OpenCredo
A TRIFORK COMPANY

Data Mess → Data Mesh:

Navigating People, Process & Platform

Nicki Watt - CEO & CTO
@techiewatt

goto; conferences

About Me

Nicki Watt
CEO & CTO

@Techiewatt

OpenCredo
A TRIFORK COMPANY

- Cloud Native Architectures
- Platform Engineering
- Data & Data Mesh Solutions
- Digital Transformation

Vault · Kubernetes & CNCF Ecosystem · golang · Google Cloud · Terraform · amazon web services · kafka · Azure · neo4j · python · Java

What does it take to move towards acquiring a half decent data mesh solution these days?

What does it take to move towards acquiring a half decent data mesh solution these days?

Especially when you feel you are starting with a data mess

Agenda

1. What is data mesh?
2. Considerations for the journey
3. Tracing a journey from mess → mesh
4. Evolutionary considerations

5. Conclusion

1 What Is Data Mesh?

What Is Data Mesh?

A decentralized sociotechnical approach to share, access, and manage analytical data in complex and large-scale environments - within or across organizations.

Zhamak Dehghani -
Data Mesh: Delivering Data-Driven Value at Scale

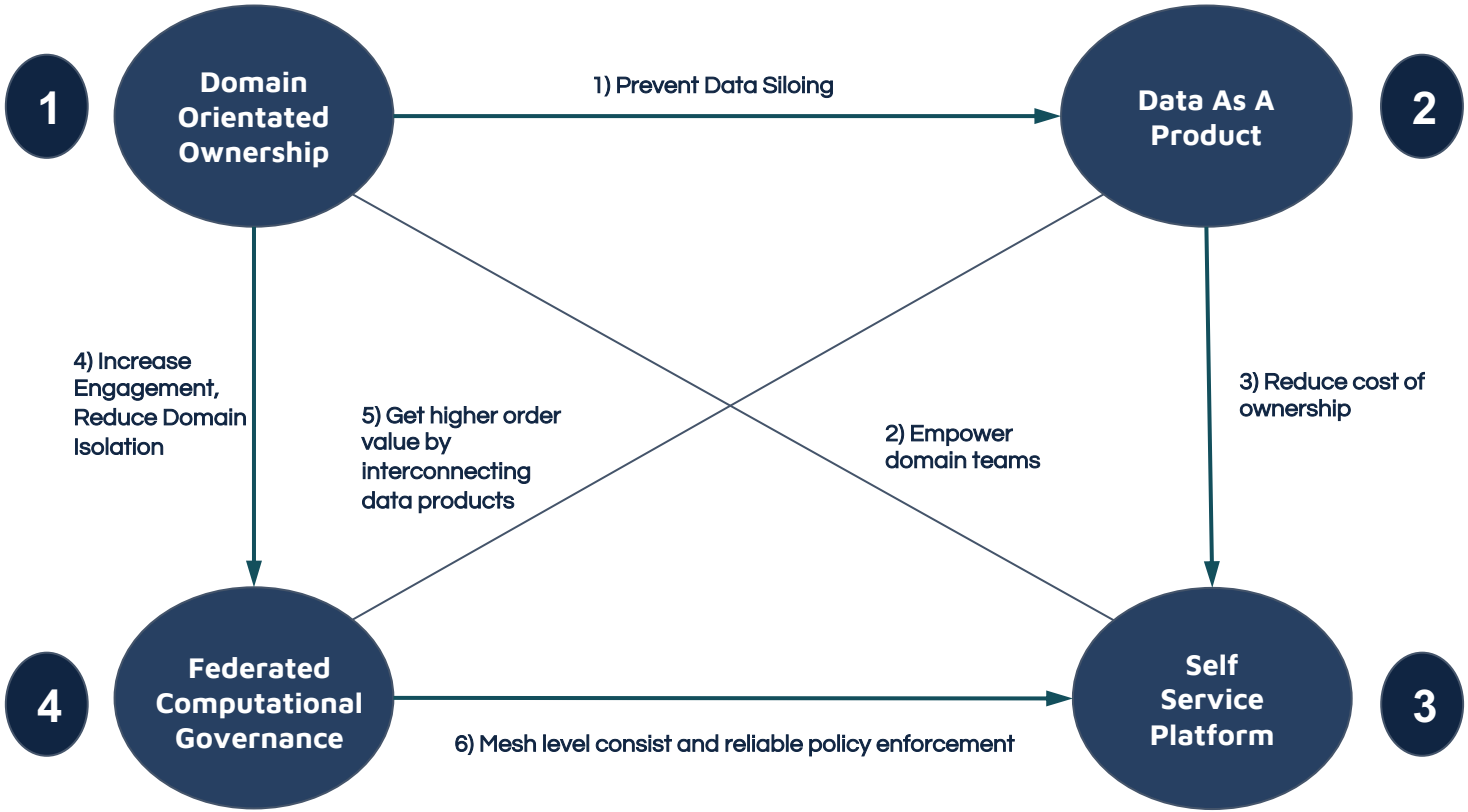
What Is Data Mesh?

