

# Exercises

---

1. **[Paper Review]** Pick one of the following papers and write a one-page paper review. The review should summarize the paper while highlighting three strong points and three points of improvements that you can think of:
  - An Industrial Case Study on the Automated Detection of Performance Regressions in Heterogeneous Environments (*icse2017.pdf*)
  - Characterizing and Detecting Anti-patterns in the Logging Code (*icse2015.pdf*)
2. **[Programming Exercise]** Open source software engineers use a set of tools while they develop, test and maintain their software systems. For example, they use source code repositories (e.g., SVN or Git) to archive their historical revisions and use bug tracking systems (e.g., Bugzilla or Jira) to report, communicate and fix their issues. In this exercise, you are provided with the source code revision logs and bug reports of HBase (<https://hbase.apache.org/>): (a) *SVN logs*: The data is stored in the zip file named “hbaseSVNLog.zip”. There are two files corresponds to two types of svn log outputs: the text format and the XML format. For more information, you can look into (<http://svnbook.red-bean.com/en/1.7/svn.ref.svn.c.log.html>); and (b) *Jira Bug reports*: The data is stored in the zip file named “hbaseBugReport.zip”. There are many XML files inside, each of which corresponds to one bug report. For example, the bug report “bugReport\_11076.xml” corresponds to HBase bug report (<https://issues.apache.org/jira/browse/HBASE-11076>). In this exercise, you need to analyze these datasets and answer the following three questions:
  - 1) In the SVN logs, which committer submitted the largest number of code revisions?
  - 2) For the bug reports which are resolved (a.k.a., have a valid time in the <resolved> tab), what is the maximum bug resolution time?
  - 3) Most of the code revisions have also included the Jira issue ID, which the code revision addresses. For example, Revision 1595269 has the following message “*HBASE-11076 Update refguide on getting 0.94.x to run on hadoop 2.2.0+ (Gustavo Anatoly)*”. This means that this code revision tries to address the Jira issue HBASE-11076 (a.k.a., bugReport\_11076.xml). However, one Jira issues might involve more than one code revisions. Furthermore, some Jira issues might do not have any code revisions associated with it at all. Hence, your task here is to report the minimum, average and maximum number of code revisions per Jira reports.

Submit your code (if you have written any) and compose a maximum two-page summary report outlining

- your approaches to tackle the above three questions; and
- the analysis results for the above three questions.