Cassandra & HBase Comparison

郭鹏 盛大在线

新浪微博：@逖靖寒

Agenda

- Function
- Implement
- What's Next

SACC2011

Agenda

- Function
- Implement
- What's Next

SACC2011

Data Model

Standard Column Family

Keyspace	Row	Column Family	Column Name	Column Value	Timestamp
KS-1	Row-1	cfName-1	colName-1	colValue-1	12345678
KS-1	Row-1	cfName-1	colName-2	colValue-2	22345638
KS-1	Row-1	cfName-2	colName-1	colValue-1	12445678
KS-1	Row-2	cfName-1	colName-1	colValue-1	12345678
KS-2	Row-1	cfName-1	colName-1	colValue-1	12345678

Data Model

Super Column Family

Keyspace	Row	Column Family	Super Column Name	Column Name	Column Value	Timestamp
KS-1	Row-1	cfName-1	sColName-1	colName-1	colValue-1	12345678
KS-1	Row-1	cfName-1	sColName-1	colName-2	colValue-2	22345638
KS-1	Row-1	cfName-2	sColName-2	colName-1	colValue-1	12445678

SACC2011

Data Model

Table	Row	Column Family	Column Name	Column Value	Timestamp	Version
Tab-1	Row-1	cfName-1	colName-1	colValue-1	12345678	V-1
Tab-1	Row-1	cfName-1	colName-2	colValue-2	22345638	V-2
Tab-1	Row-1	cfName-1	colName-2	colValue-2	12445678	V-1
Tab-1	Row-2	cfName-1	colName-1	colValue-1	12345678	V-1
Tab-2	Row-1	cfName-1	colName-1	colValue-1	12345678	V-1

Basic API

- CRUD
- ConsistencyLevel
  - ONE
  - QUORUM
  - ALL
- Range Query：
  - Row
  - Token

SACC2011

Advanced API

- Authorization & Authentication
- Secondary Index
- Online Schema Update
- Counter
- MapReduce
- CQL(Cassandra Query Language)

```
USE WebSiteKS;
SELECT FROM Standard1 WHERE KEY = "k";
UPDATE Standard1 WITH ROW("k", COL("c", "hello!"));
```

SACC2011

Basic API

- CRUD
- Scan:
  - Row

Advanced API

- Lock
  - Put
  - checkAndPut
  - Delete
  - checkAndDelete
- Limited Online Schema Update
- Counter
- MapReduce
- Filter

```
byte[] startRow = Bytes.toBytes(rowPrefix);

Scan scan = new Scan();

Filter filter = new PrefixFilter(startRow);
scan.setFilter(filter);
scan.setStartRow(startRow);
```

