bjss

HOW TO MAKE DATA WORK FOR YOUR ENTERPRISE: FIVE KEYS TO SUCCESS

BJSS Whitepaper

CONTENTS

OVERVIEW

Organisations know there's great value in data, but unlocking that value is often a challenge. In this white paper, we present five keys to delivering successful results with data, based on BJSS' experience of delivering successful data programmes in the private and public sectors. Adopting these practices will help any organisation exploit data faster and with greater confidence.

INTRODUCTION

Organisations are rightly focused on collecting, curating, analysing and – most importantly – acting on data. But data is really hard to do well. Making the most of data spans a wide range of disciplines, technologies and processes. Today's leaders of data teams are trying to build a modern data platform, manage data as an asset and deliver value through analytics, visualisation and democratisation. These are all valid aims which build a compelling vision of the data-driven organisation. But they are also really difficult to achieve – especially in combination.

They're big tasks. They take a lot of time. If you get them wrong, you can be set back months or years. We have seen failed big data platforms go by the wayside, sometimes minimised in their usage and sometimes written off completely – and often superseded by the next great hope, tempered by learnings from the last time around. In the worst cases, these bruising experiences erode confidence and lead to reduced investment in data.

In 2016, Gartner estimated that 60% of big data projects fail. A year later analyst Nick Heudecker revised that failure rate to 85%[i]. We believe the reasons for these failures include the complexity of the technologies used, the habit of defining over-large projects with big bang delivery dates, lack of co-creation with the business, and separation between the user, analytics and IT communities.

The peak of the dysfunction came with the invention of the data lake. While data lakes offer incredibly powerful capability, teams got wrapped up in the technology. They forgot to put business value and simplicity first. Hadoop, the technology that founded the data lake, is highly complex; hard to build and operate, with many options and nuances. Add a slew of new products and capabilities required to manage the data, the need to model data differently from how you used to, and a less than perfect data warehousing capability – with hindsight, it's not a recipe for success.

The profession has matured in the last two years. There's been a lot of learning and some key technology changes. In this white paper, we present five keys to delivering successful results with data within your business. These insights are based on what we have learned from delivering data programmes for a wide range of organisations in the private and public sectors.

Our insights are organised into five themes: delivering value, thinking holistically, steel threading, serverless technology and DataOps. As DataOps is a new, emerging discipline, we expand on this fifth key to success in some detail. We also look at how multi-disciplinary teams can be created to accelerate your success and make it sustainable.

Applying any of these recommendations will help an organisation achieve its data goals. Putting them all into practice together is a game-changer, bringing repeatable, efficient delivery of data with reduced risk on a solid technology foundation designed for users and the business.

