Designing Data Science

Rohan Raghavan
Manager, EYC3

Gavin Seewooruttun
Head of Advanced Analytics, EYC3
As Asia-Pacific’s largest advanced analytics advisory member firm, EYC3 has engaged with numerous organisations, across industries, assisting them on their journey of Data Science enablement. Setting up a successful Data Science capability is no small feat, and even harder, is to ensure continued business value generation from it. Through our work in industry, we have observed a wide range of Data Science initiatives, varying in size, structure and industry.

Our key insight is that to be successful, organisations need to answer four key questions. These questions revolve around mutually exclusive approaches to Data Science capability:

1. A specialised operations team or one that is integrated?
2. A centralised capability or specialists embedded in business units?
3. A top-down or a bottom-up approach to opportunity identification?
4. 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 realisation from their Data Science capability.
Specialised support model
In a specialised 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 organisation
- 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?
Centralised 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 organisation-wide view is adopted in prioritising the initiatives to be pursued. This structure requires a process to promote and prioritise 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 optimisation 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 of 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 organisation?
- What is the level of overlap between analytics demands from business units?
- What is the appetite for Data Science across the business?
**A top-down approach**

With this approach, the search for Data Science opportunities starts with business strategy, for example, to become a customer-centric organisation. 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 prioritised 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 organisation
- 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 operationalisation of Data Science solutions
- Focus on solving specific problems
- Promotes awareness at the operational level of the potential of Data Science

**Questions to guide this decision**

- Is the roadmap for Data Science agreed across the organisation?
- 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 realisation through the operationalisation 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: Data Science skills; Big Data technologies; and machine learning algorithms or models. In here we have focused on strategy: what are the big questions that Data Science organisations can ask to set them up for success? We have suggested four key dilemmas that these organisations must resolve:

- Specialised or integrated
- Centralised or embedded
- Top-down or bottom up
- Sequential or iterative

There is no universal answer to these questions, and the most successful Data Science organisations answer these questions differently as they move up the Data Science maturity curve.

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

© 2015 Ernst & Young, Australia.
All Rights Reserved.

APAC No. AUNZ00000569
M1528664
ED None

This communication provides general information which is current at the time of production. The information contained in this communication does not constitute advice and should not be relied on as such. Professional advice should be sought prior to any action being taken in reliance on any of the information. Ernst & Young disclaims all responsibility and liability (including, without limitation, for any direct or indirect or consequential costs, loss or damage or loss of profits) arising from anything done or omitted to be done by any party in reliance, whether wholly or partially, on any of the information. Any party that relies on the information does so at its own risk. Liability limited by a scheme approved under Professional Standards Legislation.

eyc3.com
ey.com/analytics

Contact details:
analytics@eyc3.com

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