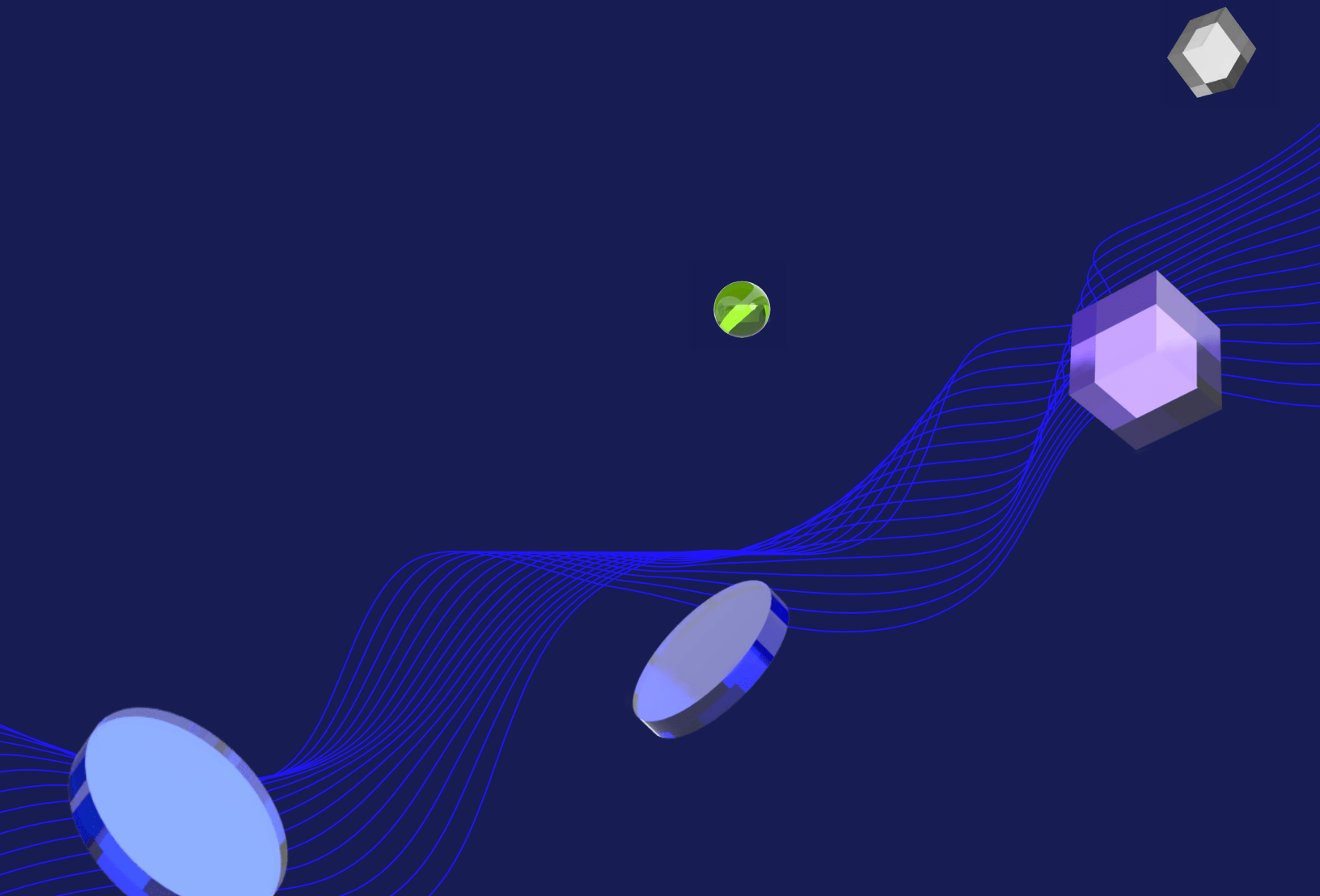U UDACITY

SCHOOL OF DATA SCIENCE

Data Architect

Nanodegree Program Syllabus

Overview

In this program, learners will plan, design, and implement enterprise data infrastructure solutions and create the blueprints for an organization's data management system. Learners will 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, learn how to apply the principles of data governance to an organization's data management system.

💡 Learning Objectives

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

⌛ Estimated Time

4 months at 10hrs/week*

Skill Level

Advanced

Prerequisites

A well-prepared learner should have knowledge of:

- Relational database management systems or foundational database skills
- Intermediate Python
- Intermediate SQL
- Batch processing and stream processing frameworks
- Operating systems, including UNIX, Linux, and MS Windows
- Basics of ETL/data pipelines

Required Hardware/Software

Learners need access to the internet and a 64-bit computer.

*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 5-10 hours per week working through the program, you should finish within the time provided. Actual hours may vary.