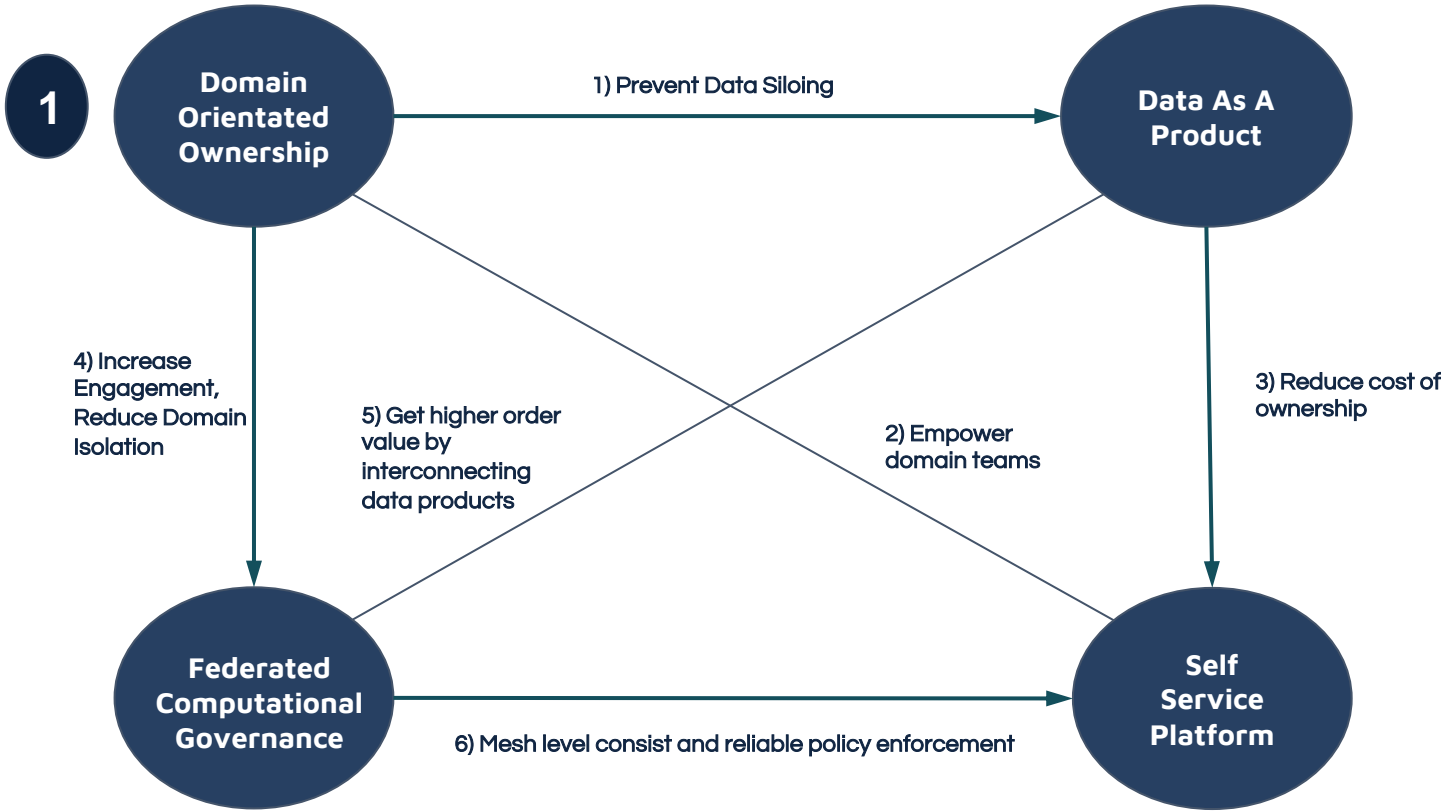
New paradigm & way of operating created to address challenges faced by large orgs trying to gain data driven insights from data at scale

