

LESSON HANDOUT

The Data Analyst role in context

Understanding the hierarchy of an organisation's Data Science needs is a great method of putting the role of a data analyst in context of both:

- The field of data science, and;
- Other roles in data science such as the Data Scientist and Data Engineer.

The Hierarchy of Data Science needs

DATA SCIENCE HEIRARCHY OF NEEDS

WITHYOU
WITHME

OPTIMISE/ LEARN

AI, Deep
Learning

MODEL/ TEST

Experimentation,
Machine learning

ANALYSE/ AGGREGATE

Analytics, Aggregation, Segment,
Metrics

TRANSFORM

Cleaning, Data preparation, Classification,
Anomaly detection

MOVE/ STORE

Data pipeline, Migration, Data storage, ETL

COLLECT

Logging, Sensors, Survey , Instrumentation, User generated
data, IT Infrastructure

The hierarchy of needs is a clever tool because it reinforces that, whilst AI and machine learning, at the top of the pyramid, are very powerful, an organisation must achieve all of the rungs below it to employ these techniques correctly.

Let's examine the data science hierarchy from an organisational perspective, starting at the bottom.

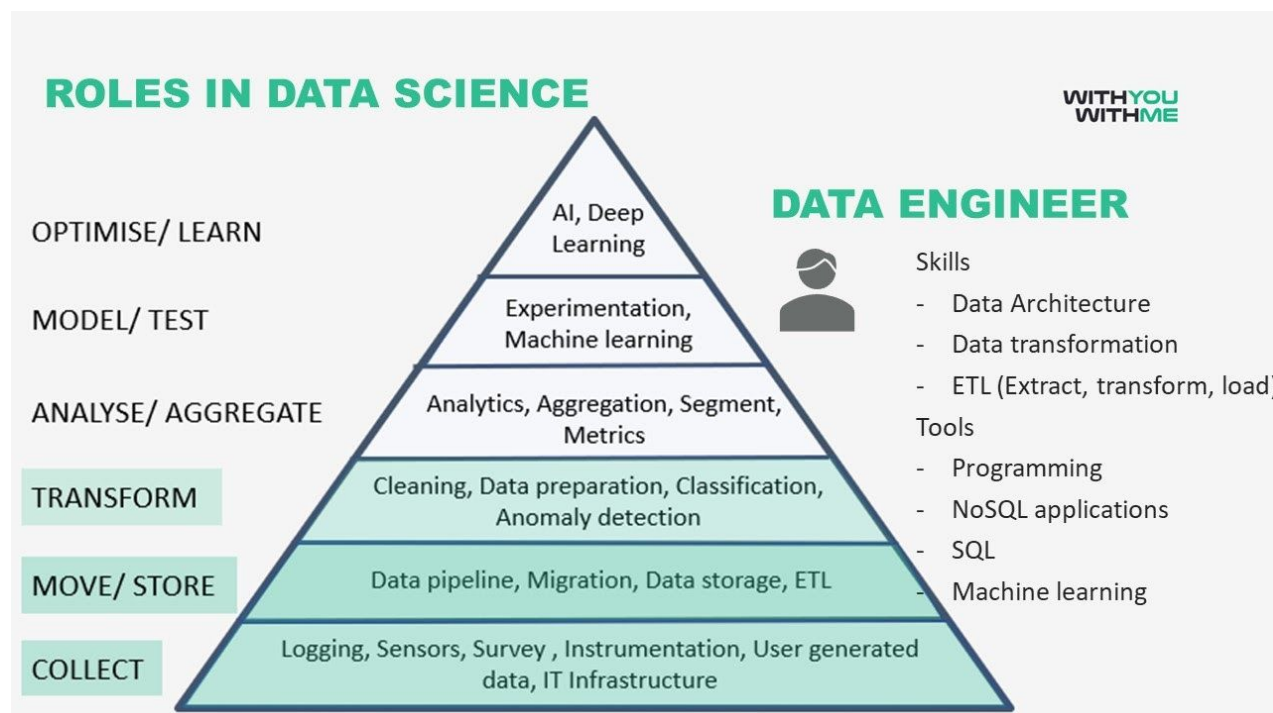
- **Collect.** This stage refers to the need for data to be captured, whether collecting data from machines through instrumentation, sensors, and logging or collecting data from humans through user-data or surveys.