THE KEYS TO SUCCESS

Delivering Value

Thinking Holistically

Steel Threading

Serverless Technology

DataOps

The Keys to Success
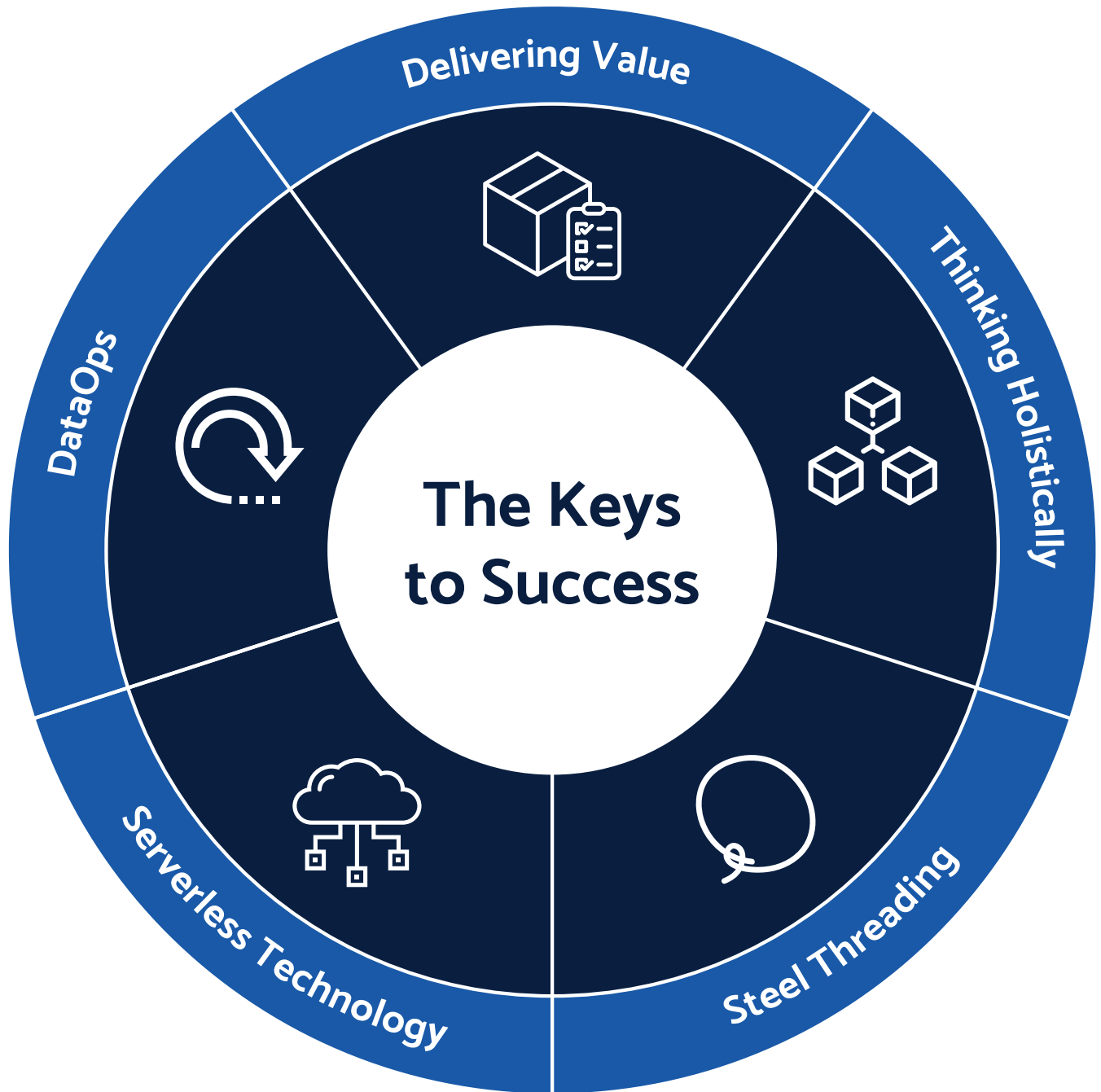
One: Focus on delivering value throughout

Organisations invest time and attention in data because they want to generate value from it. The great potential benefit to the organisation often prompts teams to embark on large, slow initiatives aimed at producing perfect results. It's not unusual to see data platform deliveries go over a year without ever being made available to a user. This is a dangerous as well as unnecessary approach. You won't know how users will react to what you are building or learn from how they use it. You also restrict your ability to generate immediate value in the business and force users to find alternative routes to their goals.

This is not to say that teams should abandon their professional standards and resort to quick and dirty solutions. Faster and more frequent delivery can be balanced with data management rigour and effective, user-relevant data modelling. Access to data can be provided with incremental management, similar to the incremental development of system features, thereby enabling the onboarding of users as early as possible.

There is often a tendency to delay or even block delivery of data science use cases until all the data has been fully managed into a data platform. In some use cases, this is a valid approach; for example, those which are sensitive to privacy concerns or require data sources with particularly poor quality.

But there is often a wide range of use cases that can progress successfully and return great value and learning to your business without waiting for the completion of the platform.

We believe you can and should support the undertaking of analytical projects outside of an in-development platform or managed data initiative where timelines and backlogs will not effectively align. This is especially important in the early stages of your data journey. The idea that you need to spend many months getting data into a fit state before you can reap the rewards of data science is a fallacy.

Focusing on delivering value in this way leads to a parallel approach to data management, data platform and data exploitation. Our experiences with this approach lead to a further recommendation: we suggest that enterprise-wide data science platforms are only put into place once several data science projects have been successfully implemented into production. This allows time to learn how the data science team works best, what tools and technologies they see success with, and how you can best support them with platform components that improve their effectiveness.

