

Coprocessors: Uses, Abuses & Solutions

Talk Data To Me @ General Assembly
May 10, 2018

Esther Kundin & Amit Anand
Senior Software Developers

TechAtBloomberg.com

About the Speakers

- Esther Kundin
 - Senior Software Developer
 - Lead architect and engineer
 - Machine Learning and Text Analysis
 - Open Source contributor
- Amit Anand
 - Senior Software Developer
 - Hadoop Services / Infrastructure team
 - Deployments/Tooling
 - Open Source contributor

Outline

- HBase Architecture Review
- Introduction to Coprocessors
- Coprocessor Uses
- Development Abuses and Solutions
- Deployment Abuses and Solutions
- Takeaways
- Q&A

HBase Architecture Review

TechAtBloomberg.com

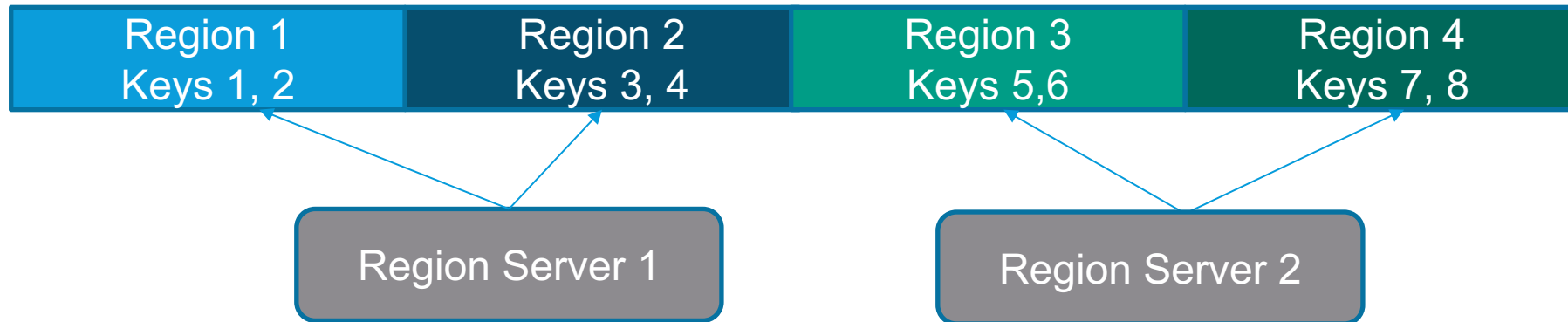
© 2018 Bloomberg Finance L.P. All rights reserved.

Bloomberg

Engineering

HBase Concepts

- HBase is the Hadoop database, a distributed, scalable, big data store
- Keys (RowKeys)
- Regions
- Region Servers



