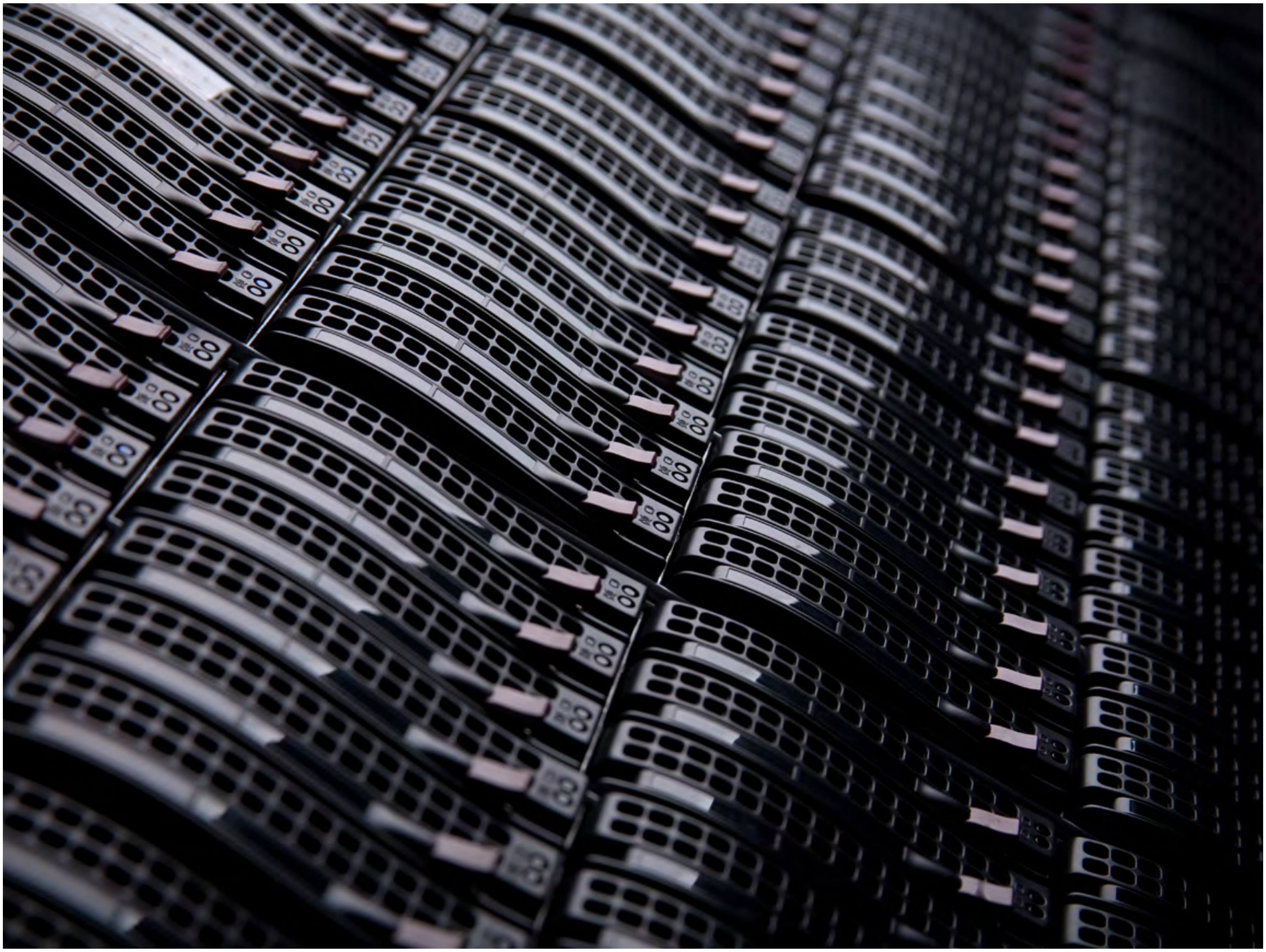

# The Modern Data Strategy: Feeding the Machine

September 2021

IBM Services Data Strategy Whitepaper

IBM

# Contents

## Aligning to the future

Technological advancement is providing better automation capabilities every day, and competitive pressures are driving companies to take advantage of them. In this environment businesses are gradually moving along a journey towards the cognitive enterprise, where intelligent automation pervades operations.[1]

Automation is founded in data, and effective and appropriate data strategies are essential if these journeys are to be optimised. This paper examines the shape of a modern data strategy, aligned to the construction of a cognitive enterprise.

