

Enabling data-driven decisions within your organisation

AMIDO[®]

Foreword

Although data has always been a core part of any project I've been involved with, my big data journey started around ten years ago. I have always enjoyed being part of business transformation, seeing organisations grow through technology, especially the data centric parts. I am attracted to the problems that data at scale present and luckily I ended up working on finance data systems and massive scale grid computing during emergence of big data.

When I first started on my big data journey, working with pioneering on-prem data platforms like Hadoop, IMB grid systems and bespoke data platforms, big data and analytics incurred such high levels of technical understanding to do even the most basic operations, that our analytics options were reasonably limited. While for the first time analytics could span massive data sets, the experimentation cycle was long and costly. It was an exciting time, however the failure rate was high and there were few businesses who could afford to do this kind of work.

Over the last five years during the cloud revolution, especially with the upsurge in virtualisation and containerisation, big data tooling has advanced exponentially. Working closely with Microsoft on projects like Azure Batch, HD Insight and Power BI, I have observed the advent of commoditisation of massive scale data systems, making this kind of project accessible to a much wider range of businesses.

I feel we are in a very interesting period in data and insight right now. Where 5 years ago the investment to incorporate even simple linear machine learning models was massive, in many cases I can help clients spin up thought experiments overnight. While I still feel there are some significant hurdles to overcome in data and analytics, the technical parts are becoming reasonably well solved, so that only leaves the most exciting part:

changing businesses through insight.

Written by:

Ian Cowley

Principal Consultant Data Engineer, Amido

Illustrated by:

Aly McKenzie

Senior Consultant Business Strategist, Amido



Failed data projects are common

In a 2016 survey by Gartner, they estimated 60% of big data projects failed. They followed that up a year later by saying that was probably conservative and it's more like 85%. The data-platform company, Snowflake Computing's CEO, Bob Muglia told the analytics publication Datanami;



I can't find a happy Hadoop customer. It's sort of as simple as that. The number of customers who have actually successfully tamed Hadoop is probably fewer than 20 and it might be fewer than ten.

- Bob Muglia, Snowflake Computing



The goals of data-oriented projects follow a similar trend. Typically, these projects attempt to use the powerful and valuable data companies generate to create predictive, insight driven decisions and products. However, we frequently see the shortcomings of the data platform causing teams to become disillusioned and begin working around the platform, finding adhoc ways of acquiring data, resulting in inconsistent analytics and reporting and duplication of effort.

Working in this adhoc way often leads to business units being unable to scale out their analytics capability and it can become incredibly difficult to integrate analytics into production products and services.

In this paper we highlight some of the common pitfalls for data programmes and the structural and cultural changes you can apply to avoid these problems and accelerate

better business outcomes.



The shortcomings of the data platforms tend to follow similar themes:



Lack of trust in data quality

Out of sync or missing data or lack of consolidation of data.



Poor data literacy

Teams are unable to understand, digest and or contribute to the data platform and structures, often caused by a lack of visibility of both data structures and versioning.



Governance and compliance hurdles

Governance policies and guidance can be difficult to access and understand, which in turn prevents teams including valuable data in their analytics practices. A common problem is visibility of data sensitivity, creating latency onboarding data streams, resulting in business units being unable or unwilling to engage with the data platform.



Difficulty promoting analytics and data pipelines into production

Many business units, development and data teams are now used to including analytics in their day to day or even in advanced scenarios utilising machine learning capabilities. Unfortunately, there is often a lack of process and tooling available to promote features into a production environment, resulting in work being held at the proof of concept stage for long periods of time or in some cases becoming out of date before it is utilised.

Solution summary

At Amido, we work with large organisations across a range of sectors, including retail, social housing and the public sector, to help them use data to drive business outcomes including growth, operational efficiency and innovation. Our successes in supporting them to become data and insight driven have often come by applying a different organisational and data centric approach.