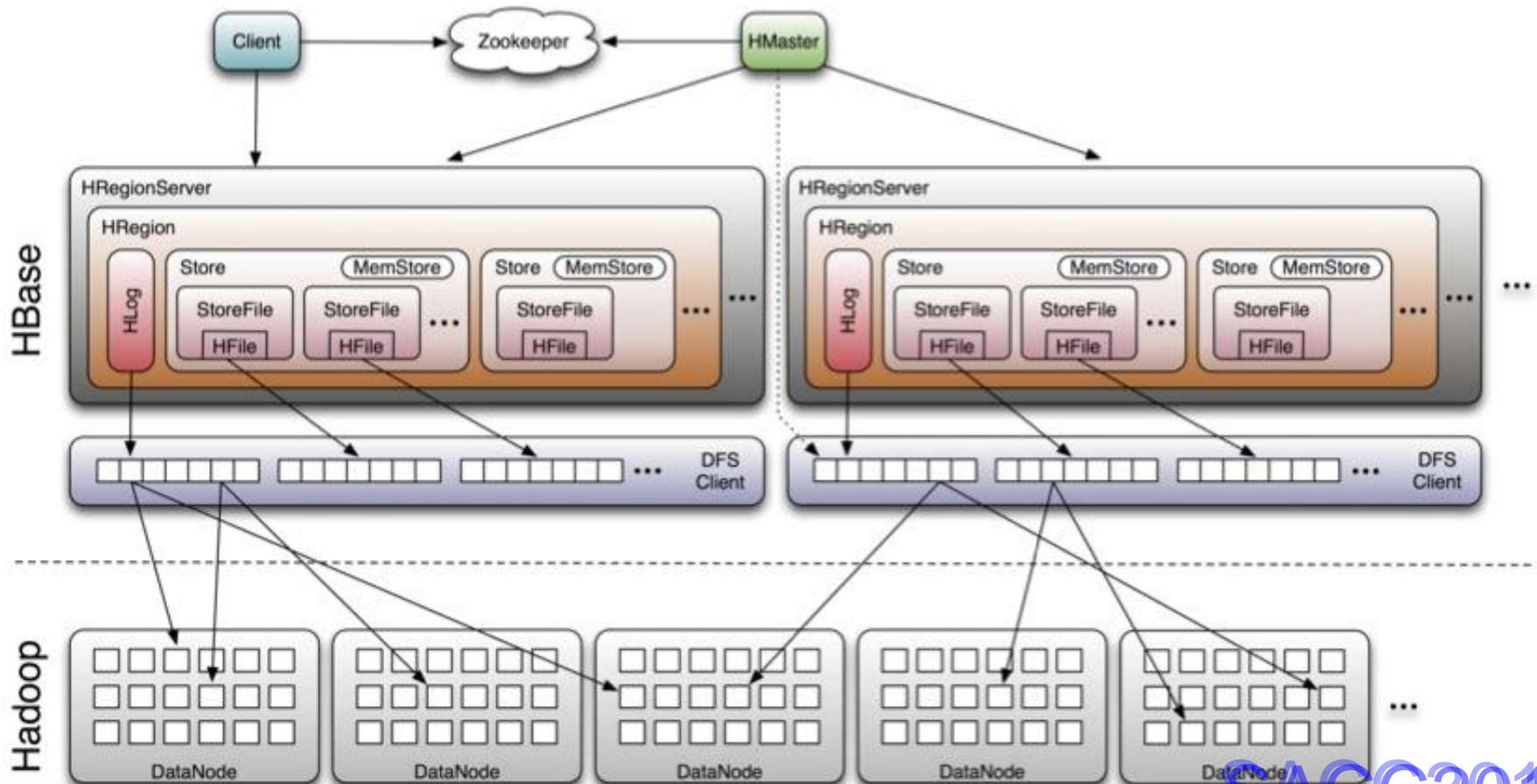SACC2011

Agenda

- ◘ Function
- ◘ Implement
- ◘ What's Next

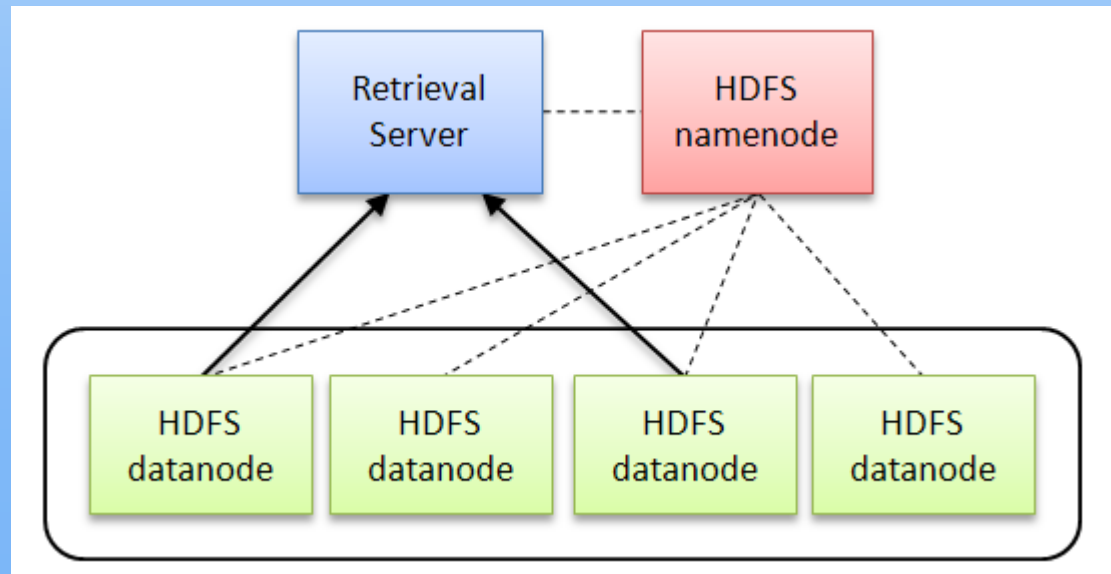SACC2011

System Architecture

Cassandra

RAM

System CF

APP CF-1

APP CF-N

Memtable

Memtable

SSTable

SSTable

Commitlog

Commitlog

Local Disk

Cassandra Server 1

...