We call this approach 'Nail it then scale it'. We recently used it at a major UK retail bank to deliver machine learning solutions into production and support the definition and delivery of a bank-wide platform for data and analytics.
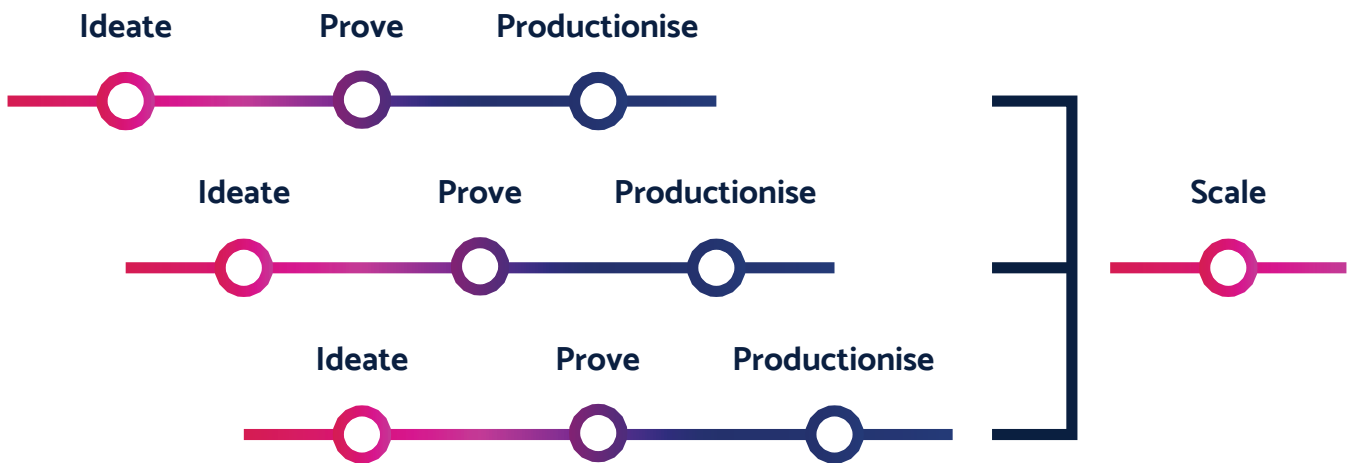
Two: Think holistically: data as a business service

A successful data programme considers data holistically. The scope encompasses considering a platform for managing the data, the processes for managing and governing the data, information and cyber security, data valuation and condition, analytics, BI and visualisation. It also includes how you exploit the data to support your business strategy.

Unsuccessful data programmes can fail on any or all of these factors. Widely discussed reasons for failure include complicated technology, unrealistic expectations, lack of communication between IT and business, and skills gaps. We would also add operational complexity to this list.

With this in mind, we believe that it is key to success that data is delivered as a service to the business. It's not simply an IT project – how many data programmes are approached today. Rather it's the combination of all of these things wrapped in a set of services that are offered to those that use or interact with data within a business.

We call this data as a business service and have been introducing service design into our data deliveries, helping organisations implement data effectively.

When you look at data through the lens of providing a service to the business, you can start to ask:

· What questions are people in the business trying to answer? And why?

· What kind, quality, volume and frequency of data might people need to help them answer these questions?

· How will people turn data into information – that is, what context do they need around the data, and how will this be provided and understood?

· How will people turn this information into action? How will it improve outcomes and impact?

The key word in each of the above questions is "people". If you're in the fortunate position of starting a data project from scratch, don't start with the technology, start with – and speak with – people. Ask questions, frame the problems you're trying to solve and capture user needs before exploring how to solve them.

Assuming a data platform is part of your solution, it's equally important to keep a human-centred, service-oriented lens on how your design, build, launch and operate it. This means you can ask questions such as:

· How will people get data into the platform?

· How will people discover, access and use data in the platform?

· What kind of support will people need?

· How will the existence and value of the platform be communicated to people?

· Who will run and maintain the platform, and how?

A tool we use to help answer these questions is the Service Blueprint. This captures the end-to-end user journeys of people interacting with the platform, expresses their user needs at each significant step, and maps out holistically what should be in place to meet those needs from the perspectives of people, technology, process, policy and partnerships.

The great thing about a Service Blueprint is that when used as part of an agile delivery, it can act as a living document and focal point for keeping a human-centred perspective on what the team is building. It can evolve and flex as you continue to learn from users and the business, and help you respond to pivots along the way by highlighting gaps or new areas that might need to be addressed to ensure successful delivery.

