

This dataset contains commit records and issue reports for the 18 Apache projects. The commit records are packaged in **1-CommitRecords-part1.zip** and **1-CommitRecords-part2.zip**, and the issue reports are packaged in **2-IssueDatabase.zip**.

(1) The fields of a commit are described as follows:

- **Revision:** the hash code of the commit, the unique identifier of the commit.
- **Author:** the name and email address of the committer of the commit.
- **Date:** the date and time when the commit is performed.
- **Message:** the message of the commit, usually explaining what and why to make the commit.
- **----**: the marker in the commit to separate the general information of the commit and the detailed changes in the commit.
- **Added:** a source file is added in the commit. Example: “Added: src/couchdb/couch_httpd.erl|832 0” means that source file “src/couchdb/couch_httpd.erl” is added, 832 lines of code are added, and 0 lines of code are deleted.
- **Deleted:** a source file is deleted in the commit. Example: “Deleted: src/couch_inets/httpc_handler.erl|0 953” means that source file “src/couch_inets/httpc_handler.erl” is deleted, 0 lines of code are added, and 953 lines of code are deleted.
- **Modified:** a source file is modified in the commit. Example: “Modified: src/couchdb/couch_db.erl|133 86” means that source file “src/couchdb/couch_db.erl” is modified, 133 lines of code are added, and 86 lines of code are deleted.
- **Renamed:** a source file is renamed in the commit. Example: “Renamed: src/couchdb/couch_replicator_worker.erl [from src/couchdb/couch_replicator_doc_copier.erl]|48 45” means that source file “src/couchdb/couch_replicator_doc_copier.erl” is renamed into “src/couchdb/couch_replicator_worker.erl”, 48 lines of code are added, and 45 lines of code are added.

(2) In the issue database, the issues of each project are stored in a separated table. The fields of an issue report in each table are briefly explained as follows:

- **Project:** the project name.
- **Key:** the ID of the issue.
- **IssueType:** the type (e.g., bug) of the issue.
- **Status:** the status (e.g., Resolved) of the issue.
- **Priority:** the priority (e.g., Major) of the issue.
- **Resolution:** the resolution (e.g., Fixed) of the issue.
- **Created:** the created time of the issue.
- **Resolved:** the resolved time of the issue.
- **IsReopened:** Whether the issue has experienced the REOPENED status.

- (3) To replicate our study, please:
- a) first understand and follow the definitions of the data items described in Table 2 and Table 3 in the paper;
 - b) then handle the data according to Step 5 – Step 8 in Section 3.4 in the paper to obtain the data items;
 - c) and finally perform appropriate statistical tests (the Mann-Whitney U test or Chi-squared test), correct the resulting *p-values* of the tests through the Benjamini-Hochberg procedure, and calculate the effect sizes for the tests (the Pearson's Correlation Coefficient r or the odd ratio), following Section 3.5 in the paper.