Cassandra ServerN

SACC2011

System Architecture

File System

Cassandra

H·BASE
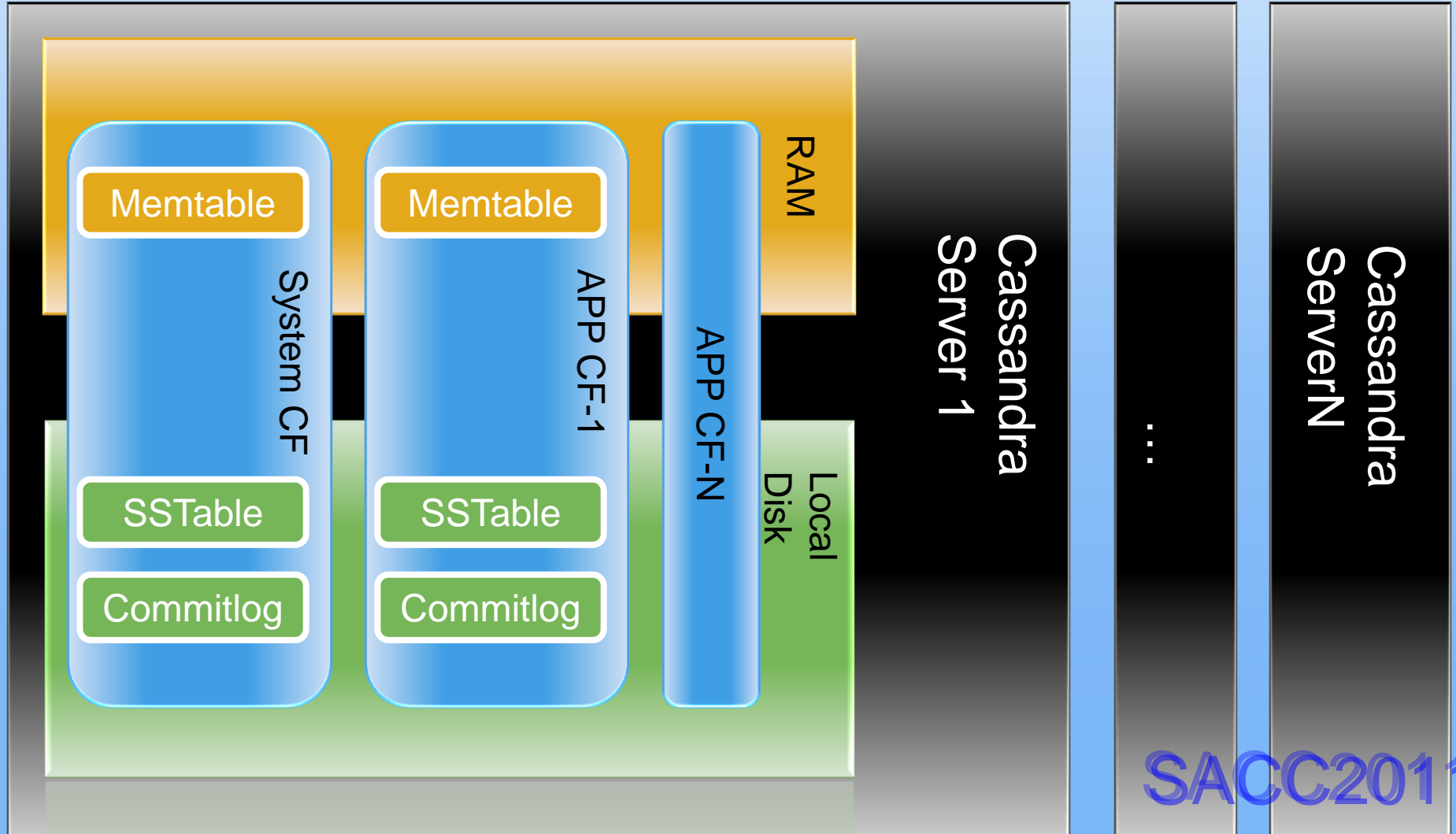
SACC2011

Consistency

 Cassandra

 HBASE

Server No.	RowKey List
Cassandra-01	key1,key2,key3,key4
Cassandra-02	key2,key3,key4,key5
Cassandra-03	key3,key4,key5,key6

Server No.	RowKey List
RegionServer-01	key1,key2,key3,key4
RegionServer-02	key5,key6,key7,key8
RegionServer-03	key9,key10,key11,key12

SACC2011

Location
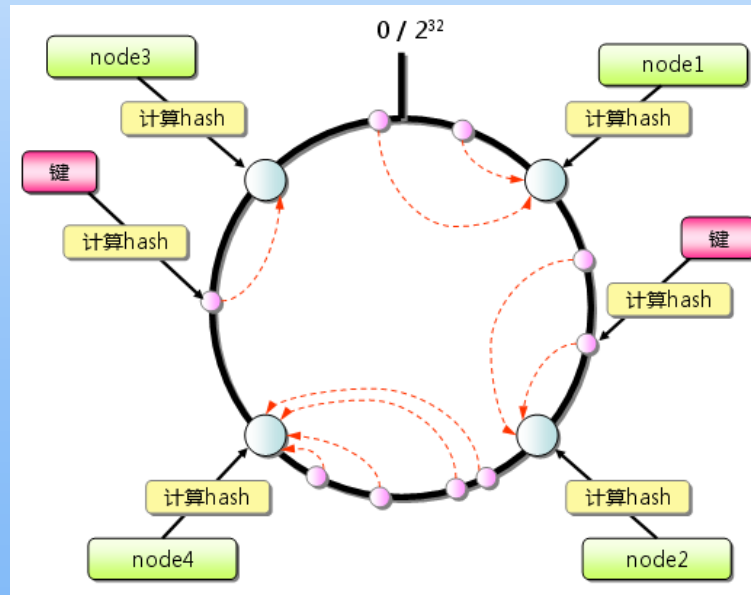
Address	Status	State	Load	Token
				11679118812112088611751366640
172.16.197.113	Up	Normal	17.56 KB	76532511552026171774714881960
172.16.197.114	Up	Normal	17.83 KB	76532511552026171774714881960
172.16.197.112	Up	Normal	22.5 KB	11679118812112088611751366640

