UDACITY

NANODEGREE PROGRAM SYLLABUS

Data Architect

Overview

In this program, you'll plan, design and implement enterprise data infrastructure solutions and create the blueprints for an organization's data management system. You'll create a relational database with PostGreSQL, design an Online Analytical Processing (OLAP) data model to build a cloud based data warehouse, and design scalable data lake architecture that meets the needs of Big Data. Finally, you'll learn how to apply the principles of data governance to an organization's data management system.

A graduate of this program will be able to:

- Build conceptual, logical and physical entity relationship diagrams (ERDs).
- Architect a physical database in PostGreSQL.
- Transform data from transactional systems into an operational data store.
- Create a data warehouse system using dimensional data models.
- Use appropriate storage and processing frameworks to manage big data.
- Design end-to-end batch and stream processing architecture.
- Establish data governance best practices including metadata management, master data management and data quality management.

Prerequisites:

- Familiarity with Relational Database Management Systems or foundational database skills
- Intermediate Python
- Intermediate SQL
- Familiarity with batch processing and stream processing frameworks
- Operating systems, including UNIX, Linux and MS Windows
- Basics of ETL/Data Pipelines

Length of Program*:
4 months

Textbooks required:
None

Frequency of Classes:
The program is flexible, self-paced with suggested project deadlines

Need Help?
udacity.com/advisor
Discuss this program with an enrollment advisor.

*The length of this program is an estimation of total hours the average student may take to complete all required coursework, including lecture and project time. If you spend about 10 hours per week working through the program, you should finish within the time provided. Actual hours may vary.

UDACITY

Course 1: Data Architecture Foundations

In this course, you will learn about the principles of data architecture. You will begin by learning the characteristics of good data architecture and how to apply them. Next you will move on to data modeling. You will learn to design a data model, normalize data and create a professional ERD. Finally, you will take everything you learned and create a physical database using PostGreSQL.

Project 1
Designing an HR Database

In this project, you will design, build, and populate a database for the Human Resources (HR) Department at the imaginary Tech ABC Corp, a video game company. This project will start with a request from the HR Manager. From there, you will need to design a database using the foundational principles of data architecture that is best suited to the department's needs. You will go through the steps of database architecture, creating database proposals, database entity relationship diagrams, and finally creating the database itself. This project is a scaled-down simulation of the kind of real-world assignments data architects work on every day.

	LEARNING OUTCOMES	
LESSON ONE	**What is Data Architecture?**	• Define data architecture characteristics • Define data governance and its role • Define scalability and flexibility in database design
LESSON TWO	**Database Framework**	• Introduction to ERDs • Develop a database schema • Understand normalization and its use cases • Learn to normalize data to the 3rd Normal Form
LESSON THREE	**Relational Data Design**	• Introduction to ERDs • Build a conceptual ERD • Build a logical ERD • Learn about cardinality and Crow's Foot notation • Build a physical ERD
LESSON FOUR	**Creating a Physical Database**	• Learn about factors that affect database performance • Learn about file and data storage solutions • Use DDL SQL to create database objects in PostGreSQL • Learn about data ingestions methods, including: ETL, Pipelines, APIs and direct feeds • Use DML SQL to populate a database with data in PostGreSQL • Use CRUD SQL commands to demonstrate proper operation of a database

UDACITY

Course 2: Designing Data Systems

In this course, you will learn to design enterprise data architecture. You will build a cloud based data warehouse with Snowflake. You will evaluate various data assets of an organization and characteristics of these data sources, design a staging area for ingesting varieties of data coming from source systems, and design an Operational Data Store (ODS). Finally, you will learn to design OLAP dimensional data models, design ELT data processing that is capable of moving data from an ODS to a data warehouse, and write SQL queries for the purpose of building reports.

