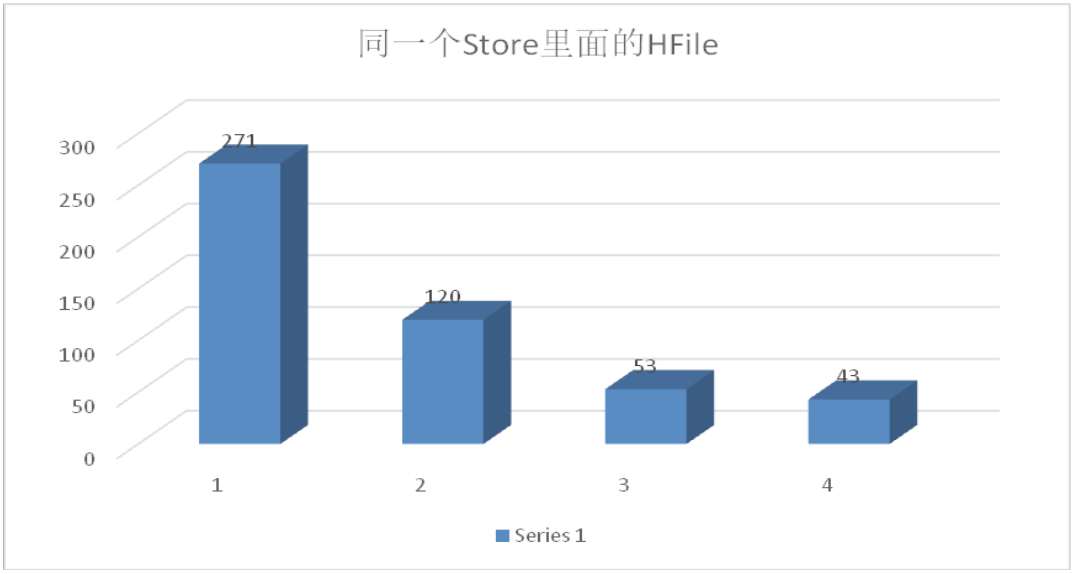

同一个Store里面的HFile
300
250
200
150
100
50
0
271
120
53
43
1
2
3
4
Series 1

图 8-14

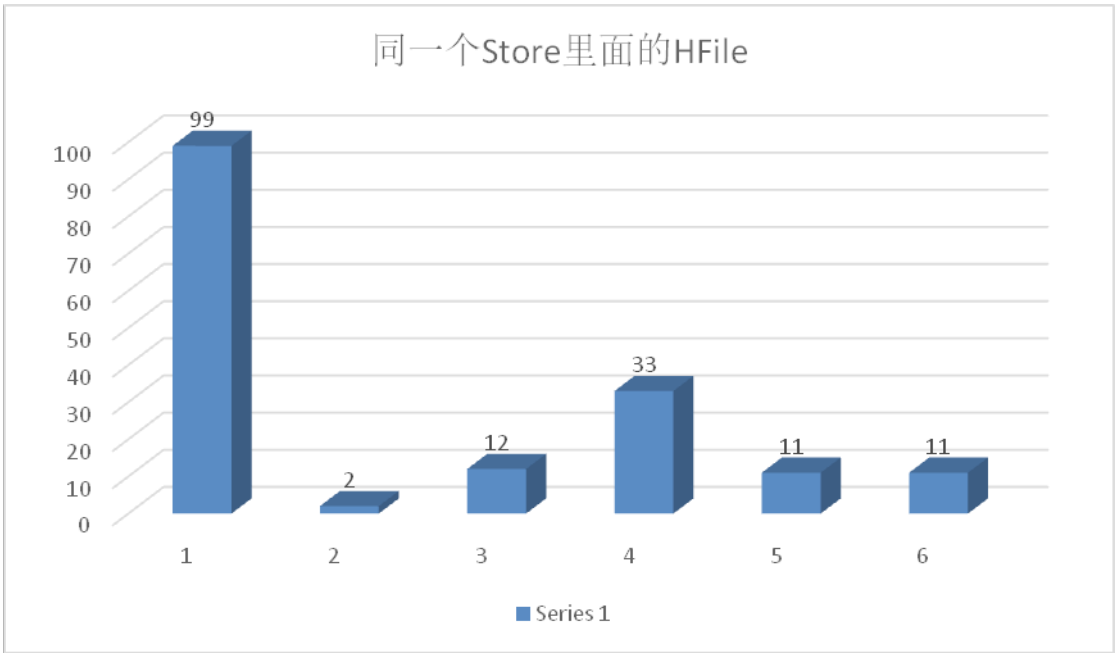

同一个Store里面的HFile
99
2
12
33
11
11
100
90
80
70
60
50
40
30
20
10
0
1
2
3
4
5
6
Series 1