Location

```
hbase(main):004:0> scan '.META.', {LIMIT => 1}
ROW                          COLUMN+CELL
 BE,,1312597904196.00716b    column=info:regioninfo, timestamp=1312597904369, value=REGION
 4aac6b3562c8fbb15afd680f    => 'BE,,1312597904196.00716b4aac6b3562c8fbb15afd680f94.', STAR
 94.                         ', ENDKEY => '', ENCODED => 00716b4aac6b3562c8fbb15afd680f94,
                             {{NAME => 'BE', FAMILIES => [{NAME => 'course', BLOOMFILTER =>
                             REPLICATION_SCOPE => '0', COMPRESSION => 'NONE', VERSIONS => '
                             > '2147483647', BLOCKSIZE => '65536', IN_MEMORY => 'false', BI
                             => 'true'}, {NAME => 'student', BLOOMFILTER => 'NONE', REPLICA
                             E => '0', COMPRESSION => 'NONE', VERSIONS => '3', TTL => '2147
                             BLOCKSIZE => '65536', IN_MEMORY => 'false', BLOCKCACHE => 'tru
                             E => 'teacher', BLOOMFILTER => 'NONE', REPLICATION_SCOPE => '0
                             SSION => 'NONE', VERSIONS => '3', TTL => '2147483647', BLOCKSI
                             536', IN_MEMORY => 'false', BLOCKCACHE => 'true'}]}}
 BE,,1312597904196.00716b    column=info:server, timestamp=1314342637415, value=hadoop02.cl
 4aac6b3562c8fbb15afd680f    :60020
 94.
 BE,,1312597904196.00716b    column=info:serverstartcode, timestamp=1314342637415, value=13
 4aac6b3562c8fbb15afd680f    70
 94.
```

SACC2011

Storage File

SACC2011

Cache

Cassandra
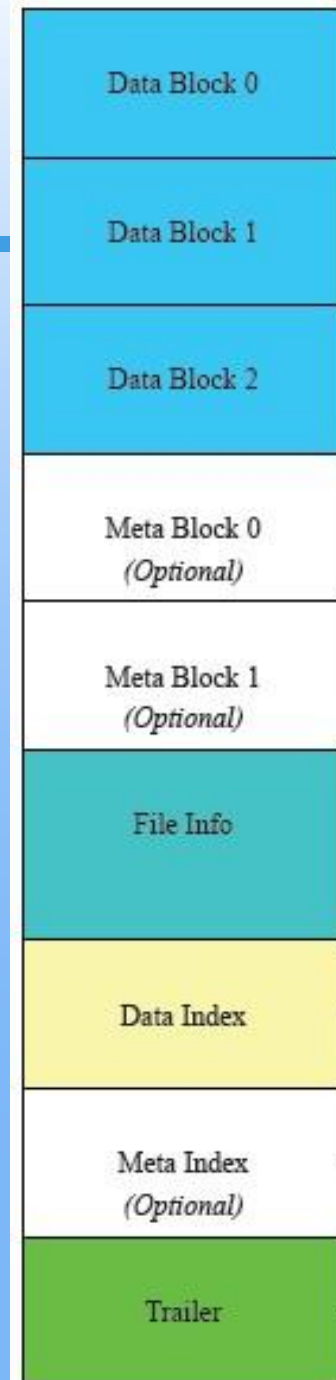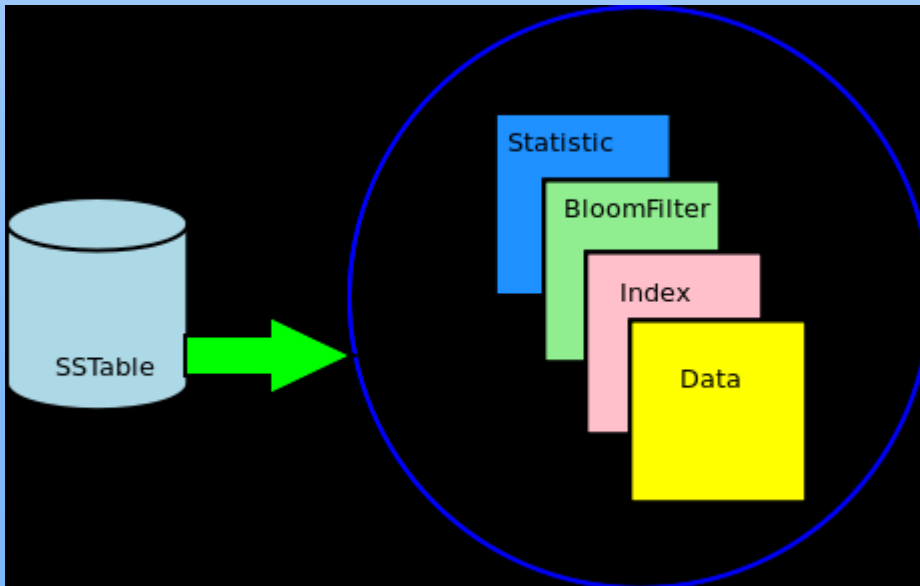
Cache	Value
RowCache	KeyIndex
KeyCache	Key/Value

