



1.http-trigger

1.1.job metadata

GCP



Cloud Function  
Crawler API

2.publish job chunks



Crawl topic  
Pub/Sub

3.topic-trigger

3.topic-trigger



Cloud Functions  
Crawl worker 1



Cloud Functions  
Crawl worker N

1.2.get status

4.upload crawl docs

4.upload crawl docs

5.status

5.status



Cloud bucket by job



Crawl jobs statuses  
Datastore