## The Role of Data Strategy

A data strategy sets out how a company should use data to support its business strategy. Historically, there have been three main strands to this:

1. supporting strategic decision making - for example: provision of reports to executive decision-making bodies;

2. supporting day-to-day, or "operational" or "transactional" decision making - for example: flagging up production-line problems to shop-floor operatives, providing information to a call-centre agent about the customer they are speaking to, presenting product information to a customer to persuade and enable them to buy, or satisfying a regulator that the business is operating safely;

3. Monetising and exchanging data - for example: selling data to third parties.

The first two of these are focused on business intelligence and analytics - which have been understood as the provision of information to human consumers, to enable them to make decisions - choices between actions. Sometimes, these choices are nuanced and may not feel like conscious decisions (for example, whether to refer to a customer as "Mx, Mr, Ms or Mrs"; whether to send a letter to one address or another, whether a regulator should investigate an anomalous result) but they are still decisions. If data does not drive a decision somewhere down the line, then it has no value.

Data strategies seek to understand decisions which can be improved by data - prioritising these improvements in line with business strategy; then establishing how to source and process that data most effectively. Many strategies focus on particular aspects of these improvements, such as data governance, or data architecture, rather than a full end-to-end solution, but fundamentally, they are supporting the same goal.

## Traditional Analytics

As we have seen, a large part of data strategy has been about the supply of data to business intelligence and analytics. Over the past two or three decades, these domains have revolutionised the enterprise, and enabled completely new business models, such as that of Uber. Businesses have been able to dramatically improve their decision-making in terms of accuracy and timeliness, making choices about how to allocate resources or to price goods. A good example of the latter is US retailer Stage Stores using predictive models to decide precisely when to apply discounts optimally, rather than simply lowering prices in season-end sales[2]. Making analytics work relies on understanding the questions that need to be answered in order to drive value-adding decisions, and then sourcing the required data that answers those questions and providing it in easily consumable formats.

Traditionally, this has been understood in terms of “Business Intelligence” - building actionable reports for users to consume. Whole industries have grown up around developing reporting toolsets and data sourcing to construct these reports. Localised team or departmental cottage industries dedicated to report building and maintenance are endemic within companies.

A growing awareness that many - perhaps most - business decisions are made without adequate data has resulted in a strong drive to empower staff at all levels to use and build reports to answer their questions.

This movement to democratise access to data has led to the concept of the "citizen data scientist" - Gartner's idea that advanced analytics should be performed by a much wider audience[3]. Supported by the growth of the newer technologies of augmented analytics that have arisen from business intelligence and analytics domains, comes the ideal that everyone can access and understand all the data they need - "data on demand" or "data as a service". This empowered, consumption-led view perhaps resonates with aspects of popular culture. A typical refrain is "You have to spread the data to everyone"[4].

To support (and promote) this demand, data strategies have traditionally had to focus on consumption as a key feature, considering how to supply data "on demand" in reports to humans.

An example might be a warehouse operative who wants to understand when to order more stock to meet requirements from manufacturing. She might construct her own report of what stock she holds and build predictions of expected stock requirements from manufacturing. To do this, she needs skills, toolsets and a supply of data.

Typical characteristics of these reports built and consumed by humans are:

- They are usually aggregate totals, albeit split by groups and categories
- They are imperfect and fault tolerant - often having data errors massaged manually by humans using tools like spreadsheets
- They are flexible and dynamic - with formats varying over time as consumers alter their questions and understanding
- They are loosely labelled - often unlabelled - relying on human readers to understand and interpret them with contextual knowledge.

While traditional analytics faces many challenges, perhaps the greatest challenge is the hard problem of conformance. This is the difficulty of achieving a sufficient degree of consistency between many different reports purporting to represent common aspects of the business. Individual reports may source data in various ways from a variety of places, or select, aggregate or present it in different formats. Most businesses suffer friction, arising from different consumers having inconsistent data and trying to match it up.

Our warehouse operative, above, might take her own copies of manufacturing data, make adjustments and interpretations and come to her own conclusions. It could be that the manufacturing department has different variations, with more up-to-date data, interpreted differently, and there may be ongoing disputes about the resulting conclusions.

