Professional Certificate Program in Data Engineering

In collaboration with IBM

Masterclasses, Exclusive Mentoring Sessions and Hackathons by IBM

simplilearn

In collaboration with:
PURDUE UNIVERSITY

Table of Contents

About the Program

Accelerate your career with this acclaimed Professional Certificate Program in Data Engineering, in collaberation with Purdue University and in collaboration with IBM. This program provides a perfect mix of theory, case studies, and extensive hands-on practice. Learners will receive a comprehensive data engineering education, leveraging Purdue's academic excellence in data engineering and Simplilearn's partnership with IBM.

This program is designed to give experienced professionals, coming from diverse backgrounds, an extensive data engineering education through a blend of online self-paced videos, live virtual classes, hands-on projects, and labs. Learners will also get access to mentorship sessions that provide a high-engagement learning experience and real-world applications to help master essential data engineering skills. You'll learn the implementation of data engineering concepts such as distributed processing using the Hadoop framework, large scale data processing using Spark, building data pipelines with Kafka, and working with databases both on on-premise, AWS and Azure cloud infrastructures.

Key Features of the Program

Joint certificate from Purdue University and Simplilearn

180+ hours of Applied Learning

14+ hands-on projects

Masterclasses from Top Purdue Faculty and Industry Experts at IBM

Eligble for Purdue Alumni Association Membership

Capstone project in 3 Domains

Curriculum aligned with Microsoft Azure (DP-203), AWS (DAS-C01), and Cloudera (CCA175) certifications

Exclusive Mentoring Sessions and Hackathons by IBM

About the Professional Certificate Program in Data Engineering in collaboration with Purdue University

Purdue University, a top public research institution, or higher education at its highest proven value. Committed to student success, Purdue is changing the student experience with a greater focus on faculty-student interaction and creative use of technology.

This Professional Certificate Program in Data Engineering in partnership with Purdue University will open pathways for you in the data engineering field, which has its presence in all the industry sectors and verticals - including manufacturing, healthcare, and ecommerce. Upon completing this program, you will receive a Certificate in collaboration Purdue University

Upon completing this program, you will receive a Purdue certification of completion.

About Simplilearn

Simplilearn is the world's #1 online bootcamp provider that enables learners through rigorous and highly specialized training. We focus on emerging technologies and processes that are transforming the digital world, at a fraction of the cost and time as traditional approaches. Over one million professionals and 2000 corporate training organizations have harnessed our award-winning programs to achieve their career and business goals.

Program Eligibility Criteria and Application Process

Those wishing to enroll in the Professional Certificate Program in Data Engineering in collaboration with Purdue University will be required to apply for admission to the program.

Eligibility Criteria

For admission to this Professional Certificate Program in Data Engineering, candidates should have:

- A bachelor's degree with an average of 50% or higher marks
- 2+ years of work experience (preferred)
- Basic understanding of object oriented programming (preferred)

Application Process

The application process consists of three simple steps. An offer of admission will be made to the selected candidates and accepted by the candidates by paying the admission fee.

STEP 1
SUBMIT AN APPLICATION

Complete the application and include a brief statement of purpose. The latter informs our admissions counselors why you're interested and qualified for the program.

STEP 2
APPLICATION REVIEW

A panel of admissions counselors will review your application and statement of purpose to determine whether you qualify for acceptance.

STEP 3
ADMISSION

An offer of admission will be made to qualified candidates. You can accept this offer by paying the program fee.

Talk to an Admissions Counselor

We have a team of dedicated admissions counselors who are here to help guide you in applying to the program. They are available to:

- Address questions related to the application
- Assist with financial aid (if required)
- Help you resolve your questions and understand the program

Program Outcomes

Gather data requirements, access data from multiple sources, process data for business needs, and store data on the cloud as well as on-premises

Gain insights into how to improve business productivity by processing Big Data on platforms that can handle its volume, velocity, variety, and veracity

Get a solid understanding of the fundamentals of the Scala language, its tooling, and the development process

Master the various components of the Hadoop ecosystem, such as Hadoop, Yarn, MapReduce, Pig, Hive, Impala, HBase, ZooKeeper, Oozie, Sqoop, and Flume

Identify AWS concepts, terminologies, benefits, and deployment options to meet business requirements

Understand how to use Amazon EMR for processing data using Hadoop ecosystem tools

Understand how to use Amazon Kinesis for Big Data processing in real-time

Analyze and transform Big Data using Kinesis Streams and visualize data and perform queries using Amazon QuickSight