Finally, when building a data platform, it can be tempting to try and boil the ocean, especially if you've done your user research and have identified lots of user needs to go after. As highlighted, we recommend a more measured approach. Aim to deliver value throughout, prioritise your backlog and consider your business objectives as well as the needs of your users. A helpful tool for this is a strategic roadmap. This allows you to plot out themed user needs, business goals and alignment with other initiatives so that you can plan your effort – and budgets – accordingly.
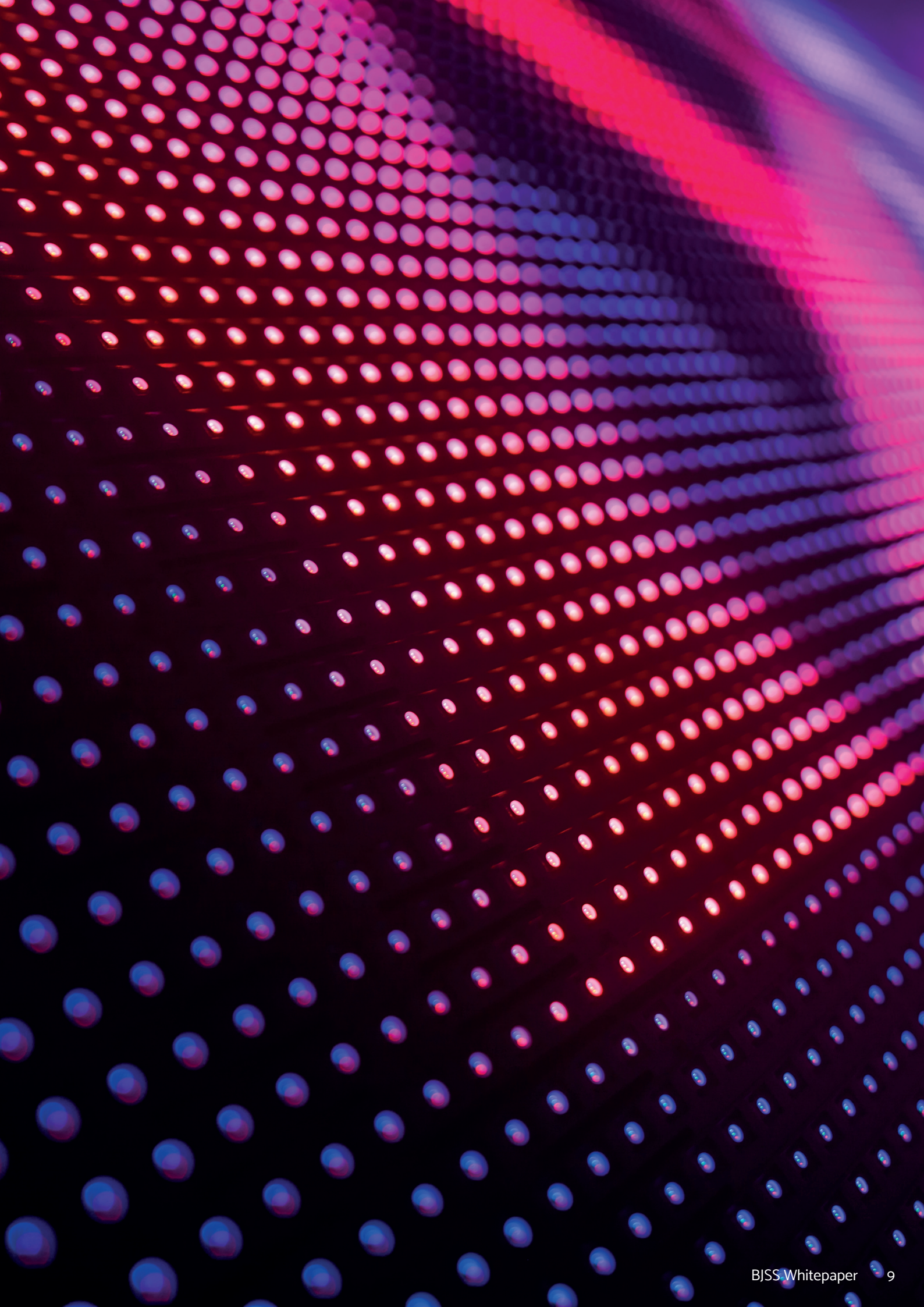
Three: Steel thread your delivery

Many enterprise data programmes take a long time to start delivering value to the business, with a linear dependency between data platform delivery and exploitation of data for business value. This echoes the traditional, waterfall approach to systems delivery which is being rapidly superseded by agile approaches. However, just as functionality can be made redundant by changes in the environment, so data value is often dependent on timing. If value from data isn't delivered to the business when it can be exploited, opportunities are lost, and the vision for data is tarnished. Interest in continued funding can quickly wane when the return on investment is not forthcoming.

We have seen great success in delivering data programmes – including platform, data management, service design and data exploitation – through the adoption of steel threading. The steel thread is a software engineering concept that's popular in agile development approaches. A system's main functionality is like a thread that runs through the system. Since all the other functionality is based on this thread, it needs to be very robust and reliable – as strong as steel.

