Designing Data Science

Rohan Raghavan
Manager, EYC3

Gavin Seewooruttun
Head of Advanced Analytics, EYC3

EY
Building a better working world

C3

Designing Data Science

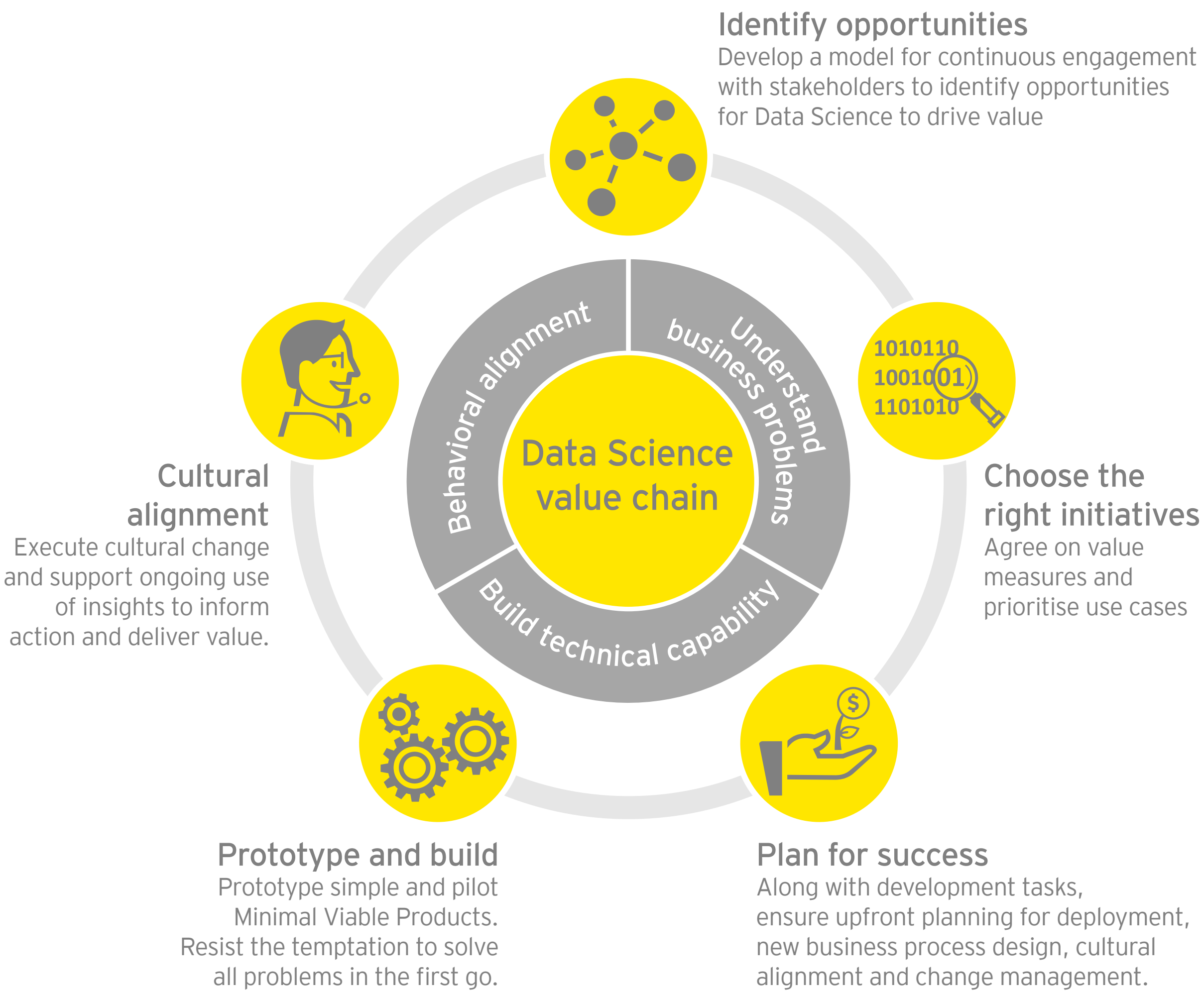
Setting up a successful Data Science capability is no small feat, and even harder, is to ensure continued business value generation from it. Technical capability is critical to success, however, embedding the insights generated into decision-making and instigating a cultural shift to a data-driven organization is the ultimate game-changer. Real value is felt when humans do something different with the insights generated - changing a decision or a process. From insight, to action, to value; Data Science must address all three to deliver true competitive advantage. As Asia-Pacific's largest advanced analytics advisory member firm, EYC3 has engaged with numerous organizations, across industries, assisting them on their journey of Data Science enablement. Here we showcase five critical questions to ask when designing a Data Science solution that will suit your organization.