Implement data storage solutions; manage and develop data processing; and monitor and optimize data solutions using Azure Cosmos DB, Azure SQL Database, Azure Synapse Analytics, Azure Data Lake Storage, Azure Data Factory, Azure Stream Analytics, Azure Databricks, and Azure Blob storage services

Apply the knowledge, skills, and capabilities gathered throughout the program to build an industry-ready product

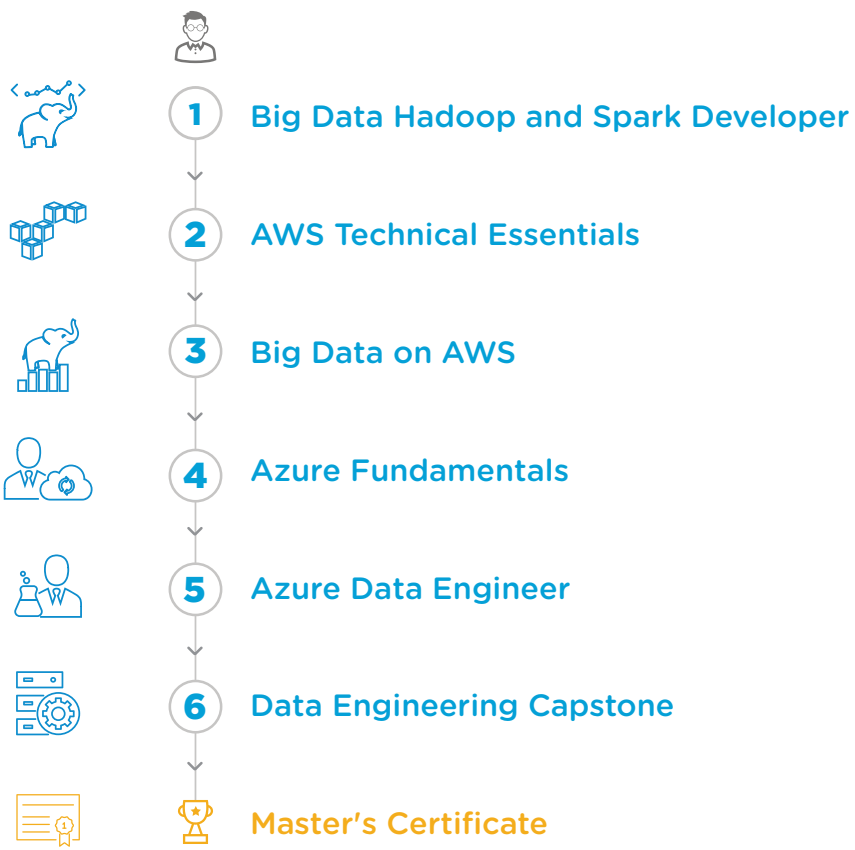
Who Should Enroll in this Program?

This program caters to graduates in any discipline and working professionals from diverse backgrounds who have basic programming knowledge. The diversity of our students adds richness to class discussions and interactions.

A data engineer builds and maintains data structures and architectures for data ingestion, processing, and deployment for large-scale, data-intensive applications. It's a promising career for both new and experienced professionals with a passion for data, including:

- IT professionals
- Database administrators
- Beginners in the data engineering domain
- BI Developers
- Data science professionals who want to expand their skill set
- Students in UG/ PG programs

Learning Path



Electives

- Python for Data Science
- PySpark Training
- Apache Kafka
- MongoDB Developer and Administrator
- GCP Fundamentals
- Java Training
- Academic master class - Data Engineering - Purdue university

Big Data Hadoop and Spark Developer

This Big Data Hadoop and Spark developer course helps you master the concepts of the Hadoop framework, Big Data, and Hadoop ecosystem tools such as HDFS, YARN, MapReduce, Hive, Impala, Pig, HBase, Spark, Flume, Sqoop, and including additional concepts of the Big Data processing life cycle. This course is aligned with Cloudera's CCA175 Big Data Certification.

Key Learning Objectives

- Learn how to navigate the Hadoop ecosystem and understand how to optimize its use
- Ingest data using Sqoop, Flume, and Kafka
- Implement partitioning, bucketing, and indexing in Hive
- Work with RDD in Apache Spark
- Process real-time streaming data
- Perform DataFrame operations in Spark using SQL queries
- Implement User-Defined Functions (UDF) and User-Defined Attribute Functions (UDAF) in Spark

Course curriculum

- Lesson 01 - Course Introduction
- Lesson 02 - Introduction to Big Data and Hadoop
- Lesson 03 - Hadoop Architecture, Distributed Storage (HDFS), and YARN