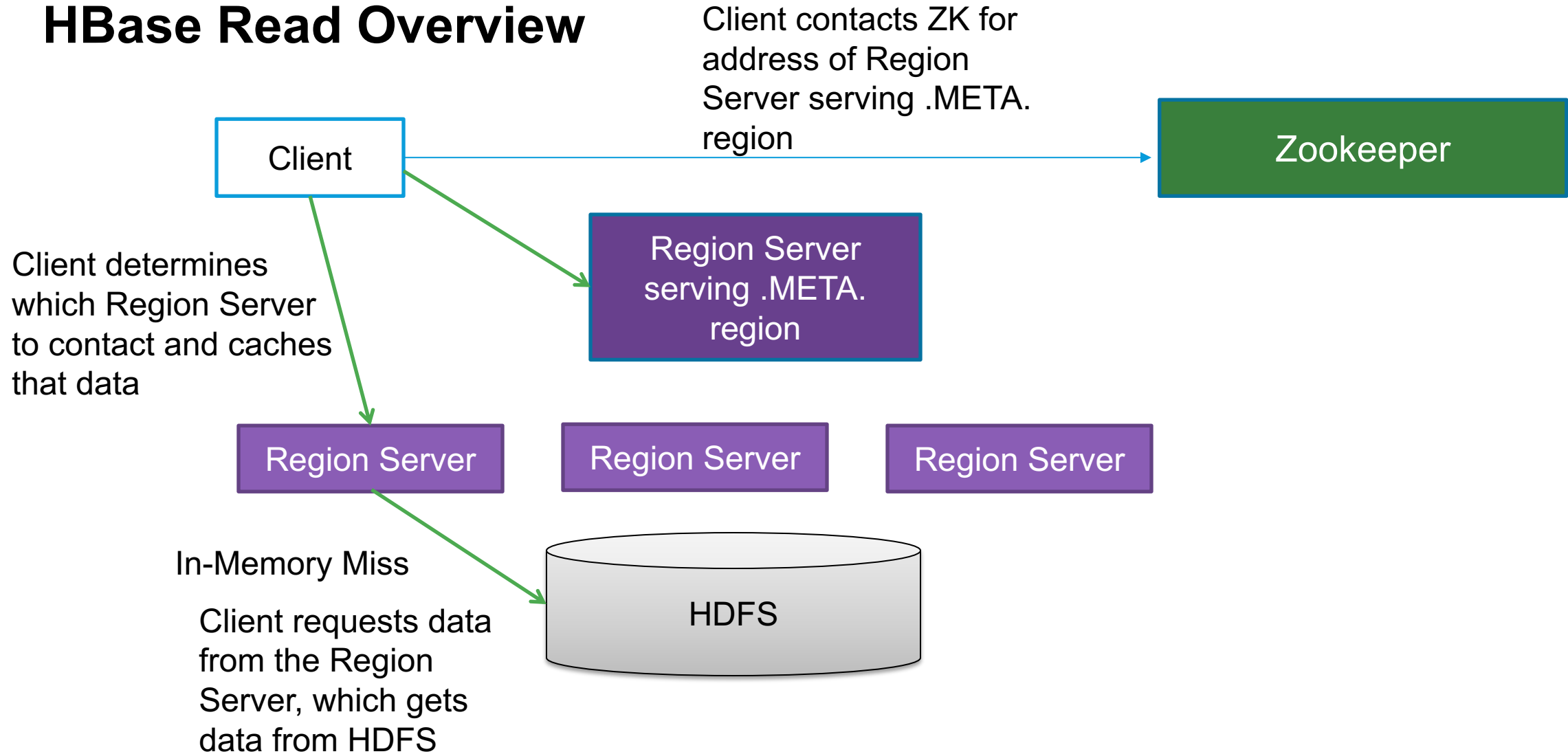
HBase daemons

- HMaster – table DDL operations
- Region Servers – data read/write operations
- Region Server with .META. table – maintains region with map of regions to Region Server
- Outside Systems:
 - Zookeeper – highly reliable distributed coordination

HMaster

- Daemon service, possibly in High Availability Mode
 - Table creation, modification, etc.
 - Assigning Regions to Region Servers
 - Splitting regions that get too big
 - Assignment saved in .META. region

