



November 2011

# Scratching your own itch

Joey Echeverria | Solutions Architect

joey@cloudera.com

@fwiffo

cloudera

# About Joey

- Solutions Architect
- 6 months
- 3+ years

# Why Apache Sqoop (incubating)?

- Data center integration
- Extensible
  - Sources
  - Destinations
- Company specific
- Open source
  - Yay!

# Getting the code

- Apache SVN
  - <https://svn.apache.org/repos/asf/incubator/sqoop/>
- Apache GIT
  - <http://git.apache.org/sqoop.git>
- GitHub
  - <https://github.com/apache/sqoop.git>

# Fork you

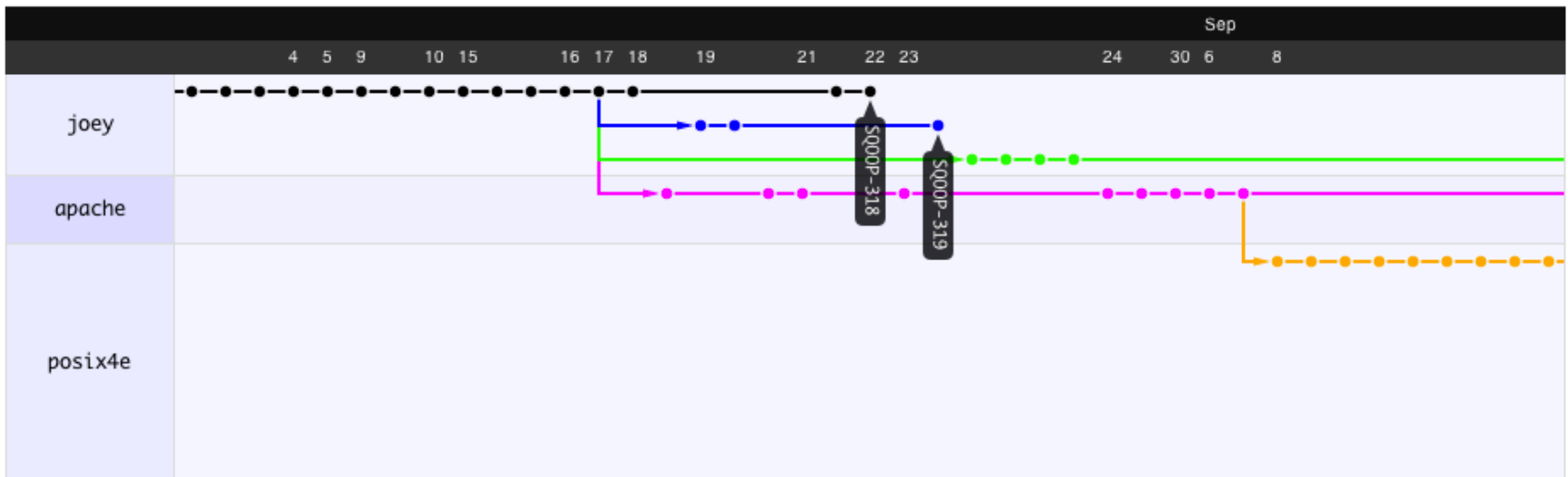
## The sqoop network graph

Keyboard shortcuts available 

All branches in the network using joey/sqoop as the reference point. [Read our blog post about how it works.](#)

Show Help

Last updated: about an hour ago



# Building the code

- Requirements
  - Ant 1.7.1
  - JDK 1.6
  - Maven
    - SQOOP-306
- Targets
  - jar
  - test
  - checkstyle

# Checkstyle

## CheckStyle Audit

Designed for use with [CheckStyle](#) and [Ant](#).

