

# Bug-Fix Dataset Description

This dataset brings information regarding the static and dynamic perspectives of 10 years (since 2009-01-01 until 2019-01-02) of bug fixing activity in 55 different projects from 9 categories (big-data, database, cloud, network-server, web-framework, security, build-management, library, and machine-learning) of Apache Software Foundation. The static perspective describes for each bug opened and fixed in this a period of time a snapshot of the bug fixing report and the changes performed in the source code. The second one brings information regarding the bug fixing activity over time, tracking the events (changes in the bug report, added comments, and performed commits) occurred since the bug was first reported until be considered fixed.

## 1. Project List

List of Target Projects				
Name	Id	Category	1st Release	Site
Hadoop Core	HADOOP	big-data	2006	<a href="http://hadoop.apache.org/">http://hadoop.apache.org/</a>
Hadoop Yarn	YARN	big-data	2012	<a href="http://hadoop.apache.org/">http://hadoop.apache.org/</a>
Hadoop HDFS	HDFS	big-data	2009	<a href="http://hadoop.apache.org/">http://hadoop.apache.org/</a>
Hadoop MapReduce	MAPREDUCE	big-data	2009	<a href="http://hadoop.apache.org/">http://hadoop.apache.org/</a>
Flink	FLINK	big-data	2015	<a href="https://flink.apache.org/">https://flink.apache.org/</a>
Spark	SPARK	big-data	2014	<a href="https://spark.apache.org/">https://spark.apache.org/</a>
Ozie	OOZIE	big-data	2012	<a href="http://oozie.apache.org/">http://oozie.apache.org/</a>
Kafka	KAFKA	big-data	2013	<a href="http://kafka.apache.org/">http://kafka.apache.org/</a>
Storm	STORM	big-data	2017	<a href="http://storm.apache.org/">http://storm.apache.org/</a>
Giraph	GIRAPH	big-data	2018	<a href="http://giraph.apache.org/">http://giraph.apache.org/</a>
Hive	HIVE	database	2010	<a href="http://hive.apache.org/">http://hive.apache.org/</a>
Cassandra	CASSANDRA	database	2009	<a href="http://cassandra.apache.org/">http://cassandra.apache.org/</a>
Lucene Core	LUCENE	database	2006	<a href="http://lucene.apache.org/">http://lucene.apache.org/</a>
HBase	HBASE	database	2015	<a href="https://hbase.apache.org/">https://hbase.apache.org/</a>
ZooKeeper	ZOOKEEPER	database	2011	<a href="https://zookeeper.apache.org/">https://zookeeper.apache.org/</a>
Derby	DERBY	database	2005	<a href="http://db.apache.org/derby/">http://db.apache.org/derby/</a>
OpenJPA	OPENJPA	database	2007	<a href="http://openjpa.apache.org/">http://openjpa.apache.org/</a>
Phoenix	PHOENIX	database	2014	<a href="http://phoenix.apache.org/">http://phoenix.apache.org/</a>
Libcloud	LIBCLOUD	cloud	2011	<a href="http://libcloud.apache.org/">http://libcloud.apache.org/</a>
jclouds	JCLOUDS	cloud	2013	<a href="http://jclouds.apache.org/">http://jclouds.apache.org/</a>
Mesos	MESOS	cloud	2013	<a href="http://mesos.apache.org/">http://mesos.apache.org/</a>
VCL	VCL	cloud	2013	<a href="http://vcl.apache.org/">http://vcl.apache.org/</a>
Helix	HELIX	cloud	2014	<a href="http://helix.apache.org/">http://helix.apache.org/</a>
Ignite	IGNITE	cloud	2015	<a href="https://ignite.apache.org/">https://ignite.apache.org/</a>
TomEE	TOMEE	network-server	2013	<a href="https://tomEE.apache.org/">https://tomEE.apache.org/</a>
Mina	DIRMINA	network-server	2006	<a href="http://mina.apache.org/">http://mina.apache.org/</a>

