

Assignment - 2

60004200107

- Q1. Write program for RDD to find cube of numbers and display greater than 10.

```
from pyspark import SparkContext, SparkConf
conf = SparkConf().setAppName("cube RDD").
    setMaster("local")
sc = SparkContext(conf=conf)
numbers = sc.parallelize([1,2,3,4,5,6,7,8,9,10])
cubes = numbers.map(lambda x: x**3)
greater_than_10 = cubes.filter(lambda x: x > 10)
print(greater_than_10.collect())
```

- Q2. Including use cases of NoSQL

- 1) E-commerce : widely used because of large volumes of unstructured data, user profiles, transaction history, product catalogue.
- 2) Social media : to store user generated content, included posts, comments, likes, real time analytics.
- 3) Gaming : online gaming to handle massive data streams, game logs, and real time gameplay data.

4. Healthcare: store patient data, medical records, diagnostic test results, and medical images.

5. Internet of things: Handle large data generated by connected devices, sensor data, device status.

Q3. Industry use cases of Apache Kafka & Apache Flink

1. Financial services: real-time processing of financial data, stock prices, trading volumes, transaction data, analyze trends.

2. Telecommunication: handle large data by mobile data, call records, text messages, location data, analyze network traffic.

3. Retail: real-time processing of customer data, purchase history, browsing behavior, recommendations. Analyze customer behaviour.

4. Healthcare: real-time processing of patient data, medical records, medical images. Analyze patient data in real time.

5. Manufacturing: For real-time monitoring of equipment performance and predictive maintenance.