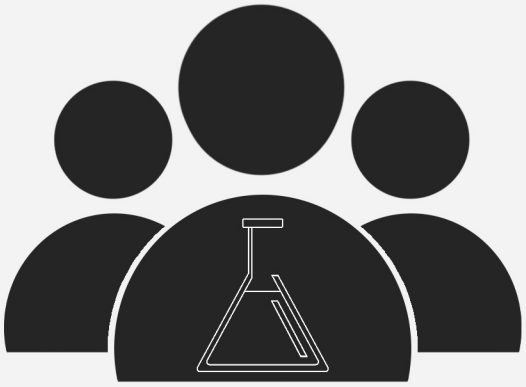
How to Create and Manage a Data Science Team.



How to Create and Manage a Data Science Team.

Data science is a relatively new, evolving & exciting data function. As this article explains, different organisations have various ways of organising their data science teams, along with tips on managing them.

Organisations increasingly see data as a valuable asset that will help them succeed now and in the future.

The value of data has been increasing in recent years, and it shows no sign of slowing down.

The first barrier to effective data and analytics is still the lack of qualified talent. Other familiar challenges include limited access to siloed data, lack of processing power, and the absence of a data strategy to help turn data into actionable information. More and more organisations are creating data science functions to lead their efforts in data mining, predictive modelling, machine learning and AI.

We have created this guide to provide best practices for creating and managing a data science team.

We have included the different ways a team can be set up, the positions it's likely to form and the executives to who a team may report to in an organisation.

Models for structuring a data science team :

Data collection, management and analysis are typically the responsibility of the chief information officer (CIO). The IT team works with business users to implement data warehouses and business intelligence (BI) systems that hold and organise data, enabling fundamental analysis and reporting. However, over the past two decades, more organisations have separated the data function into their department as internal data stores grew, supporting technologies

evolved, and data-related tasks became more differentiated and specialised.

The increasing importance of analytics to business success also drove the need for a data science team with skilled data scientists and engineers. Today, many organisations anything from a team or entire data science department providing this service. Larger organisations may have multiple teams that operate independently or in a coordinated way.

These teams are tasked with collecting and cleaning data from various sources, identifying patterns and insights, and presenting their findings to executives with actionable recommendations. Often, this involves working with internal teams, external partners, and vendors specialising in certain analysis types.

Data scientists may work in areas like marketing and sales, finance and accounting, product development, human resources (HR), customer service, operations management (OM), risk management (RM), legal affairs (LA), compliance/governance (CG), etc.

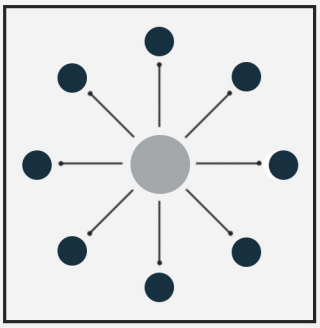
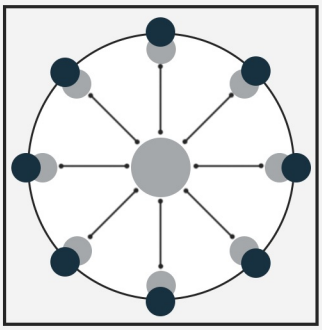
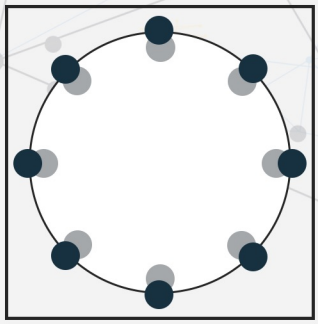
How a company structures its teams vary based on its data science program's maturity, data analytics goals, overall organisational structure and enterprise culture. However, some common data science team structure models have emerged, each with pros and cons. Team structures can be:

A Decentralised Team where members work within the individual business units they support. This allows team members to collaborate closely with business ondata science projects. This approach can hinder the strategic use of data across an organisation and require more resources than smaller companies may have available.

