Goldilocks and the 3 data mesh bears

Sydney Data Eng meetup
Thu, 1 September, 2022

Simon Aubury

/thoughtworks

Simon Aubury

Principal Data Engineer

Once upon a time ...

There were three bears, who lived together in a house of their own in a wood.

One of them was a **small bear**; one was a **middle-sized bea**r, and the other was a great, **huge bear**!

A little girl came into the house, named **Goldilocks**.

https://open.spotify.com/album/7fue7usKwgVD2YmyuJVMsA

What is data mesh?

Data mesh is a decentralized socio-technical approach to manage analytical data at scale. This is designed to solve issues arising from the gap between data consumers and producers. It resolves these by distributing data ownership, treating data as product, giving users data ownership, empowering them with self-serve tools, and federated data governance.
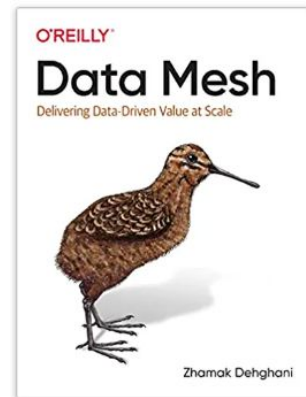
4

Principles of data mesh

Distributed domain driven architecture
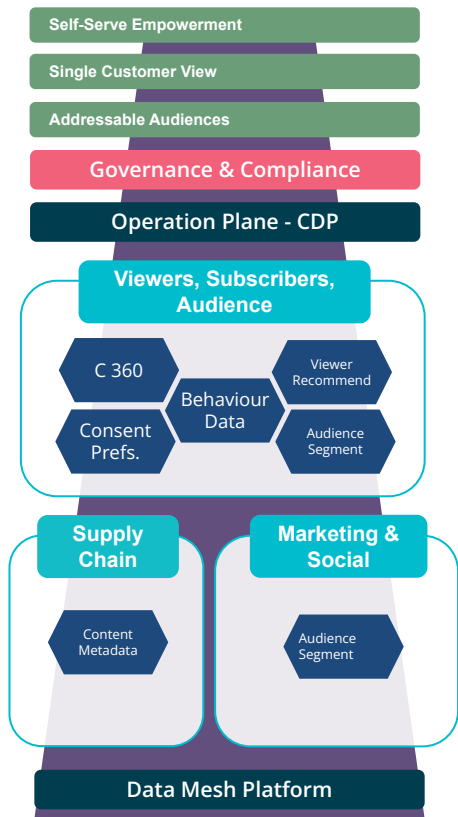
Self-serve data infra
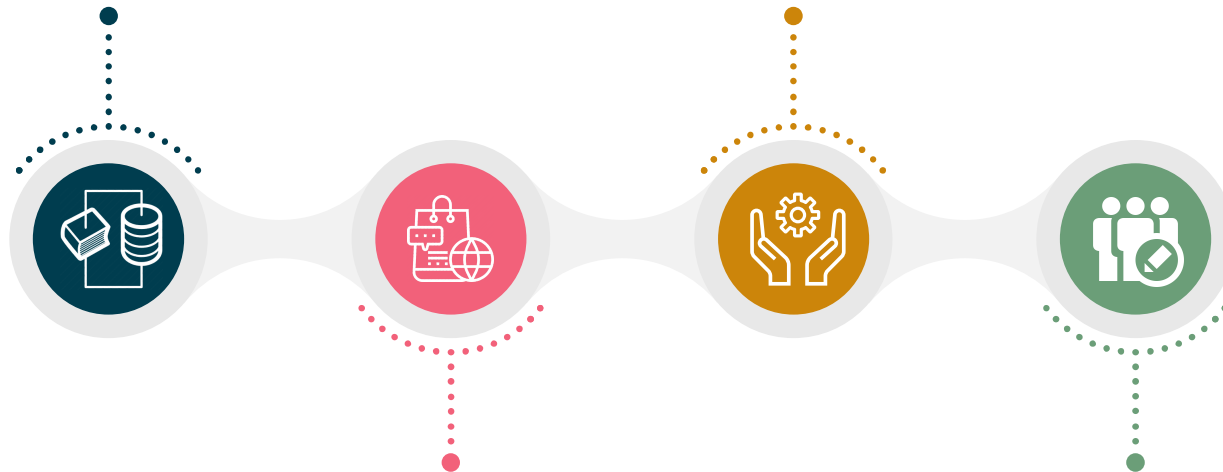
Data as product