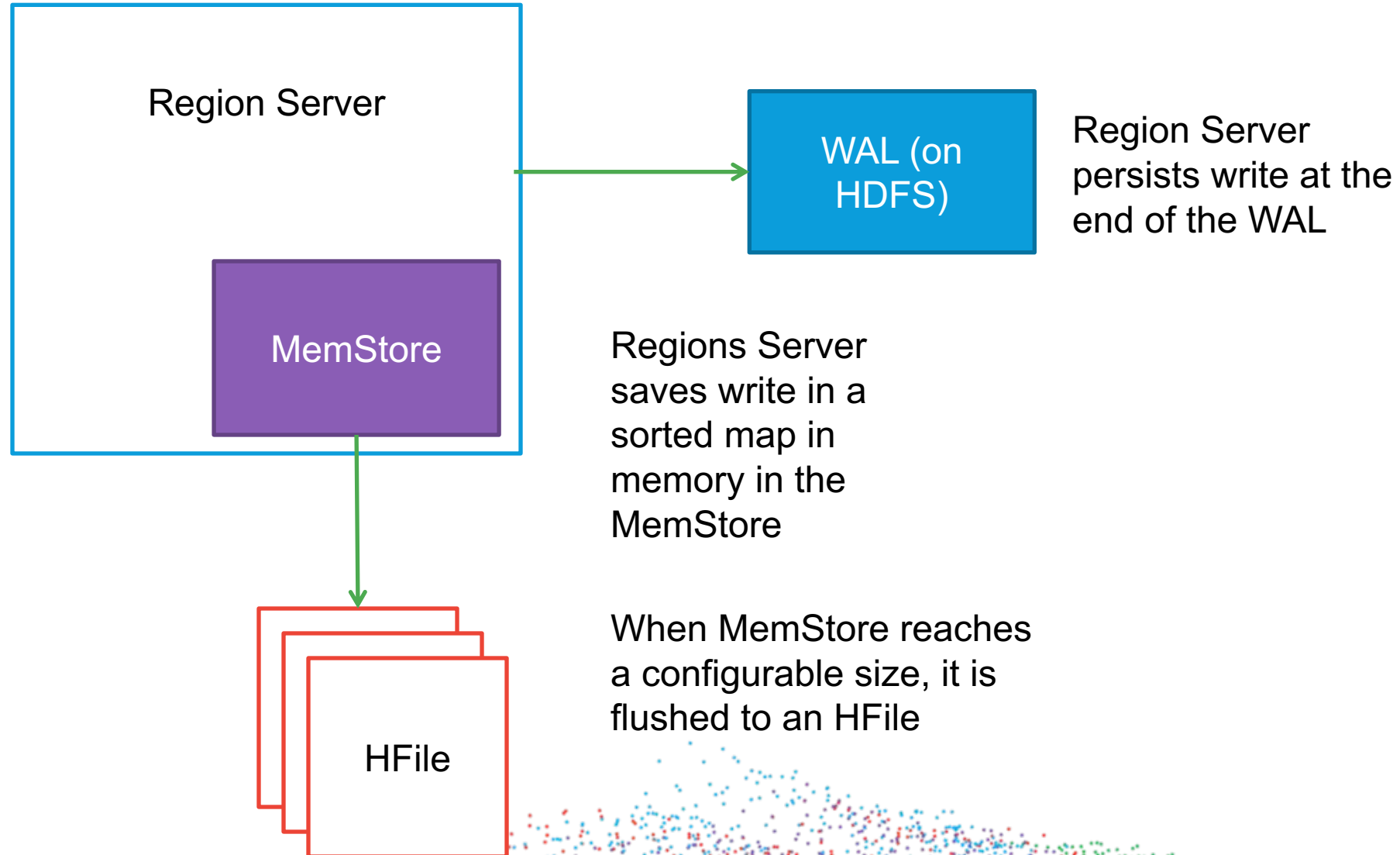
HBase Read Overview



Bloomberg

Engineering

HBase Write Overview



Bloomberg

Engineering

Introduction to Coprocessors

TechAtBloomberg.com

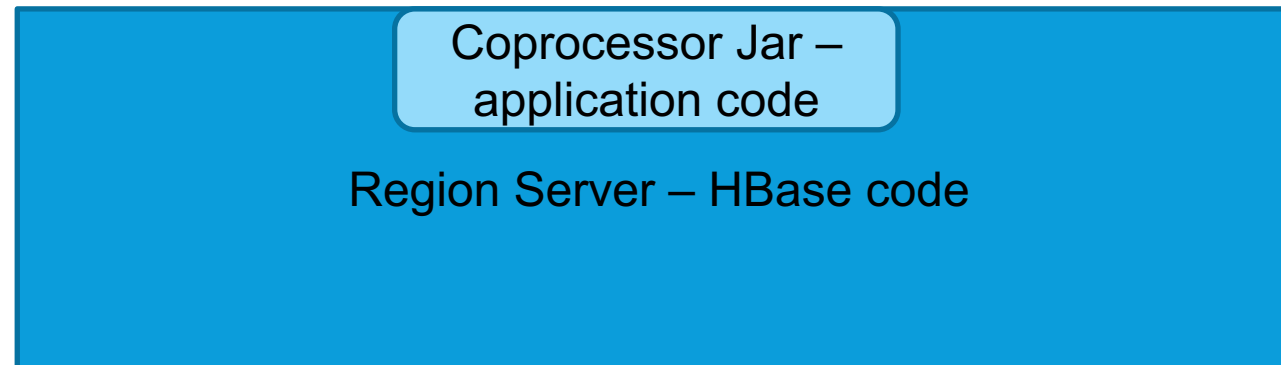
© 2018 Bloomberg Finance L.P. All rights reserved.

Bloomberg

Engineering

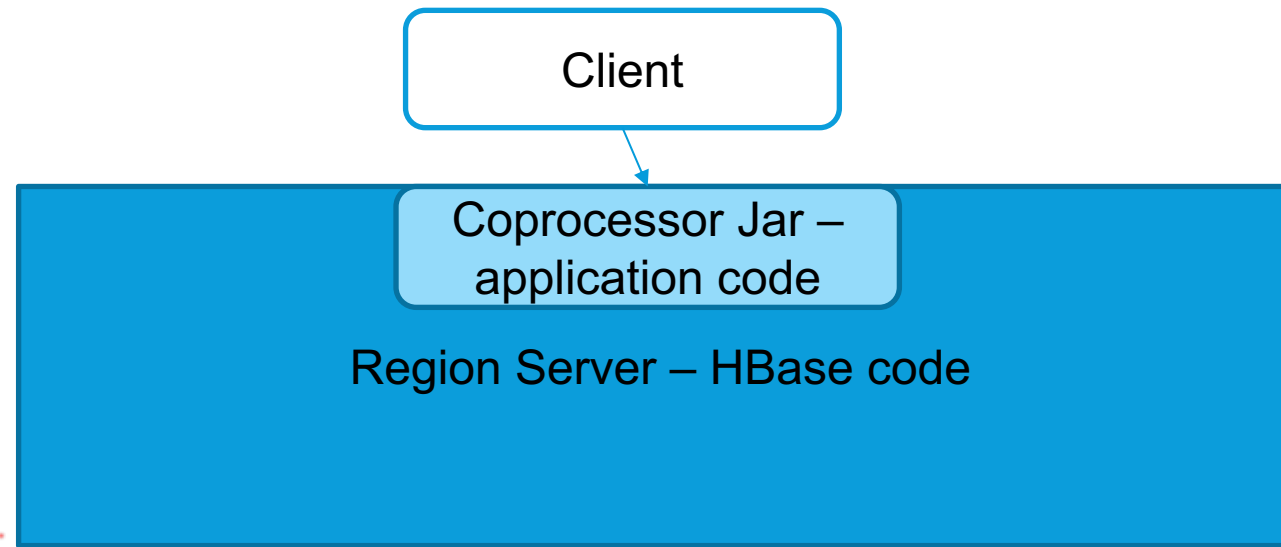
What is a coprocessor?

- Custom code written by users, run as part of the HBase system
 - Bundled Jar
 - Loaded in HBase daemon JVM
 - HMaster
 - Region Server



Endpoint Coprocessor

- Similar to Stored Procedure – triggered by client API call
- Computation at data location
 - Average
 - Summation
- Use Protobuf to specify the input/output structure of your service
 - Starting with HBase
- Must be invoked
 - `Table.CoprocessorService()`
- Implementation of service in Java
- Run the coprocessor by calling that service to do the computation on the collocated data

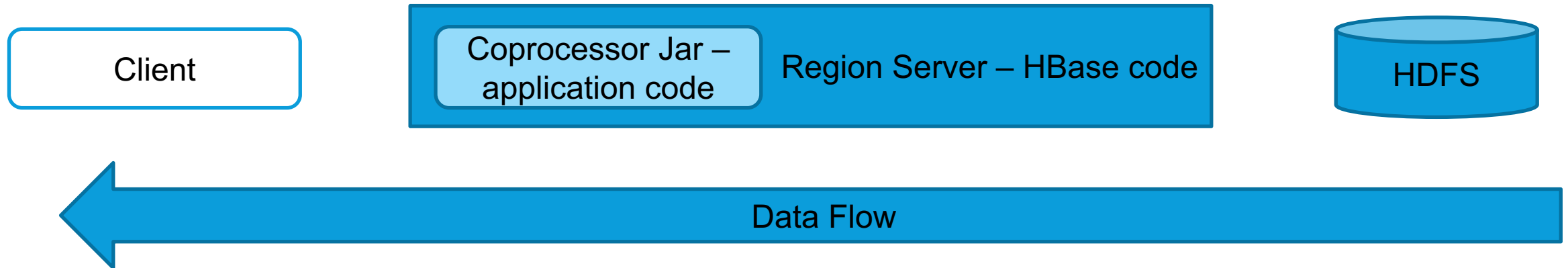


Bloomberg

Engineering

Observer Coprocessor

- Similar to trigger – triggered action happening in the daemon
- Hooks available for all HBase event types
- Canonical example - Can implement a security check on preGetOp/prePut
- Referential integrity – constraint on synthetic foreign keys
- Secondary indices – that's how Phoenix does it