- **Move/ Store.** This stage related to the storage of data. For data science to be successful, data must be accessible and in the right format for analysis. This stage might involve migrating data from different locations into a consolidated data warehouse or data lake.
- **Transform.** This rung refers to what is called cleaning the data and is where data is prepared for analysis. During this stage, organisations have to make decisions about how to classify different data.
- **Analyse/ Aggregate.** In this stage, we make value out of our data by aggregating it or summarising it into metrics that mean something to the organisation.
- **Model/ Test.** Modelling data allow organisations to ask very specific questions of it, enabled by machine learning and a scientific approach.
- **Optimise/ Learn.** Once a machine learning practice is established, organisations can apply advanced techniques such as deep learning which replicate human analytical capability.

The point of presenting this hierarchy is to make it clear that not every problem is a deep learning or machine learning problem and that there is much value to be added in all of the rungs that lead up to these advanced techniques.

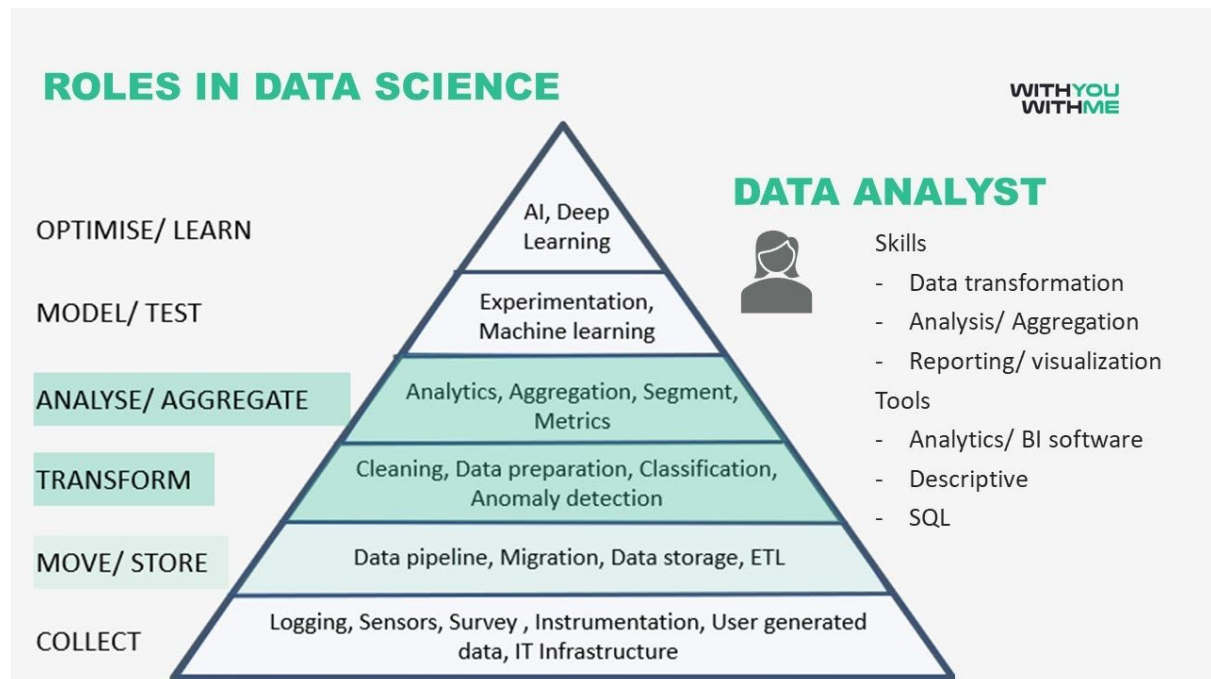
Roles in Data Science

The Data Engineer



The Data Engineer specialises in transforming data into the format required for analysis and are able to build a data pipeline where data is taken from collection through storage and into analytics platforms. The Data Engineers toolset is programming heavy and should include expertise in no SQL applications (required to deal with unstructured data). The data engineer may also use machine learning algorithms to assist in the cleaning and transformation of data.

