Big Commit Analysis

Andreas Hohler
SCG Seminar Fall 2016
Supervised by Haidar Osman
Software Evolution
Software Evolution
What’s a Big Commit?

Example of a small bug fix commit\(^1\): “FLUME-2672 NPE in KafkaSourceCounter”

`flume-ng-core/src/main/java/org/apache/flume/instrumentation/kafka/KafkaSourceCounter....  CHANGED`

```java
@@ -31,7 +31,7 @@
    "source.kafka.empty.count"

32 32
    private static final String[] ATTRIBUTES =
33 33  -    {TIMER_KAFKA_COMMIT, TIMER_KAFKA_EVENT_GET};
34 34  +    {TIMER_KAFKA_COMMIT, TIMER_KAFKA_EVENT_GET, COUNTER_KAFKA_EMPTY};
35 35
36 36    public KafkaSourceCounter(String name) {
37 37        super(name, ATTRIBUTES);
```

1) https://issues.apache.org/jira/browse/FLUME-2672
What’s a Big Commit?

Example of a big commit with

- 3’945 LOC changed
- 73 files changed
- LUCENE-6271 issue classified as Bug

2) https://issues.apache.org/jira/browse/LUCENE-6271
Dataset

Projects from GitHub

Logos for Flume, Lucene, Nutch, Hibernate, elastic, and Solr.
Demo
Analysis

Project: Flume, Issue type: Feature

Additions per commit
Analysis

Project: Flume, Issue type: Feature

Files changed per commit
Analysis

Project: Flume, Issue type: Bug

Additions per commit
Analysis

Project: Flume, Issue type: Bug

Files changed per commit
### Issue types

- **Bug**
- **Feature**
- **Improvement**
- **Refactor**
- **Documentation**

### Feature (108 results)

<table>
<thead>
<tr>
<th>Metric</th>
<th>Value</th>
<th>Description</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of issues</td>
<td>108</td>
<td>Highest additions per commit</td>
<td></td>
</tr>
<tr>
<td>Median files changed (commits)</td>
<td>5</td>
<td>Lowest additions per commit</td>
<td></td>
</tr>
<tr>
<td>Median additions (commits)</td>
<td>322</td>
<td>Highest deletions per commit</td>
<td></td>
</tr>
<tr>
<td>Median deletions (commits)</td>
<td>8</td>
<td>Lowest deletions per commit</td>
<td></td>
</tr>
<tr>
<td>Median additions (files)</td>
<td>11</td>
<td>Highest additions per file</td>
<td></td>
</tr>
<tr>
<td>Median deletions (files)</td>
<td>0</td>
<td>Lowest additions per file</td>
<td></td>
</tr>
<tr>
<td>Highest Number of files changed per commit</td>
<td>323</td>
<td>Highest deletions per file</td>
<td></td>
</tr>
<tr>
<td>Lowest Number of files changed per commit</td>
<td>1</td>
<td>Lowest deletions per file</td>
<td></td>
</tr>
</tbody>
</table>

### Additions per commit

- Bar chart showing distribution of additions per commit.