As a real example: the author's first job involved reconciling new business reports between Marketing and Actuarial functions in an insurance company. Marketing definitions of "monthly" new business were based on calendar months, with varying numbers of days. Actuarial definitions of monthly new business were based on fixed four-weekly periods. To make matters worse, the Actuarial function scientifically netted off recent cancellations and reductions in premium amounts from new business, while the Marketing team did not. The differences in numbers could be as high as 30% for some lines of business. These differences between valid, but inconsistent views of the truth caused endless debates at all levels of the company.

Data strategies need to put structures in place to help companies manage conformance. Data Governance bodies, forums, roles and the transformation to data stewardship are typical solutions, coherently managing, monitoring and approving uses and definitions of data, and escalating mismatches to decision-making bodies. It can be rapidly appreciated that some of these conformance issues can be highly political: the example quoted of the new business figures is a prime example - with Board and C-level implications.

Balancing the desire to empower data consumption, while establishing sufficient control to achieve the necessary level of conformance against technical imperatives driven by business need are key challenges for traditional data strategies.

# Automation and the rise of Embedded Analytics

As enterprises introduce more and more automation, they will progressively shift decision-making from humans to rules. Every automation activity makes decisions such as when to start; when to stop; whether to flag an error; whether to put outputs in box A or box B; whether to order more stock or not. The "Rules" may not be strict algorithms and may even be complex patterns in neural networks or other Machine Learning systems. They are often termed "AI" but they are still, fundamentally, rules. These rules replace the traditional decision-making combination of a human and a report. We can describe them as "embedded analytics", and they are the foundation of the modern, cognitive enterprise.

A further twist to the automation challenge is the development of the Internet of Things. The increasing use of connected devices - sensors and switches, and edge computing - dramatically steps up the flow of data to and from embedded (or edge) analytics, which is often built into a device itself, smartphones and tablets, both pervasive across organisations today, being prime examples.

A modern data strategy may need to be able to frame and consider the supply of data to and from embedded analytics.

At first sight, it might seem that machines will need the same data as humans to make their decisions, but the reality is likely to be different. The characteristics of feeds to machines contrast dramatically with those of human-read reports:

- They are likely to be much more granular - with machines being capable of far higher volumes of decisions and calculations than humans, at much greater levels of detail
- They are likely to be less fault tolerant, with reduced ability to massage away errors
- They are likely to be in fixed, defined formats (even if unstructured data) to feed fixed physical devices
- They need to be machine readable, with clearly defined data items.

Our warehouse operative from earlier might want to build a system which recognises license plates on incoming vehicles, cross-references them against an order database and a stock database, and directs them to particular loading bays via a system of traffic lights, opening and closing warehouse doors automatically.

This sort of embedded analytics, however, is likely to be beyond the skills of most staff.

Building these solutions will involve extensive integration of multiple systems, both inbound and outbound, and a great deal more rigour than building a report for human consumption. Safety considerations for operating physical components will likely preclude the sort of „amateur“ approaches encouraged in traditional analytics. Instead of „citizen data scientists“, modern embedded analytics will need skilled data engineers, operating under rigorous controls.

A data strategy which supports a company on its journey to automation will need to shift the emphasis of data initiatives.

The strategy will need to set out how a company can progressively construct an efficient and effective capability to feed data to and from its new, automated, systems - both internal and external - with defined, machine-readable supplies of data.

The strategy will need to set out how the company will best be able to focus data from many different sources, overcoming the challenges of different formats and technologies: sourcing, storing and standardising often disparate data. While largely technical challenges, requiring IT architectural input, these problems also involve significant questions of ownership, funding and control.

One area which might become slightly less intractable as the shift to embedded analytics progresses is conformance. The data requirements for focus and feed are likely to be much better defined than the many and various human requirements of traditional analytics. Moreover, any differences in definitions may be amenable to technical resolution, with less political escalation required. On the other hand, quality is likely to become apparent as a much bigger issue.

Embedded analytics is, in general, much less fault tolerant than human consumers. Humans are generally able to interpolate, extrapolate and smooth over wrinkles in results, whereas automated solutions today are less likely to spot or correct errors. In the case of our warehouse operative earlier, on encountering an incoming vehicle with a license plate partially obscured by mud, she may happen to know or recognise the vehicle by other features, or she may be able guess the complete plate if she knows the structure of the number. Lastly, she always has the option of going and cleaning the plate. Automated systems today are less likely to be as flexible as a human counterpart in the face of poor quality data, although this will improve over time.

