

# Data Engineer - Technical Exercise

This exercise is designed to get you acquainted with the Immuta product and its API -- something that the data engineer will be leveraging extensively. My preference is that you complete the exercise in Python, as that is the primary language that the Growth Analytics team uses for scripting and applications.

When you are finished, please send the script to me at [sbailey@immuta.com](mailto:sbailey@immuta.com). Please reach out if you have any questions or hit any stumbling blocks!

- Stephen Bailey

## Directions

First, you will spin up an Immuta trial instance as a playground.

1. Spin up a new Immuta free trial at [www.immuta.com/try](http://www.immuta.com/try)
2. Sign in and complete the "onboarding" exercises. There is an icon in the bottom left that is your guide to these 5 brief steps.

After completing the exercises, you should have access to a few data sources. These data sources should also have gone through our "sensitive data detection" process, resulting in tags applied to the data sources. For the next steps, you will need to leverage API endpoints that can be found on the instance's Swagger page:

<https://<hostname>.immuta.com/#/api>

3. Using the Python "requests" package, retrieve an authentication token from Immuta using API authentication. Example code is below:

```
import requests

session = requests.Session()
base_url = "https://my_url.com"

api_key = "<your_api_key>"
response = session.post(base_url + "/bim/apikey/authenticate",
data={"apikey": api_key})
auth_token = response.json()["token"]
session.headers["Authorization"] = f"Bearer {auth_token}"

# Example of using a Requests session to `GET`
```

```
session.get(base_url + "/dataSource/1").json()
```

4. Now, use the API to programmatically answer the following questions:
  - a. How many data sources are in your instance?
  - b. How many columns have tags on them?
  - c. What is the tag that has been most frequently applied?
5. As a thought experiment, how might you automate the extraction of this metadata into a relational database. What steps would be required? What decisions would need to be made? Please be prepared to share your thoughts in the technical interview.