Data Architecture Foundations

Learn about the principles of data architecture. Begin by learning the characteristics of good data architecture and how to apply them. Next, move on to the topic of data modeling. Learn to design a data model, normalize data, and create professional ERD. Finally, take everything you learned and create a physical database using PostGreSQL.

Course Project

Designing an HR Database

In this project, learners 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, learners will need to design a database using the foundational principles of data architecture that is best suited to the department's needs. They 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 1

What is Data Architecture?

- Define data architecture characteristics.
- Define data governance and its role.
- Define scalability and flexibility in database design.

Lesson 2

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 3

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 4

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

Learn to design enterprise data architecture and build a cloud-based data warehouse with Snowflake. Learners 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 source (ODS). Finally, 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.

Course Project

Design a Data Warehouse for Reporting & OLAP

In this project, learners 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, transform data from staging to ODS, and finally from ODS to a data warehouse system. Yelp source carries a list of businesses, restaurants, 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 1

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 2

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 3

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 4

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

Learn to help organizations with massive amounts of data, including identification of big data problems and how to design big data solutions. Learn about 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, learn about evaluating NoSQL databases, their use cases, and creating a NoSQL database with Amazon DynamoDB. Finally, learn to implement data lake design patterns and enable transactional capabilities in a data lake.

Course Project

Design an Enterprise Data Lake System

Act as a big data architect and work on a real use case faced by a medical processing company. Start by analyzing the current architecture of the company. Then understand technical and business requirements and propose a new data lake based solution to both technical and executive audiences. For technical audiences, develop a design document outlining a solution and rationale, and for the executive audience, record a short presentation pitching a solution. This is a real world scenario where learners will act as an expert data infrastructure consultant to the company and solve challenges the company is facing today. Learners will also hone their presentation skills and learn to articulate complex technical terminologies.

Lesson 1

Characteristics of Big Data

- Explain what "big data" is.
- 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 2

Ingestion, Storage & 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.

Lesson 3

NoSQL Databases

- Explain difference between SQL and NoSQL databases.
- Differentiate between ACID and CAP properties of SQL and NoSQL databases.
- Implement, create, read, write, and update NoSQL DB operations with DynamoDB.
- Create simple NoSQL data model.

Lesson 4

Scalable Data Lake Architecture

- Explain data lakes and their business value.
- Distinguish between different data formats and their applications.
- Articulate data lake design patterns and challenges.
- Explain how to enable transactional capabilities in data lakes.

Data Governance

Learn how to design a data governance solution that meets a company's needs. First, learn about the different types of metadata, and how to build a metadata management system, enterprise data model, and enterprise data catalog. Next, learn how to perform data profiling using various techniques including data quality dimensions, identify remediation options for data quality issues, and measure and monitor data quality using data quality scores, thresholds, dashboards, exception, and trend reports. Finally, learn the concept 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.

Course Project

Data Governance at SneakerPark

In this project, learners will be implementing data governance solutions for an online shoe reseller SneakerPark to better manage their data now and in the future. First, create an enterprise data model that provides a holistic view of all the data in their systems. Next, 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, 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 1

Introduction to Data Governance

- Understand data governance and its importance.

- Learn about the different disciplines of data governance.

- Understand the different stakeholders involved in data governance projects.

Lesson 2

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 3

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 4

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.

Meet your instructors.

Ben Larson

Data Architect & Analytics Consultant

Benjamin has over 15 years of experience working as a data professional in fields including medicine, telecomm, and finance, in roles ranging from data architect to data scientist and analytics consultant. He holds a PhD in decision sciences, where his research was focused on rare event detection.
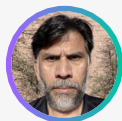
Shankar Korrapolu

CEO at OK2

Shankar Korrapolu is the cofounder and CEO of startup OK2, a cross-platform mobile gaming engine that builds games cheaper and faster without compromising quality. For over 30 years he has offered his data processing services to organizations in 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., which provides data architecture consulting and implementation services. Vijaya has extensive experience with creating architecture strategy and roadmaps, establishing frameworks and best practices, and data management.

Rostislav Rabotnik

Principal Data Architect

Rostislav is an enterprise data architect and data management leader whose expertise covers data governance, architecture, and integration practices across a diverse range of technologies. He has worked at companies of all sizes and in various industries. His musings can be found at learndataarchitecture.com.

Nicholas DeGiacomo

Data Scientist

Nick has built and managed teams of scientists and engineers for political campaigns, social media, and supply chain companies. With experiences ranging from startups to Amazon, he balances speed and scale. In his free time, Nick enjoys teaching graduate statistics courses at both Columbia and Yeshiva Universities.