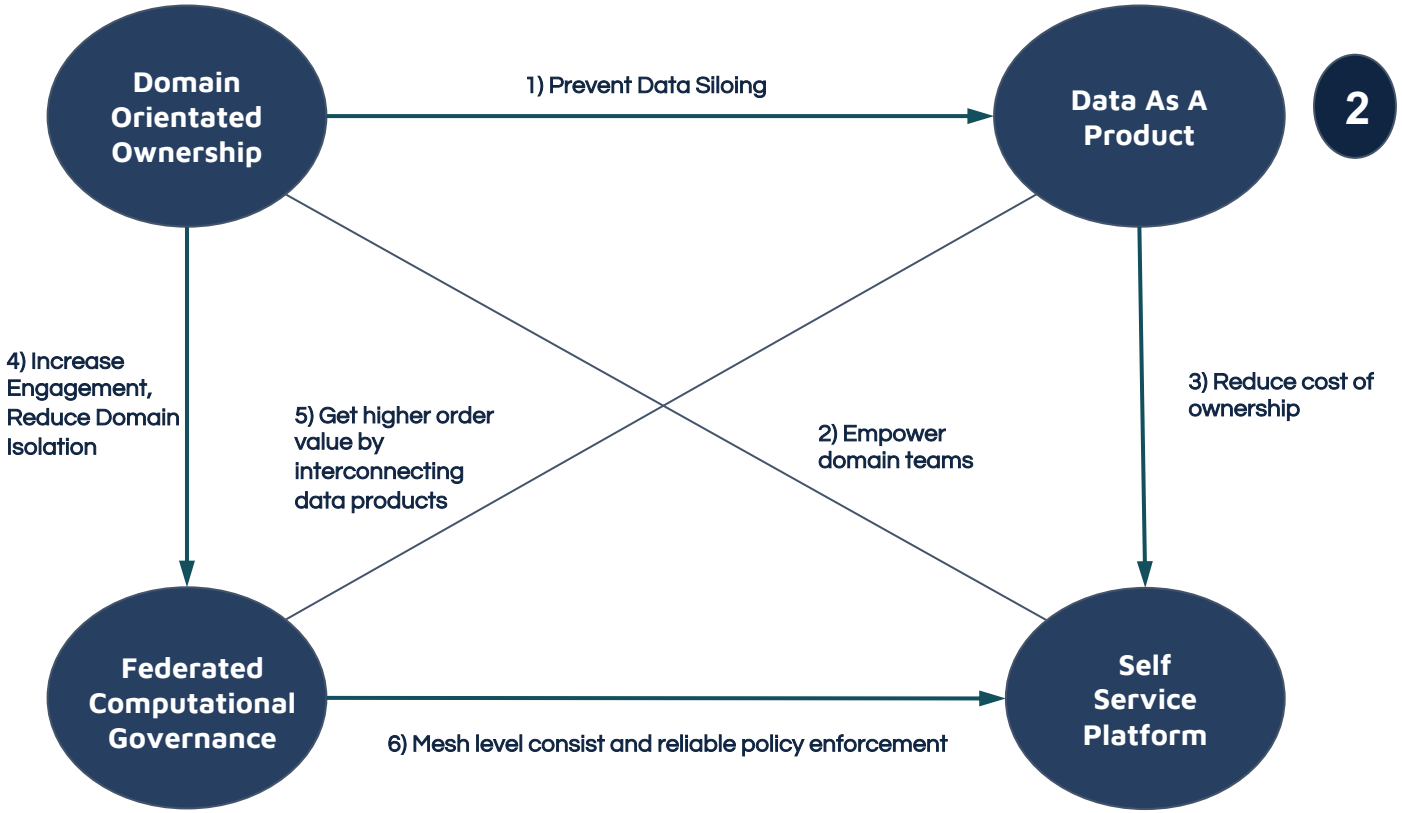
Why would I want one?

