UDACITY
FOR ENTERPRISE

SCHOOL OF DATA SCIENCE

Data Engineering with Microsoft Azure

NANODEGREE SYLLABUS

Overview

Data Engineering with Microsoft Azure Nanodegree Program

Learn to design data models, build data warehouses, build data lakes and lakehouse architecture, create data pipelines, and work with large datasets on the Azure platform using Azure Synapse Analytics, Azure Databricks, and Azure Data Factory.

Educational objectives:

A graduate of this program will be able to:

- Create relational and NoSQL data models

- Create data warehouses on the Azure cloud platform

- Work with large datasets using Spark and Azure Databricks

- Build and interact with Azure data lakes and lakehouse architecture

- Create data pipelines using Azure Data Factory and Synapse Analytics

- Develop proficiency in Spark, Azure Databricks, and Azure Databases

Program Information

TIME
4 months
Study 5-10 hours/week

LEVEL
Intermediate

PREREQUISITES
Students should have:
- Intermediate SQL programming skills
- Intermediate Python programming skills
- Familiarity with the Azure cloud platform
- Experience with Github

HARDWARE/SOFTWARE REQUIRED
There are no software and version requirements to complete this Nanodegree program. All coursework and projects can be completed via Student Workspaces and the Azure portal in the Udacity online classroom. Udacity's basic tech requirements can be found at https://www.udacity.com/tech/requirements.

LEARN MORE ABOUT THIS NANODEGREE
Contact us at enterpriseNDs@udacity.com.

Our Classroom Experience
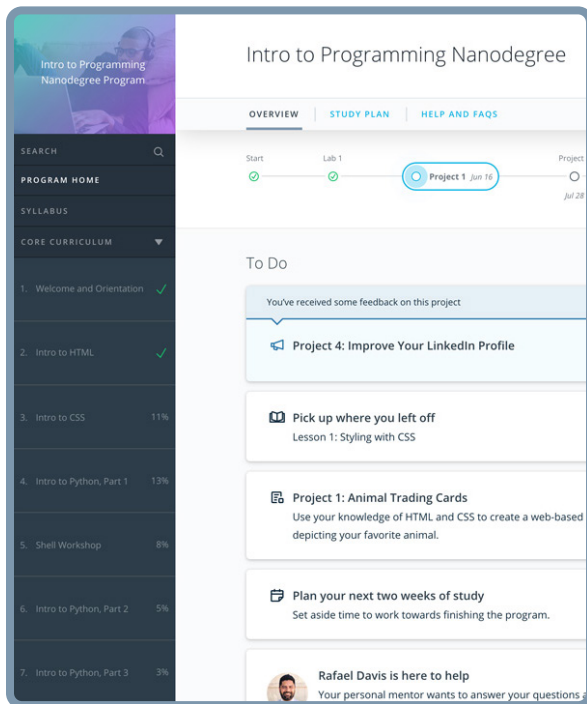
REAL-WORLD PROJECTS
Learners build new skills through industry-relevant projects and receive personalized feedback from our network of 900+ project reviewers. Our simple user interface makes it easy to submit projects as often as needed and receive unlimited feedback.

KNOWLEDGE
Answers to most questions can be found with Knowledge, our proprietary wiki. Learners can search questions asked by others and discover in real-time how to solve challenges.

WORKSPACES
Learners can check the output and quality of their code by testing it on interactive workspaces that are integrated into the classroom.

QUIZZES
Understanding concepts learned during lessons is made simple with auto-graded quizzes. Learners can easily go back and brush up on concepts at anytime during the course.