Our key insight is that to be successful, organizations need to answer five key questions.

These questions revolve around mutually exclusive approaches to Data Science capability:

1. A specialized operations team or one that is integrated?
2. A centralized capability or specialists embedded in business units?
3. Is your analytics information or transformational?
4. A top-down or a bottom-up approach to opportunity identification?
5. A sequential or iterative approach to building capability?

The answers to these questions fundamentally change the approach at each stage of the Data Science value chain.



In this paper, we provide guidance on how to answer these fundamental questions. Our intention is to help the reader ensure value realization from their Data Science capability.

Choosing a Data Science support model: specialized or integrated?

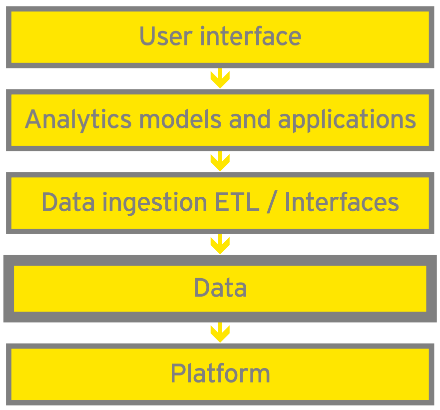
The success of Data Science requires that the applications and decision-support tools delivered be operational. Putting in place a support model to ensure that these are available and perform as needed, in particular if they support mission-critical business processes, is essential. Since Data Science solutions are realized through technology, a key question is to what extent does the organization's IT function partake in their support?

Specialized support model

In a specialized model, Data Science supports every element of the solution stack; from the front-end interfaces to the underlying technology infrastructure, and across the lifecycle of a request. This requires the Data Science team to acquire technology infrastructure and service management skills.

The merits of this approach include:

- Provides a single point of contact for users
- Enables accelerated triage of issues across the solution stack
- Assigns end-to-end accountability to a single function

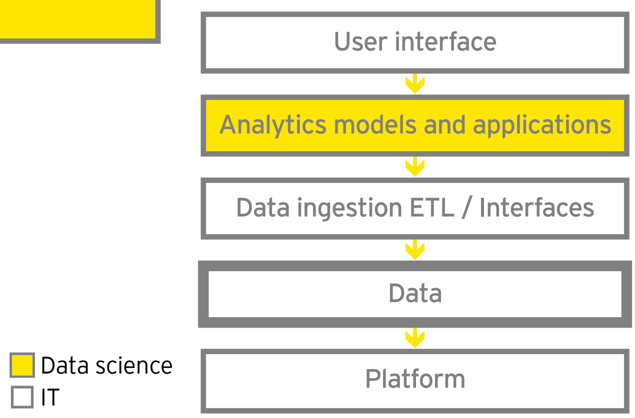


Integrated approach

In an integrated approach, the Data Science team maintains the analytics models and algorithms, while IT provides the first point of contact for requests (typically the service desk) and manages the underlying technology platform. This approach requires more detailed areas to be resolved, for example, who manages Business Intelligence reports and to what extent can Data Scientists provision IT infrastructure to provide them with flexibility if they need additional computing resources.

The advantages of this approach include:

- Leverages existing investments in resources, technology and tools that might already exist within IT
- Leverages economies of scale, reducing the overall cost of service for the organization
- Enables the allocation of highly-skilled Data Science resources to innovation
- Accelerates the resolution of issues where they may be caused due to the data interfaces between Data Science solutions and other IT systems



Questions to guide this decision

- Would Data Science be in a position to keep up with the volume of solutions likely to be generated?
- Are the solutions being provided mission critical?
- Is IT running a cost reduction program that Data Science can leverage?
- Does IT have the necessary skills in Big Data tools and technologies to be able to support them?