Mina FtpServer	FTPSEVER	network-server	2008	<a href="http://mina.apache.org/ftpserver-project/">http://mina.apache.org/ftpserver-project/</a>
Mina SSHD	SSHD	network-server	2009	<a href="http://mina.apache.org/sshd-project">http://mina.apache.org/sshd-project</a>
Mina Vysper	VYSPER	network-server	2010	<a href="http://mina.apache.org/vysper-project">http://mina.apache.org/vysper-project</a>
Camel	CAMEL	network-server	2009	<a href="http://camel.apache.org/">http://camel.apache.org/</a>
MyFaces	MYFACES	web-framework	2005	<a href="http://myfaces.apache.org/">http://myfaces.apache.org/</a>
Struts	WW	web-framework	2002	<a href="https://struts.apache.org/">https://struts.apache.org/</a>
Nutch	NUTCH	web-framework	2012	<a href="http://nutch.apache.org/">http://nutch.apache.org/</a>
Isis	ISIS	web-framework	2013	<a href="http://isis.apache.org/">http://isis.apache.org/</a>
Solr	SOLR	web-framework	2006	<a href="http://lucene.apache.org/solr/">http://lucene.apache.org/solr/</a>
Tapestry	TAP5	web-framework	2006	<a href="http://tapestry.apache.org/">http://tapestry.apache.org/</a>
Kerby	DIRKRB	security	2006	<a href="http://directory.apache.org/kerby/">http://directory.apache.org/kerby/</a>
Fortress	FC	security	2015	<a href="http://directory.apache.org/fortress/">http://directory.apache.org/fortress/</a>
Syncope	SYNCOPE	security	2012	<a href="http://syncope.apache.org/">http://syncope.apache.org/</a>
Archiva	MRM	build-management	2007	<a href="http://archiva.apache.org/">http://archiva.apache.org/</a>
Ivy	IVY	build-management	2006	<a href="http://ant.apache.org/ivy/">http://ant.apache.org/ivy/</a>
Maven	MNG	build-management	2012	<a href="https://maven.apache.org/">https://maven.apache.org/</a>
Buildr	BUILDR	build-management	2009	<a href="http://buildr.apache.org/">http://buildr.apache.org/</a>
Commons Compress	COMPRESS	library	2009	<a href="http://commons.apache.org/compress/">http://commons.apache.org/compress/</a>
Commons Collections	COLLECTIONS	library	2001	<a href="http://commons.apache.org/collections/">http://commons.apache.org/collections/</a>
Commons IO	IO	library	2004	<a href="http://commons.apache.org/io/">http://commons.apache.org/io/</a>
Commons Math	MATH	library	2004	<a href="http://commons.apache.org/math/">http://commons.apache.org/math/</a>
Commons Codec	CODEC	library	2003	<a href="http://commons.apache.org/codec/">http://commons.apache.org/codec/</a>
Commons Lang	LANG	library	2002	<a href="http://commons.apache.org/lang/">http://commons.apache.org/lang/</a>
Log4J 2	LOG4J2	library	2014	<a href="https://logging.apache.org/log4j/2.x/">https://logging.apache.org/log4j/2.x/</a>
Crunch	CRUNCH	library	2013	<a href="http://crunch.apache.org/">http://crunch.apache.org/</a>
Tika	TIKA	library	2011	<a href="http://tika.apache.org/">http://tika.apache.org/</a>
SystemML	SYSTEMML	machine-learning	2017	<a href="http://systemml.apache.org/">http://systemml.apache.org/</a>
Mahout	MAHOUT	machine-learning	2010	<a href="http://mahout.apache.org/">http://mahout.apache.org/</a>
MADlib	MADLIB	machine-learning	2017	<a href="http://madlib.apache.org/">http://madlib.apache.org/</a>

## 2. Static - Bug-Fix Dataset

Dataset location:

`./dataset/snapshot/`

Dataset file name format:

`<jira-project-id>-jira-bug-fix-dataset.csv` (only jira infos)

`<jira-project-id>-git-bug-fix-dataset.csv` (only git infos)

`<jira-project-id>-full-bug-fix-dataset.csv` (jira+git infos)

Fields Description of the Static Bug-Fix Dataset			
Field	From	Type	Description
Project	Jira	General	The parent project to which the bug belongs
Owner	Jira	General	The owner/maintainer of the project
Manager	Jira	General	The committee responsible for the project