In the data programme context, we can define a steel thread of functionality, data, service and value running end-to-end, and deliver this into production first. Using this approach, you build some of the core aspects of the technology platform, onboard key initial data sources, deliver a use case which generates business value and thereby create a platform upon which to deliver further use cases iteratively, each time building out parts of another thread.

For example, BJSS and Vanquis Bank built the core capabilities of a cloud-native data platform as well as key data sources and a use case that would both cut costs and present a foundation for further data analytics. We also implemented a DataOps culture, continuous delivery, repeatable ingest patterns and data security features to ensure data privacy. Using this steel thread approach led to the delivery of the first production use case and ROI within four months as well as a core platform and approach that was easily built upon iteratively.

Agile principles are often not applied to the delivery of data platforms, aside from certain aspects of the agile operating rhythm. We advocate taking more learnings from agile and the broader field of software delivery into data delivery – our fifth key to success, DataOps, goes into further detail on this.

Platform

Data Management

Steel Thread

Service Design

Data Exploitation
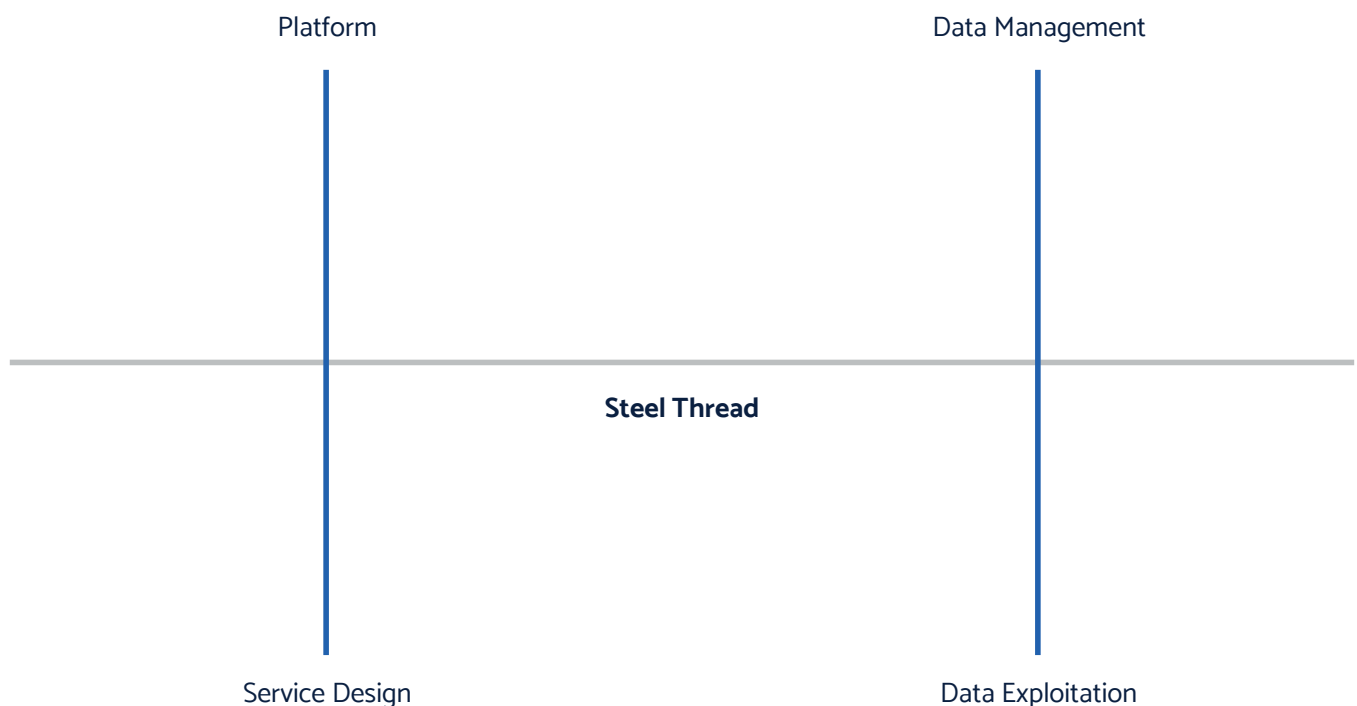
Four: Use serverless data technologies

Technology in the big data space has been genuinely difficult for years. The invention of Hadoop was a turning point for data in the enterprise. Still, it was also a huge peak in complexity – one that many significantly underestimated.

Big data technologies can be hard to integrate, understand, deliver and operate. Implementing technology in this space has required specialist engineering skills and the skills gap has been more like a chasm. This has led to many organisations asking technical staff to learn as they go, adding uncertainty to the delivery timeline and risk to the initiative.