Filters

- Similar to Observer Coprocessors for get/scan
- Can filter on any part of a row
- Many built-in filters available
- Can deploy custom filters
- Support for push-down predicates
- Different API
- Came earlier and much more limited

Filter Example

```
37 // vv CustomFilterExample
38 List<Filter> filters = new ArrayList<Filter>();
39
40 Filter filter1 = new CustomFilter(Bytes.toBytes("val-05.05"));
41 filters.add(filter1);
42
43 Filter filter2 = new CustomFilter(Bytes.toBytes("val-02.07"));
44 filters.add(filter2);
45
46 Filter filter3 = new CustomFilter(Bytes.toBytes("val-09.01"));
47 filters.add(filter3);
48
49 FilterList filterList = new FilterList(
50     FilterList.Operator.MUST_PASS_ONE, filters);
```

Snippet courtesy of <https://github.com/larsgeorge/hbase-book/blob/master/ch04/src/main/java/filters/CustomFilterExample.java>

Observer Coprocessor – Possible Triggers

- Region Server Observer
 - preStopRegionServer
 - preExecuteProcedures
 - postExecuteProcedures
 - preClearCompactionQueues
 - postClearCompactionQueues
- Region Observer
 - preGetOp
 - postGetOp
 - prePut
 - postPut
- WAL Observer
 - preWALRoll
 - preWALWrite
 - postWALRoll
 - postWALWrite

Observer Coprocessor

- Master Observer
 - Runs in HBase master
 - Create, Delete, Modify table
 - Clone, Restore, Delete Snapshots
 - Region splits and many more

Coprocessor Uses

TechAtBloomberg.com

© 2018 Bloomberg Finance L.P. All rights reserved.

Bloomberg

Engineering

Why use a coprocessor?

- Control behavior of Region Server
- Control behavior of data operations
- Server side processing
 - Filters
 - Aggregates
- Reduces pressure on the client side and network
 - Less amount of data being sent from server
- NOT for complex data analysis
 - Apache Phoenix (“We put the SQL back in NoSQL”)

Coprocessor Example

Post-Get example



Table Representation:

Key	Col1	Col2	Col3	Col4	Col5
Key1Abc	1		4	5	
Key1Def	2	2	2		
Key1Xyz			10	11	12

Coprocessor Result:

Key1	Abc-col1	Def-col2	Abc-col3	Abc-col4	Xyz-col5
Key1	1	2	4	5	12

Apache Phoenix Coprocessors

- Apache Phoenix – OLTP and operational analytics for Apache Hadoop
- Phoenix jar runs on top of the Region Server
- Phoenix coprocessors map data model to raw bits and vice versa
- Coprocessor can ensure foreign key constraints
- Write to additional tables for global mutable secondary indices
- Server-side push down of many calculations – like filter, hash join

Development Abuses and Solutions

Application Developer Perspective



TechAtBloomberg.com

© 2018 Bloomberg Finance L.P. All rights reserved.

Bloomberg

Engineering

Challenge – Exceptions

- Exceptions (other than IOExceptions) in the coprocessor bring down Region Server
 - 100% loss of service
 - Manual intervention needed bring cluster back to health
- In other cases, the coprocessor silently unloads – depends on global settings



Solution – Catch all exception

```
public final void prePut(...)
    throws IOException {
    try {
        prePutImpl(...);
    }
    catch(IOException ex) {
        // Allow IOExceptions to propagate
        // They won't cause an unload
        throw ex;
    }
    catch(Throwable ex) {
        // Wrap other exceptions as IOException
        LOG.error("prePut: caught ", ex);
        throw new IOException(ex);
    }
}
```

