**Project Description**

Assume you are a data engineer working with Google, and you have been tasked with replicating data from two different databases, Cassandra and PostgreSQL, to an Oracle database.

The Cassandra database has two tables to be replicated, "Trending\_Topics" and "Subscriptions", while PostgreSQL has three tables to be replicated, "Products",

"Google\_Certified\_Professionals", and "YouTube\_Artists". This makes a total of five tables from two different databases that need to be replicated to Oracle.

The Oracle database has the following connection details:

* Hostname: 10.197.54.90
* Username: ml\_user
* Password: g@@gle2023#
* Database: learning\_projects
* Port: 1521

The PostgreSQL connection details are:

* Hostname: 10.195.56.32
* Database name: postgre\_data
* Port: 5432
* Password: post@google
* Username: ml\_user

The Cassandra connection details are:

* Hostname: 10.97.65.12
* Username: ml\_user
* Port: 9042
* Password: cassand@google

The goal of the project is to replicate data from the source databases to the Oracle database using Rubyrep. The data should be replicated incrementally to ensure that updates are captured in real-time. No transformations are required on either of the tables, and the data should be replicated as is from the source.

**Project Requirements**

To complete this project, you will need the following:

i. Access to the source databases (Cassandra and PostgreSQL) and the target database (Oracle)

ii. Ruby installed on your machine

iii. Rubyrep installed on your machine

iv. A configuration file for Rubyrep to define the replication rules

v. A scheduler like cron to automate the replication process

**Project Steps**

Install Ruby and Rubyrep on your machine if they are not already installed.

Configure the Oracle, PostgreSQL, and Cassandra connections in the Rubyrep configuration file.

Define the replication rules in the configuration file to specify which tables to replicate and how to map the source columns to the target columns.

Test the replication process by running the Rubyrep command manually and checking the logs for errors.

Set up a cron job to automate the replication process to run at a specific time every day.

**Conclusion**

Replicating data from multiple sources to a target database is a common task in data engineering. By using Rubyrep, you can easily configure the replication rules and automate

the process to ensure that updates are captured in real-time. With the detailed steps provided in this project description, you should be able to replicate data from Cassandra and

PostgreSQL to Oracle without any issues.