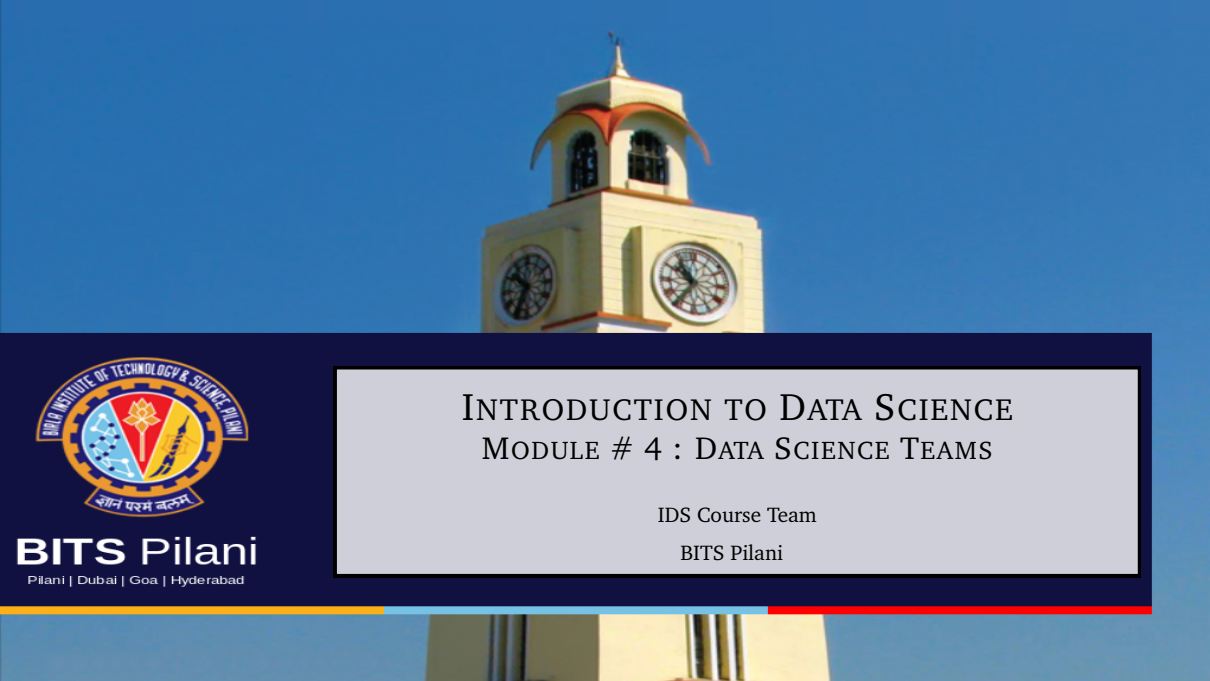


BITS Pilani
Pilani | Dubai | Goa | Hyderabad

Introduction to Data Science

Module # 4 : Data Science Teams

IDS Course Team

BITS Pilani

The instructor is gratefully acknowledging the authors who made their course materials freely available online.

Table of Contents

Data Driven Decision Making

Usecase: Airbnb

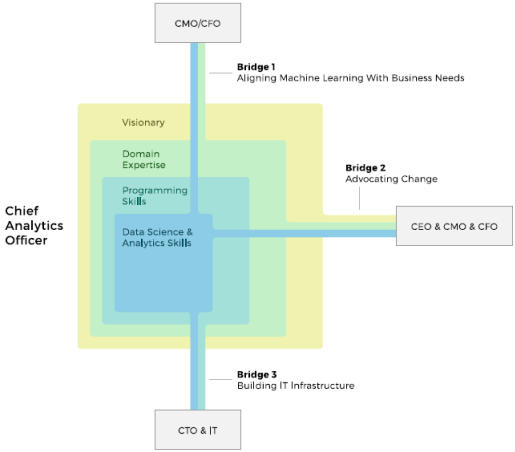
- Experiment.
  - Find ways to put data into new projects using an established Learn-Plan-Test-Measure process.
- Democratize data.
  - Scale a data science team to the whole company and even clients.
- Measure the impact.
  - Evaluate what part DS teams have in your decision-making process and give them credit for it.