Project 2
Design a Data Warehouse for Reporting and OLAP

In this project, you will design end to end data architecture, build ingestion of data from Yelp and Climatic source systems, design Operational Data Store and Data warehouse systems, and transform data from staging to ODS and finally from ODS to data warehouse system. Yelp source carries a list of businesses, restaurants, its reviews and ratings. Climatic data source keeps track of temperature and precipitation data. Both of these websites are independent sources and not related to each other. The final objective of this project is to write appropriate SQL to find the impact of weather on restaurant ratings.

		LEARNING OUTCOMES
LESSON ONE	**Enterprise Data Architecture**	• Understand importance of Data Architecture in any organization • Learn the benefits of executing a Data Architecture • Learn the business and technical artifacts required • Understand business and functional requirements • Learn how OLTP, ODS and OLAP models are being designed
LESSON TWO	**Staging Data**	• Build staging area for data ingestion • Learn to organize data assets based on schemas • Design schedules for data processing based on the requirements • Learn to manage staging area through metadata
LESSON THREE	**Operational Data Store**	• Build an integrated ER model connecting distributed data assets • Learn to design Data Dictionary and Master Data • Apply normalization rules to eliminate redundancies • Learn when to use ETL vs ELT techniques • Learn to cleanse data anomalies
LESSON FOUR	**Data Warehouse**	• Learn two OLAP modeling designs — Star and Snowflake schemas • Learn various dimensional and fact table types • Build ELT data processing from ODS to Data warehouse • Write SQL queries for the purpose of reporting

Course 3: Big Data Systems

In this course, you will learn about how to help organizations with massive amounts of data, including identification of Big Data problems and how to design Big Data solutions. You will learn about the internal architecture of many of the Big Data tools such as HDFS, MapReduce, Hive and Spark, and how these tools work internally to provide distributed storage, distributed processing capabilities, fault tolerance and scalability. Next, you will learn how to evaluate NoSQL databases, their use cases and dive deep into creating and updating a NOSQL database with Amazon DynamoDB. Finally, you will learn how to implement Data Lake design patterns and how to enable transactional capabilities in a Data Lake.

Project 3
Design an Enterprise Data Lake System

In this project, you will act as a Big Data Architect and work on a real world use case faced by a Medical Data Processing Company. The project requires you to analyze the current architecture of the company, understand technical and business requirements and propose a new Data Lake based solution to both technical and executive audiences. For technical audiences, you will develop a design document outlining your solution with rationale, and for the executive audience you will record a short presentation pitching your solution. This is a real world scenario where you will act as an expert data infrastructure consultant to the company and solve the challenges the company is facing today. You will also hone your presentation skills and learn to articulate complex technical terminologies as easy to understand and value driven objectives to company leadership.

		LEARNING OUTCOMES
LESSON ONE	**Characteristics of Big Data**	• Explain what is big data • Articulate the business value of big data • Describe the characteristics of big data • Distinguish between horizontal scaling vs vertical scaling • Describe the components of a big data ecosystem
LESSON TWO	**Ingestion, Storage and Processing Frameworks**	• Explain how distributed storage works in HDFS • Explain how distributed processing works • Explain how resources are managed in a Hadoop cluster • Distinguish between different distributed processing frameworks • Apply frameworks to appropriate use cases

	LEARNING OUTCOMES	
LESSON THREE	**NoSQL Databases**	• Explain difference between SQL and NoSQL Databases • Differentiate between ACID and CAP properties of SQL and NoSQL databases • Implement, create, read, write, update NoSQL DB operations with DynamoDB • Create simple NoSQL data model
LESSON FOUR	**Scalable Data Lake Architecture**	• Explain what is a data lake and it's business value • Distinguish between different data formats and their application • Articulate Data Lake design patterns and challenges • Explain how to enable transactional capabilities in Data Lake