HBASE

Cache	Value
BlockCache	Data Block

SACC2011

BulkLoader

Cassandra

H·BASE
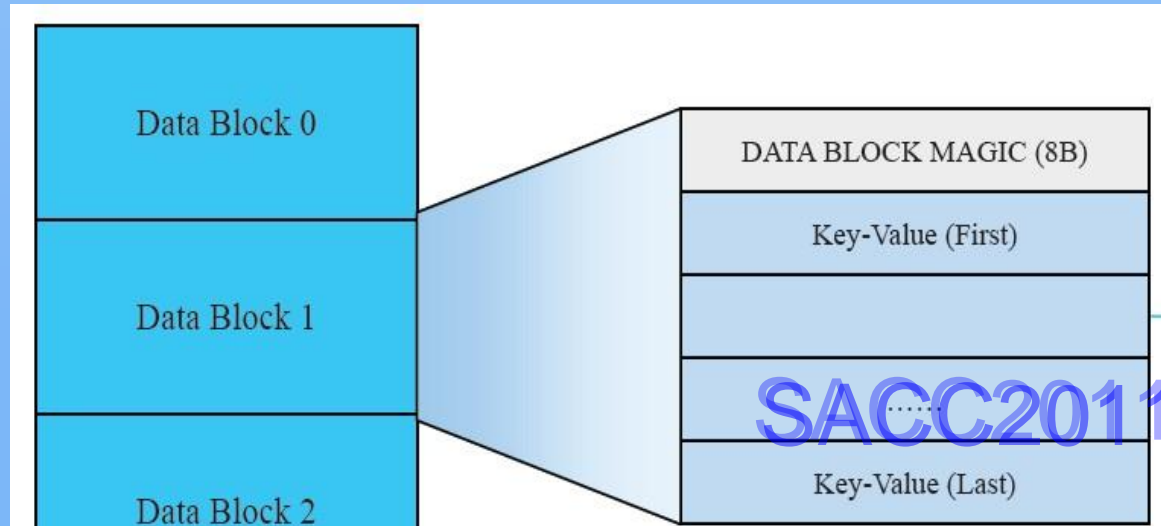
BinaryMemtable

HRegionInterface.bulkLoadHFile
- TableName
- RegionName
- CFName

nodetool cleanup

sstableloader

LoadIncrementalHFiles
- HFilePath
- TableName

SACC2011

LoadBalance

Cassandra

HBASE

nodetool move token

balancer

move regionName serverName

split regionName

SACC2011

Agenda

- ◘ Function
- ◘ Implement
- ◘ What's Next

SACC2011

Brisk

- Enhanced Hadoop and Hive distribution that
  utilizes Cassandra for all of its core services.