Choosing an organizational model: centralized or embedded?

Critical to its success is Data Science's ability to effectively engage with business stakeholders and to deliver to their needs. A typical dilemma is whether to embed capabilities within specific business units or to establish a centralized team that services needs across these business units.

Centralized team model

In this model a central function is established, with all members of the Data Science team reporting to a Chief Analytics Officer or Chief Data Scientist (CAO/CDS). The CAO/CDS ensures an organization-wide view is adopted in prioritizing the initiatives to be pursued. This structure requires a process to promote and prioritize analytics initiatives across the business that provides transparency and the right economic incentives for internal customers.

The merits of such a structure include:

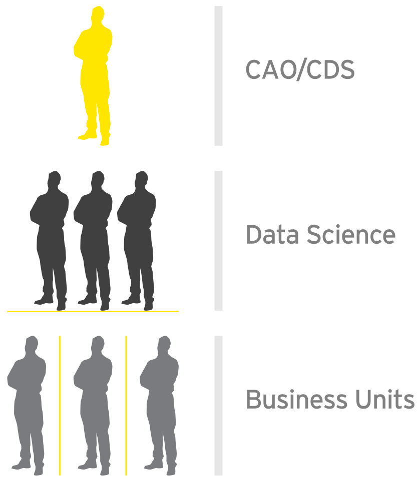
- Cross-pollination of knowledge and assets
- Enables portfolio-level optimization of resource allocation
- Focused development of capability with a degree of separation from operational pressures

Embedded team model

Under such a structure, Data Science capabilities are embedded within business units. Data Science specialists work directly for senior managers in the business. This model needs mechanisms to share best-practice and innovation across the business and for this reason may co-exist with a lean central function.

The advantages of this model include:

- Stronger integration between business units and analytics specialists reduces friction in adoption of Data Science solutions
- Improved understanding of operating context and business data improves the performance (e.g. accuracy) of Data Science solutions
- The measurement of performance (e.g. via experimental design) is easier to focus on meaningful business unit KPIs



Questions to guide this decision

- How well developed is the vision for Data Science in the organization?
- What is the level of overlap between analytics demands from business units?
- What is the appetite for Data Science across the business?

Is your analytics information or transformational?

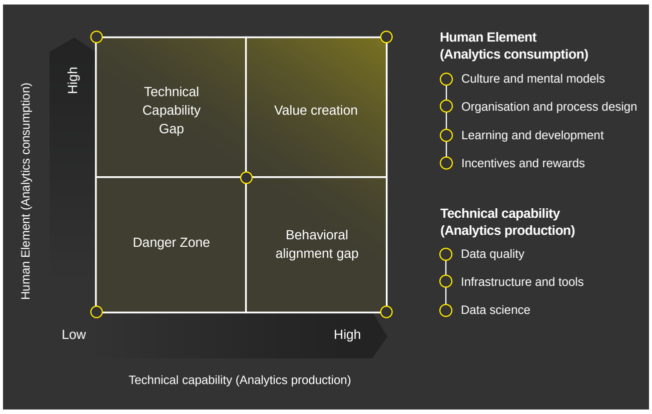
Data Science is a technical capability that generates data-driven insights. For it to deliver value and improve business outcomes however, the insight generated must be converted into action by humans. This human element is what shapes success and is critical to Data Science enablement; it will make information transformational. If there is a gap between delivering insights and taking action, the analytics will not affect change. This key question is not about choosing an approach; it's about covering both human elements of Data Science: organizational and individual.

Organizational

- **Strategy:** Analytics is central to business strategy among leading enterprises. A majority (54%) of executives with leading analytics organizations report that analytics is central to their overall business strategy, versus about 1 in 10 of respondents in the remaining lagging or learning enterprises.
- **Leadership and culture:** Excellence in big data and analytics is heavily influenced by leadership. Leading enterprises designate leaders to guide their initiatives.
- **Organization and processes:** Leading enterprises have aligned their organizations around analytics.