图 8-15

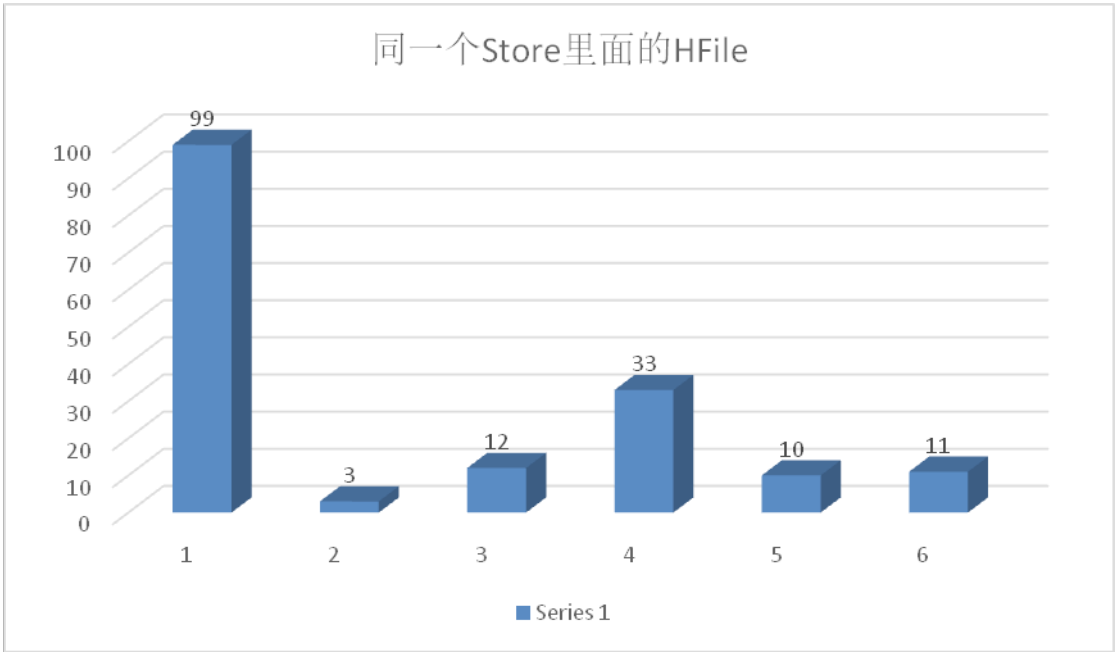

同一个Store里面的HFile
99
3
12
33
10
11
100
90
80
70
60
50
40
30
20
10
0
1
2
3
4
5
6
Series 1