Federated data governance

Data Mesh: Delivering data-driven value at scale book by
Zhamak Dehghani

Data Mesh example

Single Customer View - Single platform to capture and customer attributes & Consolidated Customer Profiles

Addressable Audiences - process to access/ store audience data and segments

Personalisation & recommendation - Segment-based personalised experience for viewers, tailored content recommendations

Experimentation - testing across teams through user-friendly tooling

Performance Measurement - High quality data with visibility and real-time data to feed analytics
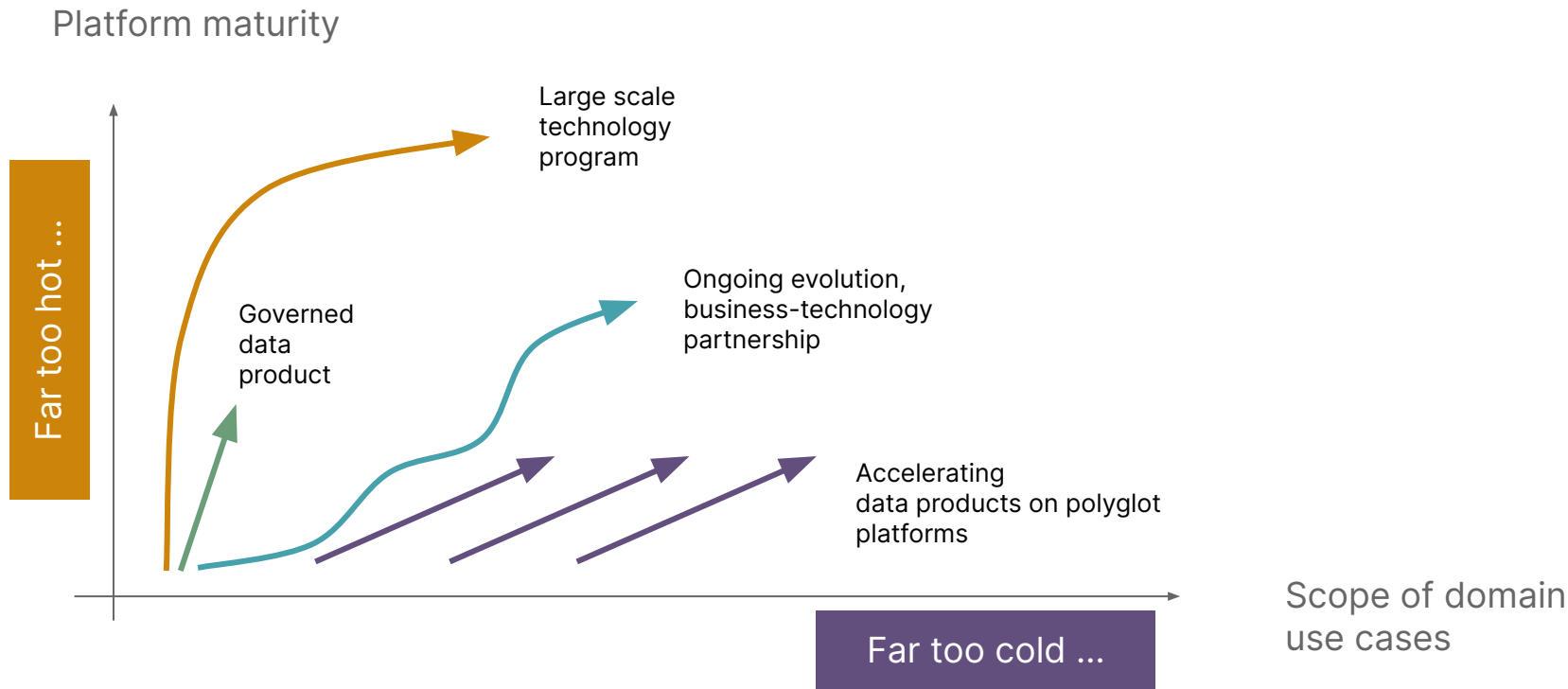
What could go wrong?