Summary	
Files	Errors
423	0

Files	
Name	Errors
<a href="#">/Users/joey/src/apache/sqoop/src/java/com/cloudera/sqoop/ConnFactory.java</a>	0
<a href="#">/Users/joey/src/apache/sqoop/src/java/com/cloudera/sqoop/Sqoop.java</a>	0
<a href="#">/Users/joey/src/apache/sqoop/src/java/com/cloudera/sqoop/SqoopOptions.java</a>	0
<a href="#">/Users/joey/src/apache/sqoop/src/java/com/cloudera/sqoop/cli/RelatedOptions.java</a>	0
<a href="#">/Users/joey/src/apache/sqoop/src/java/com/cloudera/sqoop/cli/SqoopParser.java</a>	0
<a href="#">/Users/joey/src/apache/sqoop/src/java/com/cloudera/sqoop/cli/ToolOptions.java</a>	0

# IDE

The screenshot shows an IDE window with the following components:

- Project Explorer:** Shows the 'Apache Sqoop' project structure with source packages like `com.cloudera.sqoop`.
- Members View:** Shows the methods of the `TableDefWriter` class, including `getColumnNames()`, `getCreateTableStmt()`, and `getHiveOctalCharCode(int)`.
- Code Editor:** Displays the `TableDefWriter.java` file. The code includes a conditional block for handling different codecs, with `CodecMap.LZOP` highlighted. The code is as follows:

```
197 b.append(" LINES TERMINATED BY '");
198 b.append(getHiveOctalCharCode((int) options.getOutputRecordDelim()));
199 tring codec = options.getCompressionCodec();
200 f (codec != null && (codec.equals(CodecMap.LZOP)
201 || codec.equals(CodecMap.getCodecClassName(CodecMap.LZOP)))) {
202 sb.append(" STORED AS INPUTFORMAT "
203 + "com.hadoop.mapred.DeprecatedLzoTextInputFormat");
204 sb.append(" OUTPUTFORMAT "
205 + "org.apache.hadoop.hive.ql.io.HiveIgnoreKeyTextOutputFormat"
206 else {
207 sb.append(" STORED AS TEXTFILE");
208
209
210 OG.debug("Create statement: " + sb.toString());
211 return sb.toString();
212
```
- Output Console:** Shows the compilation process with the following error message:

```
Copying 3 resources
[compiler:compile]
Compiling 338 source files to /Users/joey/src/apache/sqoop/target/classes
COMPILATION ERROR :
org/apache/sqoop/tool/VersionTool.java:[36,41] cannot find symbol
```



# Testing

- JUnit
- By hand
  - ant package
  - cd build/sqoop-1.5-incubating-SNAPSHOT
  - bin/sqoop

# Contributing back: Why?

- Next release
- Lower long-term cost
- For the community!
- It's easier than you think!

# Fears

- People won't want my patch
  - Rare
  - Small projects
- Only I need this feature
  - That's what everyone says...
- Too much process!
  - A little time now, a massive headache later

# Sqoop vs. HDFS

- Small project
  - Less contentious
  - Easier to grok the code
  - Easy deployment
- Immediate feedback
- Great place to start

# Contributing back: How?

- File a JIRA
  - <https://issues.apache.org/jira/browse/SQOOP>
- Post your patch on the issue
- File a review request
  - <https://reviews.apache.org/groups/Sqoop/>

# Posting patches

Status: Open

### Attach Files

Attachment  SQOOP-318-2.patch

No file chosen  
The maximum file upload size is 10.00 MB.

Comment

Viewable by All Users

Contributions intended for inclusion in ASF products (eg. patches, code) must be licensed to ASF under the terms of the [Apache License](#). Other attachments (eg. log dumps, test cases) need not be.

Attachment not intended for inclusion

Grant license to ASF for inclusion in ASF works (as per the [Apache License §5](#))

# The review process

- Upload your patch
- Describe your changes
- Describe your testing
- Wait for feedback
- Repeat

# What else?

- Subscribe to the mailing list
  - [sqoop-dev@incubator.apache.org](mailto:sqoop-dev@incubator.apache.org)
- Look for open issues
  - <https://issues.apache.org/jira/browse/SQOOP>
- Get hacking



# Questions?