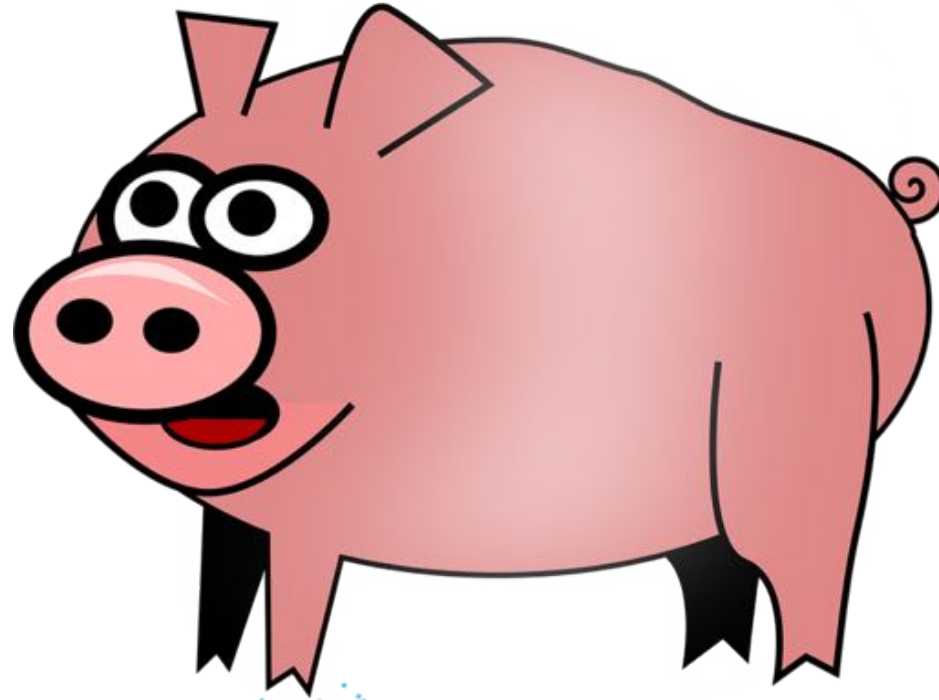
Even better – create interface code for all coprocessors

Bloomberg

Engineering

Problem – Memory hog

- Memory is shared with Region Server memory and coprocessor memory
- Memory hogging slows down Region Server



Solutions – Defensive Java code

- Profile all coprocessor code for memory usage
 - Use a generic profiler with a driver for your coprocessor
 - i.e., JProfiler
- Use common Java tricks for limiting memory usage
 - Use primitive types and underlying arrays where possible
 - Use immutable objects
 - StringBuilder vs. String concatenation

Logging and metrics tips

- Update log4j.properties file with a separate log parameter for coprocessors
- Use MDC (Mapped Diagnostic Context) context to pass parameters to all parts of the coprocessor – across pre and post operations, for example
- Profile all coprocessor code for memory usage
 - <http://www.slf4j.org/api/org/slf4j/MDC.html>
- Create an extra column in a Result to pass back an object populated with metrics

```
// Add a cell with the name of the coprocessor as the column.  
Cell diagCell =  
    CellUtil.createCell(get.getRow(),  
        DIAGNOSTIC_FAMILY,  
        Bytes.toBytes(getCoproprocessorName()),  
        System.currentTimeMillis(),  
        (byte)0x00,  
        Bytes.toBytes(true));
```

Development Abuses and Solutions

Administrator Perspective



TechAtBloomberg.com

© 2018 Bloomberg Finance L.P. All rights reserved.

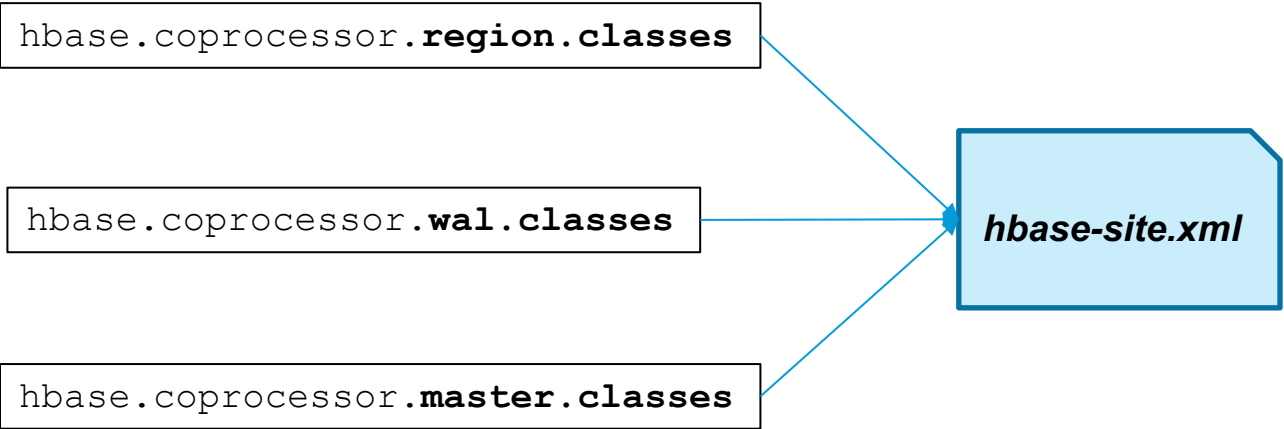
Bloomberg

Engineering

How to deploy a coprocessor?