Individual

- **Decision bias:** Analytics should be the basis for decision-making. By understanding, valuing and using the insights provided by analytics, individual decision biases can be reduced or eliminated.
- **Capabilities:** A hybrid of skill-sets is required for a successful Data Science enablement – maths whizzes, business, techies. These are not necessarily found in one person. What is not receiving enough attention is the business user who has to take action on the basis of Data Science insights. Training, easy-to-learn and easy-to-use tools need to be prioritized. Success will be dependent on creating an analytics-driven mindset and enabling business people to become better analytics consumers.
- **Incentives:** Incentives, rewards and measurement need to be aligned with the actions suggested from the analytics-based insights. Of the leading organizations surveyed by Forbes, 3 in 10 say they offer employees time away from "regular" job duties to develop or follow up on new insights they have identified.



A recent EY-Forbes Insights survey of 564 executives in large global enterprises found most still do not have an effective analytics strategy and continue to struggle with change management issues in delivering analytics-driven results (read full report here).

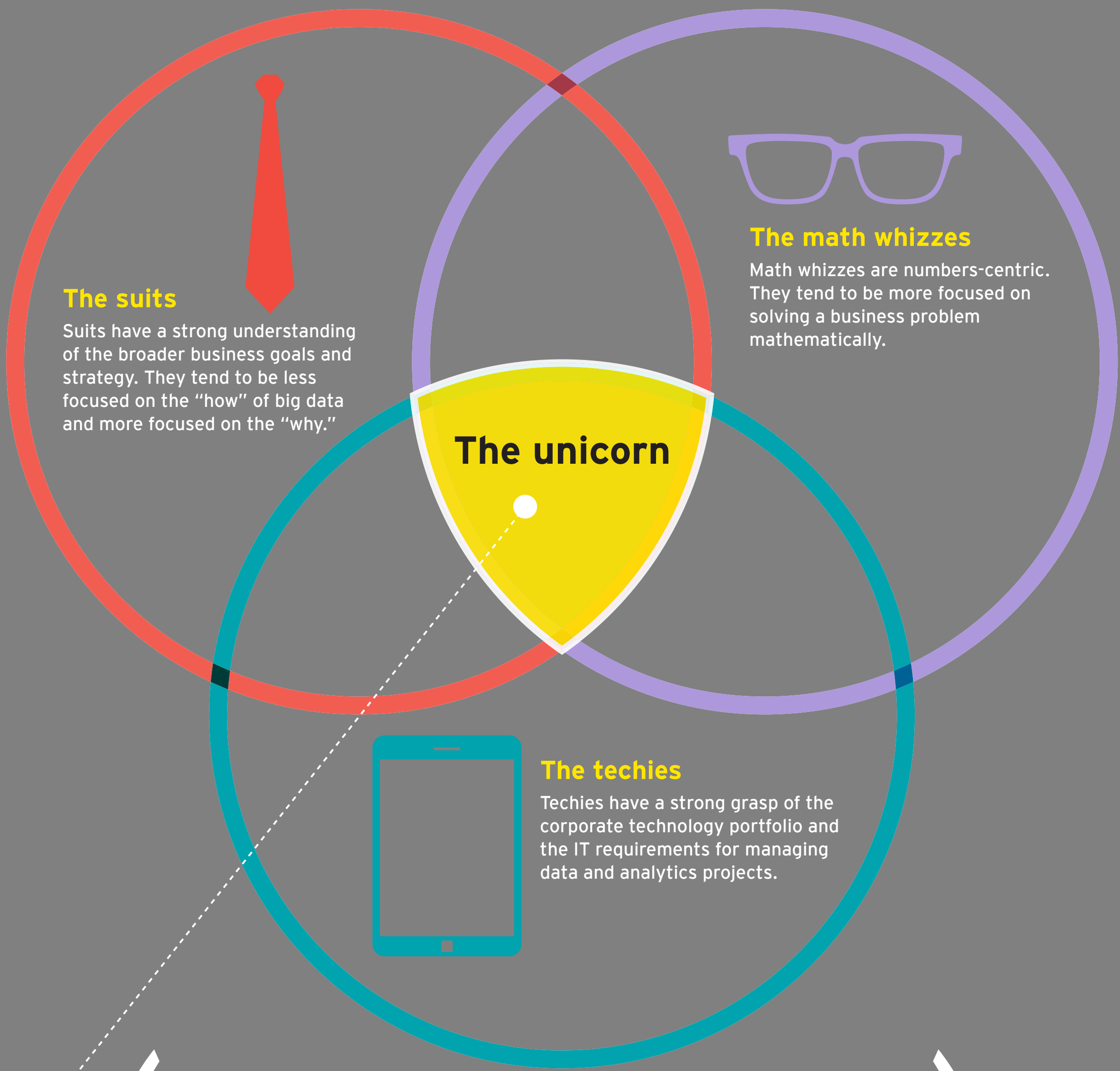
Questions to guide human alignment

- Is analytics central to your business strategy?
- Do you have someone in your C-suite driving change?
- Is your team a mix of techies, business people and maths whizzes? Or are you upskilling?
- Are you offering incentives to motivate your people to action Data Science driven insights?

Math whizzes, techies and suits

Structuring the organization for big data success

To get the most out of their analytics and big data investments, organizations need to think carefully about their mix of skills and talents — bringing together information technology, business and analytics experience.



The unicorn

It is rare to find a unicorn — someone who possesses all three of these big data skill sets. In assessing their talent needs, companies should instead strive to:

- Identify individuals who have expertise in at least two of these core focus areas
- Align big data teams to be cross-functional, bringing together stakeholders with responsibilities and experience across IT, analytics and the relevant business functions

