Building Data Science Team – Course Summary

Prepared by: Sameh Amin

Table of Contents

[1 Who is The Data Team 2](#__RefHeading___Toc149_953050409)

[1.1 The Data Engineer 2](#__RefHeading___Toc455_953050409)

[1.1.1 Responsibilities 2](#__RefHeading___Toc473_953050409)

[1.1.2 Key features 2](#__RefHeading___Toc475_953050409)

[1.2 The Data Scientist 2](#__RefHeading___Toc457_953050409)

[1.2.1 Responsibilities 2](#__RefHeading___Toc477_953050409)

[1.2.2 Key features 2](#__RefHeading___Toc479_953050409)

[1.3 The Data Science Manager 2](#__RefHeading___Toc459_953050409)

[1.3.1 Responsibilities 2](#__RefHeading___Toc481_953050409)

[1.3.2 Key features 2](#__RefHeading___Toc483_953050409)

[1.4 When you need a Data Team? 3](#__RefHeading___Toc151_953050409)

[1.4.1 If you’re a Startup company 3](#__RefHeading___Toc278_3113093047)

[1.4.2 If you’re a mid-size organization 3](#__RefHeading___Toc280_3113093047)

[1.4.3 For larger organizations 3](#__RefHeading___Toc282_3113093047)

[2 Qualifications & Skills 4](#__RefHeading___Toc153_953050409)

[2.1 Data Engineer Skills 4](#__RefHeading___Toc467_953050409)

[2.2 Data Scientist Skills 4](#__RefHeading___Toc469_953050409)

[2.3 Data Science Manager Skills 4](#__RefHeading___Toc471_953050409)

# Who is The Data Team

## The Data Engineer

### Responsibilities

Who is a person that may construct a database, deploy machine learning algorithm to production on a scale, plan and create the infrastructure.

### Key features

* Hardware and software architecture skills
* APIs and back-end development to run the machine learning algorithm online and on real-time.

## The Data Scientist

### Responsibilities

Whoe is the person that pull the data out of database, analyze it, perform experience on it and visualize it, deliver the results to the Data Science manager.

### Key features

* Python, java, SQL.
* Machine learning algorithms.

## The Data Science Manager

### Responsibilities

Whoe is the person that makes sure everybody is interacting with each other, keep things moving, recruit and build the data science team and interface with upper management and organization as same as other collaborative teams on the same level. And he advertise the Data science team, showing their capabilities and what they’re doing.

### Key features

* Leardership and managerial skills.
* Python, Java, SQL and other programming languages skills.
* Web, Desktop and Mobile applications general knowledge.
* Machine learning knowledge.
* Software and Hardware architecture skills.

# When you need a Data Team?

## If you’re a Startup company

You have to build instratcture first before hiring any data science team, making sure the date is collected the right way, data tools are available, ...

## If you’re a mid-size organization

You can start to hire data scientists. Implementing new data products. Then you can implement that in your created system. Then you can scale it up.

## For larger organizations

You have an infrastructure, you have a data science team, you have a running data projects. You may have a data science manager that organize the connections between teams, organizing the flow of work.

# Qualifications & Skills

## Data Engineer Skills

* Hardware and infrastructure knowledge, establishment and mentain.
* Different types of databases software knowledge and handson.
  + mySQL, MongoDB, ..
  + Data process on scale, Hadoop, ...
  + Computer science background.
  + Communication skills are very very important.

## Data Scientist Skills

* Statistics (Degree or online classes).
* Machine learning algorithms.
* Data Analysis skills (Pull data from source, analyze the data, clean it, do statisticals, communicate results to Data Science Manager.
* Visualize the data.
* Python or R languages, D3js for data visualization, ....
* Have a common sense of data problems, and figure out what are the right solutions to solve for each problem.

## Data Science Manager Skills

* Software and Hardware Architecture background.
* Data Science and machine learning algorithms knowledge is a very nice to have.
* To know each data science team roles and responsilibities.
* He needs to know what projects should be accepted and what can be achieved, because not all problems can be solved using Data science.
* Prioritize and set priorities for the stream of projects.
* Communicate projects plan and results to upper management.