Growth



Efficiency



Innovation

Traditionally organisations rely on a centralised analytics and a data department to provide support in innovation pipelines, products and projects. Having a centralised function often creates an analytics and insight bottleneck, resulting in business units reverting to their familiar decision making and analytic techniques. If they can't get access to and understand the data they need quickly (as is often the case, because they're sat at the bottom of another team's product backlog), then they will find other ways to solve their problems. There are no guarantees those adhoc solutions will be valid or will stay valid as the data landscape evolves and there are often real issues getting those analytics pipelines into production.

We seek to empower business units to understand and incorporate data and insights into their workflow, reducing trial and error costs and incorporating customer and process behaviour into projects and products.

In order for organisations to become truly insight-driven, using data to improve processes and products, data and analytics cannot be seen as a by-product. Instead, data and analytics must become a first-class tenet of the software design, product management and project planning.



In traditional data organisation structures, where data departments are remote, they often find they aren't able to keep pace with the data changes by each business area.

This results in low quality and reliability of data, which in turn creates a lack of trust in the data platform. Being outside of the SDLC for the team, the data function is often overlooked until late in the project, creating further latency and de-incentivising the team to engage with the data team. This is when we see business units and development teams going around the data platform to solve their analytics problems, rather than engaging with it.

What's more, a centralised data team won't have the domain knowledge to apply context and understanding to the data and consequently, often become a bottleneck in the analytics capability. Creating a federated data practice, focused primarily on moderating, monitoring, supporting and educating and less focused on implementation, creates a reduction in the bottleneck, while at the same time stabilising your data platform and enabling your business units.

In order to increase data literacy, organisations should enable business units to incorporate data and analytics into both their decision making, products and projects.

Business units and development teams should be responsible for managing their own data provision and consumption, as well as being given the authority to change their data domain freely and easily. This way the business units and development teams take the problem into their own hands, reducing the latency for data change and encouraging them to incorporate data-centric thought into their processes. Business units and development teams are best placed to apply intelligence to data to find insights to help improve their projects and products and are more likely to be able to keep pace with the constantly moving data terrain.

Another area which benefits from a de-centralised data organisation structure is in experimentation. Experimenting with data can come in various forms, whether its machine learning experiments or creating product changes for targeted demographics and measuring the success, the ability to change quickly and often, and measure the outcome is key. The length of the change and outcome measurement cycle has a direct correlation with your organisation's ability to innovate. By creating distributed data centric teams with high levels of data literacy, you encourage business units to experiment, innovate and iterate quickly, reducing this cycle time and increasing innovation within your organisation.

Finally, for any organisation, there is a steep learning curve to applying new technical, structural paradigms. To reduce risk and embed new ways of working effectively and efficiently, our advice is start small, learn quickly and apply across domain verticals.



The common pitfalls of centrally managed data programmes

1

Lack of understanding and trust in data

60%

identified data quality as a challenge

43%

had difficulty identifying data which delivers value

85%

of data projects failed to justify investment

-Gartner

In a recent Gartner survey over 60% of respondents identified data quality as a challenge to data management practices, with 43% of their surveyed audience claiming that identifying data which delivers value is a significant hurdle. With seemingly 85% of data projects failing or at least not adding value enough to justify the investment, understanding and trusting the data are common themes.

The notion of trust in and understanding of data as key to the success of a data project is something that we've seen time and again at Amido. The most successful projects tend to involve stakeholders who directly depend on data for modelling and forecasting purposes, consequently their success is inextricably linked to the data platform and their engagement is high. For example, in actuarial and finance enterprise data management solutions, the business units are unblocked directly by gaining access to new data streams, as well as being rewarded for contributing models and data to organisation. Conversely where we have seen low utilisation and adoption is where the aim of the program has been to increase democratisation of data in less data literate teams.

Nowhere more do we see the effect of this centralised data team structure than in data engineering. The life of a data engineer isn't always an easy one. They are responsible for maintaining and providing meaningful, trusted data, with little knowledge of the domain they're working in and with competing consumers and providers vying for support. This lack of domain specialist knowledge combined with the competition for resources, inevitably results in change management, versioning and release bottlenecks.