- Known as coprocessor loading
- Static loading
- Dynamic loading

Static Loading

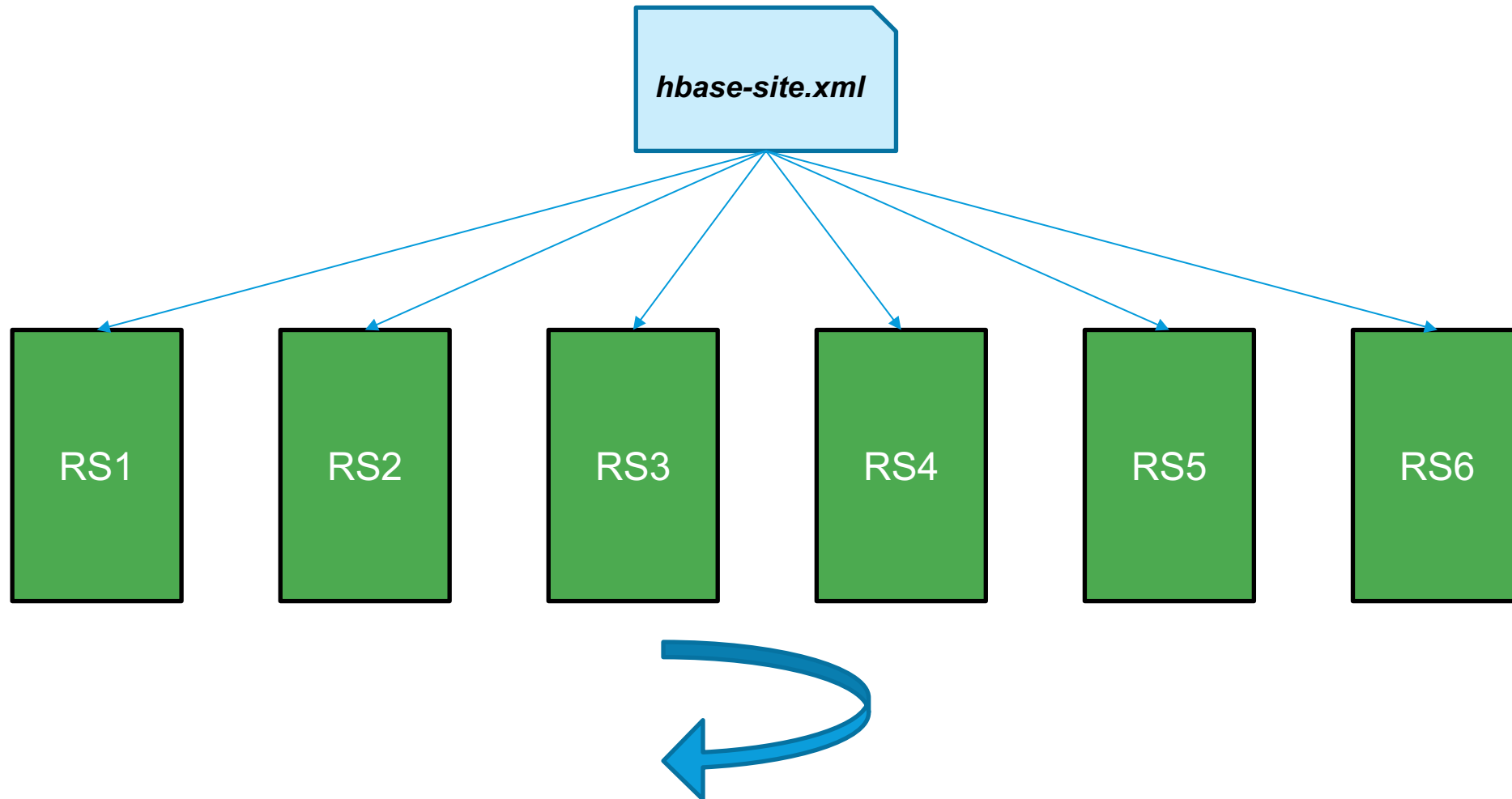


Property	Type of Coprocessor
<code>region.classes</code>	Region Observer End Points
<code>wal.classes</code>	Wal Observer
<code>master.classes</code>	Master Observer

Example:

```
<property>
  <name>hbase.coprocessor.region.classes</name>
  <value>com.bloomberg.hbase.coprocessor.endpoint.AverageEndPoint</value>
</property>
```

Static Loading



Static Loading – Key points to remember

- Active on all Region Servers
 - All the regions of all the tables
- Multiple co-processors can be deployed
 - Provide a “,” separated list
- Update HBase’s class path
 - In `hbase-env.sh`
 - Copy jar to `hbase/lib` folder
- Must restart every Region Server
 - `hbase.dynamic.jars.dir` is for filters

Static Unloading

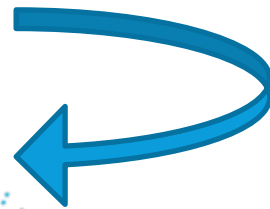
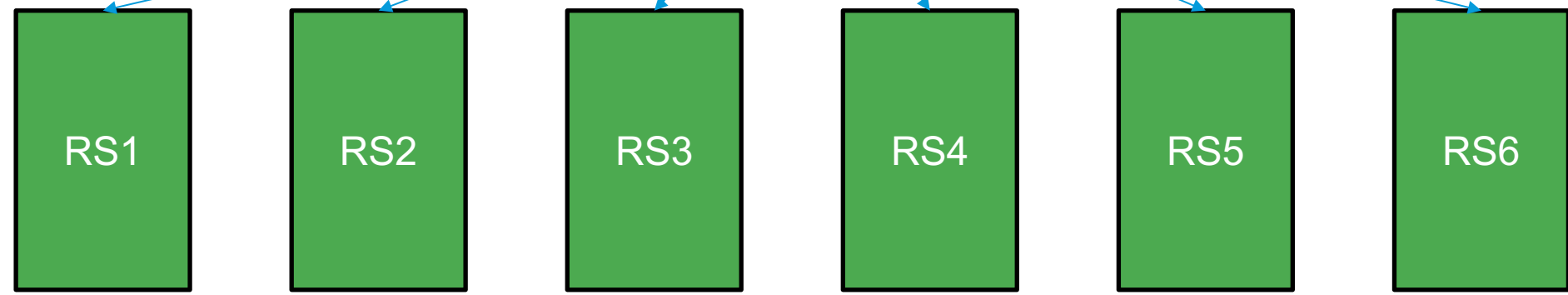
`hbase.coprocessor.region.classes`