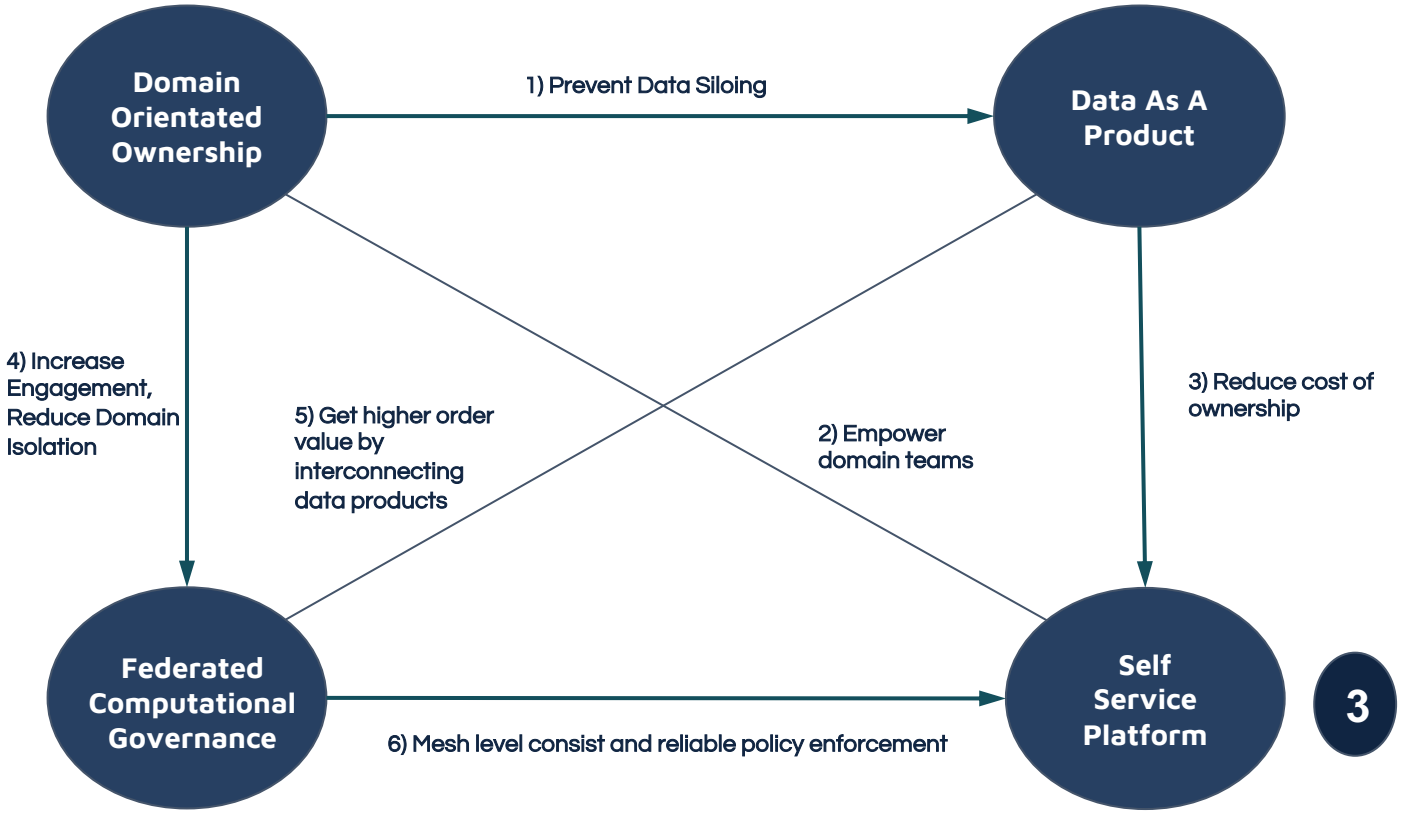
Key Objectives

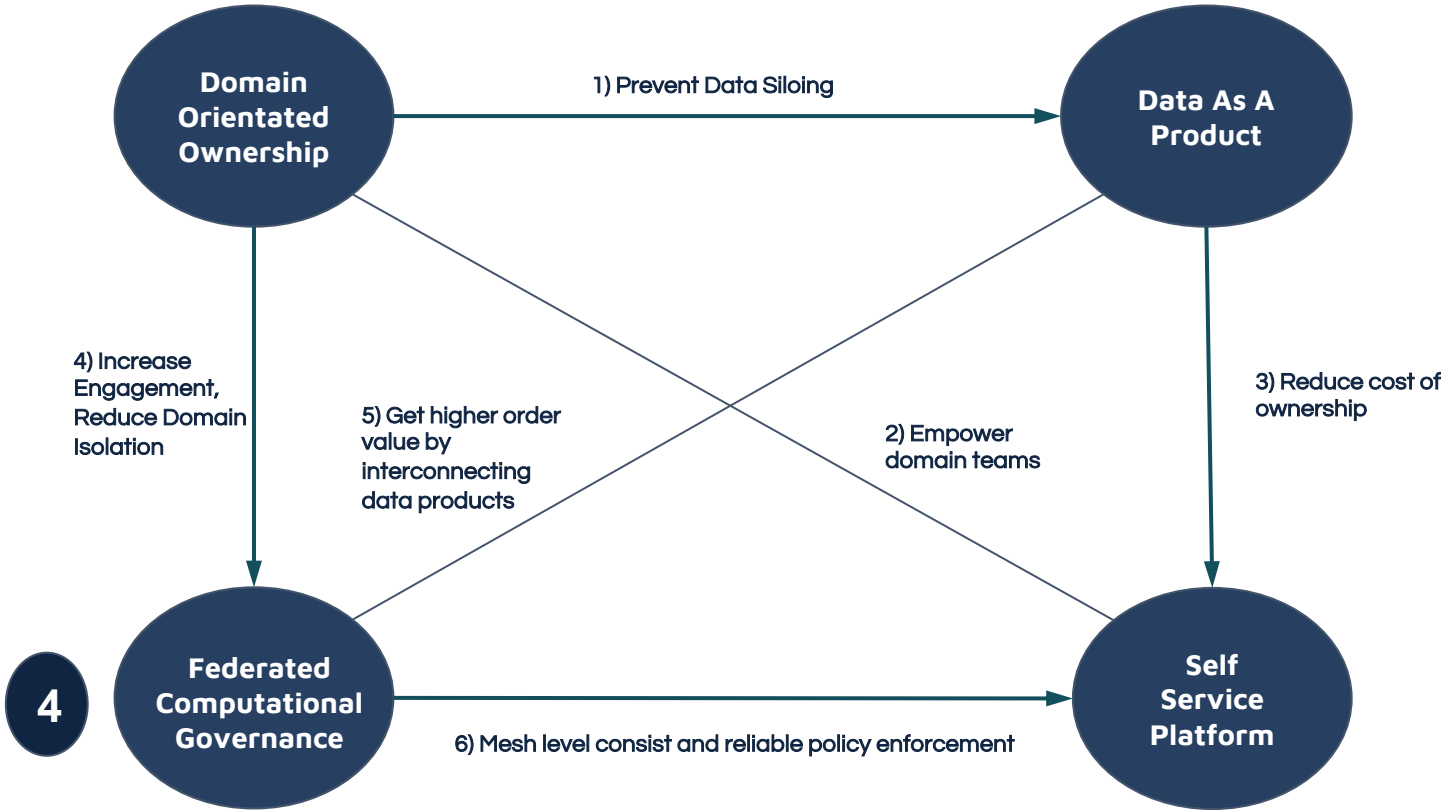
- ❑ Deal well with change and business complexity
- ❑ Remain agile especially when growing quickly
- ❑ Get better value from your data investment


1
Domain Orientated Ownership
1) Prevent Data Siloing
Data As A Product
2
4) Increase Engagement, Reduce Domain Isolation
5) Get higher order value by interconnecting data products
2) Empower domain teams
3) Reduce cost of ownership
4
Federated Computational Governance
6) Mesh level consist and reliable policy enforcement
Self Service Platform
3