CUSTOM STUDY PLANS
Create a custom study plan to suit your personal needs and use this plan to keep track of your progress toward your goal.
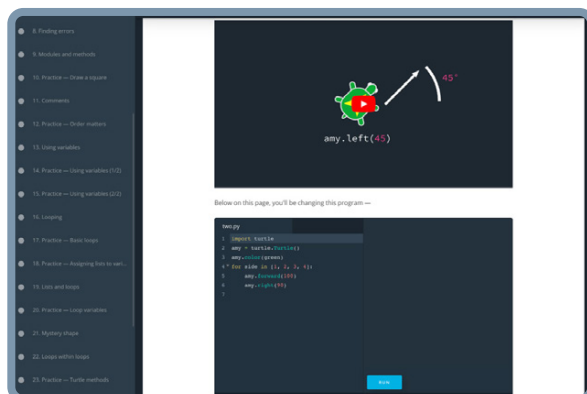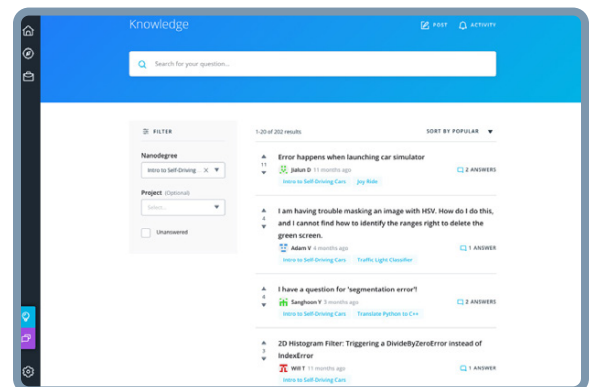
PROGRESS TRACKER
Personalized milestone reminders help learners stay on track and focused as they work to complete their Nanodegree program.

Learn with the Best

Matt Swaffer, PhD

DATA SCIENCE PRACTICE LEAD AT COGNITELL

Matt is a data science professional whose career has spanned software development, user experience design, and data visualization. He earned his PhD in the research area of cognitive psychology in human learning and is an adjunct professor teaching software design courses.

Amanda Moran

DEVELOPER ADVOCATE AT DATASTAX

Amanda is a developer advocate for DataStax after spending the last 6 years as a software engineer on 4 different distributed databases. Her passion is bridging the gap between customers and engineering. She has degrees from the University of Washington and Santa Clara University.

Vishnu (Lucky) Pamula

SR. CLOUD SOLUTION ARCHITECT AT MICROSOFT

Lucky is a data & AI evangelist with a track record of successfully helping organizations build analytics solutions. Besides his day job, he teaches as an adjunct professor, delivers lunch & learns, mentors students, and evangelizes Azure Quantum as an ambassador.

Course 1: Data Modeling

In this course, you'll learn to create relational and NoSQL data models to fit the diverse needs of data consumers. You'll understand the differences between different data models and how to choose the appropriate data model for a given situation. You'll also build fluency in PostgreSQL and Apache Cassandra.

Project 1	Data Modeling with Postgres

In this project, you'll model user activity data for a music streaming app called Sparkify. You'll create a relational database and ETL pipeline designed to optimize queries for understanding what songs users are listening to. In PostgreSQL, you will also define fact and dimension tables and insert data into your new tables.

Project 2	Data Modeling with Apache Cassandra

In this project, you'll model user activity data for a music streaming app called Sparkify. You'll create a database and ETL pipeline in both Postgres and Apache Cassandra, designed to optimize queries for understanding what songs users are listening to. For PostgreSQL, you will also define fact and dimension tables and insert data into your new tables. For Apache Cassandra, you will model your data so you can run specific queries provided by the analytics team at Sparkify.

LESSON TITLE	LEARNING OUTCOMES
INTRODUCTION TO DATA MODELING	• Understand the purpose of data modeling • Identify the strengths and weaknesses of different types of databases and data storage techniques • Create a table in Postgres and Apache Cassandra

UDACITY

Course 1: Data Modeling

LESSON TITLE	LEARNING OUTCOMES
RELATIONAL DATA MODELS	· Understand when to use a relational database · Understand the difference between OLAP and OLTP databases · Create normalized data tables · Implement denormalized schemas (e.g. STAR, Snowflake)
NOSQL DATA MODELS	· Understand when to use NoSQL databases and how they differ from relational databases · Select the appropriate primary key and clustering columns for a given use case · Create a NoSQL database in Apache Cassandra

Course 2: Cloud Data Warehouses with Azure

In this course, you will learn how to create cloud-based data warehouses and sharpen your data warehousing skills, deepen your knowledge of data infrastructure, and be introduced to data engineering on the cloud using Azure. You will start with an introduction to data warehouses and ETL, followed by an introduction to ELT and data warehouse technology in the cloud. After this you will learn about cloud data warehouse technology in Azure, including Azure Synaps Analytics.