UDACITY

Course 4: Data Governance

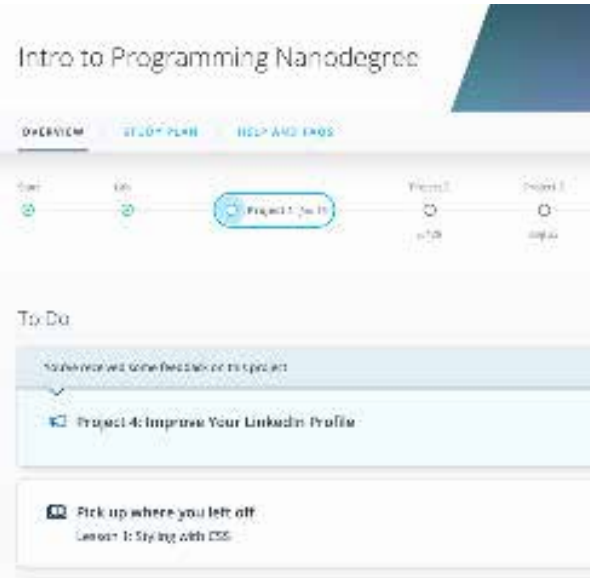
In this course you will learn how to design a data governance solution that meets your company's needs. First, you will learn about the different types of metadata and how to build a Metadata Management System, Enterprise Data Model and Enterprise Data Catalog. Next, you will learn how to perform data profiling using various techniques including data quality dimensions, how to identify remediation options for data quality issues, and how to measure and monitor data quality using data quality scores, thresholds, dashboards, exception and trend reports. Finally, you will learn the concepts of Master Data and golden record, different types of Master Data Management Architectures, as well as the golden record creation and master data governance processes.

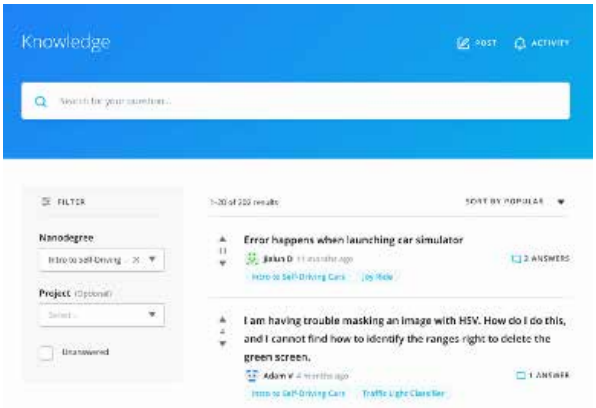
Project 4
Data Governance at SneakerPark

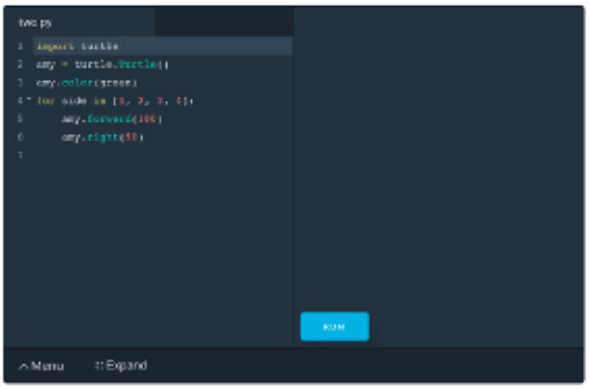
In this project, you will be implementing data governance solutions for an online shoe reseller SneakerPark to better manage their data now and in the future. First, you will create an Enterprise Data Model that provides a holistic view of all the data in their systems. Next you will document the metadata in an Enterprise Data Catalog and profile the data in their systems to identify data quality issues, suggest remediation strategies for each of these issues, and design a data quality dashboard. Finally, you will sketch out a proposed MDM implementation architecture, define a set of matching rules for the creation of customer and item master data, and define the data governance roles and responsibilities that are necessary to oversee this data governance initiative.

