Workflow:

A screenshot of a computer

Description automatically generated

GCP Services:

A close-up of a white text

Description automatically generated

We are using Mage which is an open source data pipeline tool for transforming and integrating data.

A white background with black text

Description automatically generated’Whenever we need information stored in dimensional tables, we can use foreign keys and then extract the information

Understand the data dictionary and the data before you start working with it.

Passenger count is usually in the fact table whereas for this project we can create a dimensional table

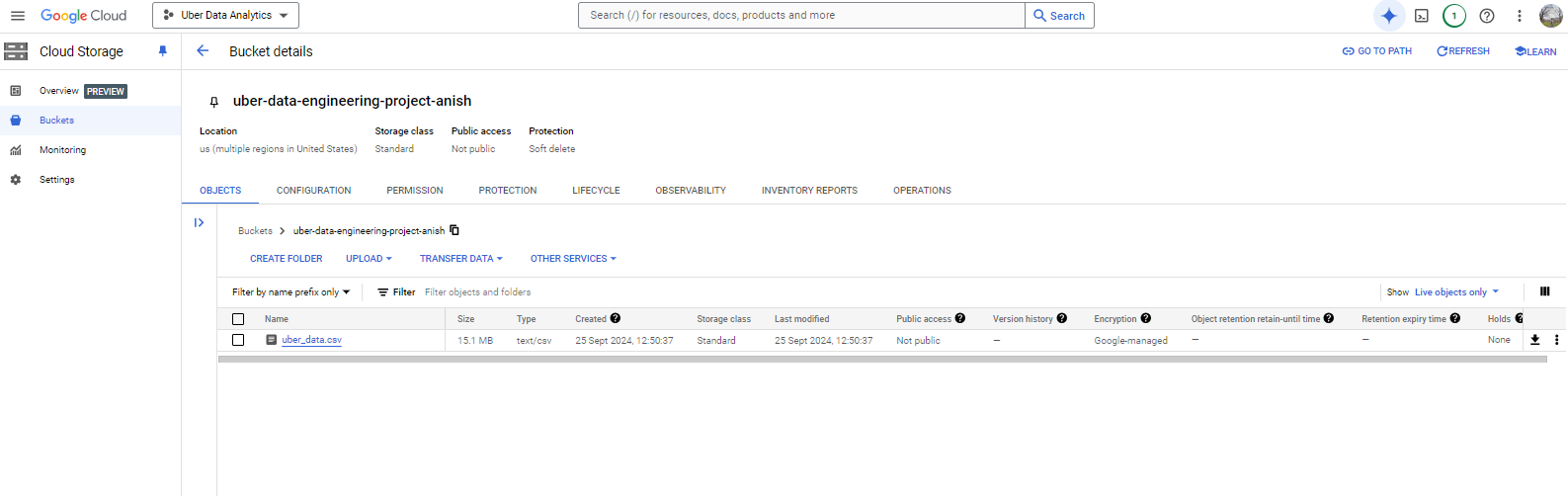
Data Model with dimensional tables and fact table:

A diagram of a computer

Description automatically generated

Using the python code we have created dimension tables and then using join we create the fact table

