Why Most Data Projects Fail

...and How to Avoid It

Go out on social media and see why people think they failed:

- Snowflake vs Databricks
- Data lake versus ...
- Python vs Java vs Rust

85%

of data projects fail

Why is this so bad?

Why can’t companies create more value with data?

Technology is just one, initially smaller but still important, part of success with data.

The Right Questions

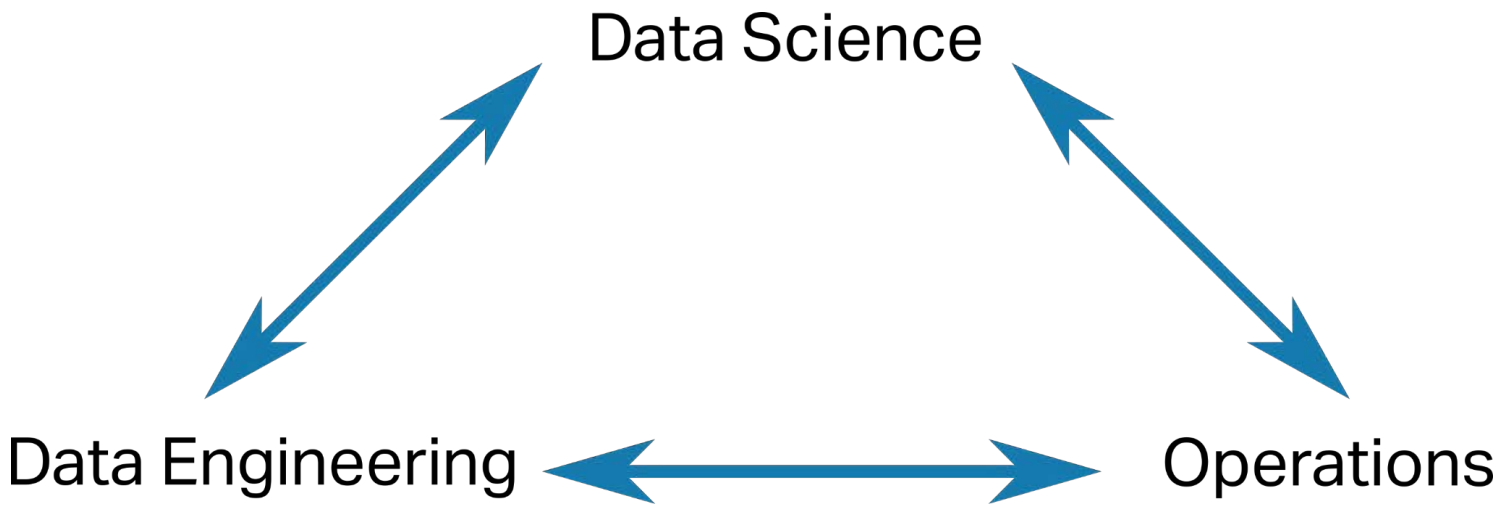
We must answer:
who, what, when,
where, and how

Who

The Right People at the Right Ratios

We need data scientists, data engineers, and operations engineers. Each person is important with the right ratio.

All Three Teams Are Required For Success



What

What is the business value?

Just saying you want AI isn't enough. There should be a clear and attainable path to value creation.