2

Inability to iterate and experiment quickly

In order to make data a part of your strategy, decision making and software, then you need secure frictionless access to data and you need to be able to experiment and iterate quickly.

Machine learning and data analytics are an experimental practice, ideally suited for a truly agile approach, requiring deep knowledge of the data and domain you're working in. You can't expect to know the answers up front or even which analytical tangents your experiments will take you in, so in order to iterate and broaden your insight generation quickly, you need rapid access to the data sets and knowledge.

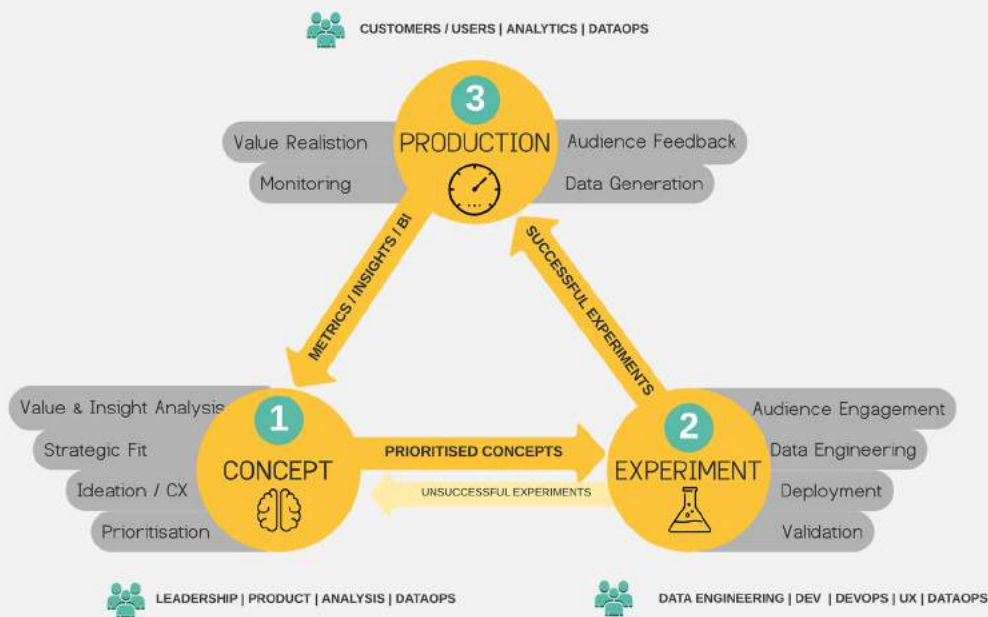
Doing this through a centralised team of non-domain specific data experts clearly lays down the groundwork for obstruction.

In order to make data a part of your strategy, decision making and software, then you need secure frictionless access to data and you need to be able to experiment and iterate quickly.

3

Continuous delivery through DataOps

A challenge that is also linked to team and structural issues is taking analytics into production. In the same Gartner survey, over half of the respondents identified deployment into business process and applications as the main bottleneck. While on the surface DataOps can appear to be like its familial cousin DevOps, it does differ, a lot. DevOps is a technical concern around the systems development lifecycles, however DataOps requires people to work together to derive maximum value from the data.