图 8-16

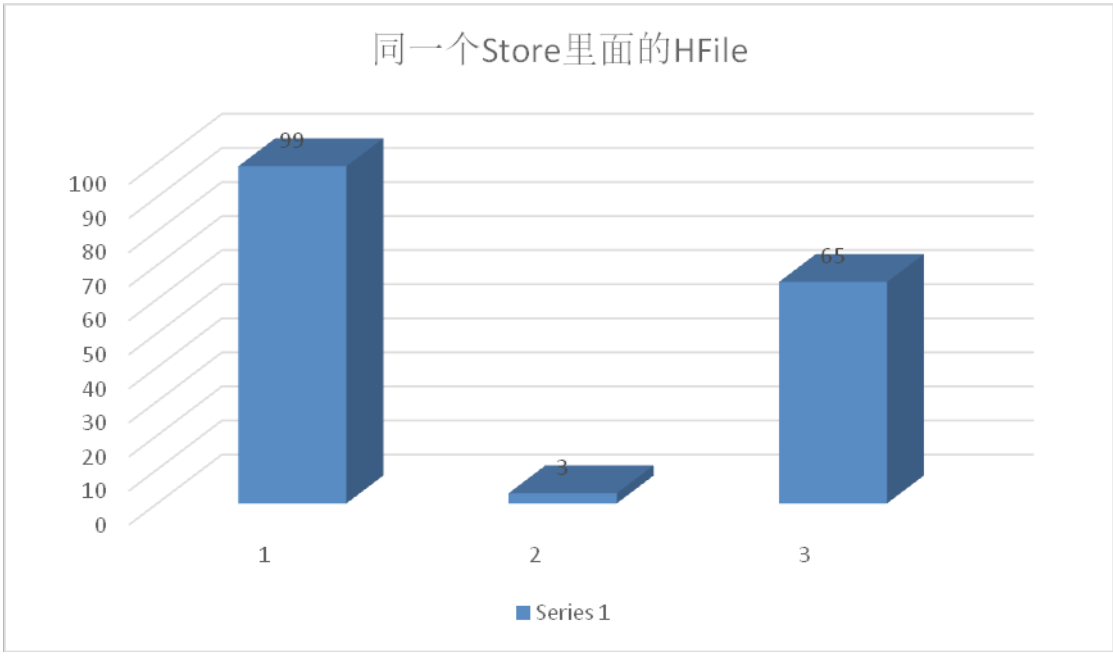

同一个Store里面的HFile
99
3
65
100
90
80
70
60
50
40
30
20
10
0
1
2
3
Series 1

图 8-17

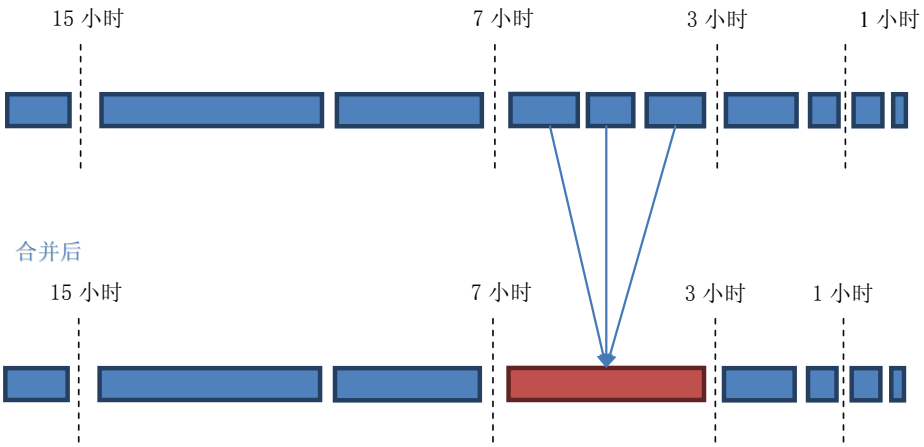

15 小时
7 小时
3 小时
1 小时
合并后
15 小时
7 小时
3 小时
1 小时

图 8-18