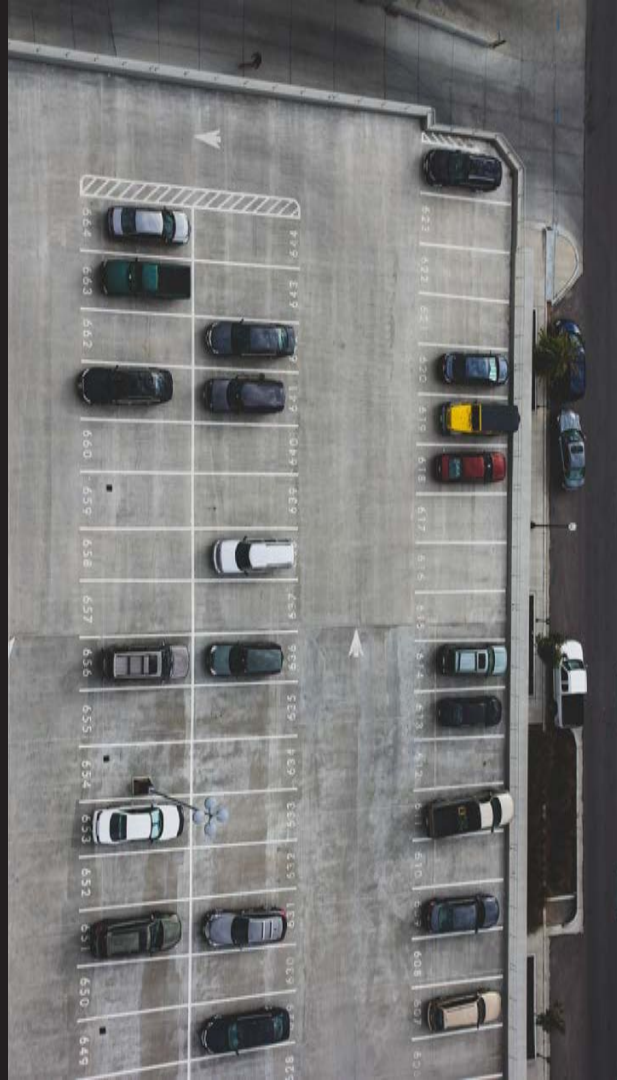
Big Data Institute

Having a data strategy isn't enough either. There needs to be a plan and execution.

When

When will you generate value?

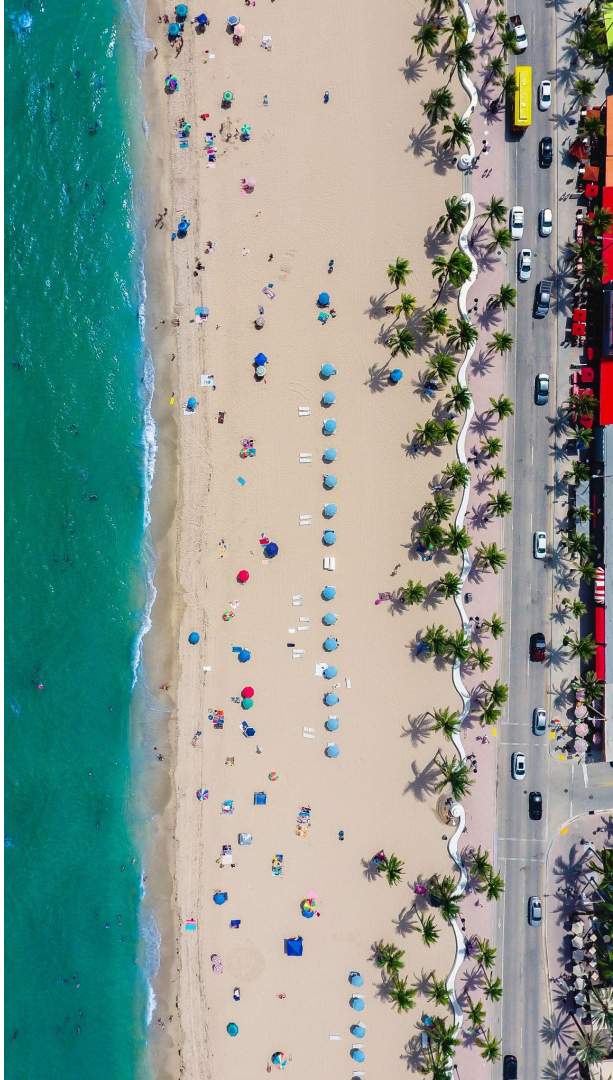
Unattainable timelines aren’t feasible and neither are “when it’s ready” timeframes.



Projects that take too long to generate value get cancelled. Projects that promise too much too soon fail to meet expectations.

Where

Data teams need to have a clear plan and architecture of where each piece will be done.



Just following a vendor’s or cloud provider’s recommendations may be using the wrong tool.

How

How will the plan be executed?

Data teams need a clear plan that they are executing. This plan needs a singular focus or the work will go in different directions.

Data teams require clear focus to not get bogged down in too many different directions.

Bonus: *Why*

Why is data valuable?

Eventually, you may be forced to justify your spending or budget. I look for a 10x ROI on the investment.