Creating a cloud storage bucket and uploading files into the bucket

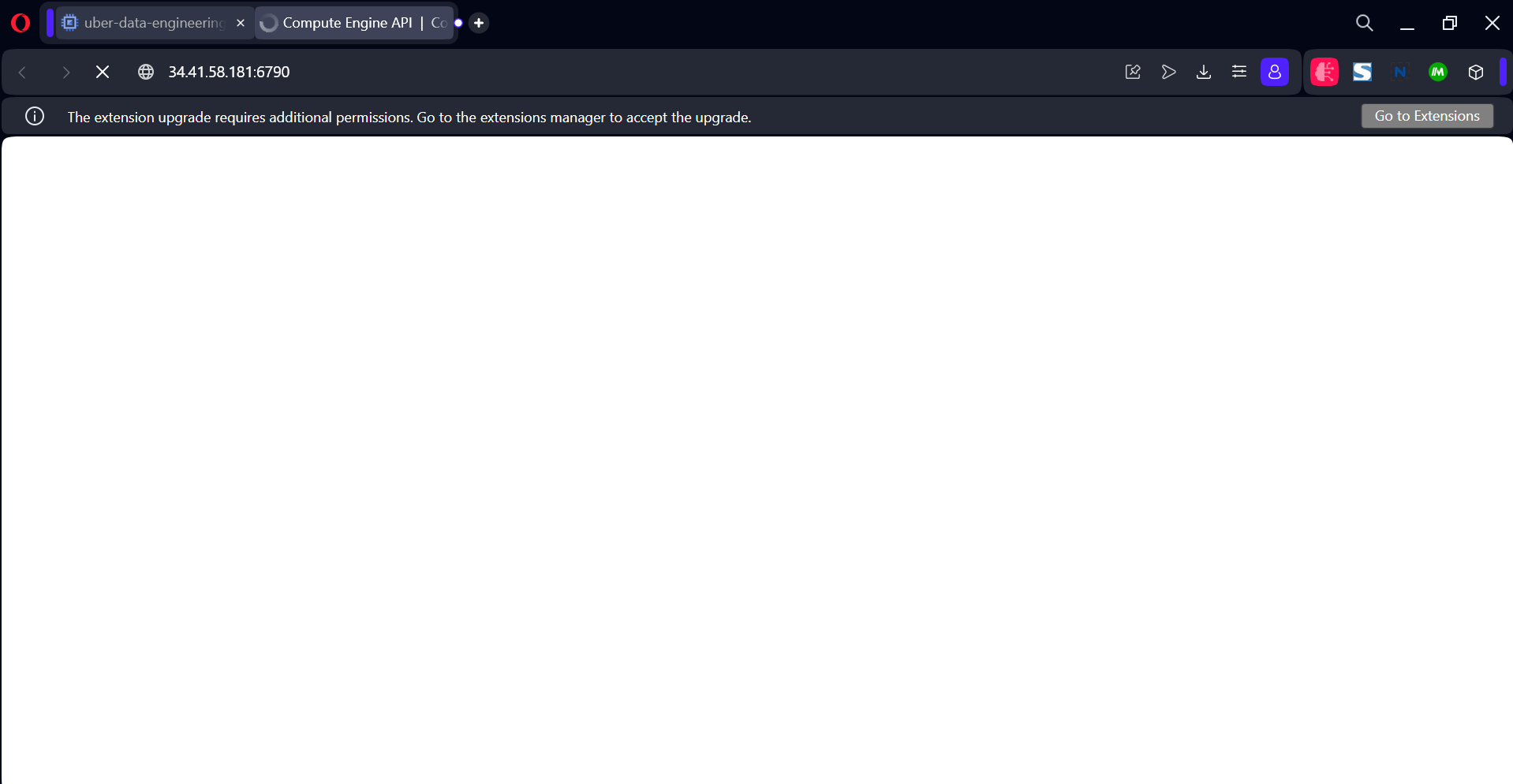


Change the [permissions to public so that we can access it from our system

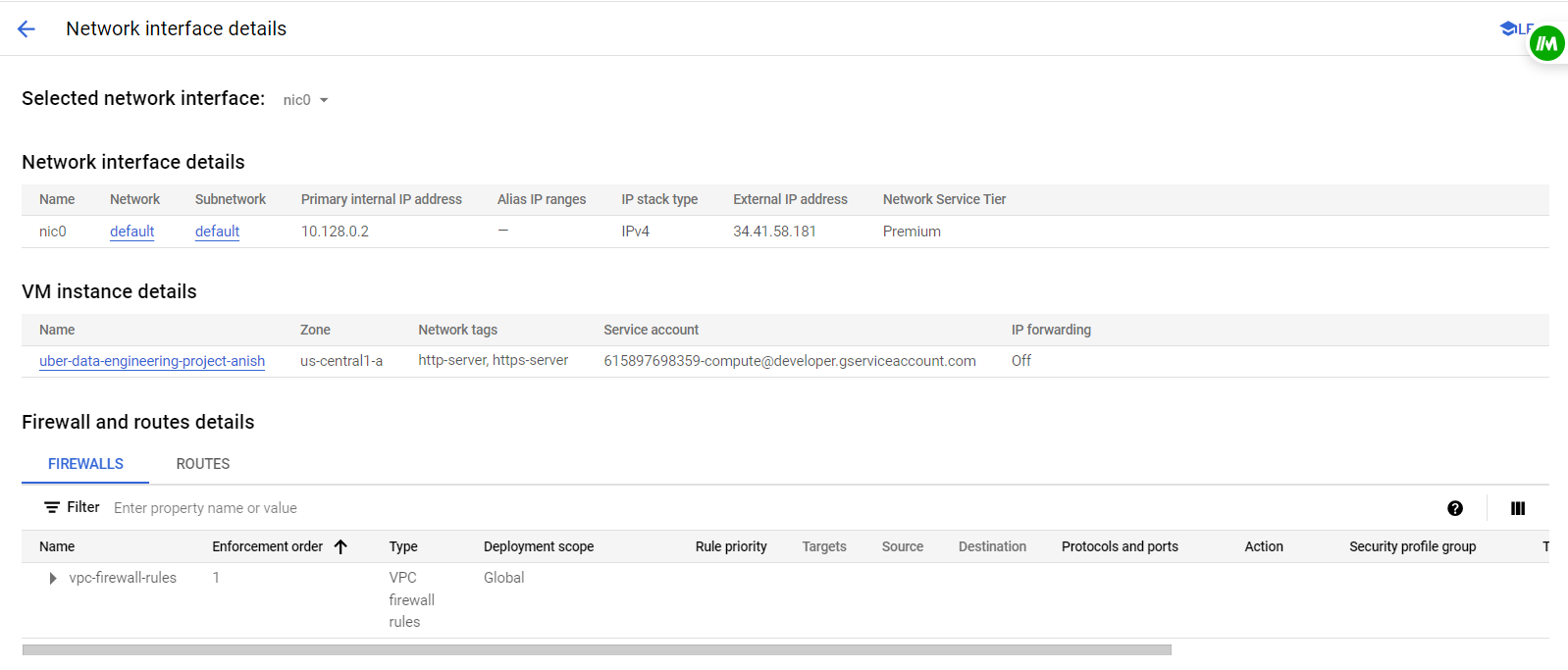
In case of AWS we had to download the ssh (pem) keys and then connect to the instance whereas google directly provides us with the SSH

We are not able to access the mage UI from this address even though mage-ai shows that it is running on localhost:6790.