Udacity's learning experience
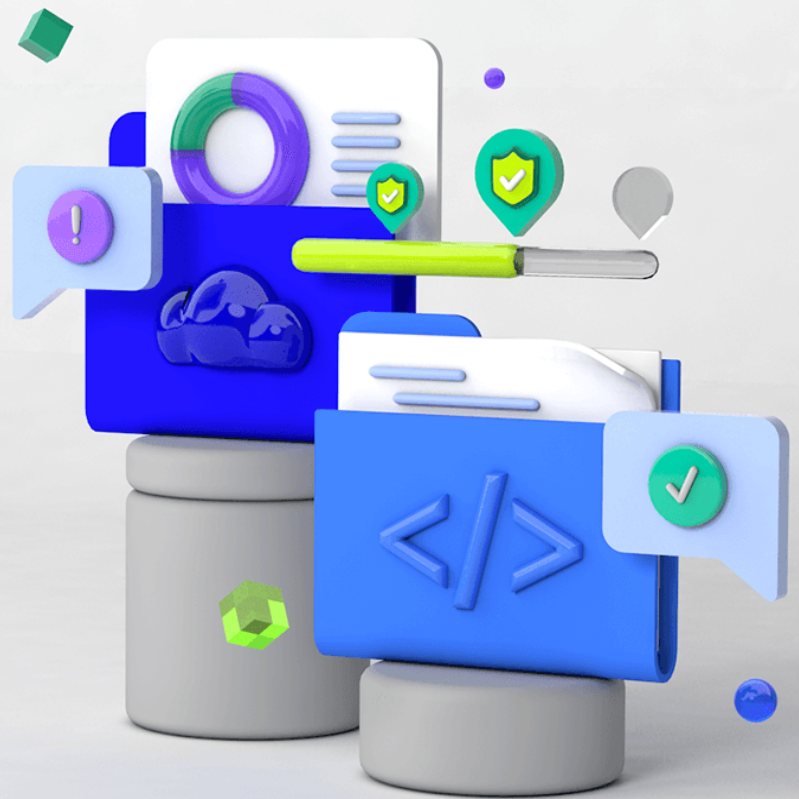
Hands-on Projects

Open-ended, experiential projects are designed to reflect actual workplace challenges. They aren't just multiple choice questions or step-by-step guides, but instead require critical thinking.

Quizzes

Auto-graded quizzes strengthen comprehension. Learners can return to lessons at any time during the course to refresh concepts.

Knowledge

Find answers to your questions with Knowledge, our proprietary wiki. Search questions asked by other students, connect with technical mentors, and discover how to solve the challenges that you encounter.

Custom Study Plans

Create a personalized study plan that fits your individual needs. Utilize this plan to keep track of movement toward your overall goal.

Workspaces

See your code in action. Check the output and quality of your code by running it on interactive workspaces that are integrated into the platform.

Progress Tracker

Take advantage of milestone reminders to stay on schedule and complete your program.

Our proven approach for building job-ready digital skills.
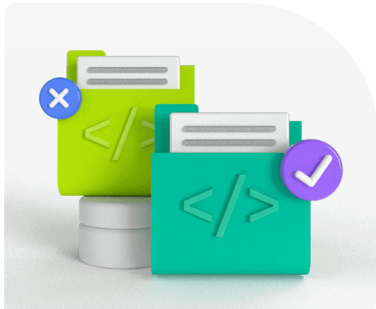
Experienced Project Reviewers

Verify skills mastery.

- Personalized project feedback and critique includes line-by-line code review from skilled practitioners with an average turnaround time of 1.1 hours.
- Project review cycle creates a feedback loop with multiple opportunities for improvement—until the concept is mastered.
- Project reviewers leverage industry best practices and provide pro tips.

Technical Mentor Support

24/7 support unblocks learning.

- Learning accelerates as skilled mentors identify areas of achievement and potential for growth.
- Unlimited access to mentors means help arrives when it's needed most.
- 2 hr or less average question response time assures that skills development stays on track.
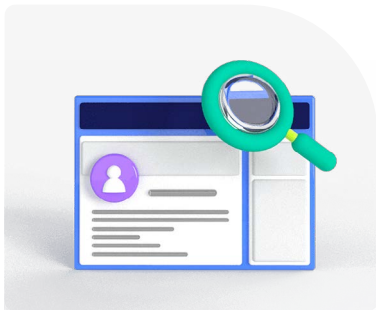
Personal Career Services

Empower job-readiness.

- Access to a Github portfolio review that can give you an edge by highlighting your strengths, and demonstrating your value to employers.*
- Get help optimizing your LinkedIn and establishing your personal brand so your profile ranks higher in searches by recruiters and hiring managers.

Mentor Network

Highly vetted for effectiveness.

- Mentors must complete a 5-step hiring process to join Udacity's selective network.
- After passing an objective and situational assessment, mentors must demonstrate communication and behavioral fit for a mentorship role.
- Mentors work across more than 30 different industries and often complete a Nanodegree program themselves.

*Applies to select Nanodegree programs only.

UDACITY

Learn more at

www.udacity.com/online-learning-for-individuals →