https://www.altexsoft.com/blog/datascience/how-to-structure-data-science-team-key-models-and-roles/

Roles in Data Science Team [1/6]

[1] Chief Analytics Officer / Chief Data Officer

- CAO, a “business translator,” bridges the gap between data science and domain expertise acting both as a visionary and a technical lead.
- Preferred skills: data science and analytics, programming skills, domain expertise, leadership and visionary abilities.



https://www.altexsoft.com/blog/datascience/how-to-structure-data-science-team-key-models-and-roles/

Roles in Data Science Team [2/6]

[2] Data analyst

- The data analyst role implies proper data collection and interpretation activities.
- An analyst ensures that collected data is relevant and exhaustive while also interpreting the analytics results.
- May require data analysts to have visualization skills to convert alienating numbers into tangible insights through graphics. (eg: IBM or HP)
- Preferred skills: R, Python, JavaScript, C/C++, SQL

https://www.altexsoft.com/blog/datascience/how-to-structure-data-science-team-key-models-and-roles/

Roles in Data Science Team [3/6]

[3] Business analyst

- A business analyst basically realizes a CAO's functions but on the operational level.
- This implies converting business expectations into data analysis.
- If your core data scientist lacks domain expertise, a business analyst bridges this gulf.
- Preferred skills: data visualization, business intelligence, SQL.

[4] Data scientist

- A data scientist is a person who solves business tasks using machine learning and data mining techniques.
- The role can be narrowed down to data preparation and cleaning with further model training and evaluation.
- Preferred skills: R, SAS, Python, Matlab, SQL, noSQL, Hive, Pig, Hadoop, Spark

https://www.altexsoft.com/blog/datascience/how-to-structure-data-science-team-key-models-and-roles/

Roles in Data Science Team [4/6]

Job of a data scientist is often divided into two roles

[4A] Machine Learning Engineer

- A machine learning engineer combines software engineering and modeling skills by determining which model to use and what data should be used for each model.
- Probability and statistics are also their forte.
- Training, monitoring, and maintaining a model.
- Preferred skills: R, Python, Scala, Julia, Java

[4B] Data Journalist

- Data journalists help make sense of data output by putting it in the right context.
- Articulating business problems and shaping analytics results into compelling stories.
- Present the idea to stakeholders and represent the data team with those unfamiliar with statistics.
- Preferred skills: SQL, Python, R, Scala, Carto, D3, QGIS, Tableau

https://www.altexsoft.com/blog/datascience/how-to-structure-data-science-team-key-models-and-roles/

Roles in Data Science Team [5/6]

[5] Data architect

- Working with Big Data.
- This role is critical to warehouse the data, define database architecture, centralize data, and ensure integrity across different sources.
- Preferred skills: SQL, noSQL, XML, Hive, Pig, Hadoop, Spark

[6] Data engineer

- Data engineers implement, test, and maintain infrastructural components that data architects design.
- Realistically, the role of an engineer and the role of an architect can be combined in one person.
- Preferred skills: SQL, noSQL, Hive, Pig, Matlab, SAS, Python, Java, Ruby, C++, Perl

https://www.altexsoft.com/blog/datascience/how-to-structure-data-science-team-key-models-and-roles/

Roles in Data Science Team [6/6]

[7] Application/data visualization engineer

- This role is only necessary for a specialized data science model.
- An application engineer or other developers from front-end units will oversee end-user data visualization.
- Preferred skills: programming, JavaScript (for visualization), SQL, noSQL.

https://www.altexsoft.com/blog/datascience/how-to-structure-data-science-team-key-models-and-roles/

Data Scientist [1/2]

- Data scientists are responsible for discovering insights from massive amounts of structured and unstructured data to help shape or meet specific business needs and goals.
- Role
  - Main objective is to organize and analyze large amounts of data, often using software specifically designed for the task.
- Responsibility
  - Chief responsibility is data analysis, a process that begins with data collection and ends with business decisions made on the basis of the data scientist's final data analytics results.

https://www.cio.com/article/3217026/what-is-a-data-scientist-a-key-data-analytics-role-and-a-lucrative-career.html

