

Open in app ↗

Sign up

Sign In



Published in datamindedbe



Kris Peeters

Follow

Jan 15, 2020 · 5 min read · Listen



Save



The data product lifecycle

Your organisation wants to dive head-first into data and AI but you don't really know where to start? Data&AI is on the radar of most C-level leaders. It's often seen as a differentiator from competition, as an enabler for more customer engagement, as a tool for cost reduction. And there is no lack of industry analysts and strategic reports that highlight the impact that data already has on the corporate world. But how to get started? We've done quite a few tours of duty in data analytics. We've reflected on what comes back across clients and here is what we learned:

1. Data products, not projects

Organisations who are successful in data have one common trait: they build data products, not data projects. What is the difference? In projects, you know what you want, you know how much you want to invest in it, you make a plan for how to build it, and you build it. Once build is done, only maintenance is needed. Building a bridge is a project. You know you want to connect two sides of a river. You get money from both communities, you hire an architect and a temporary (contractor) team and you build it. What's important in a project: Deliver predefined scope on-time, within budget.



26





<https://now.tufts.edu/articles/why-new-bridges-are-needed-cape-cod>

For products you need a completely different mindset. You are focused on business outcomes. And you give teams the freedom and responsibility to iterate towards that outcome. That fits much better with data&AI initiatives. Why? Because you often know very clearly what you want, but you don't have a clue of how to get there, even though you do recognise that data will be an important factor.

A typical example: customer satisfaction. You want that to be as high as possible. It's very clear what you want. But there are many ways to Rome. Do you want to increase customer satisfaction by:

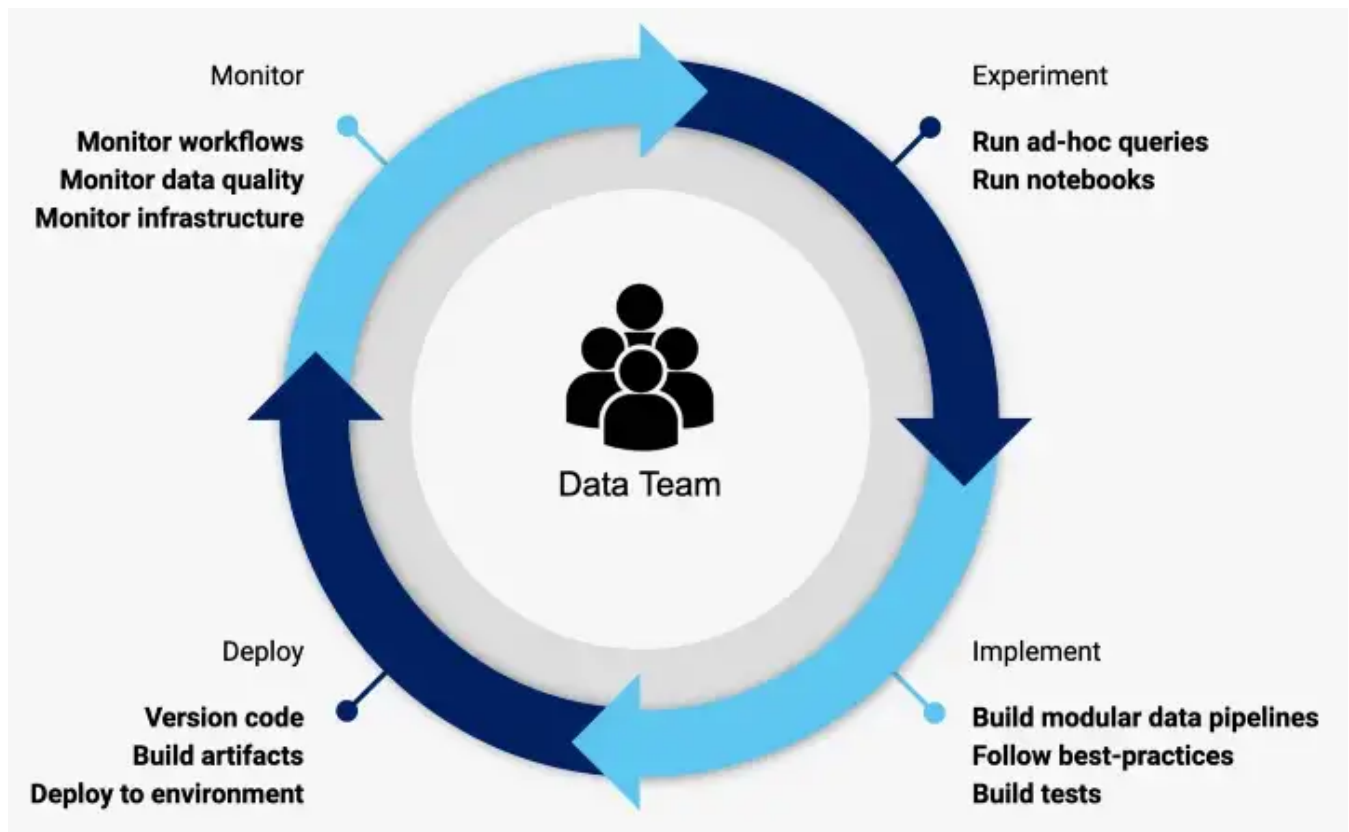
- ... finding unhappy customers who are about to churn, and send them a promotion?
- ... doing a customer segmentation project followed up by a 1-on-1 marketing strategy?
- ... building a recommendation system that can help the customer find its way in your offering?