Photo by Sandy Millar on Unsplash

🧸 *Bear #1*

Not adopting <u>all</u> principles

Photo by Luwadlin Bosman on Unsplash

Some data mesh trajectories

Platform maturity

Far too hot ...

Large scale technology program

Governed data product

Ongoing evolution, business-technology partnership

Accelerating data products on polyglot platforms

Far too cold ...

Scope of domain use cases

Foundational pillars of Data Mesh

Four interconnected &
non-negotiable

Domain-oriented Ownership

Data as a Product
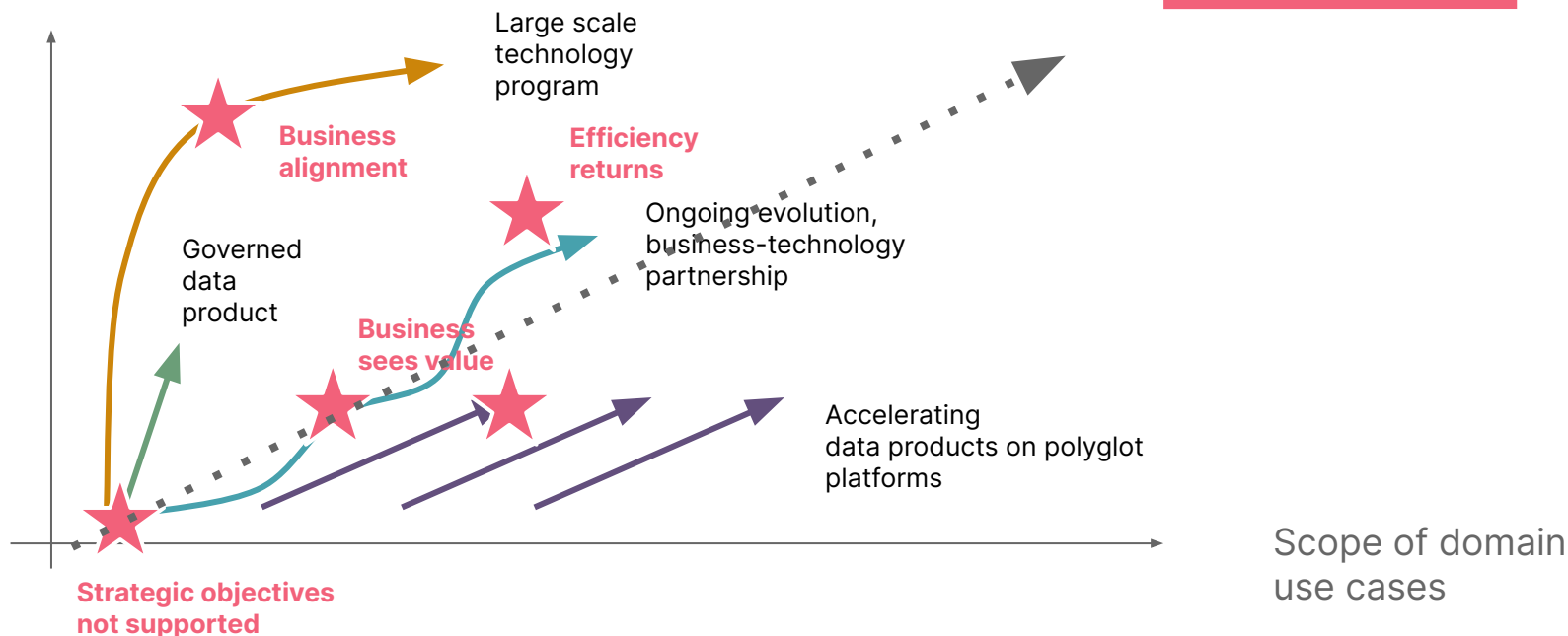
Federated Computational Governance