- Integrated with Hadoop MapReduce
  and Hive

- Providing an HDFS
  compatible storage layer powered by Cassandra

SACC2011

Application Stack
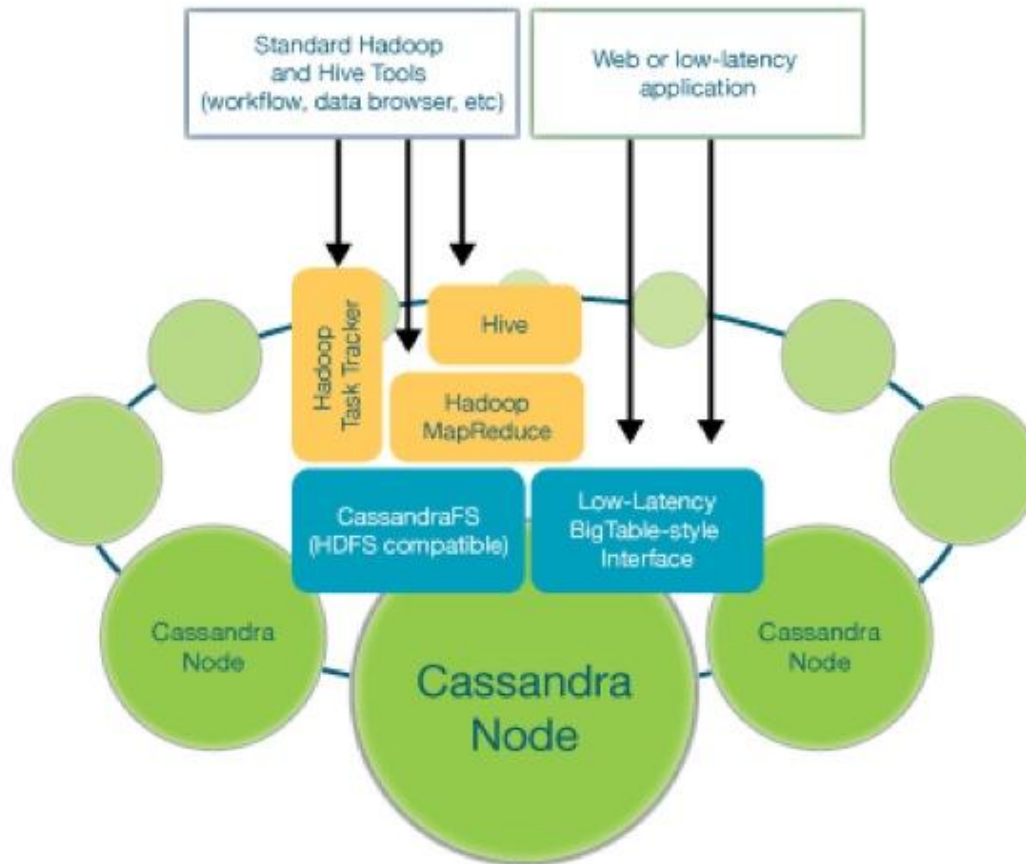
SACC2011

Deploy

Cassandra Replication Ring

SACC2011

Coprocessor
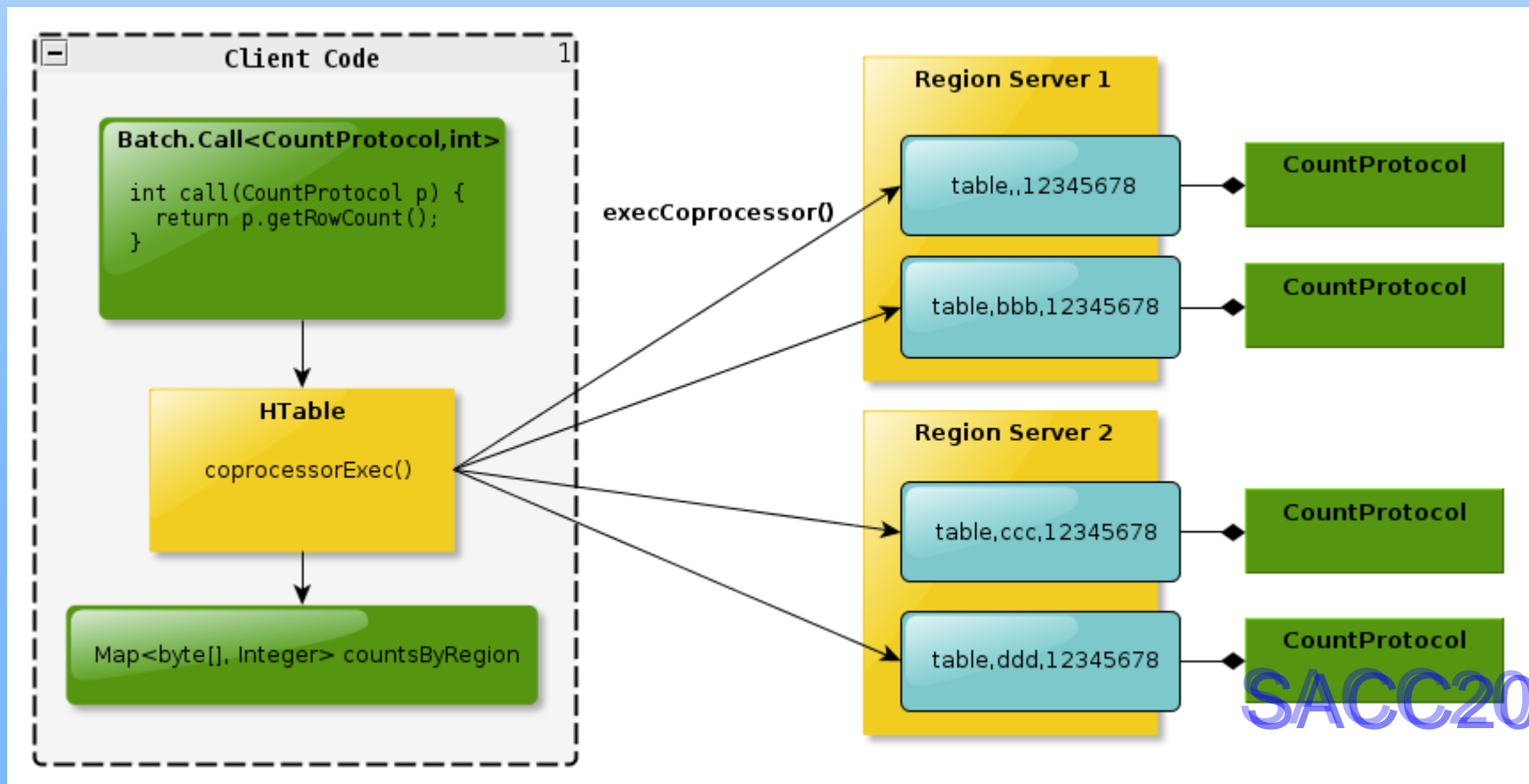
- MasterObserver
- RegionObserver
- EndPoint