1
Domain Orientated Ownership
1) Prevent Data Siloing
Data As A Product
4) Increase Engagement, Reduce Domain Isolation
5) Get higher order value by interconnecting data products
2) Empower domain teams
3) Reduce cost of ownership
Federated Computational Governance
Self Service Platform
6) Mesh level consist and reliable policy enforcement


Domain Orientated Ownership
1) Prevent Data Siloing
Data As A Product
2
4) Increase Engagement, Reduce Domain Isolation
5) Get higher order value by interconnecting data products
2) Empower domain teams
3) Reduce cost of ownership
Federated Computational Governance
6) Mesh level consist and reliable policy enforcement
Self Service Platform


Domain Orientated Ownership
1) Prevent Data Siloing
Data As A Product
4) Increase Engagement, Reduce Domain Isolation
5) Get higher order value by interconnecting data products
2) Empower domain teams
3) Reduce cost of ownership
Federated Computational Governance
6) Mesh level consist and reliable policy enforcement
Self Service Platform
3


Domain Orientated Ownership
1) Prevent Data Siloing
Data As A Product
4) Increase Engagement, Reduce Domain Isolation
5) Get higher order value by interconnecting data products
2) Empower domain teams
3) Reduce cost of ownership
4
Federated Computational Governance
6) Mesh level consist and reliable policy enforcement
Self Service Platform

2

Considerations On The Path To The Mesh

What does it take to move towards acquiring a half decent data mesh solution these days?