Maybe a combination. Or maybe something else entirely. Either way, in a product mindset, you enable your data and customer teams to work closely together to

iterate towards a solution fast, and then reliably run that solution for as long as it adds value. That is called a data product.

2. The lifecycle of a data product

Every data product goes to the following phases, whether they like it or not:



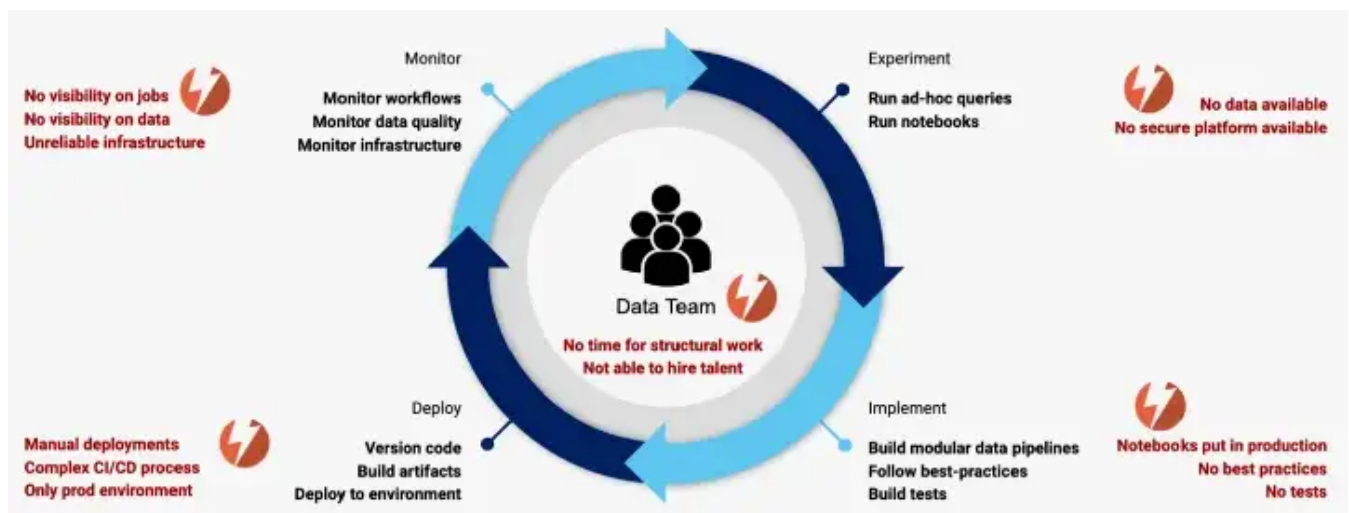
Lifecycle of a data product

- **Experiment:** Which use case are we trying to solve? Which data do we have available to solve it? I run a few notebooks, or execute a few ad-hoc queries, maybe visualise a few things in a BI tool to understand better what I'm looking for
- **Implement:** Once I know what I'm looking for, I need to implement this. That means that I want to run it regularly in batch or real-time. Mature data products are built using software engineering best practices, with version control, tests, modular code, and all that good stuff.
- **Deploy:** Ok, it works on my machine. Ship it. Data products only add value once they run in production. So we need to version our code, build artifacts and actually do deployments to development, acceptance or production accounts.

- **Monitor:** Once it's deployed, you need to monitor the performance of the data product, both from a technical and business perspective.

And then the circle starts again. You experiment again, you implement new pieces, deploy them again, and monitor the outcomes. The faster you go through this cycle, the faster you will learn and the faster you will deliver value to your stakeholders.

Based on our experience, plenty of things can go wrong in this lifecycle, as illustrated below:



Issues encountered in the data life cycle

Many organisations we see still don't have their data available on a secure platform, in a self-service mode. Often there is no actual analytics environment or the BI department hords and protects data from business users. And when there is an analytics environment, the lack of data governance turns data lakes into swamps.

Another recurring issue is that the notebook environment IS the production environment. No code reuse, no data reuse, no best practices, no tests, no version control. Nothing. You change the notebook, you change the data product. That might be fine for some innocent reports. But the more complex and customer facing your data product becomes, the less acceptable that way-of-working is.

Deployments are another pain point. Besides the production environment, there are not a lot of testing grounds for data products. This is due to data privacy, complexity of setting up an environment and complexity of CI/CD processes.

Once a product is live, the best monitoring tool that organisations have is customers complaining that the product is not available, shows incorrect data, or contains

outdated information.

And finally, and maybe the biggest painpoint of all, most companies are struggling with finding enough data talent. And once they do find them, they are not allowed to do structural improvements to this product lifecycle, because a next feature awaits. At one of our first clients, our contract was nearly terminated, because I spent time pushing our code to a git repository and integrating with a build server. “You were hired to build features, not be an operational expense”. Ouch. Luckily, I could persuade the manager to let me stay.

3. The importance of engineering and governance

I know, data scientists have the sexiest job of the 21st century. And while their genius contributions can often make a profound impact on the value being created, especially in the experimentation phase, it is not the whole story.

Proper data governance can give you a business glossary, a data catalog, data lineage, ... I know, these concepts are not sexy. But they do accelerate the build of new data products.

Proper data engineering can bring automation, scalability, security, flexibility, reliability, Again, no business sponsor gets excited when they hear these notions. But you cannot live without.

What are your own experiences building data products? Any pain points that stand out? Happy to hear your feedback and interact.

Data Engineering

Data Governance

Data Product Creation

Get the Medium app