Self-serve data platform
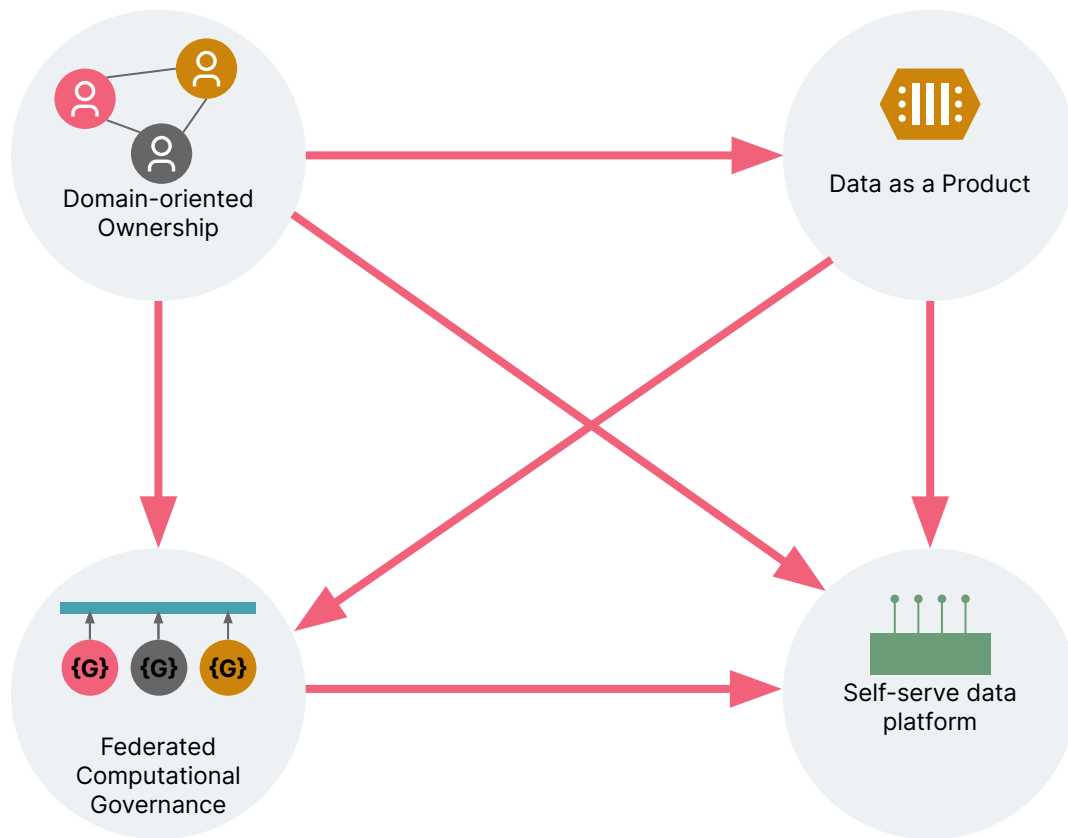
Dehghani, Z. (2022). Data Mesh: Delivering Data-Driven Value at Scale (1st ed.). O'Reilly Media.
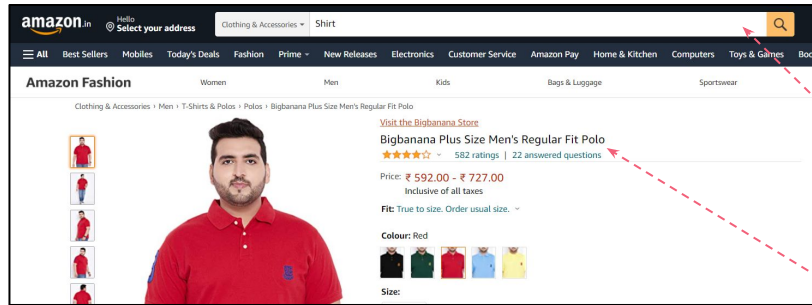
Some data mesh inflections

🧸 *Bear #2*

The right level of inclusion

Photo by Marina Shatskih on Unsplash

What is a Product?