Start by answering the questions

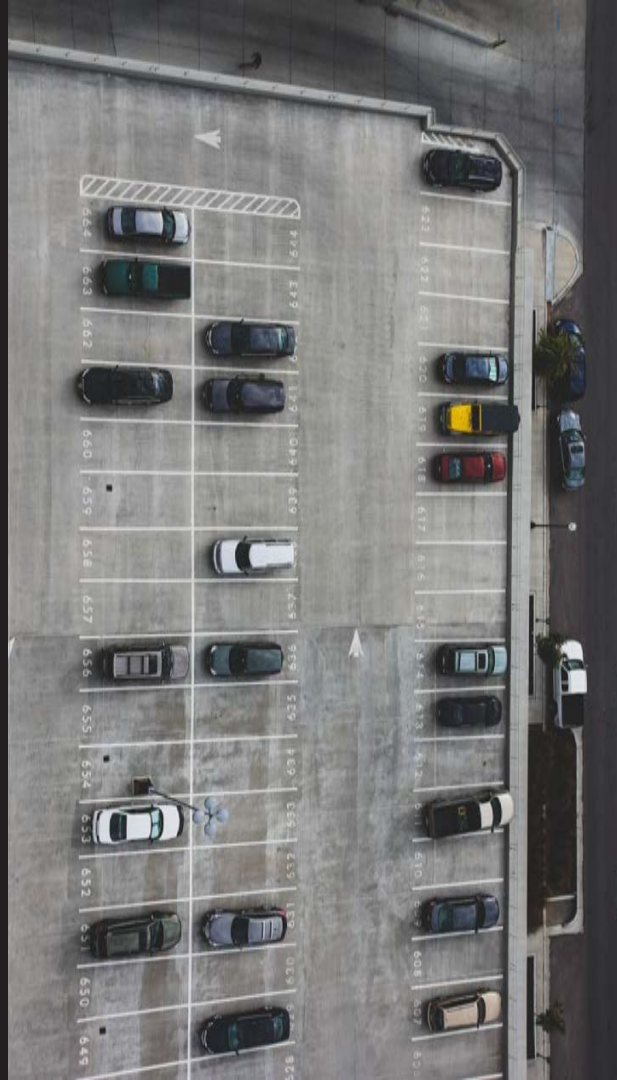
Then move on to the execution

Big Data Institute

Not all gaps are technology

Check for gaps in people and tech

AI and data projects are rarely just adding new technologies. It is an organizational, people, and skill change.



Should you get help?

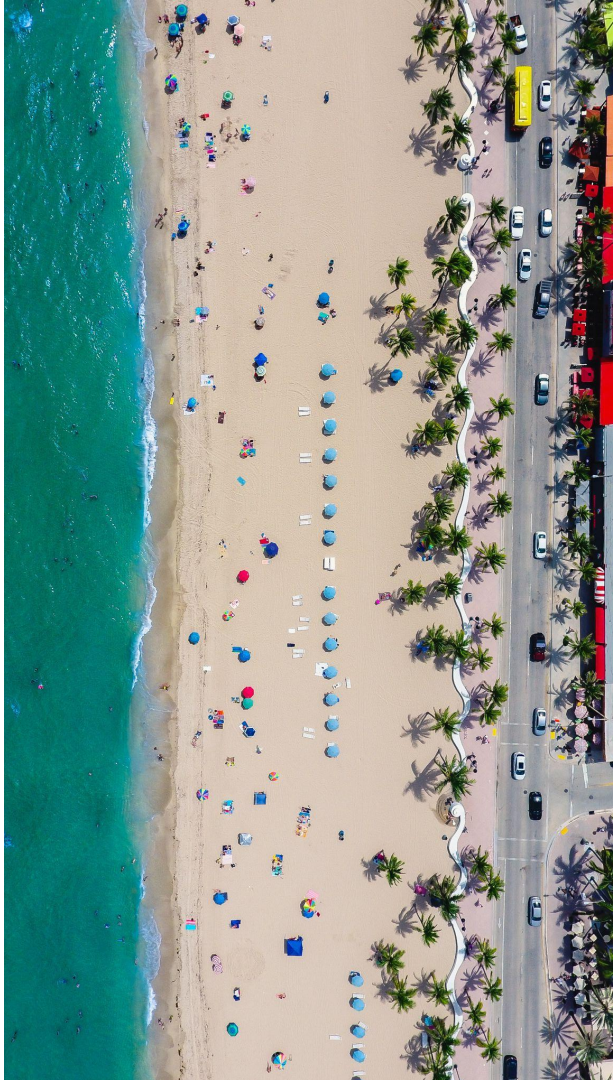
When should you get help?

Various types of help: outsourcing, technical consulting, management consulting. Problems don’t fix themselves without effort.

Most have a project going or about to start

Invert to look at failures and issues

“Invert, always invert: Turn a situation or problem upside down. Look at it backwards. What happens if all our plans go wrong? Where don’t we want to go, and how do you get there?” - Charlie Munger