- Lesson 04 - Data Ingestion into Big Data Systems and ETL
- Lesson 05 - Distributed Processing - MapReduce Framework and Pig
- Lesson 06 - Apache Hive
- Lesson 07 - NoSQL Databases - HBase
- Lesson 08 - Basics of Functional Programming and Scala
- Lesson 09 - Apache Spark Next Generation Big Data Framework
- Lesson 10 - Spark Core Processing RDD0
- Lesson 11 - Spark SQL - Processing DataFrames
- Lesson 12 - Spark MLLib - Modelling BigData with Spark
- Lesson 13 - Stream Processing Frameworks and Spark Streaming
- Lesson 14 - Spark GraphX

AWS Tech Essentials

The AWS Technical Essentials course teaches you how to navigate the AWS management console, understand AWS security measures, storage, and database options, and gain expertise in web services like RDS and EBS. This course helps you become proficient in identifying and effciently using AWS services.

Key Learning Objectives

- Understand the fundamental concepts of AWS platform and cloud computing
- Identify AWS concepts, terminologies, benefits, and deployment options to meet business requirements
- Identify deployment and network options in AWS

Course curriculum

- Lesson 01 - Introduction to Cloud Computing
- Lesson 02 - Introduction to AWS
- Lesson 03 - Storage and Content Delivery
- Lesson 04 - Compute Services and Networking
- Lesson 05 - AWS Managed Services and Databases
- Lesson 06 - Deployment and Management

Big Data on AWS

The AWS Big Data course helps you understand the Amazon Web Services cloud platform, Kinesis Analytics, AWS Big Data storage, processing, analysis, visualization, and security service, EMR, AWS Lambda, Glue, and machine learning algorithms.

Key Learning Objectives

- Understand how to use Amazon EMR for processing the data using Hadoop ecosystem tools
- Understand how to use Amazon Kinesis for Big Data processing in real-time and analyze and transform Big Data using Kinesis Streams
- Visualize data and perform queries using Amazon QuickSight

Course curriculum

- Lesson 01 - Big Data on AWS Certification Course Overview
- Lesson 02 - Big Data on AWS Introduction
- Lesson 03 - AWS Big Data Collection Services
- Lesson 04 - AWS Big Data Storage Services
- Lesson 05 - AWS Big Data Processing Services
- Lesson 06 - Analysis
- Lesson 07 - Visualization
- Lesson 08 - Security

Azure Fundamentals

The Azure Fundamentals course covers the main principles of cloud computing and how they have been implemented in Microsoft Azure. You will work on the concepts of Azure services, security, privacy, compliance, trust, pricing, and support and learn how to create the most common Azure services, including virtual machines, web apps, SQL databases, features of Azure Active Directory, and methods of integrating it with on-premises Active Directory.

Key Learning Objectives

- Describe Azure storage and create Azure web apps
- Deploy databases in Azure
- Understand Azure AD, cloud computing, Azure, and Azure subscriptions
- Create and configure VMs in Microsoft Azure

Course curriculum

- Lesson 01 - Cloud Concepts
- Lesson 02 -Core Azure Services
- Lesson 03 -Security, Privacy, Compliance, and Trust
- Lesson 04 -Azure Pricing and Support

Azure Data Engineer

The Azure Data Engineer course will focus on data-related implementation which includes provisioning data storage services, ingesting streaming and batch data, transforming data, implementing security requirements, implementing data retention policies, identifying performance bottlenecks, and accessing external data sources.

Key Learning Objectives

- Implement data storage solutions using Azure SQL Database, Azure
- Synapse Analytics, Azure Data Lake Storage, Azure Data Factory,
- Azure Stream Analytics, Azure Databricks services
- Develop batch processing and streaming solutions
- Monitor Data Storage and Data Processing
- Optimize Azure Data Solutions

Course curriculum

- Design and implement data storage
- Design and develop data processing
- Design and implement data security
- Monitor and optimize data storage and data processing

Data Engineer Capstone

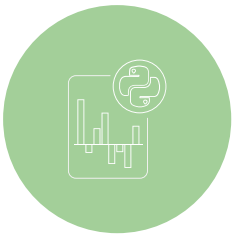
The data engineering capstone project will give you an opportunity to implement the skills you learned throughout this program. Through dedicated mentoring sessions, you'll learn how to solve a real-world, industry-aligned data engineering problem, from setting up configuration, ETL, data streaming, and data analysis to data visualization. This project is the final step in the learning path and will enable you to showcase your expertise in data engineering to future employers.

You can choose to work on projects that cover the most relevant domains (ecommerce, BFSI, video sharing) to make your practice more relevant.

