

# Data Engineering Bootcamp



Data is the lifeblood of any organization today; the data generated in the last few years accounts for 90% of all the data generated in human history. To have value, this data needs to be analyzed and turned into valuable insights; more and more organizations are turning to Data Science and Data Engineering to accomplish that goal.

## The Grand Circus Data Science Bootcamp

Turn your most valuable players into burgeoning data scientists, who can support your larger data team and goals.

Grand Circus has been up-skilling employees for over 5 years. At employers such as BCBS and Quicken Loans, we've even developed non-IT professionals into developers in as little as 12 weeks.

- The evening-based Data Engineering Bootcamp follows the up-skilling model we've already proven.
- Grand Circus will do the work of engaging and empowering your current talent, so they can in turn support your data engineering goals and teams.

## Who Can Take the Bootcamp

Employees who have a year or more programming experience and who are seeking to transition their careers into the in-demand niche of data engineering and data science, without advanced college degrees.

## Bootcamp Format

3 nights a week: Monday, Wednesday, and Thursday 6:30-8:30 pm for 26 weeks.

## Course Outline

- Unit 1: Intro to Python
- Unit 2: Descriptive Statistics & Python Libraries for Exploratory Data Analysis
- Unit 3: Storing and Manipulating Data with Structured Query Language (SQL)
- Unit 4: Java Programming
- Unit 5: Using a NOSQL database: Google Datastore
- Unit 6: Big Data Processing Systems

## Included Tools and Topics

- Python and common libraries including NumPy and pandas
- Descriptive statistics and Exploratory Data Analysis (EDA)
- Reading/writing external data and working with data from APIs
- SQL and NOSQL Databases
- Filtering, aggregating, and joining data
- Batch and real-time/streaming data processing
- Data curation, visualization, and security
- SPARK
- Hadoop
- YARN
- Hive
- Kafka
- Creating and automating pipelines

### Unit 1: Intro to Python

#### Python

Fundamental building blocks

Conditionals and loops

Ordered and Unordered Collections

Strings, Dates, and Times

### Unit 2: Descriptive Statistics & Python Libraries for Exploratory Data Analysis

#### Descriptive Statistics

Choosing the correct visualization for statistics

Data cleaning

**Python Libraries including NumPy, pandas, Datame, Matplotlib, Seaborn, GeoSpatial, Bokeh, and other popular libraries**

Creating a data-driven story

## **Unit 4: Java Programming**

## **Unit 5: Using a NOSQL database: Google Datastore**

## **Unit 6: Big Data Processing Systems**

**PObject-Oriented Programming in Java  
Java DataBase Connectivity (JDBC)**

**Google Cloud and Google Datastore  
NOSQL Queries  
Accessing Datastore through Java  
Indexing**

**Hadoop and Hadoop Distributed File System  
(HDFS)**

**YARN**

**In-memory techniques**

**Resilient Distributed Datasets**

**Apache Spark, Spark Dataframe API, Spark SQL**

**Apache Hive**

**Kafka**

**Data pipelines**