**Test 1: Live Streamed Tweets Analysis**

**Request**

Please create a simple application to:

* 1. Retrieve tweets containing keyword “python” from Twitter API (suggested package: tweepy)
  2. Plot out frequency of hashtags associated with incoming tweets streamed live from Twitter (suggested packages: pyspark, matplotlib, and seaborn)
  3. Come out another analysis that can be done on the live tweets and demonstrate how this additional analysis can be implemented

**Deliverables**

* Well commented **Jupyter notebook/Python script** named “**codepurpose\_yourname**” in github or gitlab.
* A comprehensive **readme.md** in the repository.
* A max-2-pages report named “**test1\_yourname.docx**”. In this document, you should at least explain your thinking process of designing this application and explain what insights you have got from your data analysis.

**Test 2: Data System Architecture Design**

**Request**

Please draw an architecture of a data system. The data system is able to handle both batch data and streaming data. The data system should minimally provide following services: data ingestion, data storage, data process, data catalog, and data analytics.

**Deliverables**

* An image file named “**DataSystem\_yourname**” to show the architecture designed
* A max-2-pages document named “**test2\_yourname.docx**”. Inside the document, you should at least explain the thinking process of designing this architecture, and briefly introduce all components in your data system architecture.