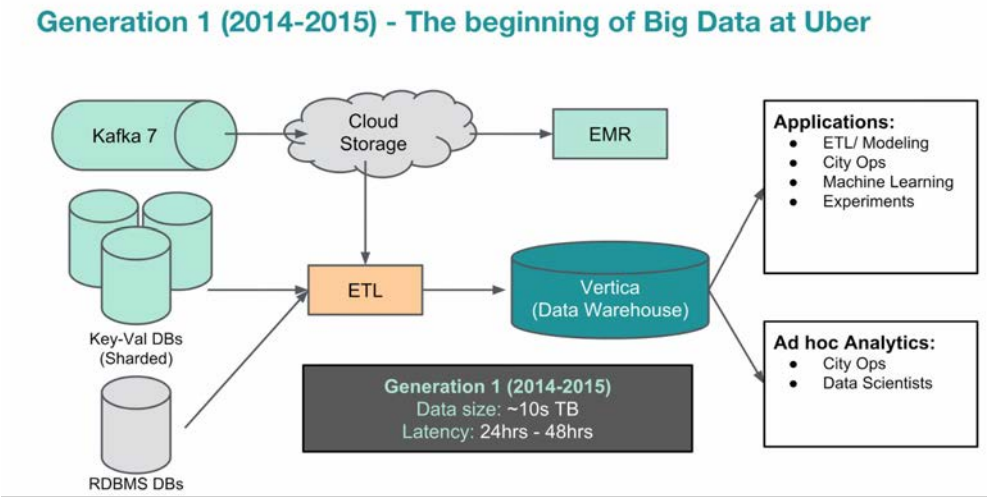


How good is a data engineer at a data science task?
How good is a data scientist at a data engineering task?

What would the business say if the project was canceled or the cluster turned off?

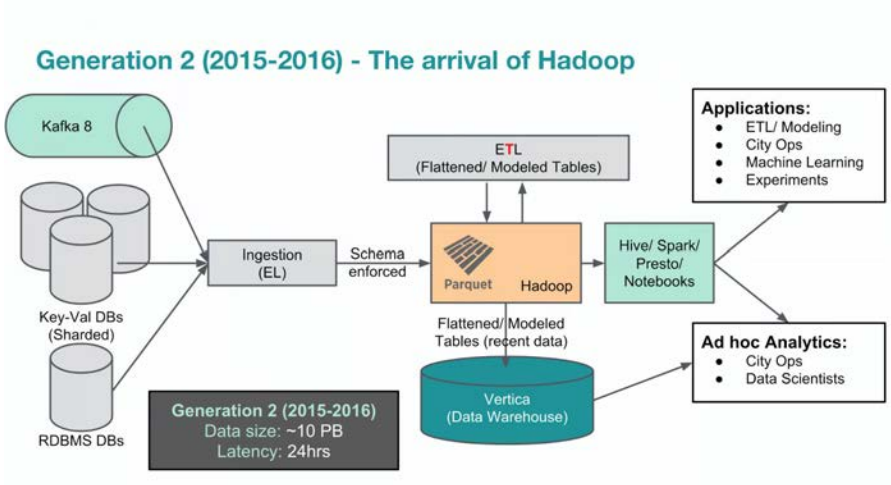
Why shouldn’t you copy someone else’s architecture?

Generation 1: The Beginning



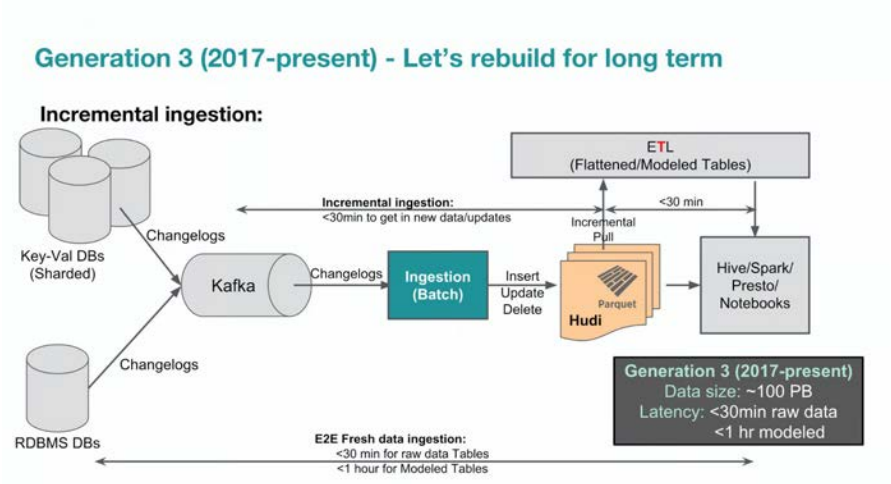
Source: https://www.uber.com/en-PT/blog/uber-big-data-platform/

