BIG DATA

Лабораторна робота #1

Бєляєва Анастасія

Дрегало Леся

Педоренко Олег

Пєнской Вадим

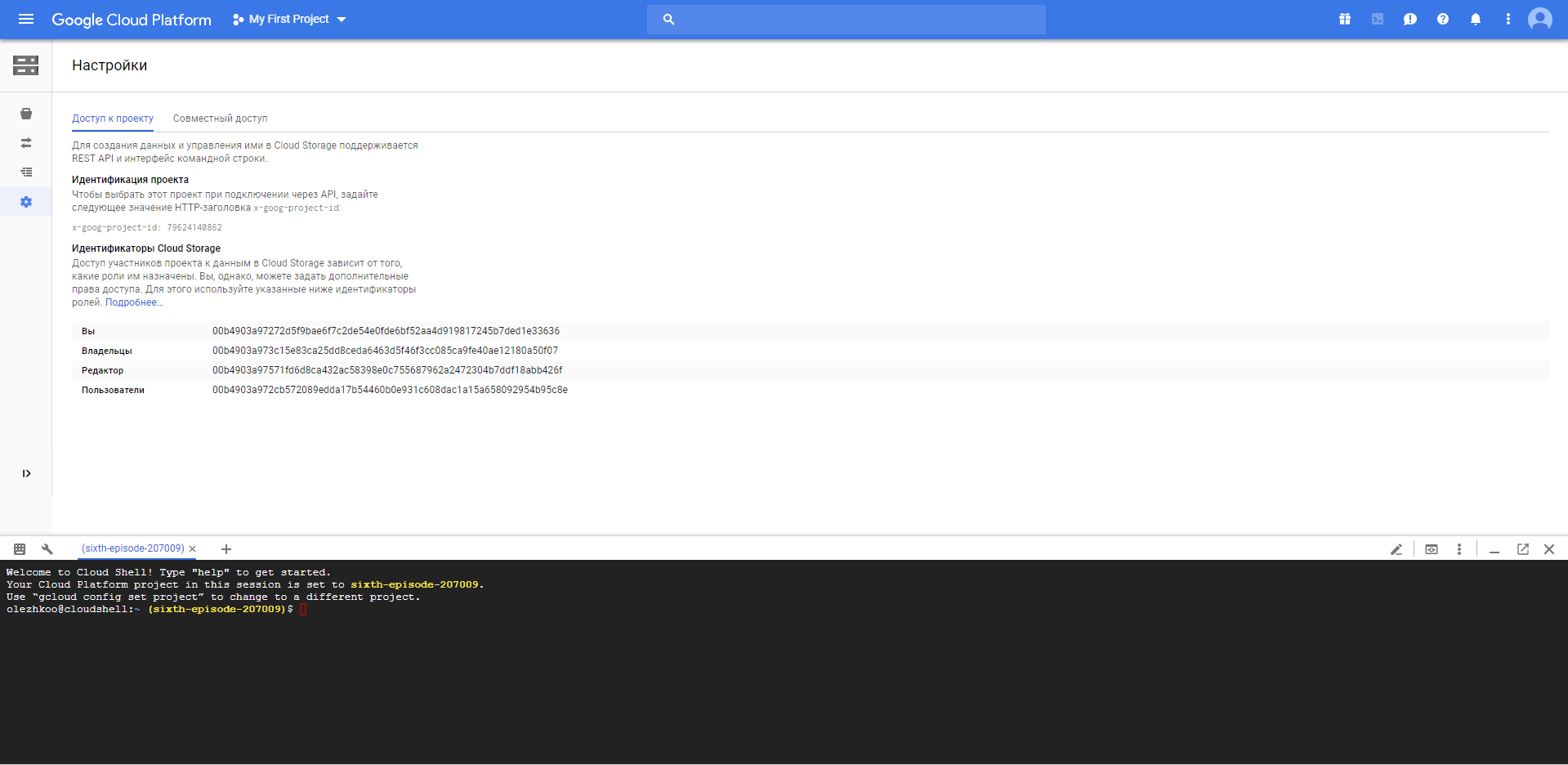
Харламбов Костянтин

Варіант завдання:

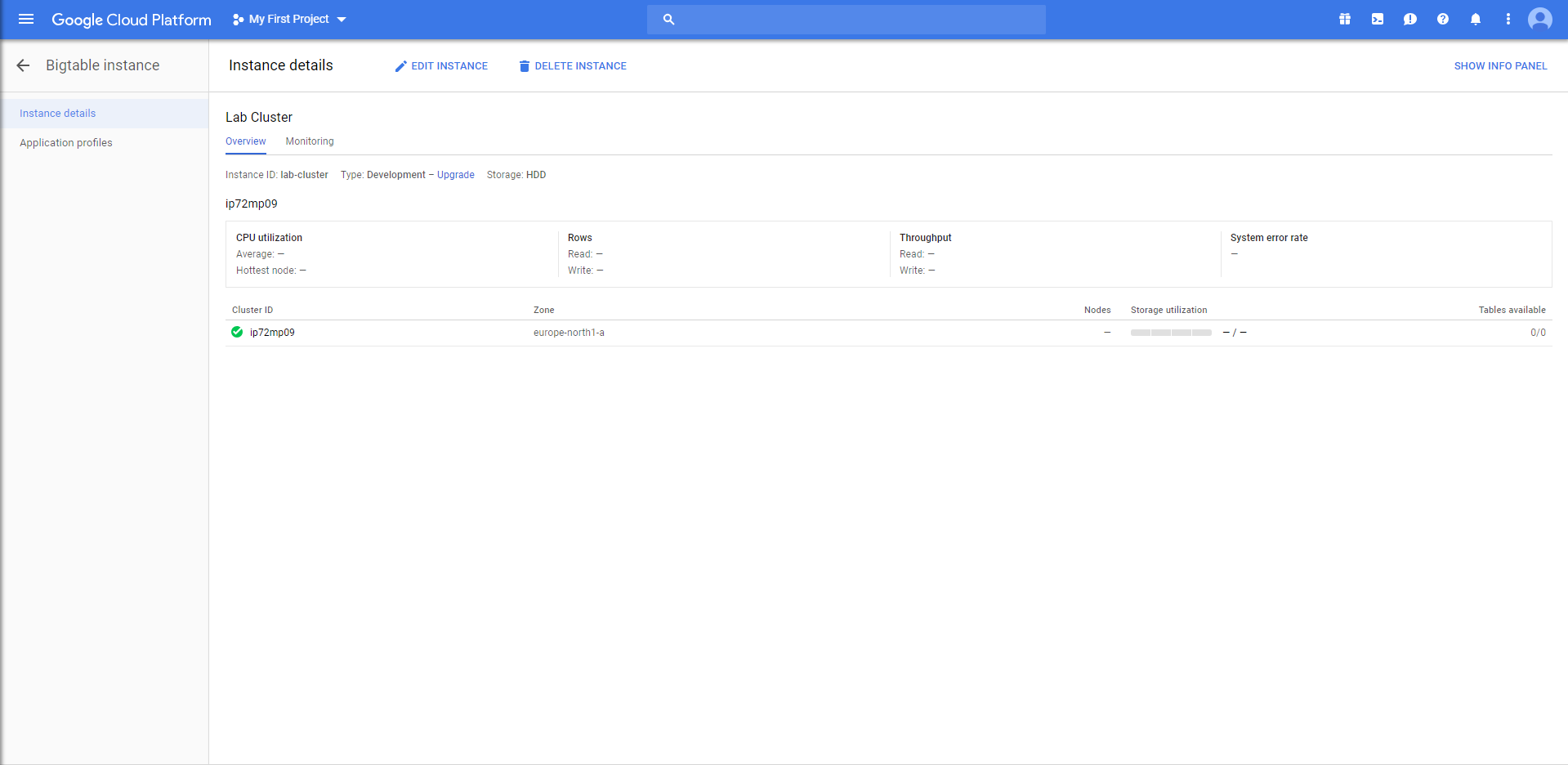
<http://data.gov.ua/passport/5fc89a6f-55b8-4ec6-95d6-38a0fdc31be1>