A Centralised Team that consolidates a data science function into the enterprise, which manages individual projects and oversees resourcing. This approach allows for an enterprise-wide strategic view and uniform implementation of analytics best practices more efficiently. However, it can limit the ability of team members to become experts in a particular area of the business. Some organisations establish a formal data science centre to house a centralised team.

A Hybrid Team approach creates a data science team who centrally manage all projects with specific business operations. This team are accountable for helping those units reach their objectives to make

data-driven decisions. In hybrid structures, a centre of excellence may also focus on promoting best practices and standards for data science. As with the decentralised model, resource constraints can be an issue.

CENTRALISED	EMPOWERED	DECENTRALISED
		
Bottleneck created through the central ownership of governance and processes. Thus limiting the ability of the business to self serve and only one direction of communication.	Innovation and Governance communities ensure that business representatives are empowered to deliver governance and processes directly to business users. Communities act as a meeting point to inform central business Intelligence team of additional business requirements and share and resolves issues following best	Full ownership delegated to business functions. Bottlenecks are removed but no benefits are realised through synergies and a fragmented and inefficient corporate function will emerge.

Data science team - Roles & Responsibilities :

Successful data science teams share common structures, roles and responsibilities regardless of the size/scale.

Small organisations or those with limited analytics needs or

early-stage data science initiatives may have a generalist handle all the required tasks. Larger entities and those with more mature programs typically include some combination of the following roles in their data science teams.

Data Scientist : Data scientists are a key member of the team, using statistical methods and machine learning algorithms to analyse data and create predictive models. They also build data products, recommendation engines and other technologies for various use cases. Data scientists typically have multiple skills in mathematics, statistics, data wrangling, data mining, coding and predictive modelling. Expect people in this role to have advanced data science degrees or graduate-level data science certifications.

Data Analyst : Data analysts are responsible for collecting and maintaining data from operational systems and databases. They use statistical methods and analytics tools to interpret the data and prepare dashboards and reports for business users. Data analysts do not have the complete skill set of a data scientist, but they can support data science efforts.

Data Engineer : This role is responsible for building, testing, and maintaining data pipelines that power a business. A data engineer uses a combination of software engineering and computer science skills to focus on the technology infrastructure, data collection, management and storage. Data engineers work closely with data scientists on tasks such as data quality, preparation (cleansing), model deployment (including setting up models in production), and maintenance.

Data Architect : Data architect are responsible for designing and overseeing the implementation of the system design and infrastructure. This role can also been assumed by a data engineer.

Machine Learning Engineer : Sometimes, this role is referred to as AI engineer: Machine learning engineers are responsible for creating, deploying and maintaining the algorithms and models needed for machine learning and AI initiatives.

In some organisations, data science teams may also include these positions.

Citizen Data Scientist : An informal role can involve business analysts, business-unit power users and other employees capable of doing their own data analytics work. Citizen data scientists are often interested in understanding or training in advanced analytics. However, the technologies they use -- for example, automated machine learning tools -- typically require little to no coding. They often work outside a data science team but may be incorporated into ones embedded in business units.

Business Analyst : Business analysts are key in supporting the work of data scientists. Data scientists are responsible for gathering, cleaning, and organizing data, as well as creating new or altering models that predict what will happen in the future.

In addition, business analysts may be attached to data science teams, which includes evaluating business processes and translating business requirements into analysis plans, areas in which they can help support the work of data scientists.

Data Translator : The term "analytics translator" is a relatively new one in the corporate world, but it refers to a very important role that's growing in popularity. The analytics translator acts as a liaison between data science teams and business operations, helping create plan projects and translate the insights from data analytics into recommended business actions. This role can often fall to the Business analyst.

Data Visualisation Developer or Engineer : They're tasked with creating data visualisations to make information more accessible and understandable for business professionals. However, data scientists and analysts may handle this role on some teams.

