

In today's IT environment, a Data Scientist requires so many skills, that it very rare to find one individual that possesses all of them. Below we take a look at data science as a whole, along with the essential toolkit needed to be a successful Data Scientist or someone who works within the Data Science team.

### What is data science?

Data science is the use of machine learning, mathematical and statistical analysis to identify correlations and relationships between data. With the main purpose being to classify, predict events, identify patterns and anomalies, whilst calculating probability and interest.

### What are the areas of focus for the data science team?

Data science covers a whole spectrum of skills (i.e. business intelligence, fields of study and technique, implementation and data relations), but can often become confused with buzzwords, such as; Big Data, decision science, Internet of things, or even hacking.

Let's take a closer look at each of these areas:

**Business Intelligence** – can be gathered via various means such as; dashboards, reports, visualisation, queries and insights.

**Study & Technique** – uses algorithms, mathematics, statistical analysis, predictive modelling, machine learning, text mining, mapping and sentiment in order to make sense of the business intelligence.

**Implementation** – would cover the applications used to filter and monitor the data, such as; distributed computing, software, open source, in-database analytics, Hadoop; and will be used for unstructured data, data cleansing, ETL and sensors.

### What skills should I be developing?

To be a good Data Scientist you should concentrate on developing the following skills and knowledge:

- Programming skills
- Communication skills
- Mathematical and statistical analysis skills
- Technical knowledge
- Domain knowledge

Whilst your strengths may lay in 1 or 2 of the areas above, it would be recommended that you develop some understanding in all areas.

### What is the Data Scientist toolkit?

The data science toolkit is vast and continually growing as new technology and techniques become available. We have listed below just some of the platforms and tools that are on the market today.

<p><b>Finding Data Platforms:</b> GreenPlum DB Hadoop SAS HPA AWS</p>	<p><b>Writing Code Editing:</b> Vi/Vim Emacs Smultron TextWrangler Eclipse Notepad++ iPython Sublime</p>	<p><b>Writing Code Languages:</b> SQL Bash Scripting C C++ C# Java Python R</p>
<p><b>Writing Code for Big Data In-Database:</b> SQL PL/Python PL/Java PL/R PL/pgSQL <b>Hadoop:</b> Pig Hive Java</p>	<p><b>Running Code:</b> pgAdmin III PSQL Pycopg2 Terminal Cygwin PuTTY WinSCP</p>	<p><b>Implement Algorithms Libraries:</b> MADlib <b>Java</b> Mahout <b>R</b> <b>Text</b> OpenNLP NLTK GP Text <b>C++</b> OpenCV <b>Python</b> NumPy Scripy Scikit Learn Pandas</p>
<p><b>Implement Algorithms Programmes:</b> Alpine Miner RStudio MATLAB SAS Stata</p>	<p><b>Results Visualisation:</b> Python Matplotlib Python Networkx D3.js Tableau GraphViz Gephi R (i.e. Lattice) Excel</p>	<p><b>Collaboration:</b> Chorus Confluence Socialcast GitHub Google Drive &amp; Hangouts</p>

We have a number of roles for Data Scientists. We would be happy to talk to you about your needs and then introduce you to some clients, who we feel would suit you. Just phone us on 020 8123 7769 or email us: rod@resourceondemand.com and one of our team will be happy to help.