# github数据集

## 数据说明

| name | type | description | payload_type |
|---|---|---|---|
| id | String | uique identity of this event instance | All |
| type | String | event instance type,event instances of different event type have different payload from GitHub API | All |
| action | String | action in the payload of event instance | All |
| actor_id | UInt64 | GitHub user id of someone who trigger this envent instance | All |
| actor_login | String | GitHub user login of someone who trigger this envent instance | All |
| repo_id | UInt64 | the unique identity of a repository | All |

| | | where this event was tragered | |
|---|---|---|---|
| repo_name | String | repository name | All |
| org_id | UInt64 | organization unique identity of this repository | All |
| org_login | String | login name of this organization | All |
| created_at | DateTime | when this event instance was generated on GitHub | All |
| created_date | Date | same as created_at but in Date format | All |
| issue_id | UInt64 | unique identity of this issue on GitHub | Issues, IssueComment, PullRequest, PullRequestRevie wComment |
| issue_number | UInt32 | the id of this issue in this repository | Issues, IssueComment, PullRequest, PullRequestRevie wComment |
| issue_title | String | titile of this issue | Issues, IssueComment, PullRequest, PullRequestRevie wComment |
| issue_body | String | body of this issue event | Issues, IssueComment, PullRequest, |

| | | | PullRequestRevie wComment |
|---|---|---|---|
| issue_labels | Nested | issue labels of the repository | Issues, IssueComment, PullRequest, PullRequestRevie wComment |
| issue_labels.name | String | issue label name | Issues, IssueComment, PullRequest, PullRequestRevie wComment |
| issue_labels.color | String | issue label color | Issues, IssueComment, PullRequest, PullRequestRevie wComment |
| issue_labels.defau lt | UInt8 | whether the label is a default label | Issues, IssueComment, PullRequest, PullRequestRevie wComment |
| issue_labels.descr iption | String | issue label description | Issues, IssueComment, PullRequest, PullRequestRevie wComment |
| issue_author_id | UInt64 | GitHub user id of this issue's author | Issues, IssueComment, PullRequest, PullRequestRevie wComment |
| issue_author_logi n | String | GitHub user login of this issue's | Issues, IssueComment, |

| | | author | PullRequest, PullRequestReviewComment |
|---|---|---|---|
| issue_author_type | String | GitHub user type of this issue's author | Issues, IssueComment, PullRequest, PullRequestReviewComment |
| issue_author_association | String | author association of issue author, like NONE, CONTRIBUTOR, COLLABORATOR, MEMBER, OWNER | Issues, IssueComment, PullRequest, PullRequestReviewComment |
| issue_assignee_id | UInt64 | assignee's GitHub user id of this issue | Issues, IssueComment, PullRequest, PullRequestReviewComment |
| issue_assignee_login | String | issue assignee's login name | Issues, IssueComment, PullRequest, PullRequestReviewComment |
| issue_assignees | Nested | all assignees of this issue | Issues, IssueComment, PullRequest, PullRequestReviewComment |
| issue_assignees.id | UInt64 | GitHub user id of the assignee | Issues, IssueComment, PullRequest, PullRequestRevie |

| | | | wComment |
|---|---|---|---|
| issue_assignees.login | String | GitHub user login of the assignee | Issues, IssueComment, PullRequest, PullRequestReviewComment |
| issue_created_at | DateTime | when this issue was created,UTC | Issues, IssueComment, PullRequest, PullRequestReviewComment |
| issue_updated_at | DateTime | when this issue was updated, UTC | Issues, IssueComment, PullRequest, PullRequestReviewComment |
| issue_comments | UInt16 | the number of issue comments | Issues, IssueComment, PullRequest, PullRequestReviewComment |
| issue_closed_at | DateTime | when this issue was closed | Issues, IssueComment, PullRequest, PullRequestReviewComment |
| issue_comment_id | UInt64 | issue comment unique identity | IssueComment |
| issue_comment_body | String | comment body in the issue comment event | IssueComment |
| issue_comment_created_at | DateTime | when this issue comment was created.UTC | IssueComment |