Category	Jira	General	The project domain category
Key	Jira	General	A unique identifier for the bug
Priority	Jira	General	The importance of the issue in relation to other issues. Blocker – Highest priority. Indicates that this issue takes precedence over all others.   Critical – Indicates that this issue is causing a problem and requires urgent attention.   Major – Indicates that this issue has a significant impact.   Minor – Indicates that this issue has a relatively minor impact.   Trivial – Lowest priority.
Status	Jira	General	The stage the bug issue is currently at in its lifecycle
Reporter	Jira	General	The person who entered the bug into the system
Assignee	Jira	General	The person wh was assigned to fix the bug
Components	Jira	General	The project component(s) to which the bug relates
SummaryTopWords	Jira	Text	1000 top most frequent words of a brief one-line summary of the bug
DescriptionTopWords	Jira	Text	1000 top most frequent words of a detailed description of the bug
CommentsTopWords	Jira	Text	1000 top most frequent words of the comments of the bug
CreationDate	Jira	Time	The date and time on which the bug was entered into Jira
ResolutionDate	Jira	Time	The date and time on which the bug was resolved
AffectsVersions	Jira	Versioning	The project version(s) for which the bug is (or was) manifesting
FixVersions	Jira	Versioning	The project version(s) in which the bug was (or will be) fixed
NoComments	Jira	Summation	The number of all comments added in the bug
FirstCommentDate	Jira	Time	The date and time on which the first comment was added in the bug report
LastCommentDate	Jira	Time	The date and time on which the last comment was added in the bug report
NoWatchers	Jira	Summation	The number of how many people are watching (interested in) the bug
NoAttachments	Jira	Summation	The number of attachments (e.g., documents, images, and screenshots) added in the bug report
FirstAttachmentDate	Jira	Time	The date and time on which the first attachment was added in the bug report
LastAttachmentDate	Jira	Time	The date and time on which the last attachment was added in the bug report
NoAttachedPatches	Jira	Summation	The number of patches added in the bug report
FirstAttachedPatchDate	Jira	Time	The date and time on which the first patch was attached in the bug report
LastAttachedPatchDate	Jira	Time	The date and time on which the last patch was attached in the bug report
InwardIssueLinks	Jira	Link	The issues on which the bug depend on. (type:inwardIssue). Ex. "the bug is blocked by issue X"
OutwardIssueLinks	Jira	Link	The issues that depend on the bug. (type:outwardIssue). Ex. "the bug blocks issue X"
HasMergeCommit	Git	Versioning	This field informs whether a bug fix had or not a merge commit

CommitsMessagesTopWords	Git	Text	1000 top most frequent words of the commit messages related to the bug
NoCommits	Git	Summation	The number of commits related to the bug
NoAuthors	Git	Summation	The number of different authors who performed commits related to the bug (The author is the person who originally wrote the code)
NoCommitters	Git	Summation	The number of different committers who performed commits related to the bug (The committer is assumed to be the person who committed the code on behalf of the original author)
AuthorsFirstCommitDate	Git	Time	The date and time on which the author performed the first commit to fix the bug
AuthorsLastCommitDate	Git	Time	The date and time on which the author performed the last commit to fix the bug
CommittersFirstCommitDate	Git	Time	The date and time on which the committer performed the first commit to fix the bug
CommittersLastCommitDate	Git	Time	The date and time on which the committer performed the last commit to fix the bug
NonSrcAddFiles	Git	SRC	The number of non-source code files added by commits related to the bug
NonSrcDelFiles	Git	SRC	The number of non-source code files deleted by commits related to the bug
NonSrcModFiles	Git	SRC	The number of non-source code files modified by commits related to the bug
NonSrcAddLines	Git	SRC	The number of non-source code lines added by commits related to the bug
NonSrcDelLines	Git	SRC	The number of non-source code lines deleted by commits related to the bug
SrcAddFiles	Git	SRC	The number of source code files added by commits related to the bug
SrcDelFiles	Git	SRC	The number of source code files deleted by commits related to the bug
SrcModFiles	Git	SRC	The number of source code files modified by commits related to the bug
SrcAddLines	Git	SRC	The number of source code lines added by commits related to the bug
SrcDelLines	Git	SRC	The number of deleted source code lines by commits related to the bug
TestAddFiles	Git	SRC	The number of source code test files added by commits related to the bug
TestDelFiles	Git	SRC	The number of source code test files deleted by commits related to the bug
TestModFiles	Git	SRC	The number of source code test files modified by commits related to the bug
TestAddLines	Git	SRC	The number of source code test lines added by commits related to the bug
TestDelLines	Git	SRC	The number of source code test lines deleted by commits related to the bug