Approach to opportunity identification: top-down or bottom-up?

Of course the investment in Data Science should be focused on high-value and low-effort opportunities first. It will deliver true value when the insights generated by the technical capability are aligned to the human element. People must change their behavior in some way; actioning the insight to improve the outcome. Should these priorities be driven from the top down or is it better to enable operational staff to decide what comes first?

A top-down approach

With this approach, the search for Data Science opportunities starts with business strategy, for example, to become a customer centric organization. This reduces the opportunity set to use cases that support the overarching business objective. Opportunities are identified from multiple sources, including other industries, and they may then be prioritized based on criteria like business value and exploitability.

The advantages of using a top-down approach include:

- Facilitates a joint vision of an analytics-enabled organization
- Enforces a strategic view of priorities to guide resource allocation
- Supports the definition of an analytics roadmap that leverages capabilities across use cases

A bottom-up approach

This approach is driven by teams who are involved in day-to-day operations, who continuously push up ideas and requests for resources up the management hierarchy. It requires the budgeting of analytics initiatives at business unit level and for the teams to be aware of available data assets. This approach involves the exploration and interrogation of data to identify a set of use cases that may be tackled using the available data. Typically a business unit would then fund a proof-of-concept and, subject to a successful outcome, develop it into an operational solution. The opportunities identified may be limited to the best practices that operational teams have exposure to, missing out on ideas or solutions developed in other business units or industries.

The merits of such an approach include:

- Smoother operationalization of Data Science solutions
- Focus on solving specific problems
- Promotes awareness at the operational level of the potential of Data Science

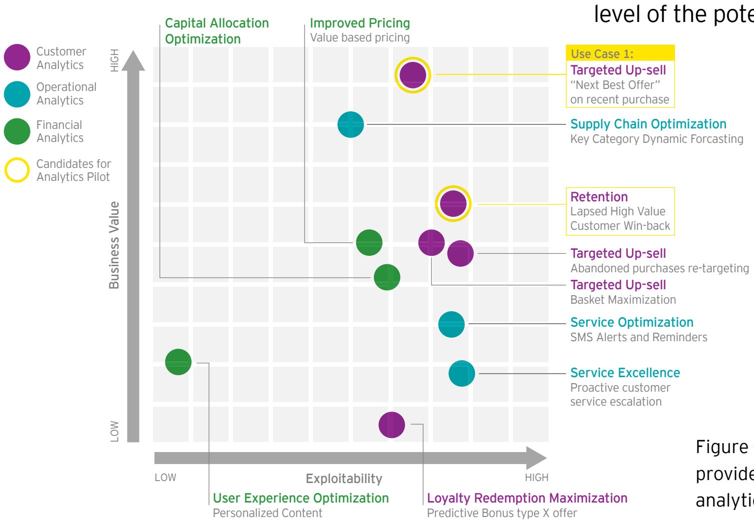


Figure 0.1: EYC3's use case matrix provides a clear path forward for your analytics projects