Generation 2: +Hadoop



Source: https://www.uber.com/en-PT/blog/uber-big-data-platform/

Generation 3: Rebuild



Source: https://www.uber.com/en-PT/blog/uber-big-data-platform/

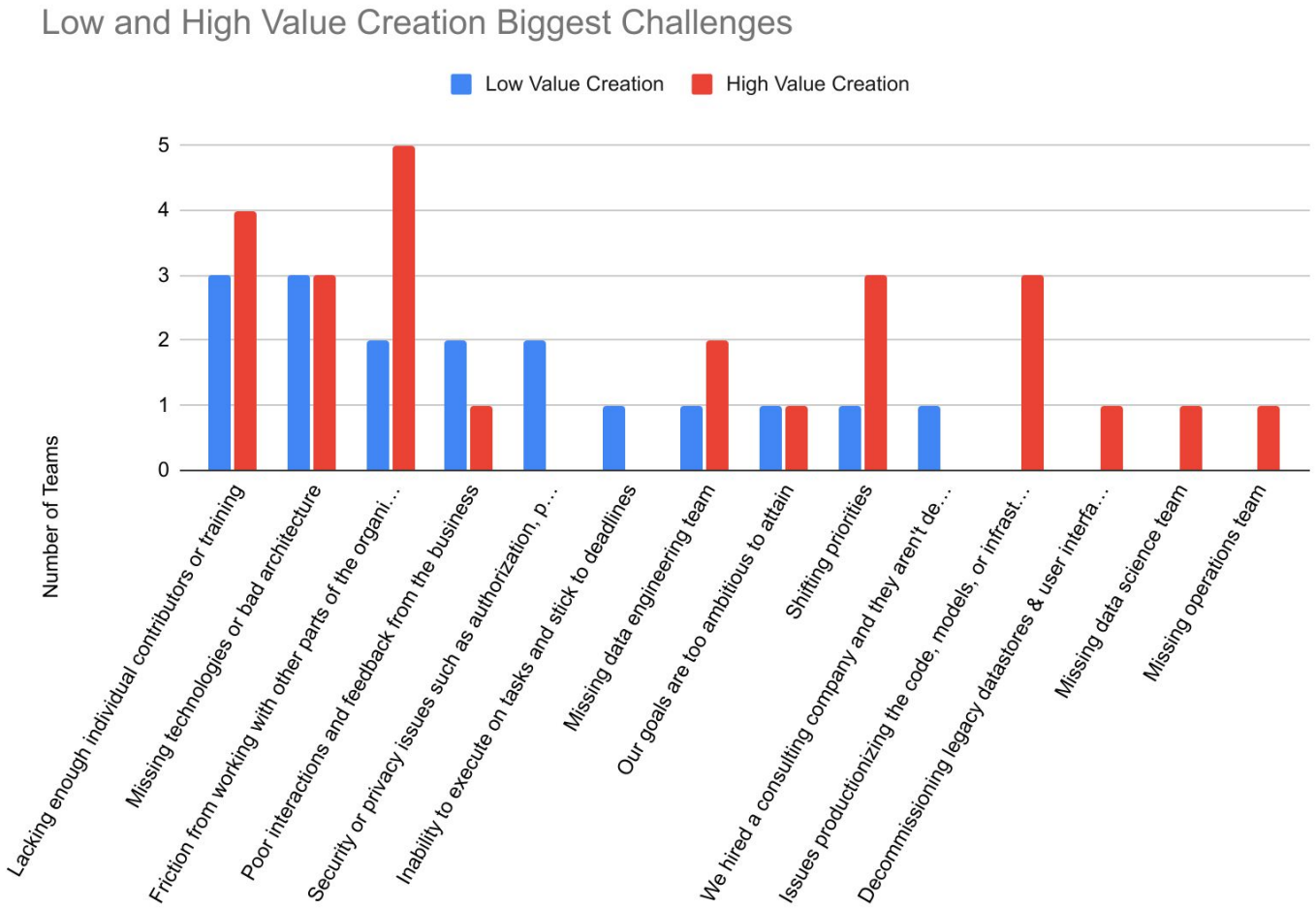
Don’t over-engineer and don’t copy. Move when necessary.

We ran our Data Teams survey in early 2023 and received 81 responses. The respondents came from all sizes of companies and maturity in big data projects.