Data Scientist [2/2]

Stitch Fix's Michael Hochster defines two types of data scientists:

- Type A stands for Analysis
  - This person is a statistician that makes sense of data without necessarily having strong programming knowledge.
  - Type A data scientists perform data cleaning, forecasting, modeling, visualization, etc.
- Type B stands for Building
  - These folks use data in production.
  - They're excellent good software engineers with some statistics background who build recommendation systems, personalization use cases, etc.

https://www.altexsoft.com/blog/datascience/how-to-structure-data-science-team-key-models-and-roles/

Data Scientist Requirements - Industry-wise

- Business
  - Data analysis of business data can inform decisions around efficiency, inventory, production errors, customer loyalty and more.
- E-commerce
  - improve customer service, find trends and develop services or products.
- Finance
  - data on accounts, credit and debit transactions and similar financial data, security and compliance, including fraud detection.
- Government
  - form decisions, support constituents and monitor overall satisfaction, security and compliance.
- Science
  - collect, share and analyze data from experiments in a better way.

https://www.cio.com/article/3217026/what-is-a-data-scientist-a-key-data-analytics-role-and-a-lucrative-career.html

Data Scientist Requirements - Industry-wise

- Social networking
  - targeted advertising, improve customer satisfaction, establish trends in location data and enhance features and services.
  - Ongoing data analysis of posts, tweets, blogs and other social media can help businesses constantly improve their services.
- Healthcare
  - Electronic medical records requires a dedication to big data, security and compliance.
  - Improve health services and uncover trends that might go unnoticed otherwise.
- Telecommunications
  - All electronics collect data, and all that data needs to be stored, managed, maintained and analyzed.
  - Data scientists help companies squash bugs, improve products and keep customers happy by delivering the features they want.

https://www.cio.com/article/3217026/what-is-a-data-scientist-a-key-data-analytics-role-and-a-lucrative-career.html

Skillset for a Data Scientist

Programming: Most fundamental of a data scientist's skill set. Programming improves your statistics skills, helps you "analyze large datasets" and gives you the ability to create your own tools.

Quantitative analysis: Improve your ability to run experimental analysis, scale your data strategy and help you implement machine learning.

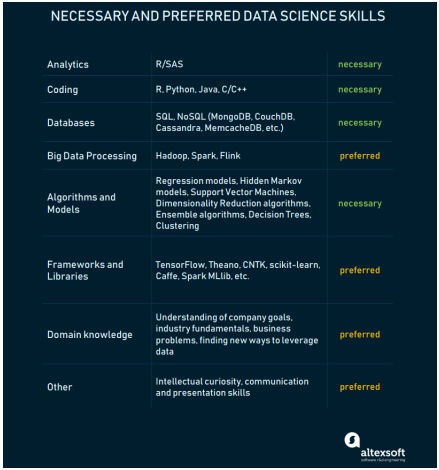
Product intuition: Understanding products will help you perform quantitative analysis. It will also help you predict system behavior, establish metrics and improve debugging skills.

Communication: Strong communication skills will help you "leverage all of the previous skills listed."

Teamwork: It requires being selfless, embracing feedback and sharing your knowledge with your team.

William Chen, Data Science Manager at Quora

Skillset of a Data Scientist



Data Science Team Building

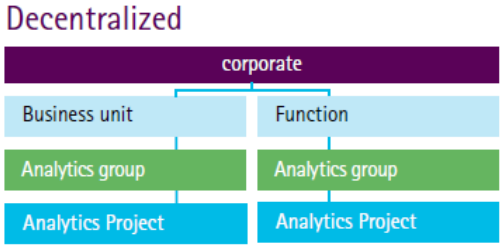
- Get to know each other for better communication
- Foster team cohesion and teamwork
- Encourage collaboration to boost team productivity and performance.

https://towardsdatascience.com/why-team-building-is-important-to-data-scientists-a8fa74dbc09b