Ideally, organisations want to learn on small, low-risk projects. But in this case, much of the learning has been undertaken on expensive programmes and platforms that are potentially critical to business success. Clearly, this represents a considerable risk and has had a large part to play in the failure of enterprise data strategy.

Recently, however, there has been a significant reduction in complexity through the use of serverless data services offered by the major public cloud providers. These services are fully managed by the cloud provider, with resources dynamically managed and allocated, allowing the developer to focus on business logic and functionality over servers and infrastructure. This can make the technology part of the data equation significantly more straightforward.

The benefits of reduced complexity in data platforms cannot be stressed enough. While there is enormous potential for value in an enterprise data platform, it is no use if you can't easily build upon it or operate it in production effectively – a trend we saw with on-premise Hadoop data lakes from vendors such as Cloudera and Hortonworks. While there are well-known tales of hugely successful big data solutions at organisations like Netflix and Uber, most enterprises do not have the engineering capability or investment needed to handle such highly complex software and data solutions. And nor, frankly, do they have requirements that demand such complexity.

There are trade-offs to be made in the use of serverless data capabilities. They often offer a reduced set of features or other trade-offs such as performance reductions in certain circumstances. Some organisations are wary of serverless data services because they don't support potential future requirements, one example being effective monitoring and workflow of data pipelines in AWS's serverless data services.