Project 3	Building an Azure Data Warehouse for Bikeshare Data Analytics

In this project, you'll create a data warehouse solution using Azure Synaps Analytics to better understand Divvy, a bike-sharing program. You'll start by importing data into Synapse Analytics, then transform the data into a star schema and view reports from Analytics to identify how much time and money is spent per ride.

LESSON TITLE	LEARNING OUTCOMES
INTRODUCTION TO DATA WAREHOUSES	• Explain how OLAP may support certain business users better than OLTP • Implement ETL for OLAP Transformations with SQL • Describe Data Warehouse Architecture • Describe OLAP cube from facts and dimensions to slice, dice, roll-up, and drill down operations. • Implement OLAP cubes from facts and dimensions to slice, dice, roll-up, and drill down. • Compare columnar vs. row-oriented approaches • Implement columnar vs. row-oriented approaches

Course 2: Cloud Data Warehouses with Azure, cont.

LESSON TITLE	LEARNING OUTCOMES
ELT AND DATA WAREHOUSE TECHNOLOGY IN THE CLOUD	• Explain the differences between ETL and ELT • Differentiate scenarios where ELT is preferred over ETL • Implement ETL for OLAP Transformations with SQL • Select appropriate cloud data storage solutions • Select appropriate cloud pipeline solutions • Select appropriate cloud data warehouse solutions
AZURE DATA WAREHOUSE TECHNOLOGIES	• Explain the benefits of Azure cloud computing services in data engineering. • Describe Azure data engineering services • Set up key Azure features • Implementing Data Warehouse on Azure with Synapse Analytics
IMPLEMENTING DATA WAREHOUSES IN THE CLOUD	• Identify components of Azure Data Warehouse Architecture • Set up Azure infrastructure using Infrastructure as Code (IaC) • Run ELT process to extract data from Azure data storage into Synapse Analytics

Course 3: Data Lakes and Lakehouse with Spark and Azure Databricks

In this course, you'll learn about the big data ecosystem and how to use Spark to work with massive datasets. You will also store big data in a data lake and develop lakehouse architecture on the Azure Databricks platform.

Project 4	Building an Azure Data Lake for Bikeshare Data Analytics

In this project, you'll build a data lake solution for Divvy bikeshare with Azure Databricks using a lakehouse architecture. You will design a star schema based on business outcomes and create a Bronze data store. Then you'll create a gold data store in Delta Lake tables and transform the data into the star schema for a Gold data store.

LESSON TITLE	LEARNING OUTCOMES
BIG DATA ECOSYSTEM, DATA LAKES, AND SPARK	• Identify what constitutes the big data ecosystem for data engineering. • Explain the purpose and evolution of data lakes in the big data ecosystem. • Compare the Spark framework with Hadoop framework • Identify when to use Spark and when not to use it • Describe the features of lakehouse architecture
DATA WRANGLING WITH SPARK	• Identify what constitutes the big data ecosystem for data engineering. • Explain the purpose and evolution of data lakes in the big data ecosystem. • Compare the Spark framework with Hadoop framework • Identify when to use Spark and when not to use it

Course 3: Data Lakes and Lakehouse with Spark and Azure Databricks, cont.

LESSON TITLE	LEARNING OUTCOMES
SPARK DEBUGGING AND OPTIMIZATION	• Troubleshoot common errors and optimize their code using Spark WebUI • Identify common Spark bugs including errors in code syntax and issues with data • Diagnose errors in a distributed cluster to correct for them.
AZURE DATABRICKS	• Set up Spark Clusters in Azure Databricks • Produce Spark code in Databricks using Jupyter Notebooks and Python scripts • Implement distributed data storage using Azure Data Storage options
DATA LAKES ON AZURE WITH AZURE DATABRICKS	• Implement key features of data lakes on Azure • Use Spark and Databricks to run ELT processes and analytics on data of diverse sources, structures, and vintages

Course 4: Data Pipelines with Azure

In this course, you'll learn to build, orchestrate, automate and monitor data pipelines in Azure using Azure Data Factory and pipelines in Azure Synapse Analytics. You'll build, trigger, and monitor data pipelines on the Azure platform for analytical workloads and run data transformations, optimize data flows, and work with data pipelines in production.