Modern data strategies will have to drive a new and increasing emphasis on nurturing and protecting data, maintaining and increasing quality levels. In many enterprises, it will be necessary to drive funded campaigns directed at building quality and data governance coherence.

Strategies should consider at a minimum two key points. First the function of „capture“ - looking at how data is acquired and ingested. For example, if an agent speaks to a customer, are they able to enter full customer details onto a reliable system of record immediately and easily, or do they end up scribbling some information on a notepad for subsequent processing? What steps does the company need to take to ease and ensure complete and accurate data capture and its ingestion into corporate systems? Fixing these essential data lifecycle components for all essential data items is likely to be a fundamental requirement for many data strategies.

The second key point is around care of data. The company needs to make sure that tools and processes are in place to enable data to be updated and curated appropriately. For example, if a customer address changes, is that change processed sufficiently rapidly, reliably and completely in all places the address is held? Any update, copy, edit or deletion can be a weak link in the data processing chain. Fixing care and capture are likely to involve significant changes in culture, training and tooling. Historical data flaws in existing stores of data, for example missing addresses, can be expensive to remediate, but resolving these is likely to be an essential step in a strategy towards automation.

## Looking Outwards

There can be strong synergies between initiatives to supply traditional analytics and initiatives to feed the new machine-centric embedded analytics - in particular, the supply of conformed data to reports can be closely aligned with the capabilities to focus and feed machines, if correctly managed.

Furthermore, development of a data framework and infrastructure to meet these needs - supplying and managing data flows - also forms the basis for supporting the third major strand of a data strategy outlined in the introduction to this paper: monetising or exchanging data.

A framework for managing data flows can act as a platform or marketplace, giving the company an opportunity to channel data to and from partners and customers in a controlled fashion - often via APIs - effectively selling their data, or services built on their data. Classic examples are Uber and AirBnB who have built entire business models around their data platforms, and a quite different and particularly interesting example is Yara, who supply farm and field data to farmers[5]. Data Platforms are likely to be crucial to a business' strategy - when platforms enter the same marketplace as traditional businesses "the platforms virtually always win"[6].

# Managing the Shift

It is evident that the shift towards automation brings new challenges in the use of data, and that modern data strategies need to support companies in assembling solutions. This shift will, in most cases, be gradual. Most businesses will continue to develop traditional analytics for many years, while gradually growing their embedded analytics capabilities alongside. Data strategies for these companies will need to propose effective solutions for managing the tension between consumption and conformance, whilst also setting out forward-looking plans to deal with the functions of feed, focus, capture and care. As we have seen it should be possible to find synergies between these areas, potentially developing further opportunities for the business to exchange data externally and improve the flow of data internally.

However, an essential part of data strategy is the realisation that not everything can be fixed at once. A modern data strategy will have to decide where to set the emphasis, and construct a roadmap for change, and this will vary case by case.

Data strategies, ultimately, must always serve the business imperatives of the company and should be closely aligned to their particular business strategy. Each company will manage the shift in their own way and plot their own course to their future.

# About the author

## Graham Olsen

Graham is a Delivery Executive with thirty years' experience in shaping and implementing data solutions in regulated industries. He helps clients transform their businesses and publishes and lectures globally on data and regulatory issues. He can be reached at graham.olsen@uk.ibm.com.

# For more information

See https://www.ibm.com/services/data-analytics

# References


1. IBM's Institute of Business Value paper "Automation and the future of work"
2. "Data Strategy" - Bernard Marr 2017
3. Gartner "Citizen Data Science Augments Data Discovery and Simplifies Data Science" - https://www.gartner.com/en/documents/3534848
4. The Economist Intelligence Unit's paper "Driving a data centric culture: a bottom-up opportunity".
5. "Yara and IBM launch an open collaboration for farm and field data to advance sustainable food production." IBM press release. January 23, 2020. https://newsroom.ibm.com/2020-01-23-Yara-and-IBM-launch-an-open-collaboration-for-farm-and-field-data-to-advance-sustainable-food-production; "Yara and IBM join forces to transform the future of farming." IBM press release. April 26, 2019. https://newsroom.ibm.com/2019-04-26-Yara-and-IBM-join-forces-to-transform-the-future-of-farming
6. Harvard Business Review: "Pipelines, Platforms, and the new rules of strategy".





IBM United Kingdom Limited 76/78 Upper Ground
South Bank London SE1 9PZ

Produced in the United Kingdom
September 2021