But for the price, you'll get significantly reduced operational complexity, reduced engineering complexity, faster delivery times, and fewer bugs and operational issues. While decisions around technology and cloud services should be made on a case-by-case basis with all considerations taken into account, we recommend weighting the benefits of serverless data services very heavily. We would go so far as to recommend a default assumption against starting a data programme today without making it fully cloud-native.
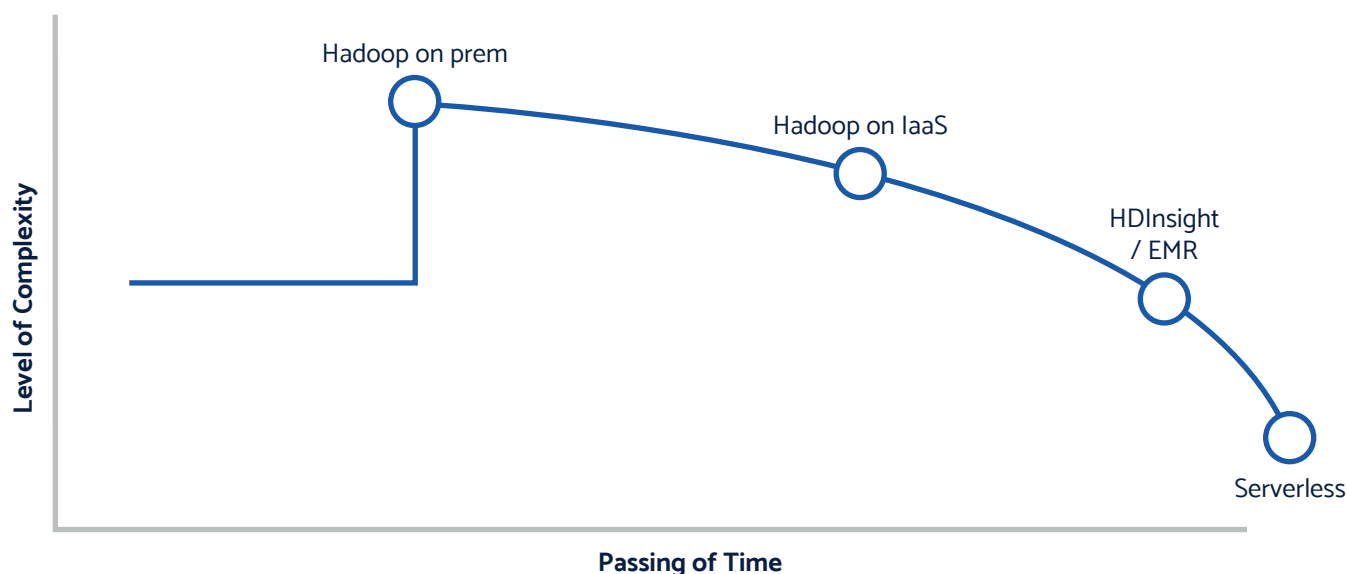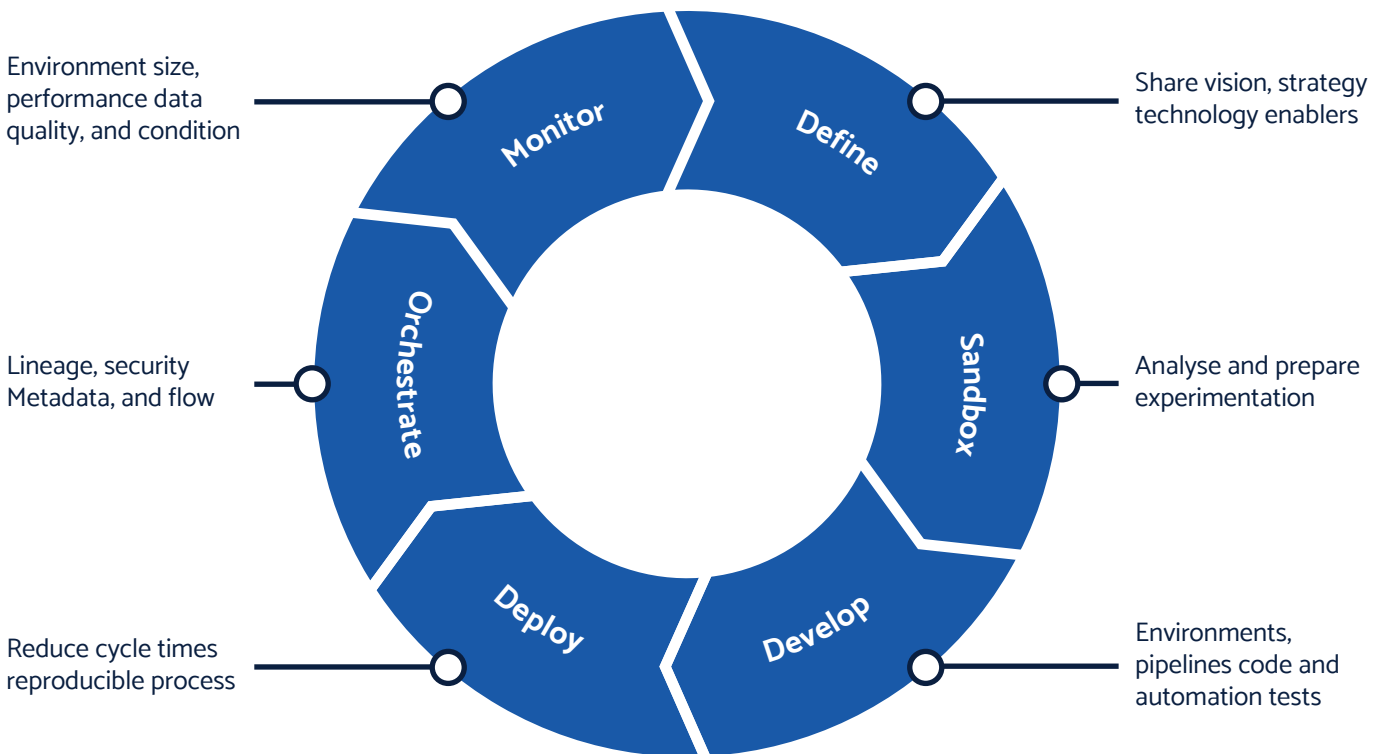
Five: Apply engineering rigour with DataOps

Considering that data initiatives are generally large, strategically important programmes, it's surprising how little engineering rigour tends to be applied to the technical delivery aspects of data. This is especially puzzling for those of us who come from a software engineering background.

Delivery of effective data programmes is more akin to software development than it has ever been, and there is much that can be learned from how we deliver software. Continuous integration and delivery, secure development, automated testing, monitoring, alerting and infrastructure as code are all applicable here.

Transfer of best practice can benefit not just the development of capability but also the management of data pipelines and delivery of data science. By applying these techniques effectively, you will see far higher levels of quality, repeatability and effectiveness all around.

We call this approach to merging the cultures of data science and software engineering DataOps, on the model of DevOps. DataOps makes the research/experimental cycle compatible with enterprise-strength technology delivery. Users of data get the ease of access, flexibility, control and fast time to market they need. At the same time, the organisation ensures proper governance, security, maintainability and affordability.