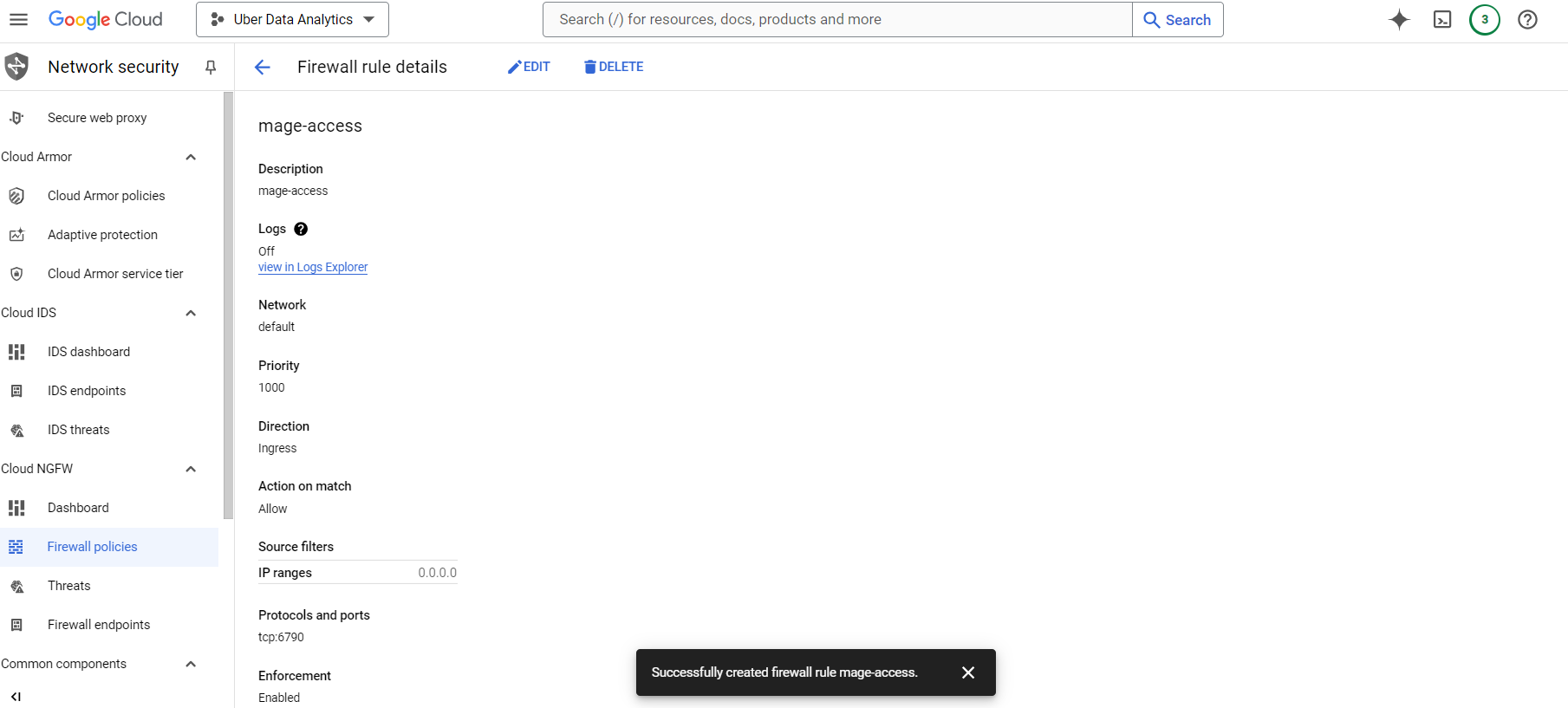
Also here localhost means the public IP of the VM instance you have started and not the local host of your computer



For this we need to specifically tell our instance that it should allow the request from this port. As you can see below the current firewall doesn’t mention any incoming traffic from 6790