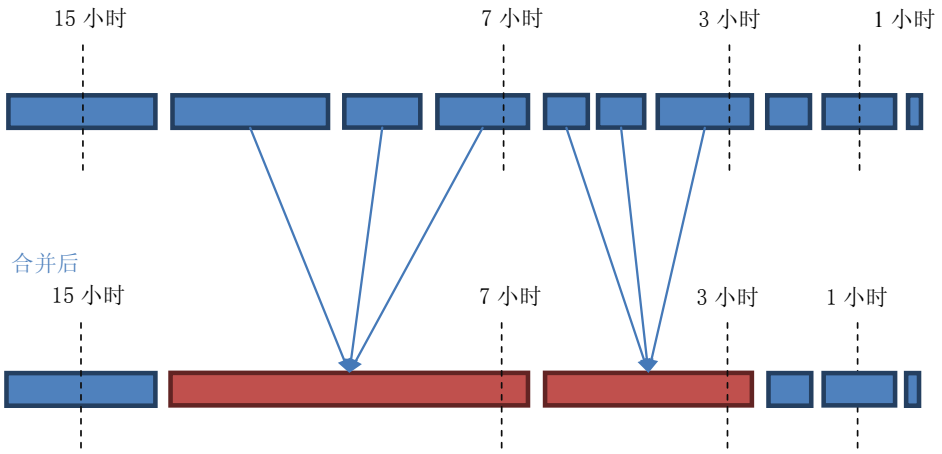


图 8-19

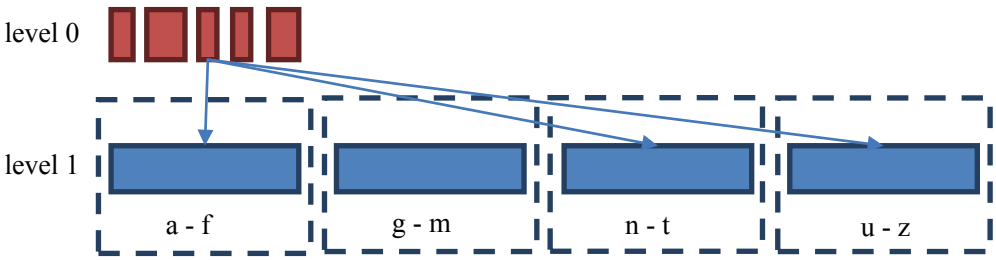


图 8-20

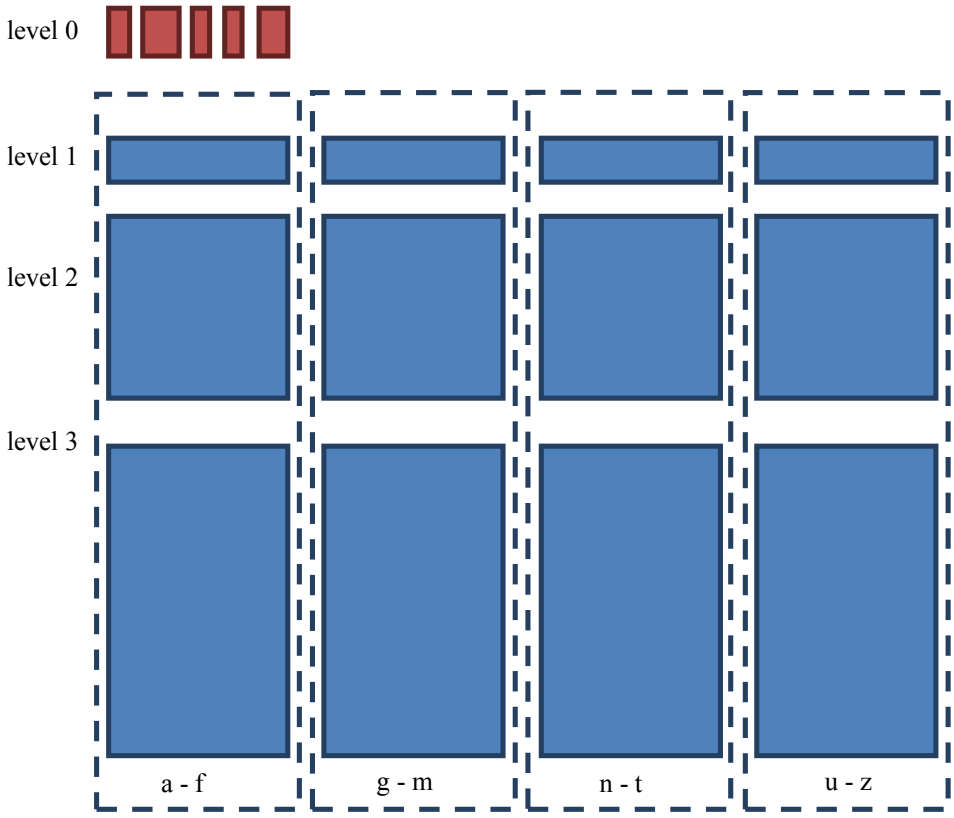


图 8-21

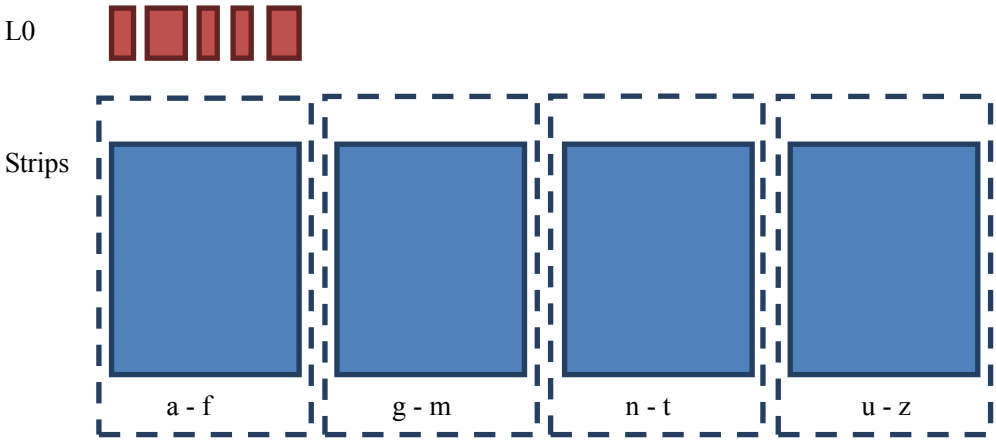


图 8-22

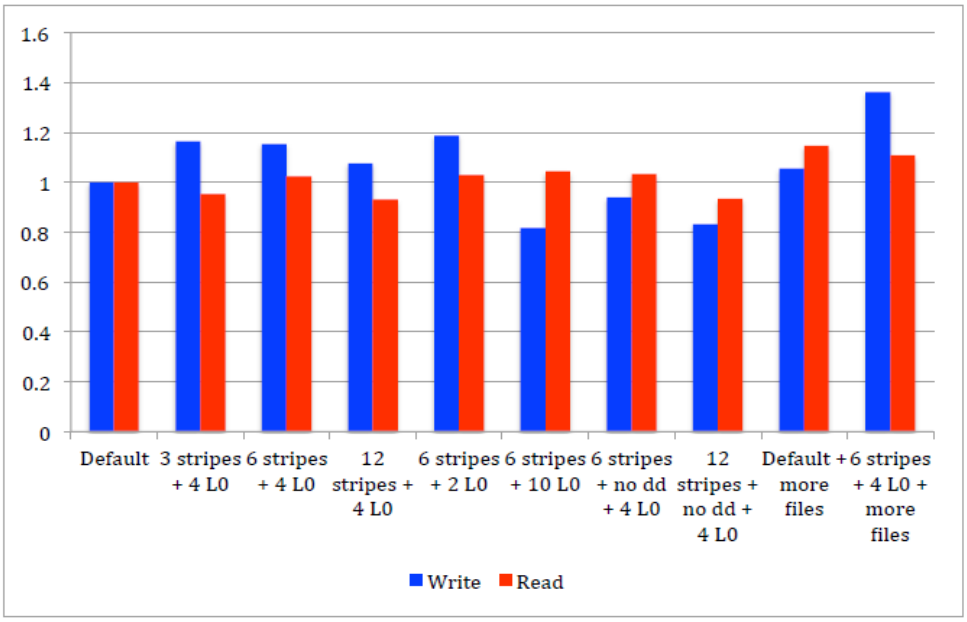

1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
Default
3 stripes + 4 L0
6 stripes + 4 L0
12 stripes + 4 L0
6 stripes + 2 L0
6 stripes + 10 L0
6 stripes + no dd + 4 L0
12 stripes + no dd + 4 L0
Default + more files
6 stripes + 4 L0 + more files
Write
Read