"Regardless of sector or industry, data science teams need to be strong in three core areas: mathematical, technology and business acumen. It is rare to find a single person that excels in all three. Often companies will have someone fluent in two of three, and then the rest of the team is built around that, filling in the gaps to ensure the team is strong in all three."

- Simon Meacher, Managing Director, EngagingData

Management and Oversight :

A data science team will be managed and overseen by either a lead data scientist, data science manager, director of data science, or similar managerial position.

The reporting structure for teams similarly varies. Generally, organisations assign a C-level executive or high-ranking functional manager to oversee the data science team.

In many cases a Chief Data Officer will oversee the data science function.

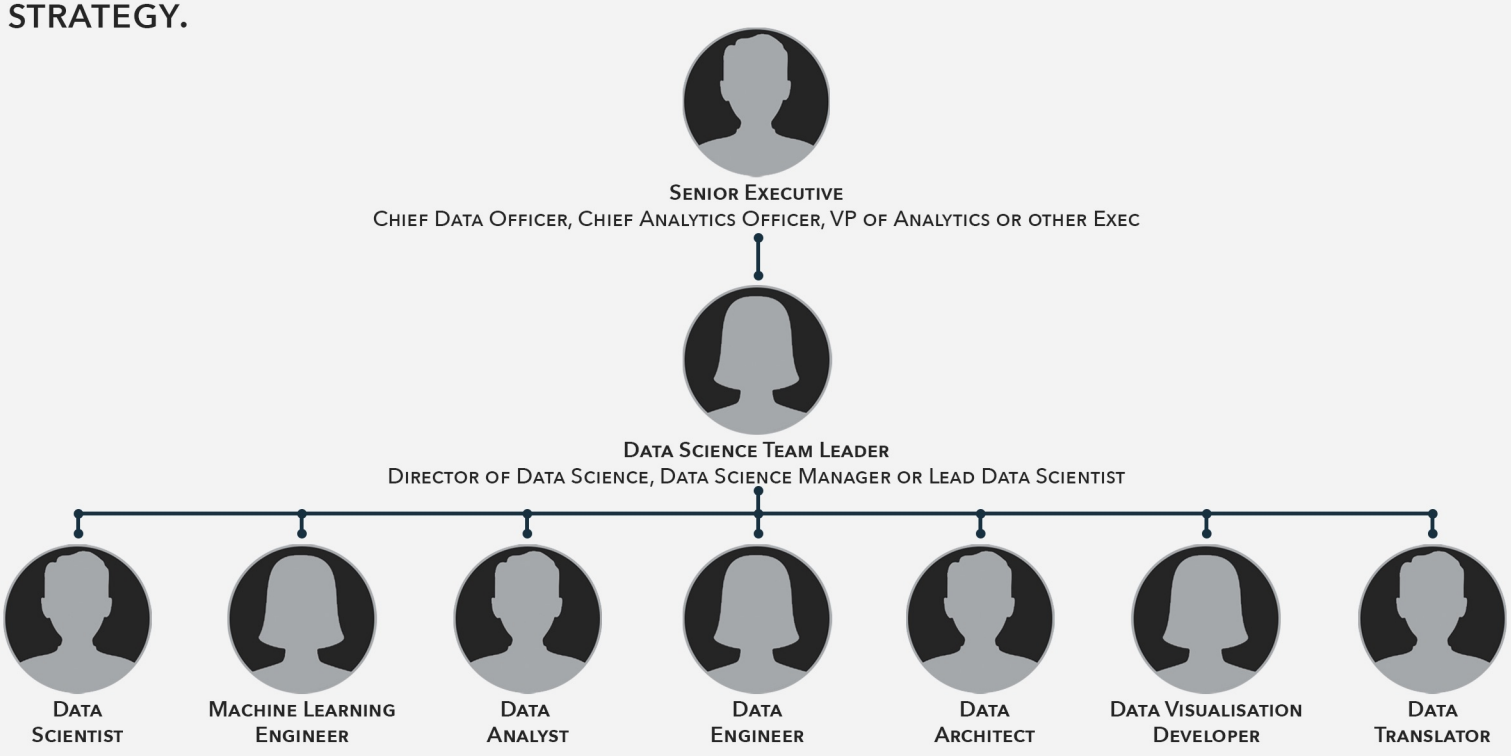
In 2002, Capital One created the first CDO position within the financial services industry. Since then, the CDO role has grown in popularity.