`hbase.coprocessor.wal.classes`



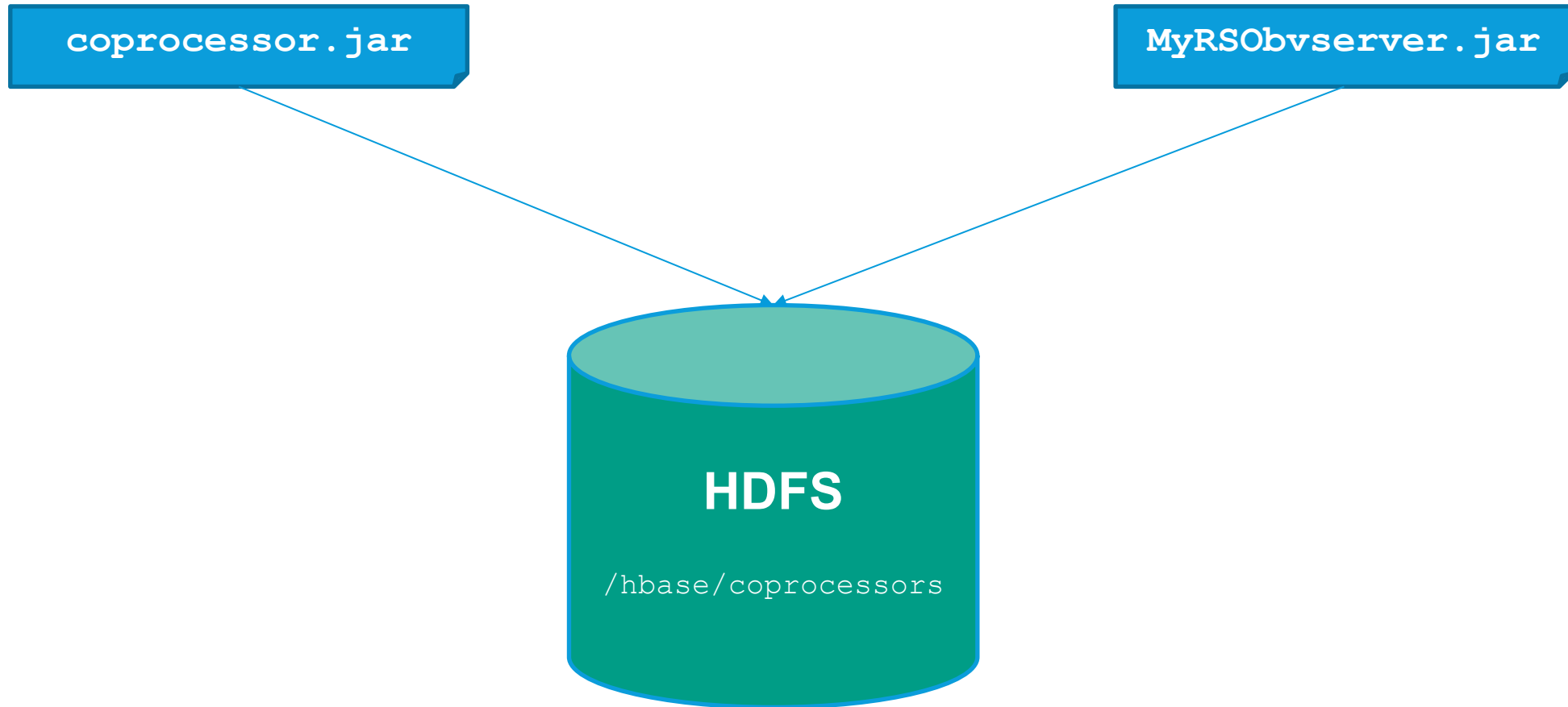
`hbase.coprocessor.master.classes`



Dynamic Loading

- Modification to `hbase-site.xml` is not required
- Loading is per table basis
 - Only available to the table they are loaded for
- Known as Table coprocessors
- Loaded via HBase shell
 - Admin API via Java
- JAR is stored on shared location
 - Usually HDFS
- Preferred method at Bloomberg

Dynamic Loading



Dynamic Loading: HBase shell – Admin/Developers

 The picture can't be displayed.

 The picture can't be displayed.

```
hbase(main):003:0> enable 'users'
```

Source: <http://hbase.apache.org/book.htm>

TechAtBloomberg.com

© 2018 Bloomberg Finance L.P. All rights reserved.

Bloomberg

Engineering

Dynamic Unloading: HBase shell – Admin/Developers

The picture can't be displayed.

```
hbase> alter 'users', METHOD => 'table_att_unset', NAME => 'coprocessor$1'
```

```
hbase(main):003:0> enable 'users'
```

Source: <http://hbase.apache.org/book.htm>

TechAtBloomberg.com

© 2018 Bloomberg Finance L.P. All rights reserved.

Bloomberg

Engineering

Dynamic Loading: Java API – Developers

```
TableName tableName = TableName.valueOf("users");
Path path = new Path("hdfs://<namenode>:<port>/user/<hadoop-user>/coprocessor.jar");
Configuration conf = HBaseConfiguration.create();
Connection connection = ConnectionFactory.createConnection(conf);
Admin admin = connection.getAdmin();
admin.disableTable(tableName);
HTableDescriptor hTableDescriptor = new HTableDescriptor(tableName);
HColumnDescriptor columnFamily1 = new HColumnDescriptor("personalDet");
columnFamily1.setMaxVersions(3);
hTableDescriptor.addFamily(columnFamily1);
HColumnDescriptor columnFamily2 = new HColumnDescriptor("salaryDet");
columnFamily2.setMaxVersions(3);
hTableDescriptor.addFamily(columnFamily2);
hTableDescriptor.addCoprocessor(RegionObserverExample.class.getCanonicalName(), path,
Coprocessor.PRIORITY_USER, null);
admin.modifyTable(tableName, hTableDescriptor);
admin.enableTable(tableName);
```

Source: <http://hbase.apache.org/book.htm>

TechAtBloomberg.com

© 2018 Bloomberg Finance L.P. All rights reserved.