Project 5	Data Integration Pipelines for NYC Payroll Data Analytics

The City of New York would like to develop a data analytics platform using Azure Synapse Analytics to analyze how the city's financial resources are allocated and how much of the city's budget is being devoted to overtime.

You have been hired as a data engineer to create high-quality data pipelines that are dynamic, can be automated, and can be monitored for efficient operation.

The source data resides in Azure Data Lake and you will build pipelines using Azure Data Factory for historical and new data to be processed in a NYC data warehouse in Azure Synapse Analytics.

LESSON TITLE	LEARNING OUTCOMES
AZURE DATA PIPELINE COMPONENTS	• Create and configure Azure data pipeline components • Create pipelines and associated components in Azure Data Factory or Azure Synapse • Configure linked service and dataset pipeline components • Choose integration runtimes for data pipelines
TRANSFORMING DATA IN AZURE DATA PIPELINES	• Create and trigger Mapping data flows and Azure pipeline activities to transform and move data • Transform data in Azure Data factory and synapse pipelines with data flows • Debug, trigger, and monitor pipeline activities containing data flows • Develop pipelines in multiple ways in Azure Data Factory and Synapse Pipelines • Integrate Power Query in Azure Pipelines

Course 4: Data Pipelines with Azure, cont.

LESSON TITLE	LEARNING OUTCOMES
AZURE PIPELINE DATA QUALITY	• Use common techniques optimize Azure data pipelines for data quality and flow • Manage data changing over time in pipeline data flows • Explain strategies and tools for data governance in Azure data pipelines
AZURE DATA PIPELINES IN PRODUCTION	• Implement production aspects of Azure data pipelines • Add Parameters to data pipelines in Azure Data Factory or Synapse Pipelines • Create pipeline objects programmatically • Automate data pipeline deployment with Azure DevOps or Github

Our Nanodegree Programs Include:

Pre-Assessments

Our in-depth workforce assessments identify your team's current level of knowledge in key areas. Results are used to generate custom learning paths designed to equip your workforce with the most applicable skill sets.

Dashboard & Progress Reports

Our interactive dashboard (enterprise management console) allows administrators to manage employee onboarding, track course progress, perform bulk enrollments and more.

Industry Validation & Reviews

Learners' progress and subject knowledge is tested and validated by industry experts and leaders from our advisory board. These in-depth reviews ensure your teams have achieved competency.

Real World Hands-on Projects

Through a series of rigorous, real-world projects, your employees learn and apply new techniques, analyze results, and produce actionable insights. Project portfolios demonstrate learners' growing proficiency and subject mastery.

Our Review Process

Real-life Reviewers for Real-life Projects

Real-world projects are at the core of our Nanodegree programs because hands-on learning is the best way to master a new skill. Receiving relevant feedback from an industry expert is a critical part of that learning process, and infinitely more useful than that from peers or automated grading systems. Udacity has a network of over 900 experienced project reviewers who provide personalized and timely feedback to help all learners succeed.

Vaibhav
UDACITY LEARNER

"I never felt overwhelmed while pursuing the Nanodegree program due to the valuable support of the reviewers, and now I am more confident in converting my ideas to reality."

now at
CODING VISIONS INFOTECH

All Learners Benefit From:

| Line-by-line feedback for coding projects | Industry tips and best practices | Advice on additional resources to research | Unlimited submissions and feedback loops |

How it Works

Real-world projects are integrated within the classroom experience, making for a seamless review process flow.

- Go through the lessons and work on the projects that follow
- Get help from your technical mentor, if needed
- Submit your project work
- Receive personalized feedback from the reviewer
- If the submission is not satisfactory, resubmit your project
- Continue submitting and receiving feedback from the reviewer until you successfully complete your project

About our Project Reviewers

Our expert project reviewers are evaluated against the highest standards and graded based on learners' progress. Here's how they measure up to ensure your success.

900+	1.8M	3	4.85 /5
Expert Project Reviewers	**Projects Reviewed**	**Hours Average Turnaround**	**Average Reviewer Rating**
Are hand-picked to provide detailed feedback on your project submissions.	Our reviewers have extensive experience in guiding learners through their course projects.	You can resubmit your project on the same day for additional feedback.	Our learners love the quality of the feedback they receive from our experienced reviewers.

UDACITY
FOR ENTERPRISE