Lessons Learned from Open-source Activities

Chiwan Park (chiwanpark@apache.org)

2015. 11. 17
Outlines

1. Speaker Introduction
2. Global Open Frontier Program
3. Projects
4. Open-source Software as a Studying Method
5. Lessons Learned from Open-source Activities
6. Summary
Speaker

- A undergraduate student majoring Geology and Computer Science
- Contributing to some open-source software projects for 2 years
  - A committer of Apache Flink (since Jun. 2015)
- Interested in large-scale data processing, scalable machine learning, and graph processing
- Supported by Global Open Frontier program (since Dec. 2013)
Global Open Frontier Program (http://devlab.oss.kr)

• Since Dec. 2013
• Biggest supporting program for open-source contributors in Korea
• 36 open-source contributors are supported by this program.
  • Apache Flink – Platform for distributed stream and batch processing
  • PacketNgin – Real-time O/S for networking application on x86 arch
  • Haroopad – Markdown editor supporting multi-platforms
  • AxisJ – Javascript UI library
  • UrQA – Crash reporting library for mobile
  • … others
Global Open Frontier Program (cont.)

- Supports
  - Grant-in-aid
  - Office for contributors
  - Venue for seminars
  - Mentoring
  - Cloud environment
  - Books
Flamingo (http://github.com/OpenCloudEngine/flamingo2)

- UI Front-end for Apache Hadoop Eco-system
- Users can use many Apache Hadoop components without CLI.
  - Hadoop MapReduce
  - HDFS
  - Apache Hive, Apache Pig
  - Apache Spark
  - ⋮ more
Apache Flink (http://flink.apache.org)

- An open-source platform for distributed stream and batch processing
- The core of Flink is a distributed streaming dataflow engine.
- Flink provides programming abstractions to deal data easily.
  - DataSet API (for batch)
  - DataStream API (for streaming)
- Flink provides also libraries for machine learning and graph processing.
Open-source Software as a Studying Method

- Many students want to know where the knowledge from class is used.
- Small projects during the class are insufficient to experience real world.
- Contributing to OSS could be nice method to meet real world.
- The students can learn the followings from OSS.
  - How to use version control software (such as git, hg, and svn)
  - How to discuss with colleague
  - How to write documentation for software
  - … more
- Only few schools in Korea started contributing to OSS in their classes.
  - NHN Next
  - Kookmin University
  - Jeju University
  - … more
Lessons Learned from Open-source Activities

• Advanced git usage
  • Cherry-pick, Rebase, Squash commits, Edit
• Sending patch to project
  • If the project uses GitHub, use pull request feature.
  • Or make patch file and send it.
• How to discuss based on mailing list
• How to deal with issues using issue tracker
Lessons Learned from Open-source Activities (cont.)

- Attending international conferences
  - Apache: Big Data Europe 2015
    (http://events.linuxfoundation.org/events/apache-big-data-europe/)
  - ApacheCon: Core Europe 2015
    (http://events.linuxfoundation.org/events/apachecon-core-europe/)
  - Flink Forward 2015 (http://flink-forward.org)
- From the conferences, I learned how people work with open-source.
  - How much time they spend to contribute open-source project
  - How they use open-source software in their products
  - What differences between open-source and open-development
  - … more
Lessons Learned from Open-source Activities (cont.)

Apache: Big Data Europe 2015 and ApacheCon: Core Europe 2015
Lessons Learned from Open-source Activities (cont.)

Apache: Big Data Europe 2015 and ApacheCon: Core Europe 2015
Lessons Learned from Open-source Activities (cont.)

• Hive Query Editor in Flamingo 1.x
  • Large-size data transfer method between client and server
    • Thrift Protocol or JDBC
• Fixing memory bug in Flink runtime (FLINK-2076)
  • How does distributed system manage there memory efficiently?
  • How can I implement hash equi-join in distributed manner?
• Improving CSV file reader for Flink (FLINK-1512, FLINK-2061, FLINK-2569)
  • Automatic type extractor using Java Reflection API
  • Reducing memory usage by avoiding object creation
Lessons Learned from Open-source Activities (cont.)

- Adding Scala 2.11 support to Flink (FLINK-2200, FLINK-2767)
  - Binary incompatibility between Scala 2.10 and Scala 2.11
  - Build configuration by Maven properties
- Contributions to FlinkML (Machine learning library based on Flink)
  - How to implement machine learning algorithm in distributed manner
  - How to make the algorithm scalable
Summary

• Because of good practices and nice community in open-source software, contributing to open-source software is very good method to study for students.
• We need more programs or classes to teach how open-source works to students.
  • Supporting program such as Global Open Frontier Program would be good.
• I hope that many schools include contributing to open-source software into their classes.
  • Newbies will be helpful to grow open-source communities.
Thank you for listening!