This role initially focused on data governance, management and security functions. More recently, CDOs have alsotaken on responsibilityfor data science, analytics and AI.

Other organisations have created a chief analytics officer (CAO) role tooversee their data science and analytics teams.

Hybrid roles exist, combining the CDO and CAO responsibilities into a chief data and analytics officer role.

The head of a data science team my be subject to matrix reporting, allowing the role to report to a different executive; for example, the COO, CFO or CIO, or a position such as director of analytics, business intelligence director, head of business data or director of data and strategy.



How Data Scientists work with Business Users :

Organisations within all industries arerecognised the need to become data-driven & see it as a key to remaining competitive & set up the data science team to collaborate with business teams to:

- Understand the business problem or questions that the team want to answer.
- Set and articulate the objectives for using the data.
- Plan how to apply the knowledge to make decisions and take action.

Once they understand, data science teams cannot merely present their findings. They work with the business teams to understand the insights gained from the data and how that information can shape product and service offerings, marketing campaigns, supply chain management and other critical parts of business processes/operations to support company goals, such as higher revenue, increased efficiency and better customer service.

"In my experience, Data science teams need to work closely with the business. Without the using the wealth of knowledge about the data from the business, the data scientist will struggle to provide value from the data."
- Carl Richards, Head of Consulting, EngagingData

Tools that a Data Science Team needs :

Dozens of tools, ranging from data visualisation and reporting software to advanced analytics, machine learning and AI technologies, enable data science teams' work. The number and combination of technologies needed can be unique to each team, based on its goals and skill levels.

The following is a list of commonly used data science tools that includes both commercial and open source technologies:

- statistical analysistools, such as SAS and IBM SPSS;
- machine learning frameworks and libraries, including TensorFlow, Weka, Scikit-learn, Keras and PyTorch;
- Data Science platforms from various vendors that provide diverse sets of capabilities for analytics, automated machine learning, and workflow management and collaboration;

programming languages, in particular, Python, R, Julia, SQL, Scala and Java;

- Jupyter Notebook and other interactive notebook applications for sharing documents that contain code, equations, comments and related information;
- Data Visualisation tools, such as Tableau, QlikView, Power BI, D3.js and Matplotlib;
- AWS, Azure, Google BigQuery, Hadoop, Snowflake, Spark and other analytics engines and big data platforms;
- Cloud Object Storage services and NoSQL databases; and
- The Kubernetes container orchestration service for deploying analytics and machine learning workloads in the cloud.

Best Practices for Managing a Data Science Team :

Executives and team leaders seeking to build and mature their data science programs should consider the following best practices for managing their teams.

Seek out workers with a range of business, interpersonal, and technical skills to help ensure that the team can meet organisational objectives.

Create a culture of learning and innovation that challenges team members and encourages them to bring new thinking to business problems and issues.

Promote analytics projects that encourage close collaboration between the data science team and the business units they support.

Evaluate team members at least partly on the business successes and work drives. Create a mentorship program to help advance the skills of junior team members, and do ongoing training to ensure that all workers stay current on key data techniques and technologies.

Talent retention programmes will help to keep data scientists, who are in high demand and experienced ones have plenty of job opportunities.

Summary :

The world is changing, and data science is one of the most powerful tools for that change.

Data science is more than just crunching numbers; The ability to create a great data science function within your company will benefit your organisation and future proof it's ability to change with its strategic objectives.

By embracing data science as an integral part of your business, you can ensure that your organisation is agile enough to keep up with changes in technology and consumer behaviour.