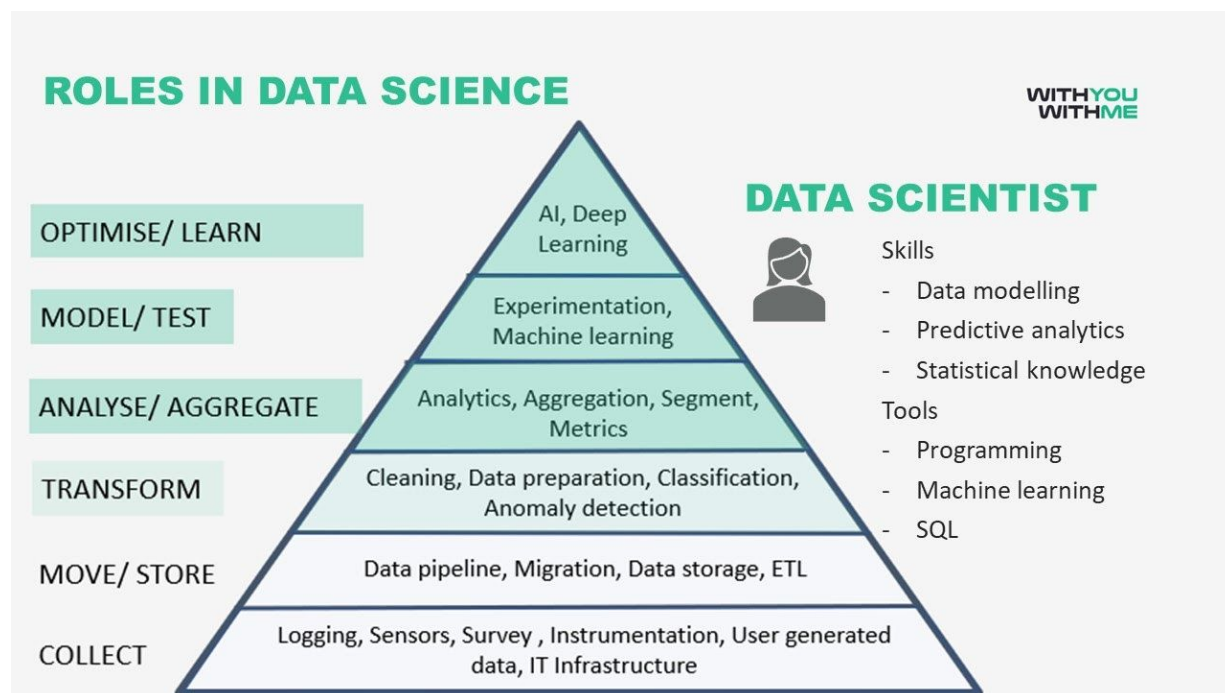
The Data Analyst



The Data Analyst normally sits firmly in the centre of the pyramid and is concerned with making the immediate value of data. The Data Analyst will typically do this through aggregation (summarising it), segmentation (dividing it into meaningful groups) and by building metrics (calculations and ratios that mean something to the organisation). The Data Analyst typically communicates this value to the organisation through visualisation and reporting.

The Data Analyst must also be able to transform extract and transform data but the extent to which this is required will depend on the organisation. This is because large organisations can afford to stovepipe whereas small organisations (where resources are scarce) are likely to require Data Analysts to complete a broader scope of work.

The Data Scientist



The Data Scientist uses techniques such as machine learning to ask more specific questions of the data. Where a Data Analyst might identify a trend and a broad solution to a problem, the Data Scientist can also provide specific advice on how to fix it. For example, a Data Analyst might identify that a product's sales are declining and that a recent price hike is responsible. The Data Scientist can go one step further and model the data to produce an exact recommend price for that product.