Especially when you feel you are starting with a data mess

1. Recognise the reality of your situation

2. Ask the right questions

3. Choose the right *approach* (and tools) for the job

Recognise the reality of your situation

You are here

But you want to be …

Data Mesh

You are here

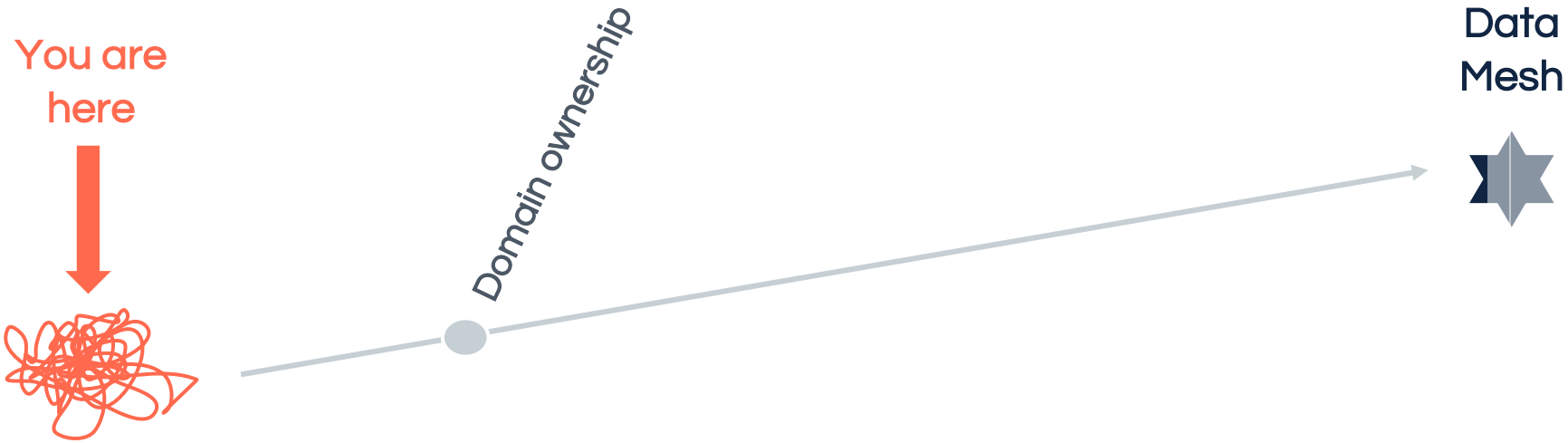
Data Mesh

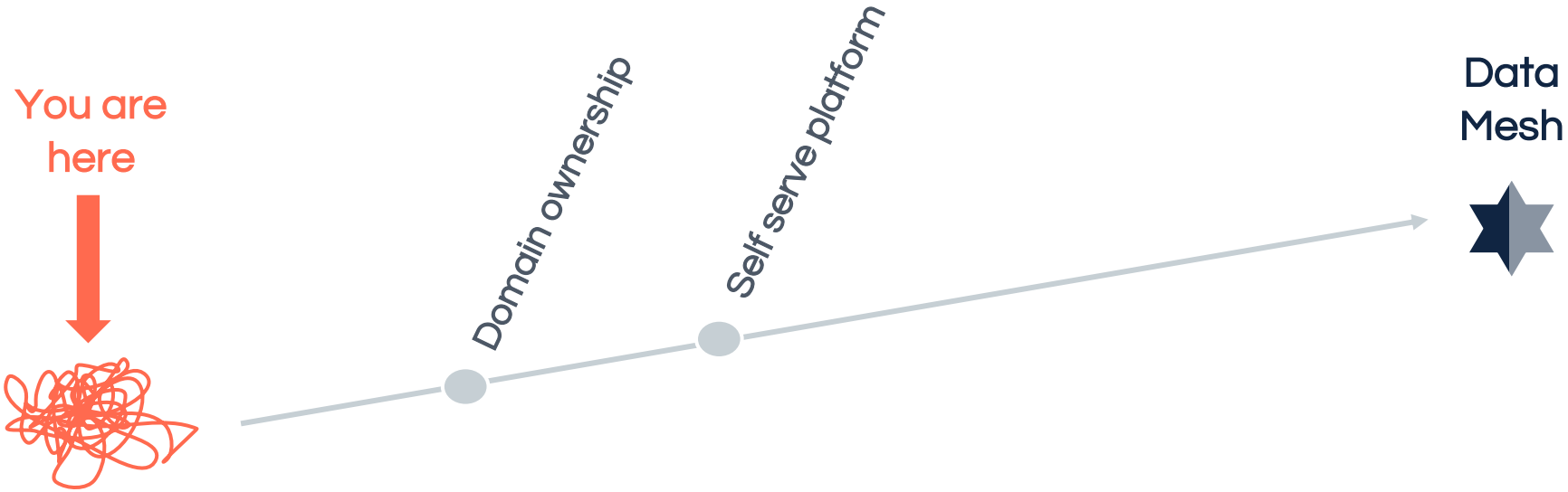
But you want to be …

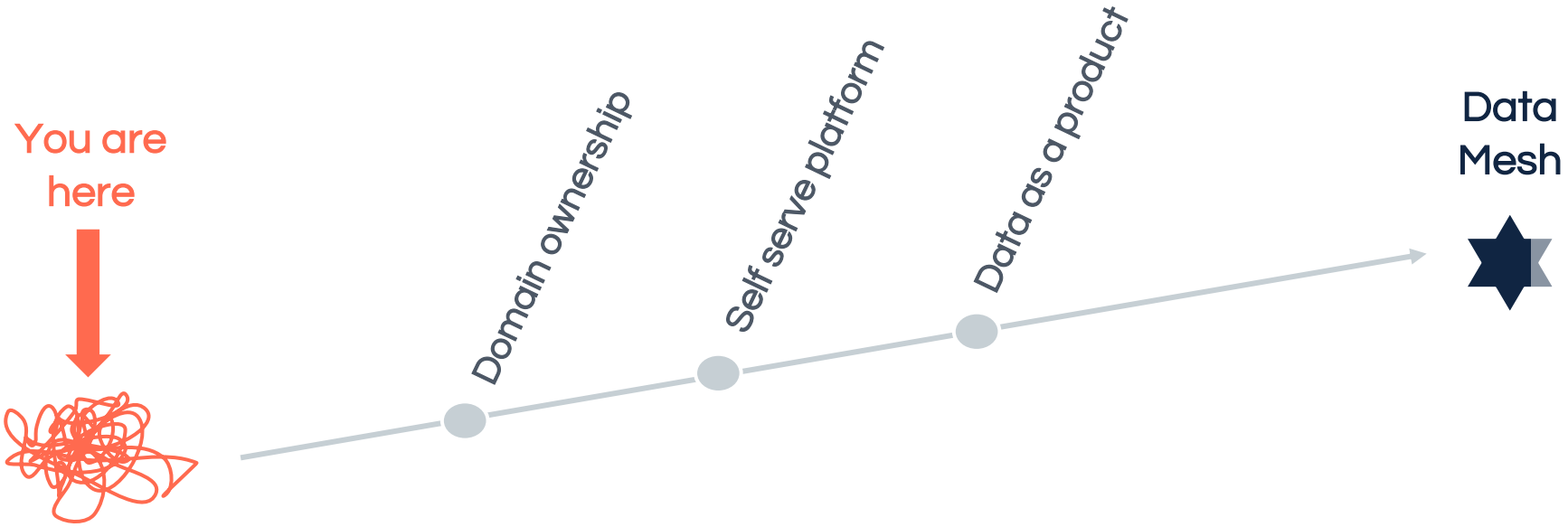
Reality Check #1
Existing life is not as good as it can be