		LEARNING OUTCOMES
LESSON ONE	**Introduction to Data Governance**	• Understand what is Data Governance and its importance • Learn about the different disciplines of Data Governance • Understand the different stakeholders involved in Data Governance projects
LESSON TWO	**Metadata Management**	• Understand the different types of metadata • Understand the components and capabilities of Metadata Management System • Create conceptual and logical Enterprise Data Models • Create an Enterprise Data Catalog
LESSON THREE	**Data Quality Management**	• Perform data profiling using various techniques using data quality dimensions • Identify remediation options for data quality issues • Measure data quality using data quality scores and thresholds • Monitor data quality using dashboards, exception and trend reports
LESSON FOUR	**Master Data Management**	• Understand the concepts of master data and golden record • Understand different types of Master Data Management Architectures • Create a golden record using various match and merge techniques • Understand data governance processes for authoring, monitoring and approval of master data

Our Classroom Experience







REAL-WORLD PROJECTS

Build your skills through industry-relevant projects. Get personalized feedback from our network of 900+ project reviewers. Our simple interface makes it easy to submit your projects as often as you need and receive unlimited feedback on your work.

KNOWLEDGE

Find answers to your questions with Knowledge, our proprietary wiki. Search questions asked by other students and discover in real-time how to solve the challenges that you encounter.

STUDENT HUB

Leverage the power of community through a simple, yet powerful chat interface built within the classroom. Use Student Hub to connect with your technical mentor and fellow students in your Nanodegree program.

WORKSPACES

See your code in action. Check the output and quality of your code by running them on workspaces that are a part of our classroom.

QUIZZES

Check your understanding of concepts learned in the program by answering simple and auto-graded quizzes. Easily go back to the lessons to brush up on concepts anytime you get an answer wrong.

CUSTOM STUDY PLANS

Work with a mentor to create a custom study plan to suit your personal needs. Use this plan to keep track of your progress toward your goal.

PROGRESS TRACKER

Stay on track to complete your Nanodegree program with useful milestone reminders.

Learn with the Best

Ben Larson

DATA ARCHITECT / ANALYTICS CONSULTANT

Benjamin Larson, Ph.D. has over 15 years of experience working as a data professional in fields including medicine, telecommunications, and finance. His roles have included data architect, data scientist, and analytics consultant. He holds a Ph.D. in Decision Sciences, where his research was focused on rare event detection, using machine learning to detect credit card fraud.

Shankar Korrapolu

CEO AT OK2

Shankar Korrapolu is the co-founder and CEO of startup OK2, an innovative cross-platform mobile gaming engine that helps build games 50% cheaper and 50% faster without compromising the quality. For the past 30 years, he offered his enterprise data processing services to many organizations in Wall Street, investment banking, pharma, government and education sectors.

Shrinath Parikh

SENIOR DATA ARCHITECT

Shrinath is an entrepreneur and Data Architect passionate about helping enterprise companies transform and engineer their big data analytics applications on Cloud. He has worked with AWS, Google and Microsoft cloud platforms, has over 15 certifications and an MS in Computer Science from The University Of Texas at Dallas.

Vijaya Nelavelli

FOUNDER & PRINCIPAL DATA ARCHITECT

Vijaya is the Founder and Principal Data Architect for Great View Data Corp, where she works with clients like Wayfair, Ironwood Pharmaceuticals, Teradyne and National Grid.. She holds a Masters Degree in Computer Science and has 20 years of experience in Data Architecture and Data Management.

Rostislav Rabotnik

PRINCIPAL DATA ARCHITECT

Rostislav is an Enterprise Data Architect and Data Management Leader whose expertise covers a wide range of data governance, architecture, and integration practices across a diverse range of technologies. He has worked at companies of all sizes and in a variety of industries. His musings can be found at learndataarchitecture.com.