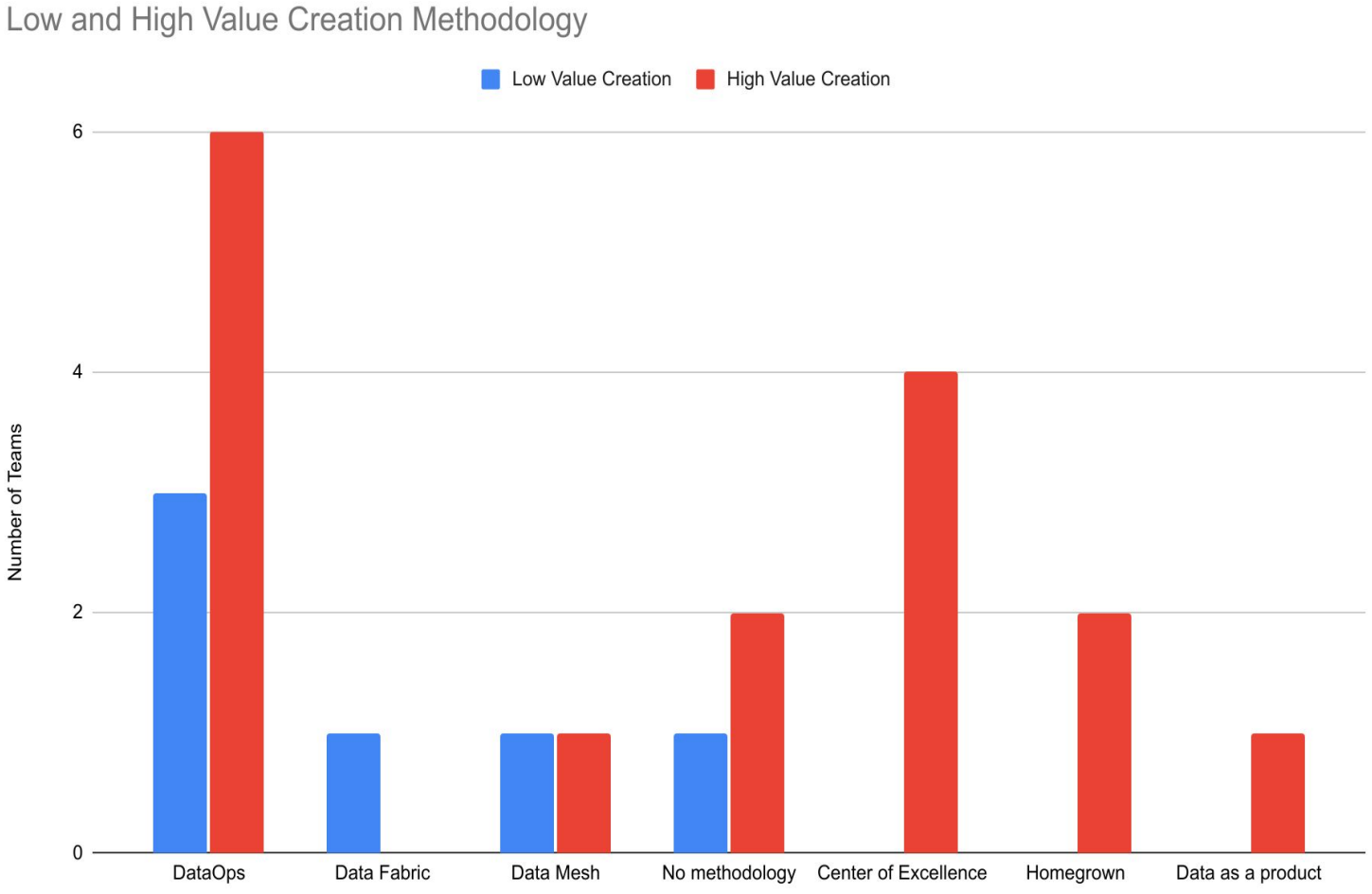
https://tiny.jesse-anderson.com/survey2023