MasterObserver
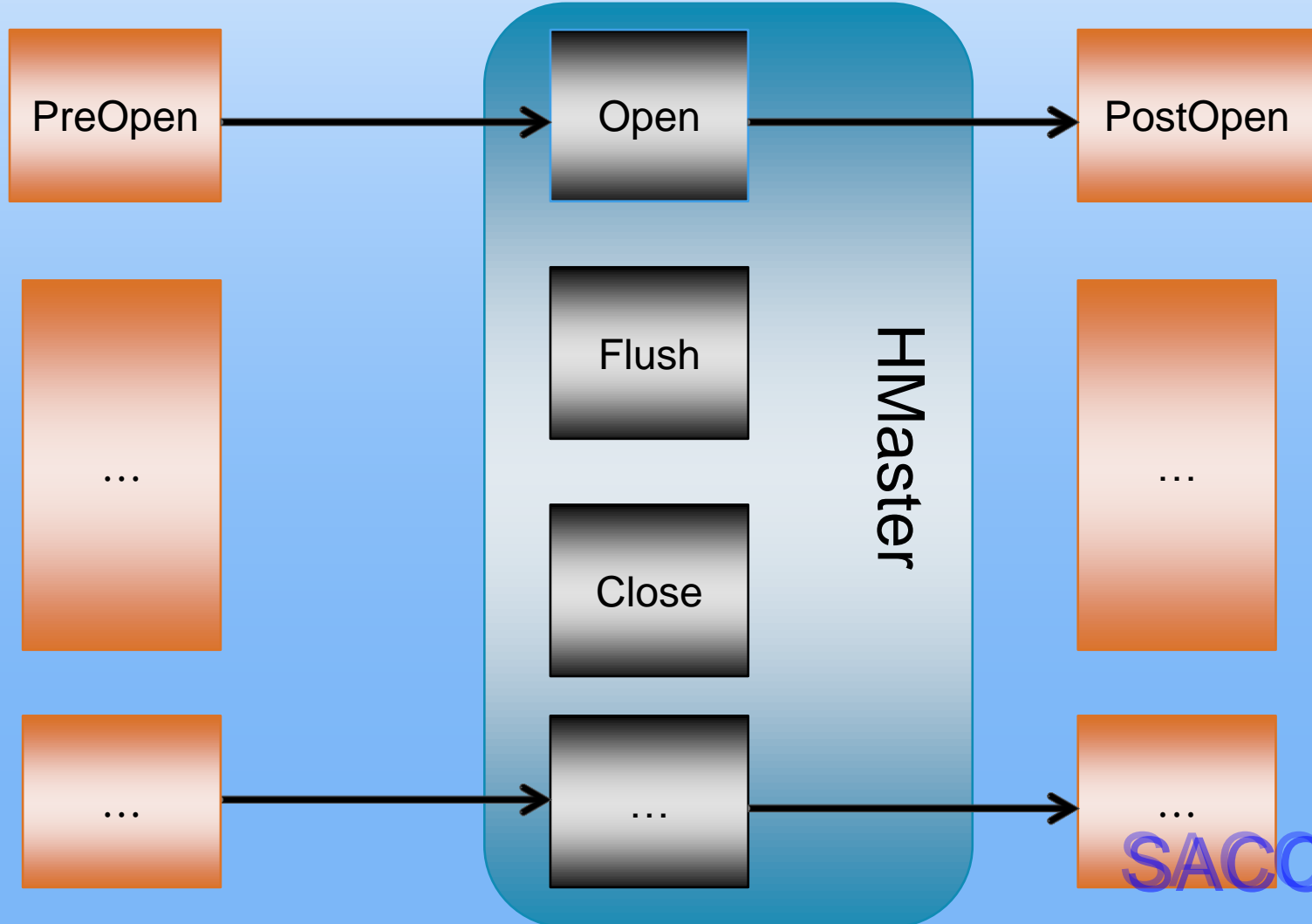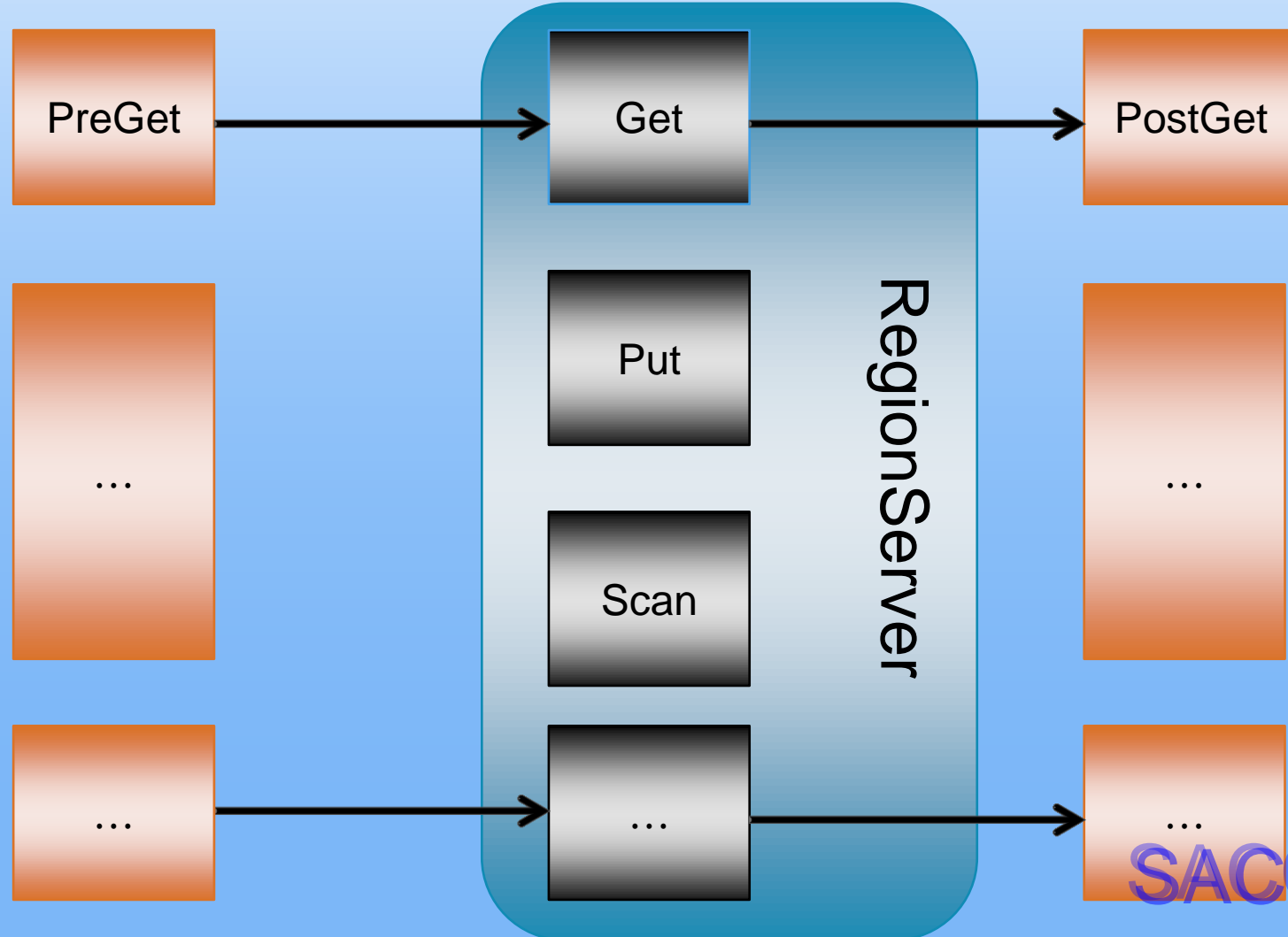
EndPoint

Define your own extensions to HBase RPC transactions exchanged between clients and the region servers.

Applications

- HBase access control: provides basic access control for HBase

- Column aggregate: SQL-like sum(), avg(), max(), min(), etc.

- Region level indexing

- HBase table, region access statistic

SACC2011

使用场景

- Key Value
- Key Scan
- Column动态添加
- 排序
- 二级索引

Cassandra适用场景

- 从零开始
- 快速开发
- 快速部署
- 简单的安全认证

HBase适用场景

- 已有Hadoop集群
- 要求数据一致性
- 高效的数据平衡
- 服务端代码执行

SACC2011

Thanks!

盛大在线
WWW.SDO.COM

SACC2011