- A product is goods or services that can be offered to a market to satisfy customer needs and wants

- A product has a bundle of tangible and intangible characteristics

Self Describing

Discoverable

Addressable

Trustworthy

Federated data governance

Enterprise Domain Model (Conceptual)

maintains — Enterprise Data Architect — activate domain — Data Governance Office

inform   collaborate   identify

Domain Engineer — collaborate — Data Owner

authoritative sources

approve

Kafka data hub
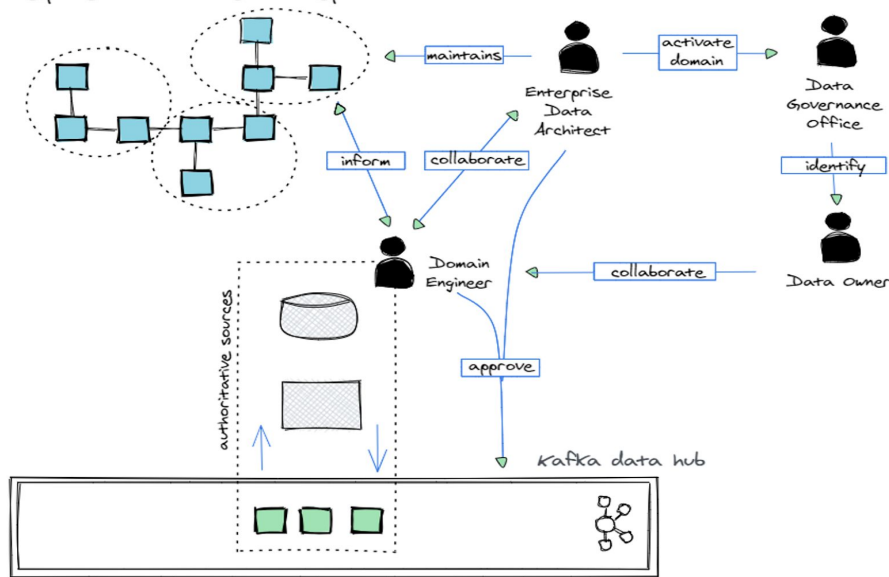
"Enterprise Data Model"

Data Catalog = Technical Metadata + Ownership Information + Data Lineage + Business Glossary

Data Quality = DQ Rules visualisation + Self-serve capability to create rules

Data ownership= Personas + DG Committee + Cadences + Org Incentives

Goldilocks moment for Data Products

Just enough centralised governance

Discoverable

Addressable

Self Describing

Trustworthy

🧸 *Bear #3*

Implementation considerations

Photo by Oxana Lyashenko on Unsplash

Data mesh implementation: top down

Advantages:

- Clear visibility

- Ability to choose right milestones

- Organisation wide change management program is easy when there is buy in from top

Considerations:

- Large organizational impact

- Big commitment

- Organisation maturity required

Setting up enterprise domains & data culture
Change Management, Organisational Incentivisation, Ownership policies, Domain standards, Governance policies, et al
Either set up all domains or choose one domain depending on business situation.

Self serve infra platform
Self serve data platform with infrastructure, product and pan enterprise data discovery capabilities

Federated data governance
Federated DG= Data Quality + Data Catalog + Data Ownership
Data Catalog Technical Metadata + Ownership Information + Data Lineage + Business Glossary
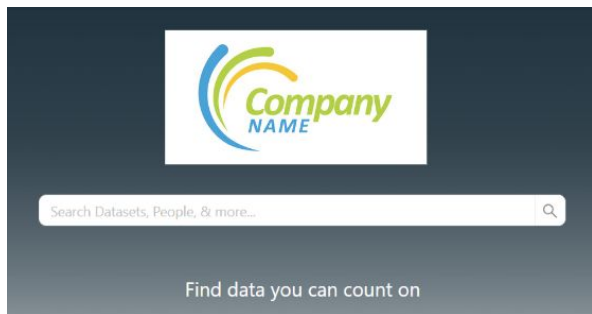Data Quality = DQ Rules visualisation + Self-serve capability to create rules
Data ownership= Personas + DG Committee + Cadences + Org Incentives

Data as product
Data assets discoverability, interoperability, addressability, self-description, trust, security. Contract definition of product, Self serve capability to easily and uniformly create data products.

Domain Onboarding

Data mesh implementation: bottom up

Advantages:

- Iterative process: test & learn
- Ability to show results
- Change management in small steps
- Nascent maturity orgs can also adopt this approach

Considerations:

- Ability to scale up the program
- Leadership engagement
- Long roadmap

Setting up enterprise domains & data culture

Change Management, Organisational Incentivisation, Ownership policies, Domain standards, Governance policies, et al Either set up all domains or choose one domain depending on business situation.