So we create a firewall access policy

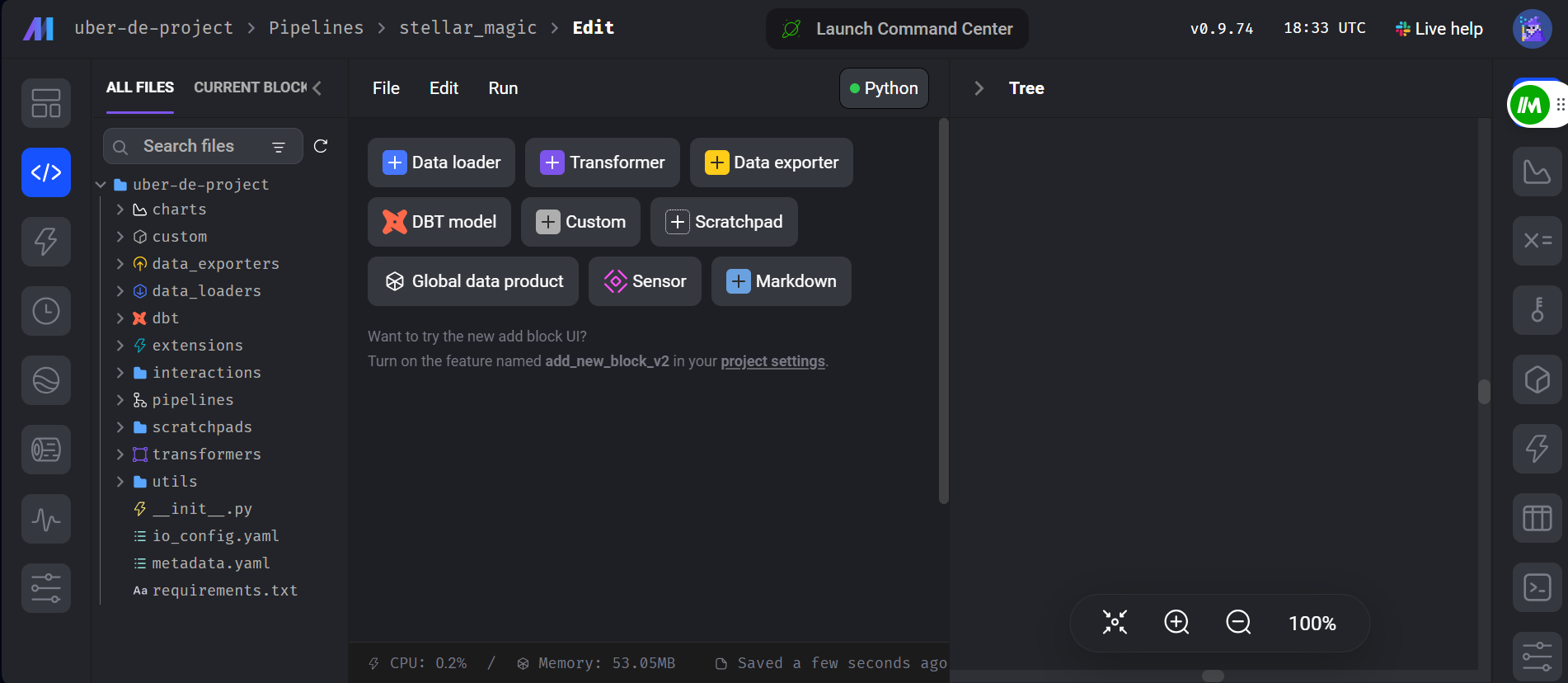


One common mistake is not giving the range in IP. So instead of 0.0.0.0 change it to 0.0.0.0/0 for it allow traffic from any source

A screenshot of a computer

Description automatically generated

Create a standard pipeline and since we have a unique link for the data in the buckets we can choose extracting data from API:



After finishing the transformations in Mage, we need to pass the multiple dataframes of fact and dimensional tables into the loader function and load this data into Bigquery and for that we pass it into a dictionary

The transformation code:

1. Extracting the data from an API and converting it into the data frame
2. Passing it to the transformation block so that we convert it into different dimensional tables and fact table

In Google Cloud Platform (GCP), **Compute Engine** and **VM Instance** are related but refer to different concepts:

1. **Compute Engine**:
   * **Compute Engine** is the service offered by GCP that allows you to run virtual machines (VMs) on Google’s infrastructure. It provides the underlying platform for managing, configuring, and deploying virtual machines. Essentially, Compute Engine is the overarching service that encompasses all the tools, APIs, and functionalities related to virtual machines in GCP.
2. **VM Instance**:
   * A **VM Instance** is an individual virtual machine that you create and run within the Compute Engine service. It's a specific instance of a virtual machine that you can configure with your choice of operating system, hardware (e.g., CPU, memory), storage options, and network settings. When you launch a VM instance, it is hosted on Compute Engine.

Create a dataset in big query after connecting mage to big query

A screenshot of a computer

Description automatically generated

Enter the details in the dataset details in the exporter

If you don’t find any modules, start another window of the instance by clickling the SSH button and then use pip install