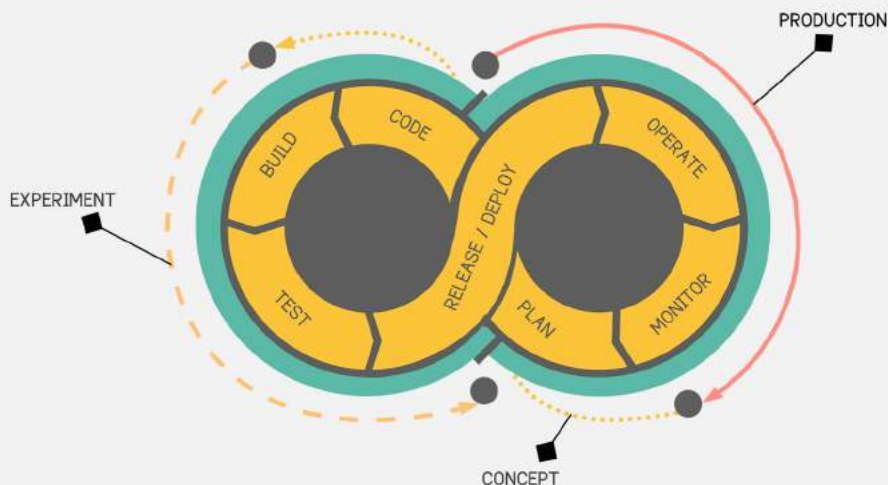
DataOps, especially when embedded into your business units, reduces friction to changes in analytics and data pipelines. DataOps should be able to take the analytics and insights created on the shop floor and promote and integrate them quickly into your production environment, making that data available to the rest of your organisation, reducing the cycle time for innovation and change. Failure to invest in DataOps will slow down the innovation and change cycle, which in turn reduces engagement and increases feedback latency.

Outline of solution

If we go back ten years, most IT departments were separated into sub-departments, with one sub-department being operations. Operations were the gatekeepers for software and releases. Operations had the authority to block or release features and the responsibility to support the platform in the middle of the night when the system went down. On the other hand, development teams and business units had the responsibility for getting features into production, without the authority to do so. These functional silos and the balance of authority and responsibility frequently created friction, reducing co-operation between the teams, creating a blame culture and negatively incentivised fiefdoms.

Then came the advent of feature teams and DevOps. Suddenly feature teams, working together through all of the sprint ceremonies, design and discussions included the work necessary for repeatable, reliable releases. In fact, the very best teams advocate releasing several times a day, reducing the risk payload. Releases require a cursory sign off from a centralised team, with a view to ensure governance and policies have been adhered to.

DATAOPS IN THE CONTEXT OF DEVOPS



This second wave of development team structure enabled business units and development teams to have new levels of agility, able to change priorities, release features and experiment at rates previously unthinkable. Data is experiencing a similar structural arch.

If we go back ten to fifteen years, data fell into two categories.

The first being regular data projects and platforms which contained transaction processing and management information systems. And the second being a Big Data system.



Regular data projects & platforms



Big Data System

Big data was initially a scary capability, requiring a large team of highly specialised developers, capable of distributed reporting via on premise grid computing, focused on CPU harvesting and optimising I/O and CPU utilisation. As big data became data, companies began to see the value in the data within their organisation and how powerful analytics and machine learning could be for them, often creating a central team of highly specialised data experts, the centralised structure that brought us to the problems we're seeing today.

With the huge rise in cloud technologies and with the entry level for big data tooling threshold being much lower, the data capability can follow the QA and DevOps function federating within business unit and development teams.

At Amido, we see the data capability as being a data practice or centre of excellence.

The data practice is not responsible for implementation concerns i.e. maintaining data structures, data quality or reporting inputs, rather they are responsible for data governance, moderation, support and maintenance of the platform and tooling. Development teams/business units become responsible for maintaining and utilising their data contributions and consumption.

The data team also need to build the platform and additional features, focusing on overarching concerns like optimisation, governance tooling, transparency and availability.

An immediate benefit of this approach is that the data units/development teams become experts of their own domain or rather bounded context.

They can freely change, experiment, incorporate and iterate at their own pace without the latency relying on an external team for support, enabling them to make better business decisions, faster, and accelerate innovation in order to stand out against competitors.

By removing the central data team from the responsibilities of implementation, the communication becomes direct between the business experts, which has the benefit of improving data literacy and understanding. By incentivising communication between the different data domains, each department increases their data vocabulary, slowly but surely becoming organisational information experts.

With support from the federated data engineers, the development teams/business units gain the frictionless capability to manage their own data changes and releases without the delay of waiting for other departments to keep in step.