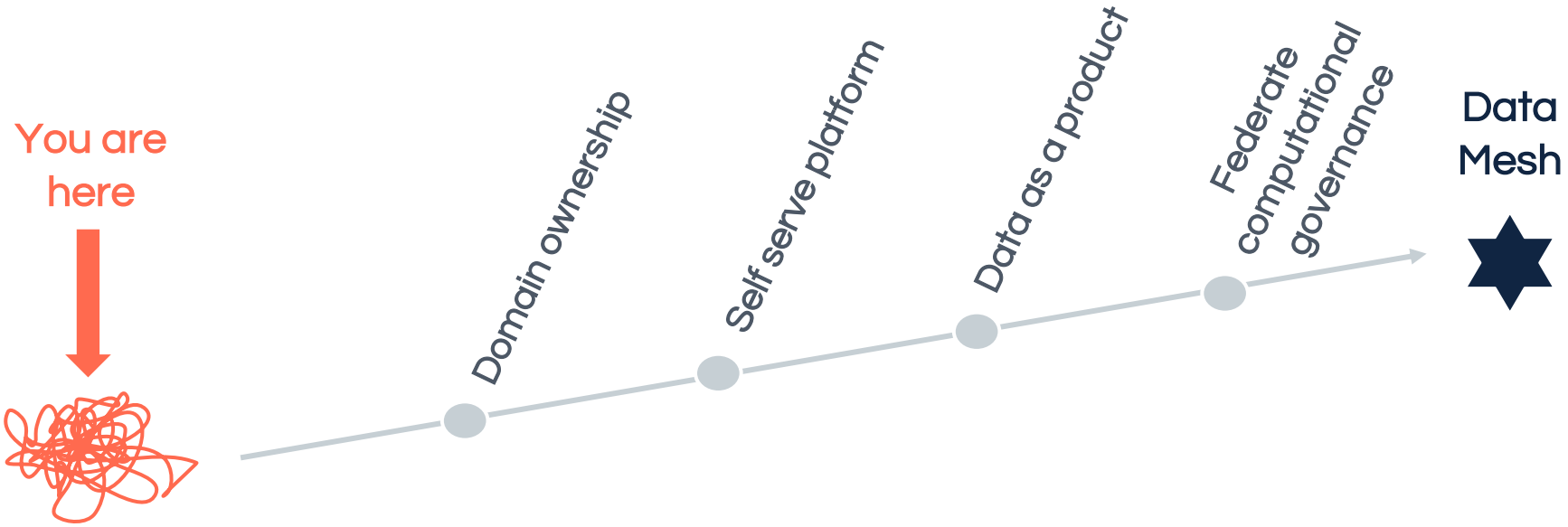
You are here

Data Mesh


You are here
Domain ownership
Data Mesh


You are here
Domain ownership
Self serve platform
Data Mesh


You are here
Domain ownership
Self serve platform
Data as a product
Data Mesh


You are here
Domain ownership
Self serve platform
Data as a product
Federate computational governance
Data Mesh

You are here

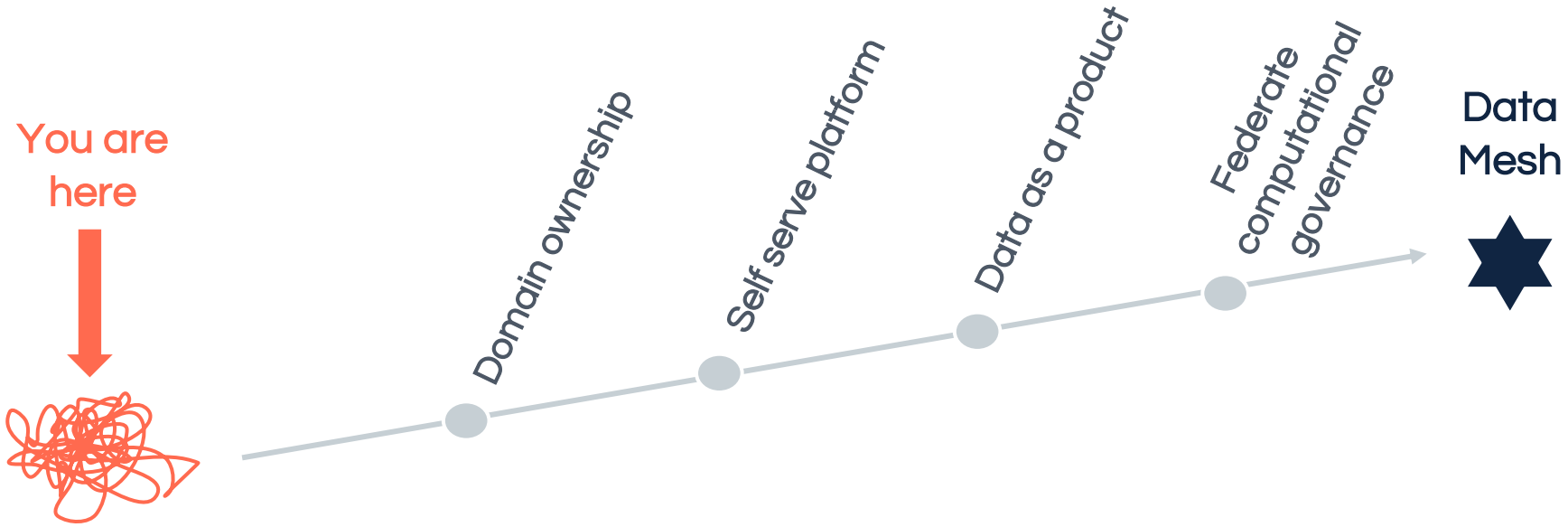
Domain ownership

Self serve platform

Data as a product

Federate computational governance

Data Mesh



Reality Check #2
Moving to data mesh is not a simple n-step process that fits nice and neatly on a project gantt chart