Environment size, performance data quality, and condition

Lineage, security Metadata, and flow

Reduce cycle times reproducible process

Monitor

Orchestrate

Deploy

Define

Sandbox

Develop

Share vision, strategy technology enablers

Analyse and prepare experimentation

Environments, pipelines code and automation tests

RECOMMENDATIONS FOR IMPLEMENTING A DATAOPS CULTURE

Automation

Automating data pipeline activity will increase velocity and repeatability, helping to overcome one of the common issues with data platforms: long lead times to onboard new data sources resulting in unavailability of data when users need it. Patternised ingest with full automation, testing and monitoring will significantly simplify the task of ingestion. When implemented successfully, this can reduce data in-gest from days or weeks to hours or minutes.

Automating testing, deployment, and infrastructure will enable you to implement continuous integration and continuous delivery for your data platform. While this is similar to the approach taken in software development, there are some interesting quirks that can make this a challenge in data platforms – for example, automating testing and deployment of code contained in notebook form.

Security and compliance

Data residing in data lakes and warehouses can be of high value to an attacker. We recommend that you implement security as early as possible in your SDLC and ensure your data engineering team are well versed in common exploits, and secure development approaches.

You should design the platform to be secure from the outset. Access issues are more easily overcome while defining the security model than when retrofitted to an existing model. Additionally, this allows your team to run effective, incremental threat modelling sessions, identifying security risk early and often. The outputs from this approach will aid the implementation of appropriate security controls.

Finally, use cloud-native capabilities to ensure regulatory compliance. Develop blueprints, templates and compliance-as-code which can transform one-time InfoSec sign-off into continually monitored, business-as-usual compliance. This provides an early warning system, saving you from potential data breaches

Assurance

Availability of test data is paramount. Without enough high-quality test data, it is challenging to effectively test your pipelines, performance and analytics, increasing the risk of unexpected issues when switching to production workloads.

A testing strategy designed to work in production through isolation of data is an excellent investment. This will enable you to run high-quality end-to-end testing with production data, significantly reducing the risk of unexpected outcomes in live operation.

Peer review paired working, and collaborative sessions, including security and design, should also be used to provide further quality assurance.

Operational monitoring

Like DevOps, DataOps is as much about operating a data platform and data flows as it is developing. Identification of metrics required to operate the platform and KPIs required to operate the business will allow you to set up operational monitoring, logging and alerting early on. Information gained from trend analysis, combined with feedback from users, can form further backlog stories for improvements.

FITTING TO A T: HOW TO HELP YOUR PEOPLE USE THE KEYS TO SUCCESS

The five keys to success which we outline here cover a broad range of concerns from technology choices through to cultural change. Enabling your people to use these keys effectively requires a new way of thinking about skills, collaborative working and complementary roles. Finding – and making – T-shaped people, is a great way to make this happen.

T-shaped people have a core area of expertise, along with secondary capabilities in many other areas. They are specialists who can also generalise. In the agile development world, T-shaped people may be multi-skilled developers with differing major skills.

At BJSS we strongly believe in the benefits of T-shaped people across all activities – including data projects. In fact, we believe that for successful enterprise-scale data solutions in production, it is becoming mandatory to support multi-disciplinary learning and experience. Development of more broadly skilled data professionals will expedite delivery of production-grade solutions while supporting a highly collaborative, educational team environment which positions individuals to succeed, builds a strong team delivery ethic and creates an enjoyable, productive working atmosphere.

Data scientists often come from an academic and research background, which is a boon to the data science experimentation process. But for data science to deliver value in the enterprise, it must go far beyond experimentation into production use and be integrated into other business systems and processes. We have found that by having data scientists working on production solutions closely with those from other technical disciplines, they can pick up further skills and capability that hugely benefit the productionisation of data science solutions. We recommend placing data scientists alongside roles such as software engineering, platform and testing in rounded, multi-disciplinary teams to encourage cross-pollination of skills and approaches.

This approach aligns with our recommendations to deliver a steel thread and focus on value throughout. To ensure success, it is vital to deliver data science outputs in tandem with aspects of the data platform. When goals and objectives are set for a close-knit, multi-disciplinary team, rather than separate functional teams or individuals, we see success rates increase dramatically.

CONCLUSION

The value of data can be unlocked by delivering value, thinking holistically, steel threading, serverless technology and DataOps. Adopting these practices will help any organisation exploit data faster and with greater confidence. These are practical steps which support each other and can be implemented incrementally without impeding the flow of benefit to the business or causing major reorganisations. Organisations which follow this path will find their data capabilities maturing rapidly and contributing clear, recognised advantages to everyone in the business. With the right keys, the door is open to a brighter future.

NOTES

i   **85% of big data projects fail, but your developers can help yours succeed, Matt Asay, TechRepublic, 10 November 2017**

Visit: bjss.com/data/

Contact: info@bjss.com