Хід виконанання роботи

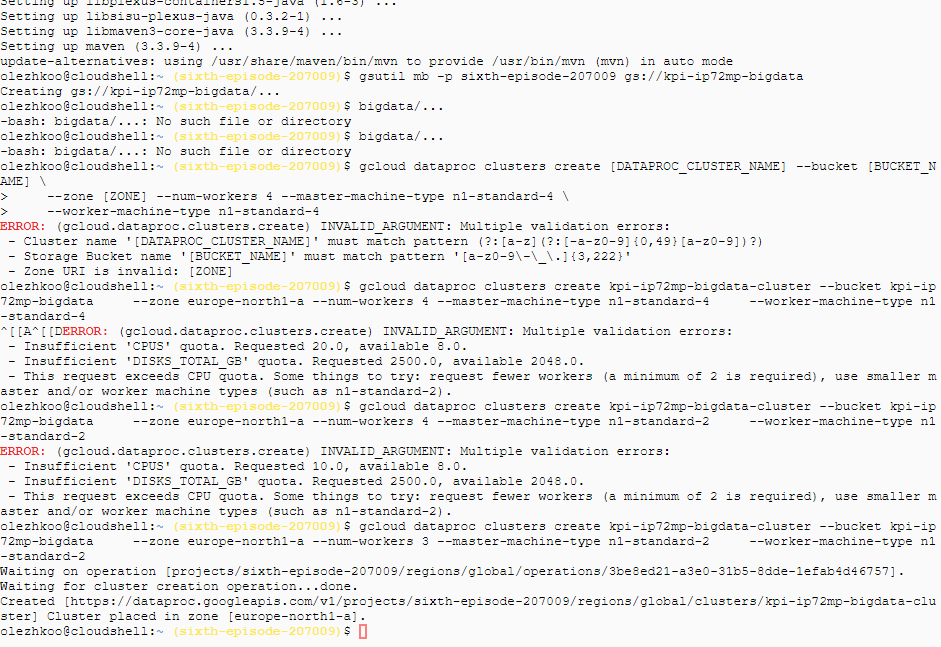
1. Налаштувати середовище на Google Cloud
   1. Створено акаунт на Google Cloud



* 1. Створено Google Bigtable Instance



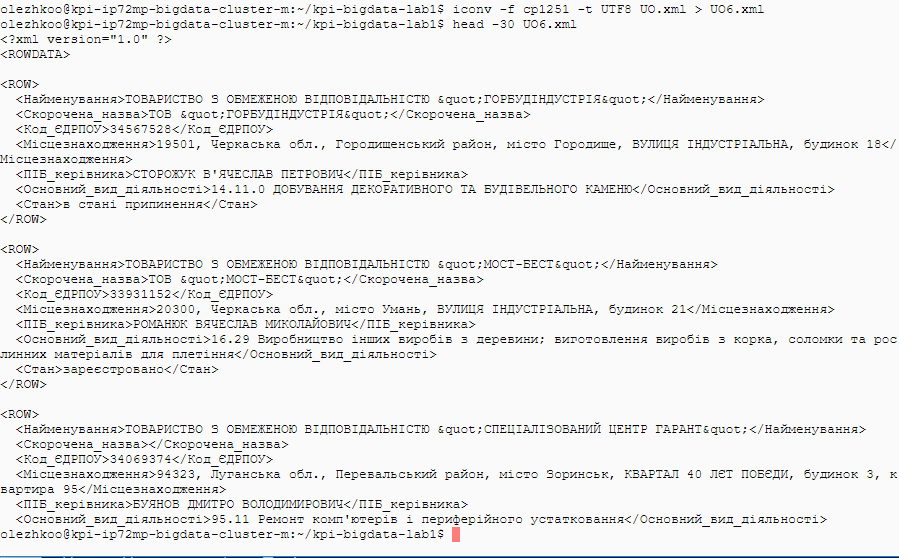
* 1. Створено Google Dataproc Cluster



1. Виконати завдання 1 (завантажити дані)
   1. Завантажив файл. Він виявився у форматі XML



Перекодував файл.



Так як файл у форматі xml, треба завантажити СерДе для xml

* 1. Підключитись до кластера по ssh (через браузер)
  2. Завантажив дані на кластер
  3. Завантажив дані у таблицю…