图 8-23

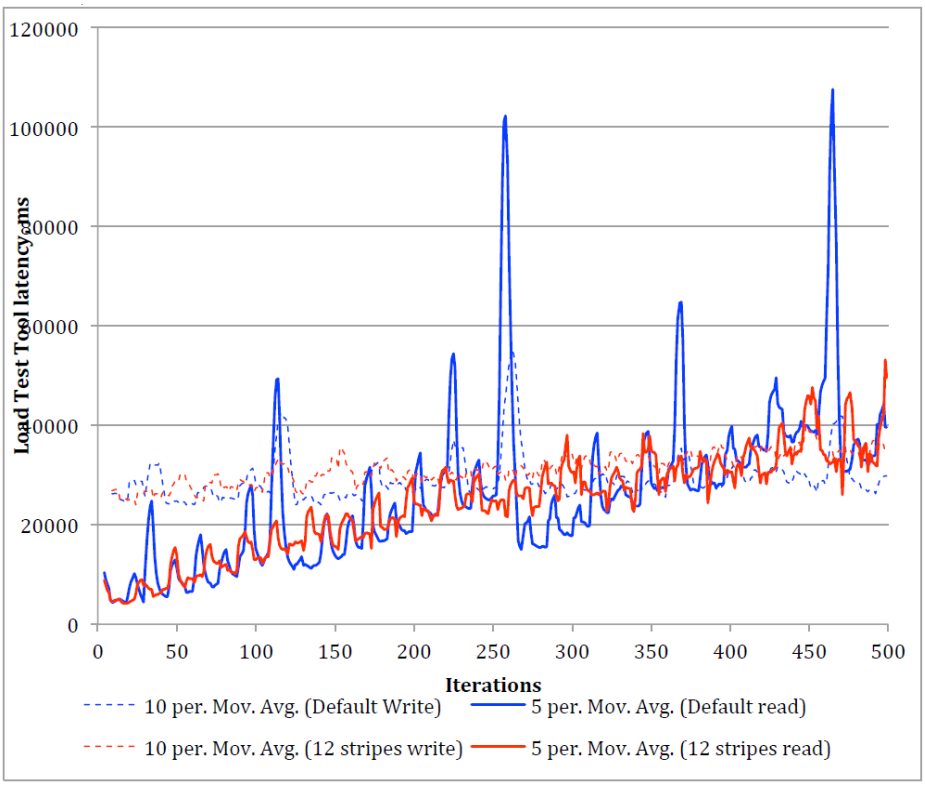

120000
100000
80000
60000
40000
20000
0
Load Test Tool latency, ms
0
50
100
150
200
250
300
350
400
450
500
Iterations
10 per. Mov. Avg. (Default Write)
5 per. Mov. Avg. (Default read)
10 per. Mov. Avg. (12 stripes write)
5 per. Mov. Avg. (12 stripes read)