All Our Nanodegree Programs Include:

EXPERIENCED PROJECT REVIEWERS

REVIEWER SERVICES

- Personalized feedback & line by line code reviews
- 1600+ Reviewers with a 4.85/5 average rating
- 3 hour average project review turnaround time
- Unlimited submissions and feedback loops
- Practical tips and industry best practices
- Additional suggested resources to improve

TECHNICAL MENTOR SUPPORT

MENTORSHIP SERVICES

- Questions answered quickly by our team of technical mentors
- 1000+ Mentors with a 4.7/5 average rating
- Support for all your technical questions

PERSONAL CAREER SERVICES

CAREER SUPPORT

- Resume support
- Github portfolio review
- LinkedIn profile optimization

Frequently Asked Questions

PROGRAM OVERVIEW

WHY SHOULD I ENROLL?

This program was designed to help you take advantage of the growing need for skilled data talent. Companies are struggling to find the people they need to build data infrastructure and organize massive inflows of data, thus presenting you with a big opportunity to land a high-paying role designing cutting edge data infrastructure solutions. By the time you graduate, you'll be prepared to create the data infrastructure for an entire organization -- all the data management, infrastructure, systems, models, implementation and requirements for data engineers.

WHAT JOBS WILL THIS PROGRAM PREPARE ME FOR?

The goal of the Data Architect Nanodegree program is to prepare learners for advanced data professional roles, primarily as Data Architects. These senior level roles design blueprints for an organization's overall data management system including database structure, systems, models and other components and oversee the integration, centralization, maintenance and protection of data sources. These Data Architect skills can also be transferrable to other data and engineering roles at any company looking to manage or scale their data infrastructure.

HOW DO I KNOW IF THIS PROGRAM IS RIGHT FOR ME?

This is a more advanced program that is intended for data and engineering professionals who already have experience with core programming and data engineering skills and concepts and are looking to further advance their career into data architecture.

ENROLLMENT AND ADMISSION

DO I NEED TO APPLY? WHAT ARE THE ADMISSION CRITERIA?

No. This Nanodegree program accepts all applicants regardless of experience and specific background.

WHAT ARE THE PREREQUISITES FOR ENROLLMENT?

The Data Architect Nanodegree program is designed for students with intermediate Python and SQL skills, and basics of ETL/Data Pipelines including:

- Familiarity with Relational Database Management Systems or foundational database skills
- Intermediate Python
- Intermediate SQL
- Familiarity with batch processing and stream processing frameworks
- Operating systems, including UNIX, Linux, and MS Windows
- Basics of ETL/Data Pipelines

FAQs Continued

IF I DO NOT MEET THE REQUIREMENTS TO ENROLL, WHAT SHOULD I DO?

Depending on what skills you need to develop, Udacity has a range of courses and Nanodegree programs to help you prepare for the Data Architect Nanodegree program. We recommend our Data Engineer Nanodegree Program as one such option, but encourage you to browse the catalog for specific programs that meet your needs.

TUITION AND TERM OF PROGRAM

HOW IS THIS NANODEGREE PROGRAM STRUCTURED?

The Data Architect Nanodegree program is comprised of content and curriculum to support four projects. We estimate that students can complete the program in four months, working five to ten hours per week. Each project will be reviewed by the Udacity reviewer network. Feedback will be provided, and if you do not pass the project, you will be asked to resubmit the project until it passes.

HOW LONG IS THIS NANODEGREE PROGRAM?

Access to this Nanodegree program runs for the length of time specified above. If you do not graduate within that time period, you will continue learning with month to month payments. See the Terms of Use and FAQs for other policies regarding the terms of access to our Nanodegree programs.

CAN I SWITCH MY START DATE? CAN I GET A REFUND?

Please see the Udacity Program Terms of Use and FAQs for policies on enrollment in our programs.

SOFTWARE AND HARDWARE

WHAT SOFTWARE AND VERSIONS WILL I NEED IN THIS PROGRAM?

There are no software and version requirements to complete this Nanodegree program. All coursework and projects can be completed via Student Workspaces in the Udacity online classroom.