

# Big Data Analysis

## Overview

This course provides an in-depth overview of the choices you have in processing Big Data. It introduces Big Data, the types of data you might have, approaches to working on and processing the data, and the capabilities, strengths, and weaknesses of those approaches.

Topics covered include:

- NewSQL Databases
- NoSQL Overview
- Hadoop and MapReduce
- Apache Pig and Hive
- Apache Storm and Spark
- MongoDB, HBase

## AUDIENCE

This course is suitable for: Data engineers, analysts, architects, data scientist, software engineers, and technical managers who want a quick introduction to big data processing tools and choices.

## PREREQUISITES

Some knowledge of database technologies is helpful.

## Course Outline

Upon completing this course, students will be able to:

- Understand what Big Data is
- Know the difference between “data-at-rest” and “data-in-motion”
- Understand what map-reduce / Hadoop is, and what it can do
- Be aware of query technologies for easily querying with Hadoop (e.g. Hive, Pig, and others)
- Understand what NoSQL databases are and what they can do
- Become familiar with the choices in the NoSQL landscape
- Understand the strengths and weaknesses of different NoSQL technologies
- Be well-informed on your choices in Big Data processing, and evaluate them for your needs