Worst and Best

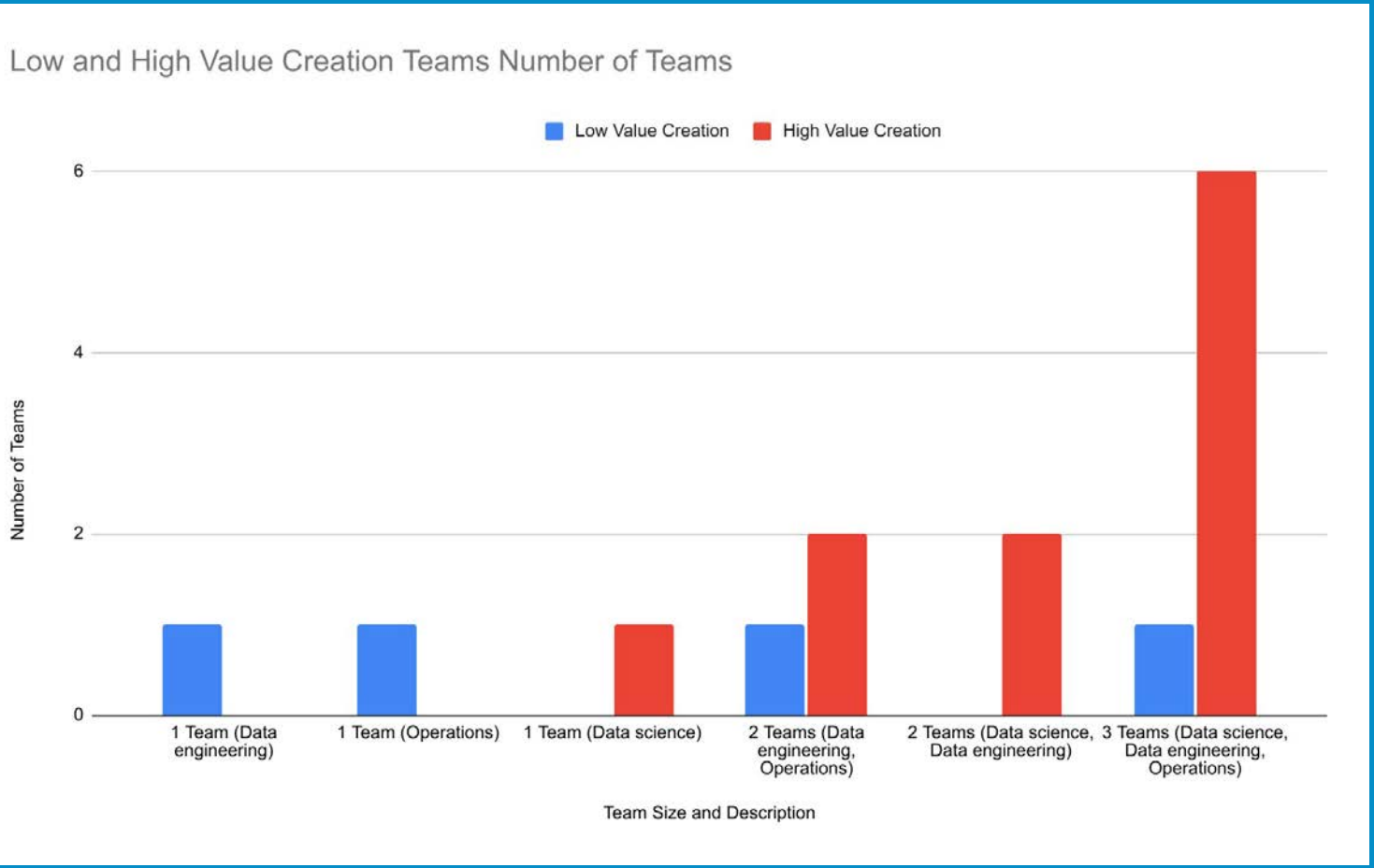
I looked for the worst and best value generation to see what they did.



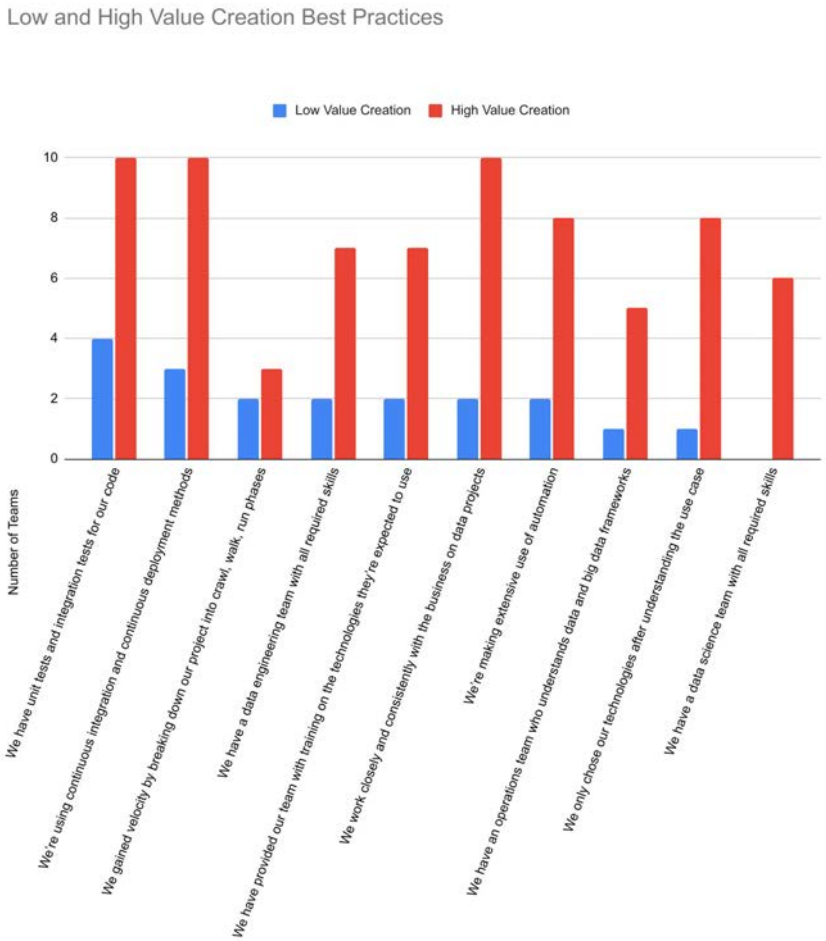
Both value creation levels face similar challenges. High-value faces more advanced. Low-value faces early mistakes.