By incorporating the data engineering and analytics capability in your development / business unit teams, especially during the sprint ceremonies, planning and design, it encourages data centric thinking. Much like we have seen with feature squads including QA and DevOps, where planning how a change is tested and released through different environments and to different customers is part of the doctrine of feature design, we are now in a position to apply this to data. An embedded data engineer and or analytics capability within the team, during design and even implementation time keeps a data centric voice nearby, encouraging fact-based decision making. A data engineer is more likely to promote databased decisions and integrating data into features.

There is no such thing as a free lunch.

and much like the waves of software structural approach we have experienced before there may be a learning curve in adopting this approach. You're going to need structures, versioning strategy and a strong governance program in place to effectively communicate and version data change. We have found being able to distribute the responsibility and deliver in an agile cadence, with the responsibility residing in the development/business unit teams, makes the problem a more manageable one. There needs to be a period of encouraging and evangelising the approach and it will inevitably feel to the development teams that they are being slowed down, so there needs to be top down tolerance built into delivery timelines for teams who are onboarding to the data practice.



One common misconception is that DataOps is DevOps for data. DataOps attempts to automate analytics and data, reducing the analytics cycle time.

In data and therefore DataOps the flow, monitoring and orchestration of analytics and data pipelines takes a different skillset and involves different steps in the cycle. The concerns of a big data platform involve a different focus on scale, robustness and a constantly moving terrain. There is a direct correlation between data teams who contain DataOps capability (even if that is with data engineers) and their ability to apply analytics cycles into production.

The final puzzle to solve is funding.

If we continue to view data management as an IT problem, with IT holding the keys to the data kingdom, then there is the opportunity for negative incentives. Our approach has been to suggest that a large percentage of the data funding resides within the development teams/business units.



Any organization that designs a system (defined more broadly here than just information systems) will inevitably produce a design whose structure is a copy of the organization's communication structure

- Melvin Conway



When it comes to data, much like other areas of the software terrain, Conways' Law is critical to the outcome. If you centralise the effort and funding you will end up with a centralised platform, which is abstract, has broad capabilities and will solve some but not all problems for all departments.

Conversely distributing the funding and capabilities gives teams the autonomy and independence to invest directly in their priorities. There is still a need for a central data budget, for platform improvements and data feature upgrades, however most of the budget, especially when it comes to implementation needs to reside within the business units, reflecting the fact that data is a business problem, not a technical one.



Summary

Data, and certainly big data, is a comparatively young technology practice and there is still a lot of work to do. Maintaining a centralised data capability can create a bottleneck for change and innovation. If you want to create data-centric thinking within your organisation, where business units and development teams integrating data and analytics into their processes, projects and products, then confluence of data and business must be core to your thinking. Some of the key features of this approach are:



Create a federated data department

Who are not distracted by implementation concerns and instead concentrate on supporting and enabling your business units will enable your teams to grow in data literacy and increase their insight centric approach to problems.



DataOps is key to success

One of the common hurdles businesses are experiencing is taking analytics into production. DataOps as a practice differs from DevOps and requires a specialised focus.

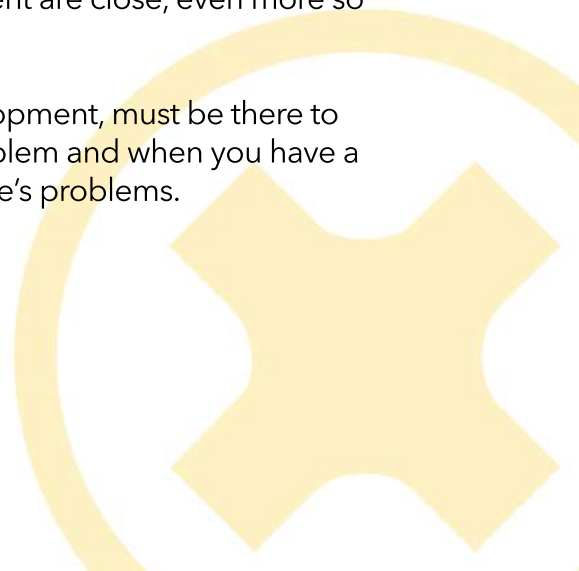


Agile approach

It's important with this, as with any other restructuring to take an agile approach. To scale horizontally and apply your lessons in iterations. Our advice would be start small, iterate and learn quickly.