Elective Course

Python for Data Science

The Python for Data Science course, carefully crafted by IBM helps you understand how to integrate Python using the PySpark interface. This course enables you to write your Python scripts, perform fundamental hands-on data analysis using the Jupyter-based lab environment, and create your data science projects using IBM Watson.



Pyspark

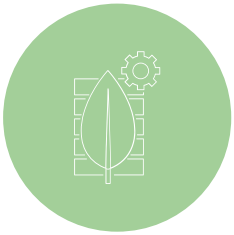
This PySpark course provides an overview of Apache Spark, the open-source query engine for processing large datasets, and how to integrate it with Python using the PySpark interface. You will learn to build and implement data-intensive applications as you dive into the world of high-performance machine learning and leverage Spark RDD, Spark SQL, Spark MLlib, Spark Streaming, HDFS, Sqoop, Flume, Spark GraphX, and Kafka.

Apache Kafka

In this Apache Kafka course, you will learn the architecture, installation, configuration, and interfaces of Kafka open-source messaging. You will gain a fair understanding of basics of Apache ZooKeeper as a centralized service and develop the skills to deploy Kafka for real-time messaging.

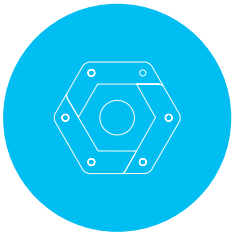
MongoDB Developer and Administrator

This course provides you an in-depth knowledge of NoSQL, data modeling, ingestion, query, sharding, and data replication.



GCP Fundamentals

In GCP Fundamentals course, you will learn how to analyze and deploy infrastructure components such as networks, storage systems, and application services in Google Cloud Platform. This course covers IAM, networking, and cloud storage and introduces you to the flexible infrastructure and platform services provided by Google Cloud Platform.



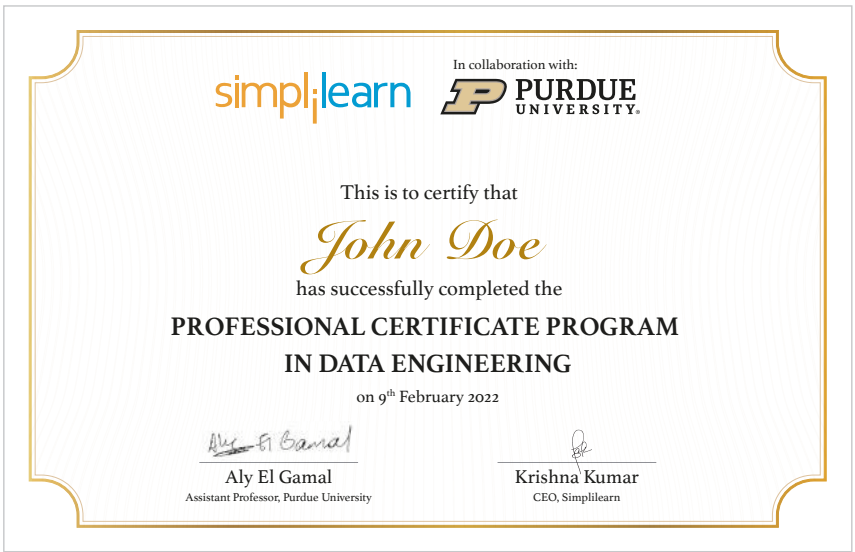
Java Training

This Java course covers the concepts of Java, from introductory techniques to advanced programming skills and provides you with the knowledge of Core Java 8, operators, arrays, loops, methods, and constructors in JDBC and JUnit framework.

Master class - Data Engineering - Purdue university

Attend an online interactive Masterclass and get insights into the world of data engineering.

Certificate



Upon completion of this Professional Certificate Program in Data Engineering in collaberation with University, you will receive a program completion certifcate from Purdue University and Simplilearn. You will also receive certificates from Simplilearn for the courses in the learning path. These certificates will testify to your skills as a data engineering Expert.

Program Endorsers

Aly El Gamal

Assistant Professor, Purdue University

Aly El Gamal has a Ph.D. in Electrical and Computer Engineering and M.S. in Mathematics from the University of Illinois. Dr. El Gamal specializes in the areas of information theory and machine learning and has received multiple commendations for his research and teaching expertise.

simplilearn

USA

Simplilearn Americas, Inc.
201 Spear Street, Suite 1100, San Francisco, CA 94105
United States
Phone No: +1-844-532-7688

INDIA

Simplilearn Solutions Pvt Ltd.
53/1 C, Manoj Arcade, 24th Main, Harlkunte
2nd Sector, HSR Layout
Bangalore - 560102
Call us at: 1800-212-7688

www.simplilearn.com