



| Entity                | Description   | Raw data entity | Num Items  |
|-----------------------|---|-----------------|------------|
| —                     | Events on repositories.   | events          | 43,090,195 |
| projects              | Project repositories.   | repos           | 1,326,900  |
| users                 | Github users.   | users           | 793,855    |
| project_members       | Users with commit access to the referenced project.   | repo_collabs    | 983,629    |
| organization_members  | List of members in an organization.   | org_members     | 34,924     |
| commits               | A list of all commits on Github. The <code>project_id</code> field refers to the first <code>project</code> this commit has been added to.          | commits         | 29,978,291 |
| project_commits       | List of all commits to a project.   | —               | —          |
| commit_parents        | Commits that are parents to a commit.   | —               | —          |
| commit_comments       | Code review comments for a commit.  | commit_comments | 126,697    |
| watchers              | users that have starred (was watched) a project   | watchers        | 7,744,619  |
| followers             | users that are following another user.  | followers       | 1,797,343  |
| issues                | Issues that have been recorded for a project.   | issues          | 2,326,069  |
| issue_events          | Chronologically ordered list of events on an issue.   | issue_events    | 4,085,294  |
| issue_comments        | Discussion comments on an issue   | issue_comments  | 2,886,006  |
| pull_requests         | List of pull requests for <code>base_repo</code> . Requests originate at head <code>head_repo/commit</code> and are created by <code>user_id</code> | pull_requests   | 1,144,251  |
| pull_request_comments | Discussion comments on a <code>pull_request</code>  | pullreq_comnts  | 2,228,894  |
| pull_request_history  | Chronologically ordered list of events on a <code>pull_request</code>   | —               | —          |

Fig. 1. Schema entities, their description, the corresponding raw data entities and the number of raw data items (Feb 15, 2013).