Questions to guide this decision

- Is the roadmap for Data Science agreed across the organization?
- What is the level of resistance to adoption of Data Science solutions?
- How aware are operational teams of the opportunity to leverage data for continuous improvement?

Choosing an approach to builds: sequential or iterative?

The nature of the work undertaken by Data Science teams can vary greatly; from foundational projects to establish underlying technology platforms to ad-hoc modelling and insights discovery activities. As a result, the approach to execution needs to be tailored to the nature of the activity and also to the business environment in which Data Science operates. A key debate in this regard is whether to use a sequential, sometimes called "waterfall", approach to delivering Data Science solutions or adopt an iterative and incremental approach (there are several iterative methods such as Agile, DevOps, etc.).

The sequential approach

The sequential approach is a structured project delivery framework split into phases for requirements gathering, development, implementation, testing and maintenance. It relies on detailed requirements gathering and planning upfront. Following the requirements gathering phase, the need for business involvement diminishes until the time when the resulting solution is ready for acceptance and deployment.

The strengths of this approach include:

- Forces a thorough understanding of the user requirements, costs and timelines upfront
- Suited to foundational projects in which activities are executed once, for example, the implementation of technology platforms
- May provide greater certainty where the timescales of Data Science initiatives need to align with those of other programs, subject to appropriate contingency being allocated

The iterative approach

Iterative and incremental approaches are typically used to execute on ideas where the objective is clear but the enabling solutions may not be well defined. Projects are divided into cycles or sprints, each involving similar activities but incorporating additional features into solutions. The objective at the end of each sprint is to have a working product that can be transitioned into operations. In practice a number of sprints may be bundled into a release that then can be deployed to the business. Frequent interaction between the Data Science and business teams is mandatory to make this approach work.

The advantages of using an iterative approach include:

- Enables rapid evaluation of ideas
- Promotes accelerated benefits realization through the operationalization of Minimal Viable Products
- Maintains team motivation and stakeholder buy-in by providing frequent success

Questions to guide this decision

- Are you enhancing or establishing new capabilities?
- How much time do key business stakeholders have to work on the project?
- How stable are the requirements and timelines for the project?

Much of the literature on Data Science focuses on tactics and technical capability. We focus on strategy and aligning the technically generated insight with the humans who must take action to deliver value. These five key questions will help you set Data Science up for success in your organization:

- specialized or integrated
- centralized or embedded
- information or transformational
- top-down or bottom up
- sequential or iterative.

There is no universal answer to any of these questions; and successful Data Science organizations will often answer each differently as they evolve their data-driven decision making.

If you would like to learn more please contact us at: analytics@eyc3.com

EY | Assurance | Tax | Transactions | Advisory

About EY
EY is a global leader in assurance, tax, transaction and advisory services. The insights and quality services we deliver help build trust and confidence in the capital markets and in economies the world over. We develop outstanding leaders who team to deliver on our promises to all of our stakeholders. In so doing, we play a critical role in building a better working world for our people, for our clients and for our communities.

EY refers to the global organization, and may refer to one or more, of the member firms of Ernst & Young Global Limited, each of which is a separate legal entity. Ernst & Young Global Limited, a UK company limited by guarantee, does not provide services to clients. For more information about our organization, please visit ey.com.

APAC No. OC00000472
M1629486
ED None

eyc3.com
ey.com/analytics

Contact details:
analytics@eyc3.com

EYC3 helps client organizations become more 'information intelligent' by embedding analytics and enterprise intelligence into organizations from the executive to the front line.