图 8-24

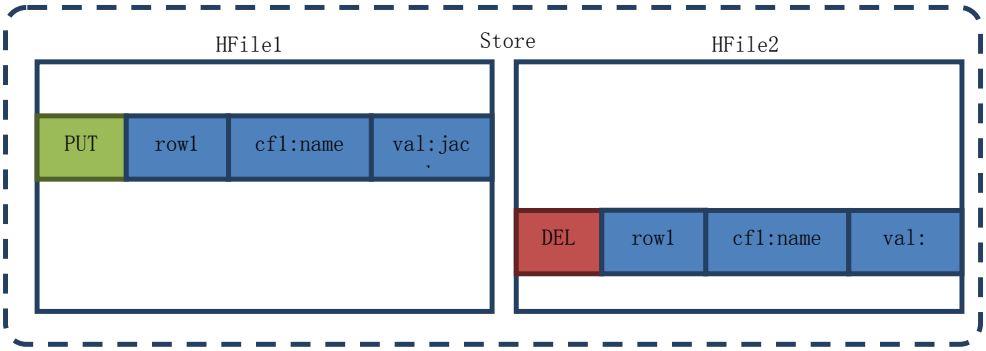


图 8-25

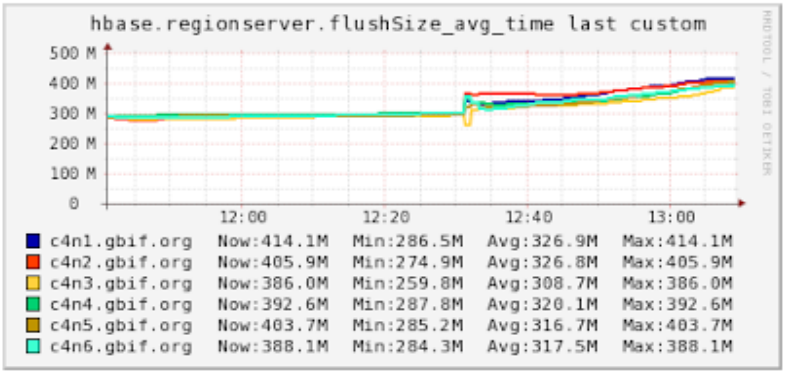


Figure 1: Average memstore flush size over time

图 8-26

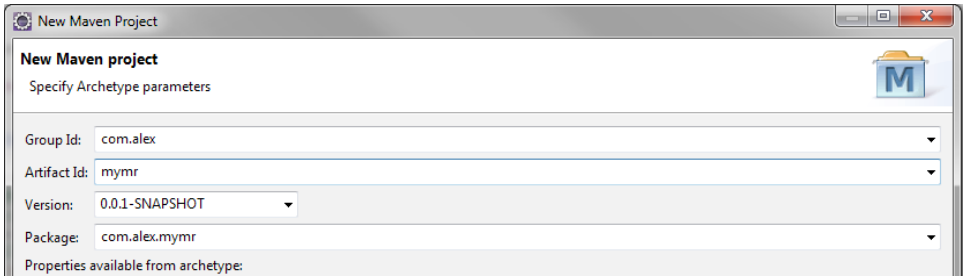


图 9-1

接下来，我们去修改 pom.xml 文件，增加对 hbase-server 的依赖。

```
<dependency>
    <groupId>org.apache.hbase</groupId>
```

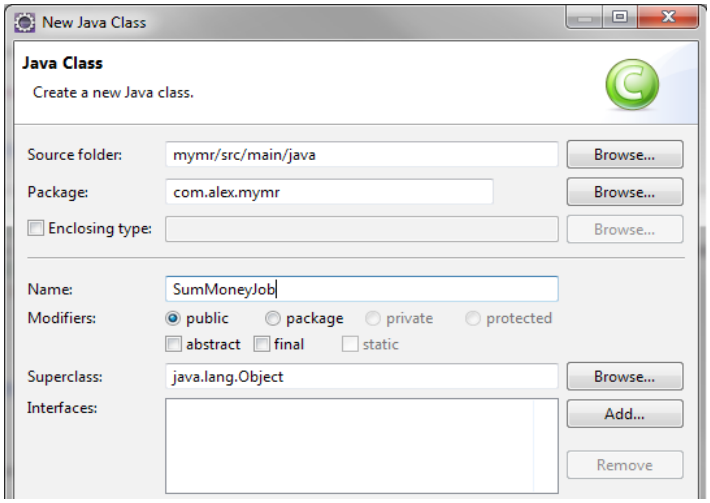

New Java Class
Java Class
Create a new Java class.
Source folder:
mymr/src/main/java
Browse...
Package:
com.alex.mymr
Browse...
Enclosing type:
Browse...
Name:
SumMoneyJob
Modifiers:
public
package
private
protected
abstract
final
static
Superclass:
java.lang.Object
Browse...
Interfaces:
Add...
Remove

图 9-2

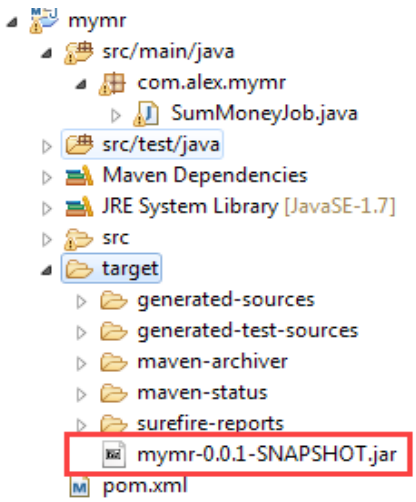

mymr
src/main/java
com.alex.mymr
SumMoneyJob.java
src/test/java
Maven Dependencies
JRE System Library [JavaSE-1.7]
src
target
generated-sources
generated-test-sources
maven-archiver
maven-status
surefire-reports
mymr-0.0.1-SNAPSHOT.jar
pom.xml

图 9-3