

About the Product:

Atlan Collect is a data collection platform that is being used by customers in 50+ countries in more than 200 organizations. Its features include team management, multilingual forms, and offline data collection. Our customers use Atlan Collect to power their most critical activities—from governments delivering vaccines to small business owners managing their daily inventory, to an international zoo monitoring a rare wildlife species.

The Challenge

Atlan Collect has a variety of long-running tasks that require time and resources on the servers. As it stands now, once we have triggered off a long-running task, there is no way to tap into it and pause/stop/terminate the task, upon realizing that an erroneous request went through from one of the clients (mostly web or pipeline).

Examples:

- **Example 1:** A support user starts a baseline upload for an organization. The baseline upload had 100000 rows in the CSV file uploaded, and the support user - right after clicking the “Start” button - realizes that he seemed to have wrongly exchanged data between two adjacent columns. Subsequently, the support user would need to wait until the current request completes, which can be in the magnitude of hours, and since we don't allow two concurrent baseline uploads currently, this implies he would not be able to upload the correct file till this task completes. Once done, the user would need to remove those 100,000 entries, and finally upload the correct file.
- **Example 2:** The pipeline sends an export request to Atlan Collect to update the data on dashboards. In one of the incidents triggered manually, the “from date” in the export request was (by mistake) set as “2015-08-01”, while it was meant to be set as “2017-08-01”. As a result, the export started unnecessarily for 20 million rows, whereas setting the right dates would have resulted in only 2 million rows. There is currently no way for me to stop the previous export, so that time and resources are concentrated upon this request to be completed.
- **Example 3:** A user on the web dashboard wanted to create teams in bulk. There were 10,000 teams to be created which would take an hour or more. They observed that mapping phone numbers to teams had been mistakenly shifted down ten rows, due to which inappropriate managers would be added to teams. Now the user has no way to stop the request.

We want to offer an implementation through which the user can now stop the long-running task at any given point in time, and can choose to resume or terminate it. This will ensure

that the resources like compute/memory/storage/time are used efficiently at our end, and do not go into processing tasks that have already been stopped (and then to roll back the work done post the stop-action)

Create Rest API endpoints where the problem mentioned in the above three examples can be solved and package the solution in a docker image that can be deployed on Kubernetes directly.

You can try out the product at <https://atlan.com/collect/> to get an understanding of the above problems.

****Please use any file with dummy data.**

All the best! :)