### 3. Dynamic - Bug-Fix Changelog Dataset

Dataset location:

./dataset/changelog/

Dataset file name format:

<jira-project-id>-bug-fix-changelog-dataset.csv

Fields Description of the Bug-Fix Change Log Dataset			
Field	From	Type	Description
Project	Jira	General	The parent project to which the bug belongs
Manager	Jira	General	The owner/maintainer of the project
Category	Jira	General	The project domain category (big-data, database, cloud, network-server, web-framework, security, build-management, library, and machine-learning)
Key	Jira	General	A unique identifier for the bug
Author	Jira	General	The people who made the comment
ChangeDate	Jira	Time	The date and time when the change was made in the bug report
Field	Jira	Location	The bug report field where the change happened
From	Jira	Text	The content stored in the field before the change
To	Jira	Text	The content stored in the field after the change

## 4. Dynamic - Bug-Fix Comments Log Dataset

Dataset location:

./dataset/comment-log/

Dataset file name format:

<jira-project-id>-bug-fix-comment-log-dataset.csv

Fields Description of the Bug-Fix Comments Log Dataset			
Field	From	Type	Description
Project	Jira	General	The parent project to which the bug belongs
Manager	Jira	General	The owner/maintainer of the project
Category	Jira	General	The project domain category (big-data, database, cloud, network-server, web-framework, security, build-management, library, and machine-learning)
Key	Jira	General	A unique identifier for the bug
Author	Jira	General	The people who made the comment
CreationDate	Jira	Time	The date and time when the comment was added in the bug report.
Content	Jira	Text	1000 top most frequent words of the added comment

## 5. Dynamic - Bug-Fix Commits Log Dataset

Dataset location:

./dataset/commit-log/

Dataset file name format:

<jira-project-id>-bug-fix-commit-log-dataset.csv

Fields Description of the Bug-Fix Commits Log Dataset			
Field	From	Type	Description
Project	Jira	General	The parent project to which the bug belongs
Manager	Jira	General	The owner/maintainer of the project
Category	Jira	General	The project domain category (big-data, database, cloud, network-server, web-framework, security, build-management, library, and machine-learning)
Key	Jira	General	A unique identifier for the bug
CommitHash	Git	Versioning	The hash code of the commit related to the bug
IsMergeCommit	Git	Versioning	This field informs whether a the commit is or not a merge commit
Author	Git	General	The name of the person who originally wrote the code to fix the bug
AuthorDate	Git	Time	The date and time when the author performed the commit
Committer	Git	General	The name of the person who committed the bug fix code on behalf of the original author
CommitterDate	Git	Time	The date and time on which the committer performed the commit
CommitMessageTopWords	Git	Text	1000 top most frequent words of the commit message
FileName	Git	SRC	The name of the file affected by the commit
FilePath	Git	SRC	The path of the file affected by the commit
ChangeType	Git	SRC	The type of the change made in the file (Added, Deleted, Modified, or Renamed)
IsSrcFile	Git	SRC	This field informs whether a the modified file is or not a source code file
IsTestFile	Git	SRC	This field informs whether a the modified file is or not a source code test file
AddLines	Git	SRC	The number of lines added from the modified file
Dellines	Git	SRC	The number of lines deleted from the modified file
NoMethods	Git	SRC	The number of methods in the source code file
LoC	Git	SRC	The number of lines of code of the modified source code file
CyC	Git	SRC	The cyclomatic complexity of the modified source code file
NoTokens	Git	SRC	The number of tokens of the modified file