| | | | |
|---|---|---|---|
| issue_comment_updated_at | DateTime | when this issue comment was updated,UTC | IssueComment |
| issue_comment_author_association | String | author association of issue comment author, like NONE, CONTRIBUTOR, COLLABORATOR, MEMBER, OWNER | IssueComment |
| issue_comment_author_id | UInt64 | GitHub user id of the issue comment author | IssueComment |
| issue_comment_author_login | String | GitHub user login of this issue comment author | IssueComment |
| issue_comment_author_type | String | GitHub user type of this issue comment author | IssueComment |
| pull_commits | UInt16 | how many commits in this pull request | PullRequest, PullRequestReviewComment |
| pull_additions | UInt16 | additions in this pull request | PullRequest, PullRequestReviewComment |
| pull_deletions | UInt16 | deletions in this pull request | PullRequest, PullRequestReviewComment |
| pull_changed_files | UInt32 | how many files changed in this pull request | PullRequest, PullRequestReviewComment |
| pull_merged | UInt8 | whether this pull | PullRequest, |

| | | request is merged | PullRequestRevie wComment |
|---|---|---|---|
| pull_merged_com mit_sha | String | the merged commit SHA of the pull request | PullRequest, PullRequestRevie wComment |
| pull_merged_at | DateTime | when this pull request was merged | PullRequest, PullRequestRevie wComment |
| pull_merged_by_i d | UInt64 | GitHub user id of someone who merged this pull request | PullRequest, PullRequestRevie wComment |
| pull_merged_by_lo gin | String | GitHub user login of someone who merged this pull request | PullRequest, PullRequestRevie wComment |
| pull_merged_by_t ype | String | GitHub user type of someone who merged this pull request | PullRequest, PullRequestRevie wComment |
| pull_requested_re viewer_id | UInt64 | first requested reviewer id | PullRequest, PullRequestRevie wComment |
| pull_requested_re viewer_login | String | first requested reviewer login | PullRequest, PullRequestRevie wComment |
| pull_requested_re viewer_type | String | first requested reviewer type | PullRequest, PullRequestRevie wComment |
| pull_review_comm ents | UInt16 | the number of pull review comments when this event log generated on GitHub | PullRequest, PullRequestRevie wComment |

| | | GitHub | |
|---|---|---|---|
| repo_description | String | repository description | PullRequest, PullRequestReviewComment |
| repo_size | UInt32 | reository size when this event log generated from GitHub | PullRequest, PullRequestReviewComment |
| repo_stargazers_count | UInt32 | how many people star this project on GitHub | PullRequest, PullRequestReviewComment |
| repo_forks_count | UInt32 | how many people fork this repository when this event log generated from GitHb | PullRequest, PullRequestReviewComment |
| repo_language | String | the main programming language of this repository | PullRequest, PullRequestReviewComment |
| repo_has_issues | UInt8 | whether this project enables Issues feature on GitHub | PullRequest, PullRequestReviewComment |
| repo_has_projects | UInt8 | whether this project enables Project feature on GitHub | PullRequest, PullRequestReviewComment |
| repo_has_downloads | UInt8 | whether this project enables Downloads feature on GitHub, deprecated | PullRequest, PullRequestReviewComment |