Bare minimum self-serve capabilities to support source oriented data products

Federated data governance (DC + DQ+ DO)

1 Data Catalog =Technical Metadata + Ownership Information + Data Lineage + Business Glossary

2 Data Quality = DQ Rules visualisation + Self-serve capability to create rules

3 Data ownership= Personas + DG Committee + Cadences + Org Incentives

Domain onboarding

Data as product

1 Source oriented products

2 domain oriented products

3 Fit for purpose or Consumer oriented products

Domain onboarding

Full scale self serve infra platform (remaining parts)

Self serve data platform with infrastructure, product and pan enterprise data discovery capabilities

Domain onboarding

Data mesh implementation: hybrid

Just right ???

Advantages:

- Ability to show results across
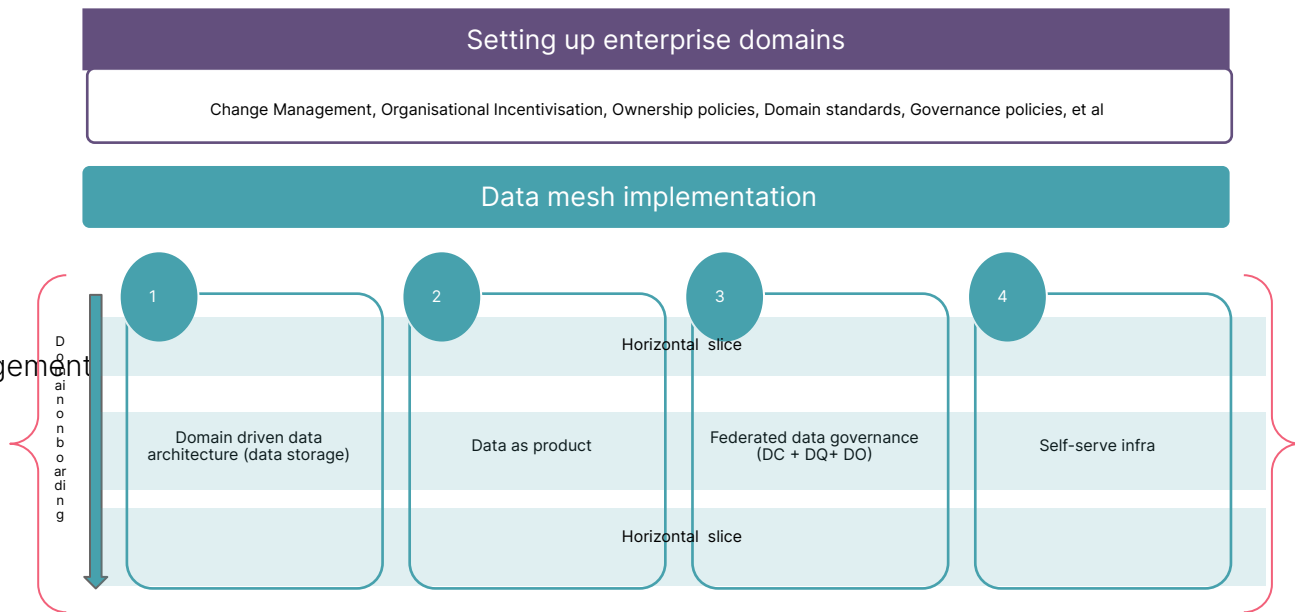- Faster approach

Considerations:

- Complex program management
- Leadership commitment
- Massive and rapid change management

Setting up enterprise domains
Change Management, Organisational Incentivisation, Ownership policies, Domain standards, Governance policies, et al

Data mesh implementation

Domain onboarding

1	2	3	4
		Horizontal slice	
Domain driven data architecture (data storage)	Data as product	Federated data governance (DC + DQ+ DO)	Self-serve infra
		Horizontal slice	

🧸

In summary

/thoughtworks

Photo by Chris Nagahama on Unsplash

Fairy tail ending?

Adopt all four pillars of Data Mesh

- Four interconnected & non-negotiable

Include everyone

- Appreciate their needs

Organisational appropriate adoption

- Top down, bottom up - but show results

Photo by ___drz___ on Unsplash

Thanks - questions?

/thoughtworks

Photo by Nathan Dumlao on Unsplash