In agile we see the best results when business and development are close, even more so when they are co-located.

Data, much like the evolution of other areas of software development, must be there to enable the business. Data and especially big data is a big problem and when you have a monolithic data platform, it can be very tricky to solve everyone's problems.



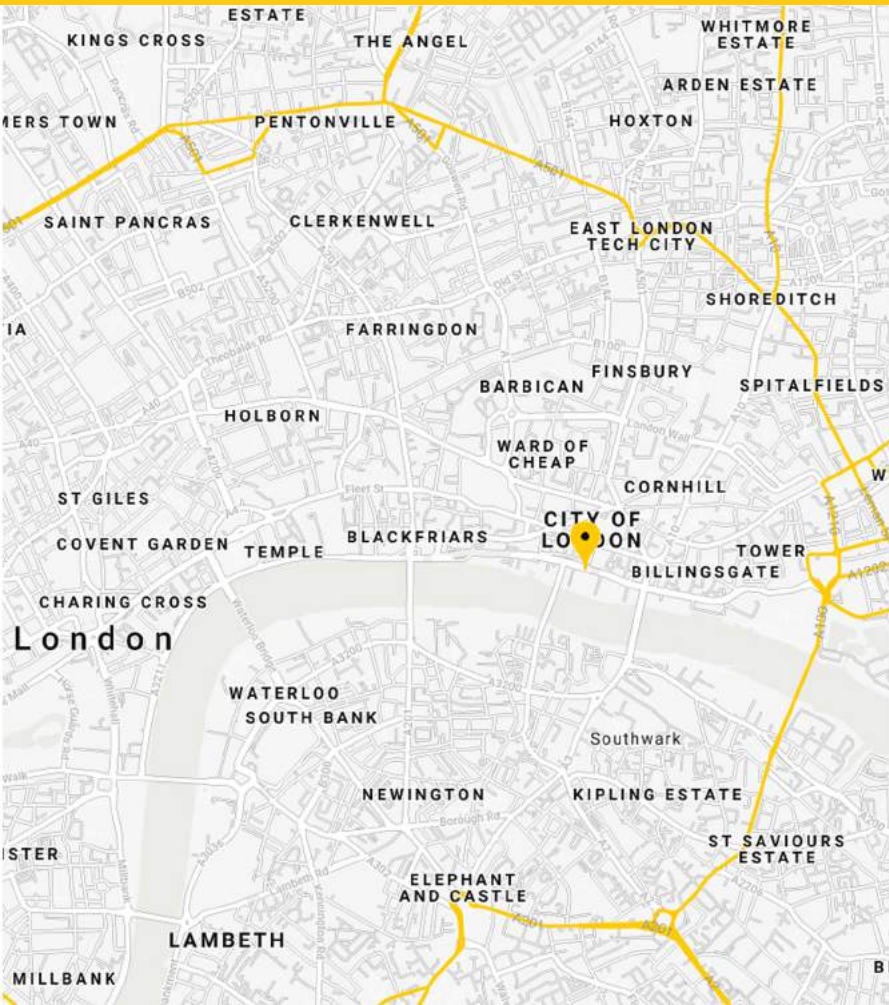
CLLOUD-NATIVE CONSULTING & IMPLEMENTATION

Smarter engineering. Better business.

Amido is the cloud-native consultancy for better business outcomes.

We help our clients think and do, delivering strategy, data platforms and software engineering that gives them the confidence to take on their toughest challenges and improve the lives of their people and customers.

If you need advice on getting more value from your data, [please get in touch.](#)



AMIDO 



enquiries@amido.com



0203 176 4690



amido.com