SEARCH 🔍

RESOURCES

CONCEPTS

✓ 1. Introduction

✓ 2. Project Details

✓ 3. Project Workspace

✓ 4. Project: Data Modeling with Apach…

Project: Data Modeling with Cassandra

A startup called Sparkify wants to analyze the data they've been collecting on so
their new music streaming app. The analysis team is particularly interested in un
songs users are listening to. Currently, there is no easy way to query the data to
since the data reside in a directory of CSV files on user activity on the app.

They'd like a data engineer to create an Apache Cassandra database which can
play data to answer the questions, and wish to bring you on the project. Your rol
database for this analysis. You'll be able to test your database by running querie
analytics team from Sparkify to create the results.

Project Overview

In this project, you'll apply what you've learned on data modeling with Apache Ca
an ETL pipeline using Python. To complete the project, you will need to model yo
tables in Apache Cassandra to run queries. You are provided with part of the ETL
data from a set of CSV files within a directory to create a streamlined CSV file to
into Apache Cassandra tables.

We have provided you with a project template that takes care of all the imports a
structure for ETL pipeline you'd need to process this data.

Knowledge
Search project Q&A

Student Hub
Chat with peers and mentors