| repo_has_wiki | UInt8 | whether this project enables Wiki feature on GitHub | PullRequest, PullRequestReviewComment |
|---|---|---|---|
| repo_has_pages | UInt8 | whether this project enables Pages feature on GitHub | PullRequest, PullRequestReviewComment |
| repo_license | String | open source license used by this repository | PullRequest, PullRequestReviewComment |
| repo_default_branch | String | default branch of this repository | PullRequest, PullRequestReviewComment |
| repo_created_at | DateTime | when this repository was created | PullRequest, PullRequestReviewComment |
| repo_updated_at | DateTime | when this repository was updated | PullRequest, PullRequestReviewComment |
| repo_pushed_at | DateTime | when this repository was pushed | PullRequest, PullRequestReviewComment |
| pull_review_id | UInt64 | unique identity of this review on GitHub | PullRequestReviewComment |
| pull_review_comment_id | UInt64 | unique identity of this review comment on GitHub | PullRequestReviewComment |
| pull_review_comment_path | String | the related review file path | PullRequestReviewComment |

| | | | |
|---|---|---|---|
| pull_review_comment_position | String | the related review file position | PullRequestReviewComment |
| pull_review_comment_author_id | UInt64 | GitHub user id of someone who made this pull review comment | PullRequestReviewComment |
| pull_review_comment_author_login | String | GitHub user login of someone who made this pull review comment | PullRequestReviewComment |
| pull_review_comment_author_type | String | GitHub user type of someone who made this pull review comment | PullRequestReviewComment |
| pull_review_comment_author_association | String | Review comment author association, like NONE, CONTRIBUTOR, COLLABORATOR, MEMBER, OWNER | PullRequestReviewComment |
| pull_review_comment_body | String | pull review comment body | PullRequestReviewComment |
| pull_review_comment_created_at | DateTime | when this pull review comment was created | PullRequestReviewComment |
| pull_review_comment_updated_at | DateTime | when this pull review comment was updated | PullRequestReviewComment |
| push_id | UInt64 | unique push id on GitHub | Push |
| push_size | UInt32 | The number of commits in the | Push |

| | | commits in the push? | |
|---|---|---|---|
| push_distinct_size | UInt32 | The number of distinct commits in the push? | Push |
| push_ref | String | the ref branch of the push | Push |
| push_head | String | the head of the push | Push |
| push_before | String | the head before this push | Push |
| push_commits | Nested | commits in this push event | Push |
| push_commits.name | String | commit author name | Push |
| push_commits.email | String | commit author email | Push |
| push_commits.message | String | commit message | Push |
| fork_forkee_id | UInt64 | GitHub user id of forkee | Fork |
| fork_forkee_full_name | String | repository full name of forked repository | Fork |
| fork_forkee_owner_id | UInt64 | GitHub user id of the forked repository's owner | Fork |
| fork_forkee_owner_login | String | GitHub user login of the forked repository's owner | Fork |
| fork_forkee_owner_type | String | GitHub user type of the forked | Fork |

| | | repository's owner | |
|---|---|---|---|
| delete_ref | String | the git ref resource | Delete |
| delete_ref_type | String | the type of git ref oject deleted in the repository. Can be branch or tag | Delete |
| delete_pusher_typ e | String | GitHub user type of the user that trigger this event. Can be user or deloykey | Delete |
| create_ref | String | the git ref resource | Create |
| create_ref_type | String | the type of Git ref object created in the repository. Can be either branch or tag | Create |
| create_master_br anch | String | the base branch of current create event | Create |
| create_descriptio n | String | the repository's current description | Create |
| create_pusher_typ e | String | GitHub user type of the user that trager this event | Create |
| gollum_pages | Nested | the pages modified in this event | Gollum |
| gollum_pages_nag | String | the page name | Gollum |

| gollum_pages.page_name | String | the page name | Gollum |
|---|---|---|---|
| gollum_pages.title | String | the page title | Gollum |
| gollum_pages.action | String | the action that was performed on the page. Can be created or edited. | Gollum |
| member_id | UInt64 | GitHub user id of member | Member |
| member_login | String | GitHub user login of member | Member |
| member_type | String | GitHub user type of user | Member |
| release_id | UInt64 | unique release id on GitHub | Release |
| release_tag_name | String | tag name of this release | Release |
| release_target_commitish | String | tag related branch head | Release |
| release_name | String | name of this release | Release |
| release_draft | UInt8 | whether this is a draft, true or false | Release |
| release_author_id | UInt64 | GitHub user id of the release author | Release |
| release_author_login | String | GitHub user loginof the release author | Release |
| release_author_type | String | GitHub user type of the release author | Release |
| release_prereleas | UInt8 | whether this is a | Release |

