



UDACITY
FOR ENTERPRISE

THE SCHOOL OF DATA SCIENCE

Data Architect



NANODEGREE SYLLABUS

Overview

Data Architect Nanodegree Program

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.

Program Information



TIME

4 months
Study 10 hours/week



LEVEL

Specialist



PREREQUISITES

Intermediate Python, SQL, and basic familiarity with ETL/Data Pipelines



HARDWARE/SOFTWARE REQUIRED

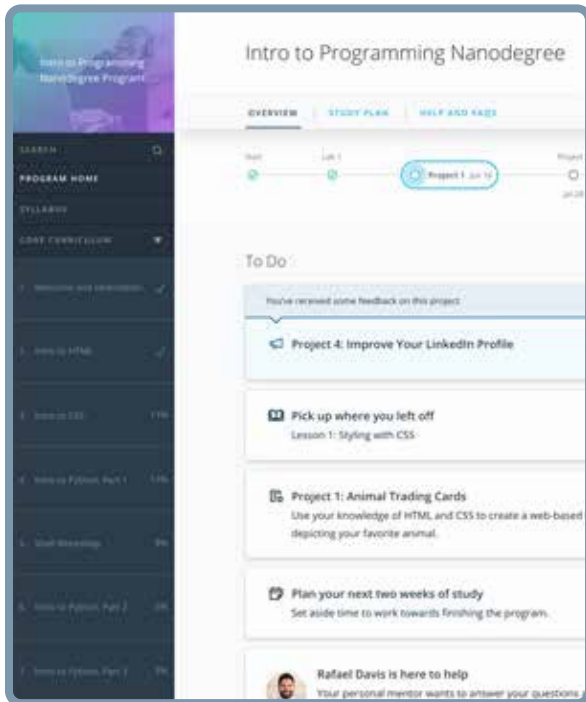
Access to the internet and a 64-bit computer.



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.

LEARNER HUB

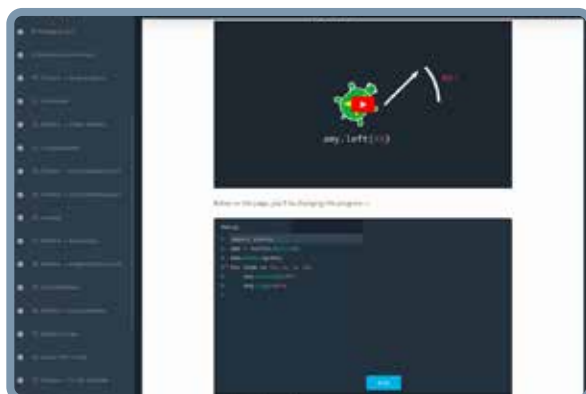
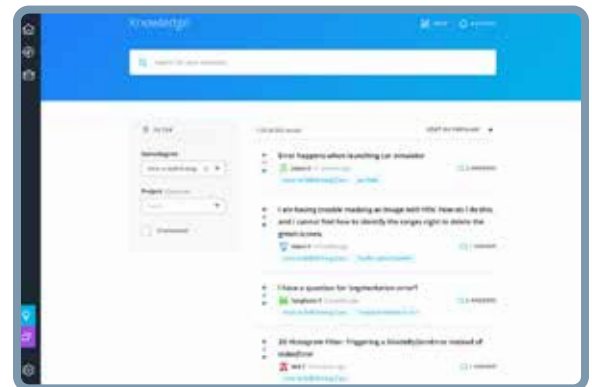
Learners leverage the power of community through a simple, yet powerful chat interface built within the classroom. Learner Hub connects learners with their technical mentor and fellow learners.

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

Mentors create a custom study plan tailored to learners' needs. This plan keeps track of progress toward learner goals.

PROGRESS TRACKER

Personalized milestone reminders help learners stay on track and focused as they work to complete their Nanodegree program.

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 Raboutnik

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.



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

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.

LESSON TITLE	LEARNING OUTCOMES
WHAT IS DATA ARCHITECTURE?	<ul style="list-style-type: none">• Define data architecture characteristics• Define data governance and its role• Define scalability and flexibility in database design
DATABASE FRAMEWORK	<ul style="list-style-type: none">• Introduction to ERDs• Develop a database schema• Understand normalization and its use cases• Learn to normalize data to the 3rd Normal Form
RELATIONAL DATA DESIGN	<ul style="list-style-type: none">• Introduction to ERDs• Build a conceptual ERD• Build a logical ERD• Learn about cardinality and Crow's Foot notation• Build a physical ERD

Nanodegree Program Overview

Course 1: Data Architecture Foundations, cont.

LESSON TITLE

LEARNING OUTCOMES

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





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

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.

LESSON TITLE	LEARNING OUTCOMES
ENTERPRISE DATA ARCHITECTURE	<ul style="list-style-type: none">• 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
STAGING DATA	<ul style="list-style-type: none">• 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

Nanodegree Program Overview

Course 2: Designing Data Systems, cont.

LESSON TITLE	LEARNING OUTCOMES
OPERATIONAL DATA STORE	<ul style="list-style-type: none">• 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
DATA WAREHOUSE	<ul style="list-style-type: none">• 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

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.

LESSON TITLE

LEARNING OUTCOMES

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

Nanodegree Program Overview

Course 3: Big Data Systems, cont.

LESSON TITLE	LEARNING OUTCOMES
INGESTION, STORAGE AND PROCESSING FRAMEWORKS	<ul style="list-style-type: none">• 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
NOSQL DATABASES	<ul style="list-style-type: none">• 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
SCALABLE DATA LAKE ARCHITECTURE	<ul style="list-style-type: none">• 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



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

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.

LESSON TITLE

LEARNING OUTCOMES

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

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

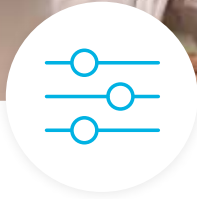
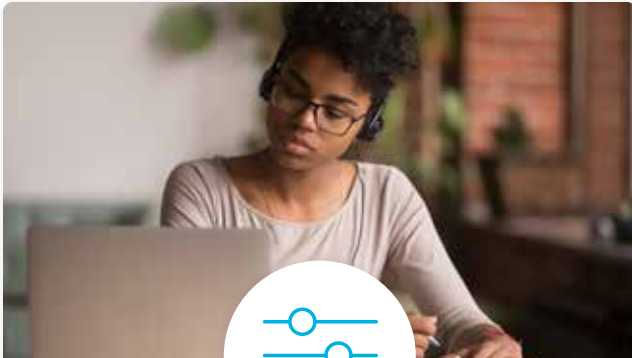
Nanodegree Program Overview

Course 4: Data Governance, cont.

LESSON TITLE	LEARNING OUTCOMES
DATA QUALITY MANAGEMENT	<ul style="list-style-type: none">• 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
MASTER DATA MANAGEMENT	<ul style="list-style-type: none">• 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 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.


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.



Expert Project Reviewers
Are hand-picked to provide detailed feedback on your project submissions.



Projects Reviewed
Our reviewers have extensive experience in guiding learners through their course projects.



Hours Average Turnaround
You can resubmit your project on the same day for additional feedback.



Average Reviewer Rating
Our learners love the quality of the feedback they receive from our experienced reviewers.



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



UDACITY
FOR ENTERPRISE

Udacity © 2020

2440 W El Camino Real, #101
Mountain View, CA 94040, USA - HQ

For more information visit: www.udacity.com/enterprise