Organisation of Data Science Team

[1] Decentralized

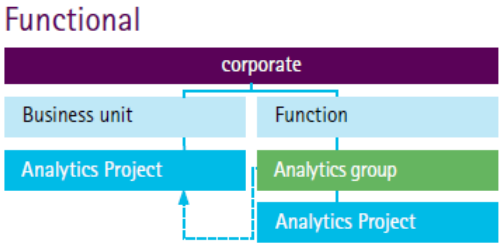
- Data scientists report into specific business units (ex: Marketing) or functional units (ex: Product Recommendations) within a company.
- Resources allocated only to projects within their silos with no view of analytics activities or priorities outside their function or business unit.
- Analytics are scattered across the organization in different functions and business units.
- Little to no coordination
- Drawback – lead to isolated teams



Organisation of Data Science Team

[2] Functional

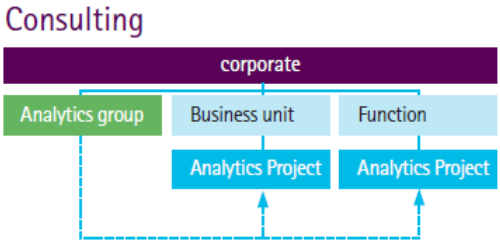
- Resource allocation driven by a functional agenda rather than an enterprise agenda.
- Analysts are located in the functions where the most analytical activity takes place, but may also provide services to rest of the corporation.
- Little coordination



Organisation of Data Science Team

[3] Consulting

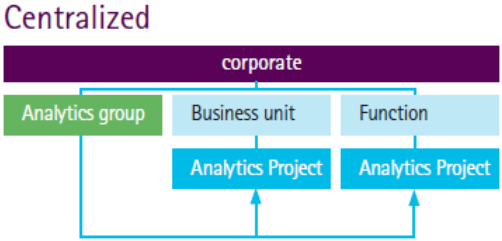
- Resources allocated based on availability on a first-come first-served basis without necessarily aligning to enterprise objectives
- Analysts work together in a central group but act as internal consultants who charge "clients" (business units) for their services
- No centralized coordination



Organisation of Data Science Team

[4] Centralized

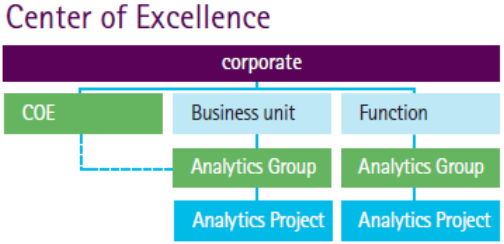
- Data scientists are members of a core group, reporting to a head of data science or analytics.
- Stronger ownership and management of resource allocation and project prioritization within a central pool.
- Analysts reside in central group, where they serve a variety of functions and business units and work on diverse projects.
- Coordination by central analytic unit
- Challenge – Hard to assess and meet demands for incoming data science projects. (esp in smaller teams)



Organisation of Data Science Team

[5] Center of Excellence

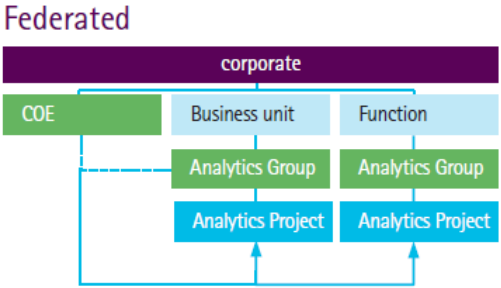
- Better alignment of analytics initiatives and resource allocation to enterprise priorities without operational involvement.
- Analysts are allocated to units throughout the organization and their activities are coordinated by a central entity.
- Flexible model with right balance of centralized and distributed coordination.



Organisation of Data Science Team

[6] Federated

- Same as “Center of Excellence” model with need-based operational involvement to provide SME support.
- A centralized group of advanced analysts is strategically deployed to enterprise-wide initiatives.
- Flexible model with right balance of centralized and distributed coordination.



- Building an Analytics-Driven Organization, Accenture
- https://www.altexsoft.com/blog/datascience/how-to-structure-data-science-team-key-models-and-roles/
- https://www.cio.com/article/3217026/what-is-a-data-scientist-a-key-data-analytics-role-and-a-lucrative-html

THANK YOU