Set-up continuous integration for Apache RYA (incubating)

GSoC 2016

Goals

- Setup continuous integration and automatic build for RYA using Jenkins (in the original proposal but already done)
- Setup automatic pre-commit checks (helpful for committers and contributors)
- Make it easier to run performance tests and compare the results for different back-ends, data sets, software configurations, hardware configurations
- Integrate performance testing in the automatic builds

Current plan for pre-commit checks

- Investigate use of Jenkins to do the build checks pre-commit, when a new pull request is made
- Investigate use of Apache Yetus test-patch for the pre-commit checks
- Investigate use of Travis CI
- Decide on solution/tools and implement it

Current plan for performance testing

- Identify queries to test performance
- Run benchmark queries
- Write Jmeter scripts
- Integrate performance tests with Rya builds

Identify queries to test performance

- Identify important queries for Rya users
  - LUBM benchmark queries
  - Other queries? (include geo, temporal, etc)
- Data sources for testing
  - LUBM data generator for LUBM queries
  - What data for other queries?
- We use these queries for benchmarking

Running benchmark queries

- By running those identified queries, we can get an idea about the current performance, eg: how many queries can run in 1 sec.
- We need to run benchmark queries for
  - Different instances that Rya use
  - Different data sizes
  - Different software configurations (indexes, …)
  - Different hardware configurations - if possible

Write Jmeter[1] scripts

- We can use Jmeter to run queries and get their success rate.

eg:10,50,100.. queries per second.

- We can determine the minimum performance on that.
- Then we write Jmeter scripts based on that performance.

[1]http://jmeter.apache.org/

Integrating performance tests

- We use those Jmeter scripts in Jenkins.
- We can test the performances continuously.
- Travis CI could be used instead of Jenkins, but not yet clear if it has any advantages

Other tools and ideas?

Automate pre-commit checks - Apache Yetus Test-Patch http://yetus.apache.org/documentation/0.2.1/precommit-basic/

Automate setting up Accumulo and Fluo - Zetten https://github.com/fluo-io/zetten

Performance testing script

BSBM driver http://wifo5-03.informatik.uni-mannheim.de/bizer/berlinsparqlbenchmark/