| release_prereleas e | UInt8 | whether this is a prerelease | Release |
|---|---|---|---|
| release_created_a t | DateTime | when this release was created | Release |
| release_published _at | DateTime | when this release was published | Release |
| release_body | String | text content body of this release | Release |
| release_assets | Nested | the assets attached to the release | Release |
| release_assets.na me | String | the name of the asset | Release |
| release_assets.upl oader_id | UInt64 | GitHub user id of the uploader | Release |
| release_assets.upl oader_login | String | GitHub user login of the uploader | Release |
| release_assets.co ntent_type | String | file type of the asset | Release |
| release_assets.sta te | String | state of the asset | Release |
| release_assets.siz e | UInt64 | size of the asset | Release |
| release_assets.do wnload_count | UInt16 | download times of the asset | Release |
| commit_comment _id | UInt64 | unique commit comment id on GitHub | CommitComment |
| commit_comment _author_id | UInt64 | GitHub user id of the commit comment author | CommitComment |
| commit_comment | String | GitHub user login | CommitComment |

| commit_comment _author_login | String | GitHub user login of the commit comment author | CommitComment |
|---|---|---|---|
| commit_comment _author_type | String | GitHub user type of the commit comment author | CommitComment |
| commit_comment _author_associati on | String | the commit comment author association, like NONE, CONTRIBUTOR, COLLABORATOR, MEMBER, OWNER | CommitComment |
| commit_comment _body | String | text content body of this commit comment | CommitComment |
| commit_comment _path | String | the commit comment file path | CommitComment |
| commit_comment _position | String | the commit comment position when shows on page, not related important | CommitComment |
| commit_comment _line | String | the commit comment file line | CommitComment |
| commit_comment _created_at | DateTime | when this commit comment was created | CommitComment |
| commit_comment _updated_at | DateTime | when this commit comment was updated | CommitComment |

# ods层数据集

更新时间2022年1月10日

github日志数据集在clickhouse和maxcompute上面都有，由opendigger项目每天先放到clickhouse，然后每月同步到maxcompute，maxcompute数据集按月进行分区，分区值为每月1号，例如

```
# 查询2015年的数据
SELECT count(*) FROM ods_github_log WHERE pt>='20150101' and pt<'20160101';
# 查询2015年1月的数据
SELECT count(*) FROM ods_github_log WHERE pt='20150101';
```

| 年 | 说明 | clickhouse | maxcompute |
|---|---|---|---|
| 2015 | 2015年github_log数据 | 212221628 | 212221627 |
| 2016 | 2016年github_log数据 | 320726236 | 320726236 |
| 2017 | 2017年github_log数据 | 412942763 | 412942763 |
| 2018 | 2018年github_log数据 | 479185525 | 479185525 |
| 2019 | 2019年github_log数据 | 605544846 | 605544846 |
| 2020 | 2020年github_log数据 | 863415606 | 863415598 |
| 2021 | 2021年github_log数据 | 998564686 | 998564683 |

| 表 | 说明 | mongoDB | maxcompute |
| --- | --- | --- | --- |
| ods_github_users | | | 10307 |
| ods_github_users | | | |

## dim层数据集

| 表 | 说明 | 数据量 |
| --- | --- | --- |
| dim_github_actor | 用户表 | 21135299 |
| dim_github_repo | 仓库表 | 213266709 |
| dim_github_org | 组织表 | 1828684 |
| dim_github_issue | issue表 | 279500657 |
| dim_github_pr | pr表 | 177141328 |