High value uses more method-ologies



High value have 2-3 teams and low value have fewer teams with 1-2 teams

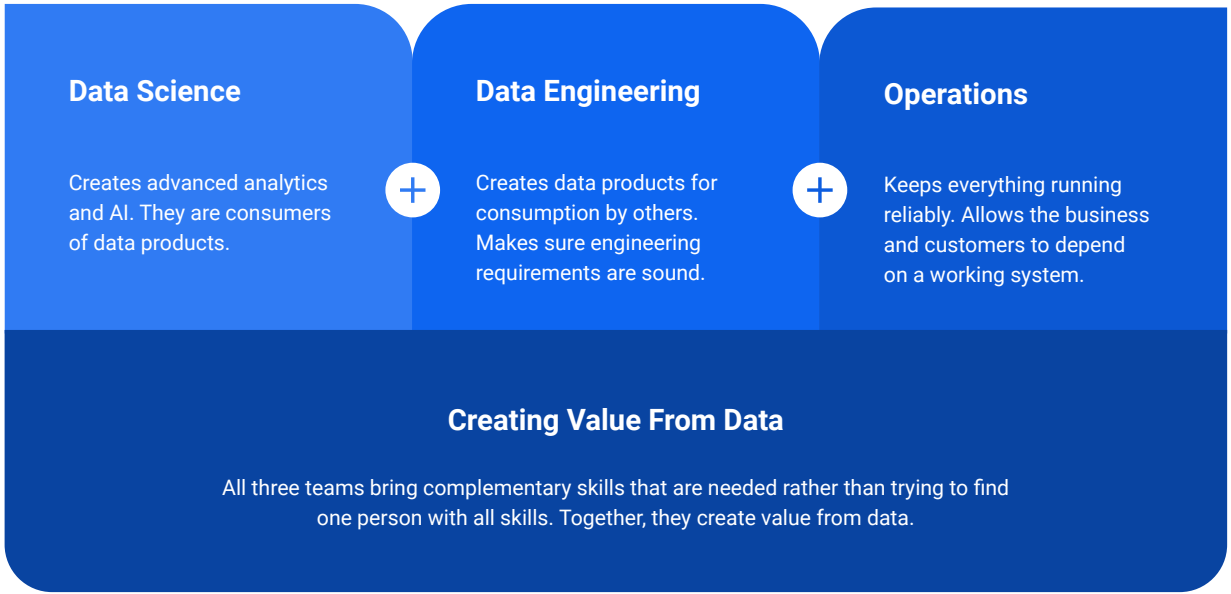


Best Practices

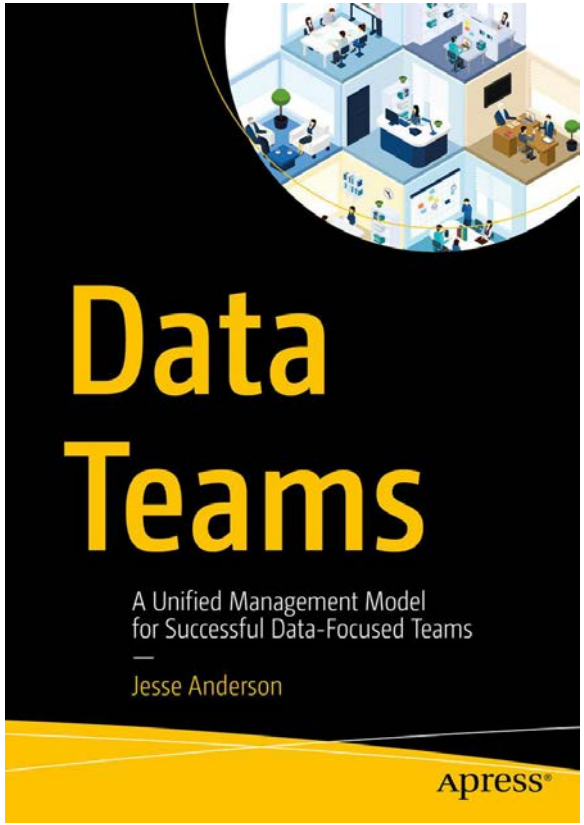
High-value creation teams use far more best practices than low-value.

Data Teams

A unified management model for success



Big Data Institute



Thank you

bigdatainstitute.io


Big Data
Institute