Bloomberg

Engineering

Dynamic Unloading: Java API - Developers

```
TableName tableName = TableName.valueOf("users");
String path = "hdfs://<namenode>:<port>/user/<hadoop-user>/coprocessor.jar";
Configuration conf = HBaseConfiguration.create();
Connection connection = ConnectionFactory.createConnection(conf);
Admin admin = connection.getAdmin();
admin.disableTable(tableName);
HTableDescriptor hTableDescriptor = new HTableDescriptor(tableName);
HColumnDescriptor columnFamily1 = new HColumnDescriptor("personalDet");
columnFamily1.setMaxVersions(3);
hTableDescriptor.addFamily(columnFamily1);
HColumnDescriptor columnFamily2 = new HColumnDescriptor("salaryDet");
columnFamily2.setMaxVersions(3);
hTableDescriptor.addFamily(columnFamily2);
admin.modifyTable(tableName, hTableDescriptor);
admin.enableTable(tableName);
```

Note: In HBase 0.96 and newer, you can instead use the `removeCoprocessor()` method of the `HTableDescriptor` class.

Source: <http://hbase.apache.org/book.htm>

Comparison

Description	Static Loading	Dynamic Loading
Changes to hbase-site.xml	Yes	No
Restart region servers	Yes	No
Jar location	Local filesystem	HDFS
Coprocessor availability	Global	Per table
Loaded via hbase shell	No	Yes
Loaded via java API	No	Yes
Read permissions	HBase	HBase
Management complexity	High	Low

Challenges

- Version compatibility can cause failures
- Rollout of non-backward-compatible coprocessor difficult
- Clean-up can be messy HBASE-14190 - Assign system tables ahead of user region assignment

Solutions

- Compile code against version deployed on the cluster
 - Keep your developers informed
 - Work with developers
 - Review their code
- Test version changes in development environment

Challenges

- User deployment is dangerous
 - HDFS Permissions Changes
 - Crashes entire cluster
 - If `hbase.coprocessor.abortonerror` is set to true
- Bringing up cluster is challenging
 - Requires manual intervention
 - Missing jar will bring down the Region Server

ERROR

```
org.apache.hadoop.hbase.coprocessor.CoprocessorHost: The coprocessor  
fooCoprocessor threw java.io.FileNotFoundException: File does not exist:  
/path/to/corprocessor.jar  
java.io.FileNotFoundException: File does not exist: /path/to/corprocessor.jar
```

Solutions

- Create a common shared directory under `/hbase (/hbase/coprocessors)`
- Use automation to deploy user coprocessor to shared location

TechAtBloomberg.com

© 2018 Bloomberg Finance L.P. All rights reserved.

Bloomberg

Engineering

Deployment Guidelines

- Enable user coprocessors
 - Set `hbase.coprocessor.user.enabled` to `true`
 - Phoenix coprocessors are treated as user coprocessor
- Set `hbase.coprocessor.enabled` to `true`
 - Keeps system coprocessors enabled (AccessController)
- Enable coprocessor white listing (HBASE-16700)
 - Set `hbase.coprocessor.region.whitelist.paths`
 - Specify each directory individually
 - Wildcards won't include subdirectories (documentation says otherwise)
 - Entire filesystem, i.e., `hdfs://Test-Laptop`, won't work (documentation says otherwise)

Deployment guidelines

- Create a bundled jar
 - Includes all coprocessors
 - Single entry in HBase classpath
- Use automation
 - Chef/Puppet/Ansible

`hbase.coprocessor.enabled`

`hbase.coprocessor.user.enabled`

`hbase.coprocessor.regionserver.classes`

`hbase.coprocessor.region.classes`

`hbase.coprocessor.user.region.classes`

`hbase.coprocessor.master.classes`

`hbase.coprocessor.wal.classes`

`hbase.coprocessor.abortonerror`

`hbase.coprocessor.region.whitelist.paths`

Takeaways

TechAtBloomberg.com

© 2018 Bloomberg Finance L.P. All rights reserved.

Bloomberg

Engineering

Recap

- Coprocessors are necessary
 - Phoenix
 - HBase security
- User coprocessors are dangerous
 - Write defensive code
 - Be careful with deployment
- Make use of HBASE-16700
- Cleanup can be messy
 - HBASE-14190 – Assign system tables ahead of user region assignment

Needed from the Community

- Story for coprocessor deployment
- Process isolation
- JMX metrics

Thank You!

Reference: <http://hbase.apache.org>

Chef Code: <https://github.com/bloomberg/chef-bach.git>

Connect with Hadoop Team: hadoop@bloomberg.net

Engineering

Bloomberg

TechAtBloomberg.com

© 2018 Bloomberg Finance L.P. All rights reserved.

We are hiring!

<https://www.bloomberg.com/careers>

Questions?

TechAtBloomberg.com

© 2018 Bloomberg Finance L.P. All rights reserved.

Bloomberg
Engineering