CAREERS IN DATA SCIENCE

It's easier than ever for technology to capture data. But what do we do with it? Almost every interaction with technology, every website we peruse, and every app we install, utilizes data. With the glut of raw data now being collected, we need professionals who can find patterns in that data and tell stories from it. Data experts are in demand in almost every field, from government security to dating apps. Millions of businesses, nonprofits, and governmental agencies rely on big data to succeed and better serve their customers and constituents.

Sample Career Paths in Data Science:

Data science roles and responsibilities can vary quite a bit. We've outlined five different data roles along with the skill sets and technical knowledge required. Keep in mind that the technologies and tools needed for many of these positions will change, but many of the soft skills will remain the same.

DATA ANALYST
Use data to tell stories about the past

Data analysts dig into large data sets to discover new insights or trends. They ensure the quality of data being scraped, triage data issues, and query databases for stakeholder requests. While they aren't inventing new algorithms, they are capable of using existing analysis tools to solve problems. Often, they are responsible for packaging the data to provide digestible takeaways in a visual or narrative form.

Skills required: Data munging and data visualization, statistics, machine learning, Excel, SQL, NoSQL, R, HTML, Python, Javascript, SQL, C/C++ , SAS, XML, Hadoop

DATA ENGINEER
Build, test, and maintain data architecture

A data engineer is usually a software developer by trade. Rather than perform data analysis, they install and compile database systems, set up disaster recovery systems and write complex queries. Data engineers work in tandem with data architects to build the working framework for data search and retrieval. They are responsible for ensuring that data flows easily from origin to destination so that it can be processed.

Skills required: Building & designing large-scale applications, database architecture, data modeling & mining, statistical modeling, regression analysis, distributed computing, Hadoop-based analytics (Hive, MapReduce, Pig, HBase), R, SAS, Python, C/C++, Ruby Perl, Java, MatLab, SQL, Cassandra, Linux, UNIX, Solaris, machine learning

DATA SCIENTIST
Use data to strategize for the future

Essentially, a data scientist creates the tools that a data analyst then uses. Data scientists tend to work on a more theoretical, research-based level than data analysts. They look at data from multiple sources and create new algorithms to solve organizations' data problems. Data scientists become experts on a wide range of skills and competencies, including but not limited to handling raw data, analyzing that data via statistical methods, and sharing their insights with peers in a captivating way. Data scientists may be given the leeway to follow their own interests and ideas. Because they have a skillset founded in research and theory, they can run their own experiments with data to find new trends and insights. They are driven by intense curiosity and an urge to ask questions and find data-driven solutions.

Skills required: Distributed computing, predictive modeling, storytelling and visualizing, linear algebra, calculus, probability, stats, machine learning, data mining, data cleaning and munging, data visualization/unstructured data techniques, SAS, SPSS, MATLAB R, Python, Java, C/C++, Hadoop Platform, SQL/NoSQL databases

DATA ARCHITECT
Design frameworks for data organization

Data architects lay the foundation for data analysts and scientists to easily retrieve the needed data for their assessments and projects. Data architects create the blueprints for data management systems to integrate, centralize, protect and maintain data sources. Along with the data engineer, the data architect builds this framework. What increasingly differentiates them from data engineers is that they then use that framework to visualize data in order to make it more digestible for data scientists and analysts to use. They often have a wide breadth of knowledge in data warehousing, database management, and data modeling.

Skills required: Applied math and statistics, data visualization and data migration, machine learning, RDMSs (relational database management systems), information management and data processing, backup/archival software, Java and Python, Perl, C/C++, NoSQL, Oracle, Linux, UNIX, Hadoop, SQL Server, Enterprise Architect, Visio

DATABASE ADMINISTRATOR
Oversee, monitor and ensure access

Database administrators design and maintain an organization's data storage facilities. They are responsible for the warehousing, arrangement, usage, presentation, and examination of an operation's database and database management software. DBA's also maintain the backup and recovery systems and track who has access to the database in addition to monitoring for any unauthorized usage. They may build new databases from scratch and figure out how to collect new types of data and integrate them into the database. They must routinely test the database and modify it to ensure it is running correctly.

Skills required: Backup & recovery, data modeling & design, data security, ERP, SQL, NoSQL, XML, Python, Java, Ruby on Rails, C#, Hadoop, ETL, BI, Puppet, relational database frameworks (IBM, Oracle, Microsoft), project management disciplines, change management