

Data Engineering Nanodegree Syllabus



Contact Info

While going through the program, if you have questions about anything, you can reach us at support@udacity.com. For help from Udacity Mentors and your peers visit the Udacity Classroom.

Nanodegree Program Info

Prerequisite Skills

A well-prepared learner is able to:

- Intermediate Python programming knowledge, of the sort gained through the Programming for Data Science Nanodegree program, other introductory programming courses or programs, or additional real-world software development experience.
- Strings, numbers, and variables; statements, operators, and expressions;
- Lists, tuples, and dictionaries; Conditions, loops;
- Procedures, objects, modules, and libraries;
- Troubleshooting and debugging; Research & documentation;
- Problem solving; Algorithms and data structures
- Intermediate SQL knowledge and linear algebra mastery, addressed in the Programming for Data Science Nanodegree program
- Joins, Aggregations, and Subqueries
- Table definition and manipulation (Create, Update, Insert, Alter)

Required Software

- Python 3.6
- Anaconda 4.7 or latest
- Jupyter Notebook 6.01 or latest
- GIT Bash 2.23 or latest
- Apache Cassandra, pip, Pandas, Numpy 3.0 or latest
- AWS Regular Account with CC

Version: 2.0.0

Length of Program: 136 Days*

* This is a self-paced program and the length is an estimation of total hours the average student may take to complete all required coursework, including lecture and project time. Actual hours may vary.

Part 1: Welcome to the Nanodegree Program

Part 2: Data Modeling

Learn to create relational and NoSQL data models to fit the diverse needs of data consumers. Use ETL to build databases in PostgreSQL and Apache Cassandra.

Project: Data Modeling with Postgres

Students will model user activity data to create a database and ETL pipeline in Postgres for a music streaming app. They will define Fact and Dimension tables and insert data into new tables.

Supporting Lessons

Lesson	Summary
Introduction to Data Modeling	In this lesson, students will learn the basic difference between relational and non-relational databases, and how each type of database fits the diverse needs of data consumers.
Relational Data Models	In this lesson, students understand the purpose of data modeling, the strengths and weaknesses of relational databases, and create schemas and tables in Postgres

Project: Data Modeling with Apache Cassandra

Students will model event data to create a non-relational database and ETL pipeline for a music streaming app. They will define queries and tables for a database built using Apache Cassandra.

Supporting Lessons

Lesson	Summary
NoSQL Data Models	Students will understand when to use non-relational databases based on the data business needs, their strengths and weaknesses, and how to creates tables in Apache Cassandra.

Part 3: Cloud Data Warehouses

Project: Data Warehouse

Students will build an ETL pipeline that extracts data from S3, stages them in Redshift, and transforms data into a set of dimensional tables for their analytics team.

Part 4: Data Lakes with Spark

Project: Data Lake

Students will build a data lake and an ETL pipeline in Spark that loads data from S3, processes the data into analytics tables, and loads them back into S3.

Project: Optimize Your GitHub Profile

Other professionals are collaborating on GitHub and growing their network. Submit your profile to ensure your profile is on par with leaders in your field.

Part 5: Data Pipelines with Airflow

Project: Data Pipelines

Students continue to work on the music streaming company's data infrastructure by creating and automating a set of data pipelines with Airflow, monitoring and debugging production pipelines

Project: Improve Your LinkedIn Profile

Find your next job or connect with industry peers on LinkedIn. Ensure your profile attracts relevant leads that will grow your professional network.

Part 6: Capstone Project

Project: Data Engineering Capstone Project

In this Capstone project, students will define the scope of the project and the data they will be working with to demonstrate what they have learned in this Data Engineering Nanodegree.



Udacity

Generated Tue May 26 09:03:50 PDT 2020