You are here

Domain ownership

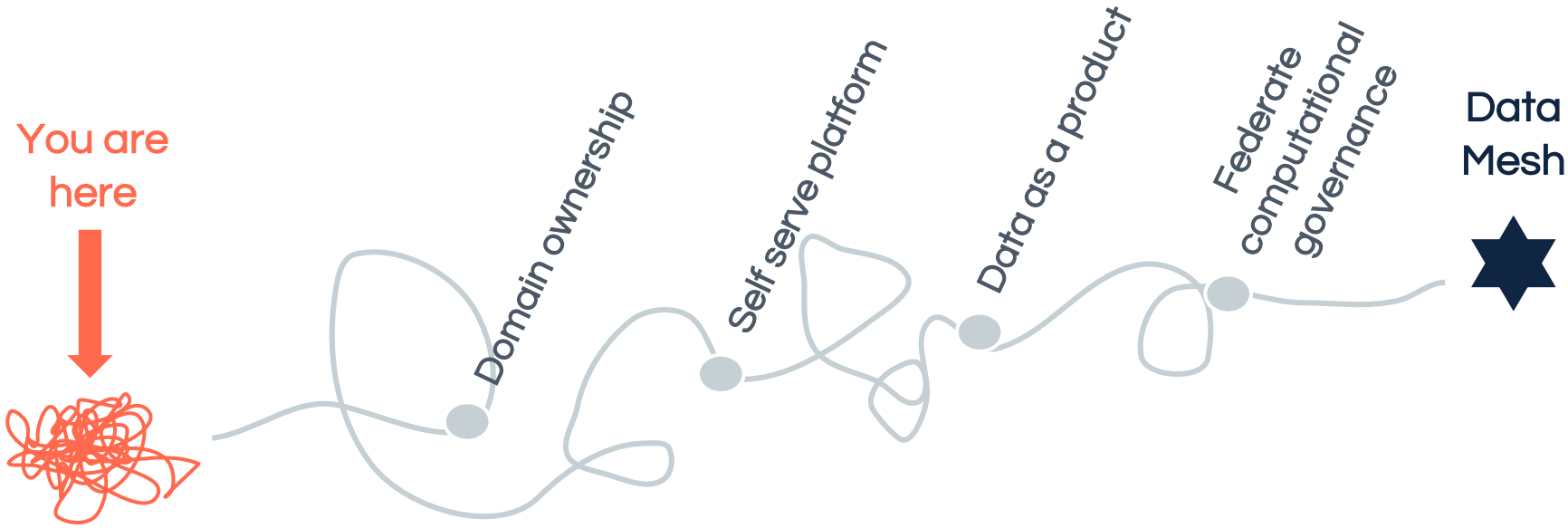
Self serve platform

Data as a product

Federate computational governance

Data Mesh

OK …
It’s going to be a journey !!!



OK …
It’s going to be a journey !!!

You are here

Domain ownership

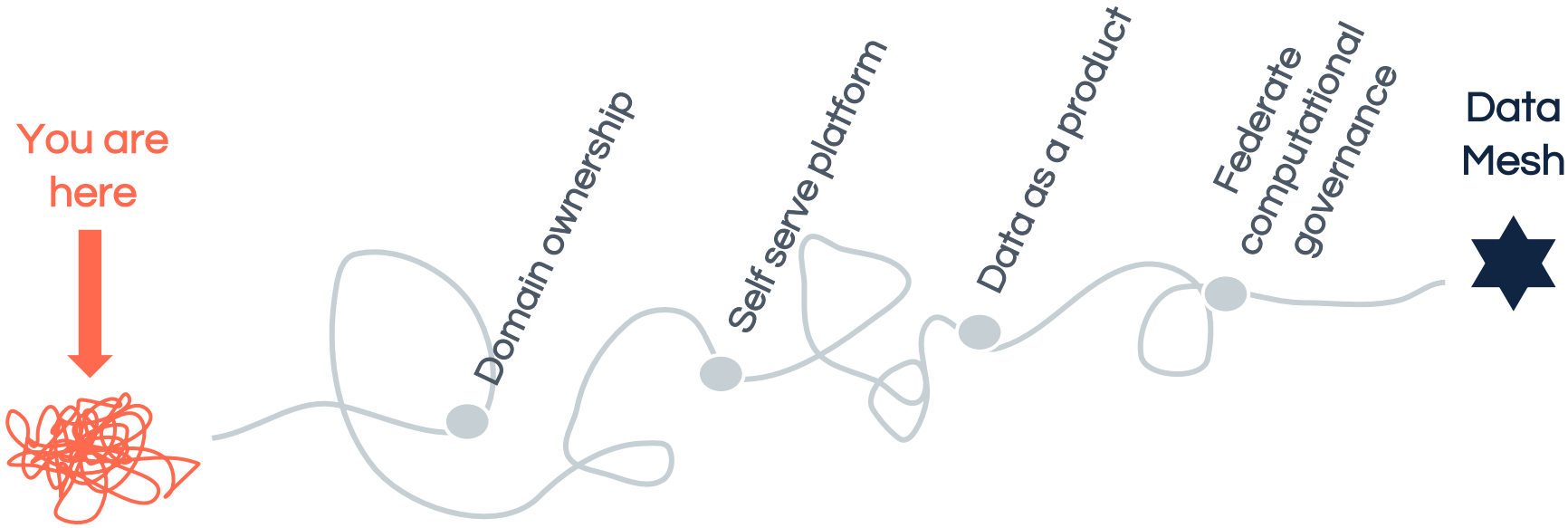
Self serve platform

Data as a product

Federate computational governance

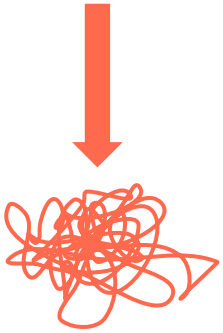
Data Mesh

OK …
It’s going to be a journey !!!



Reality Check #3
Moving to data mesh is not a journey along a single axis

You are here



Domain ownership

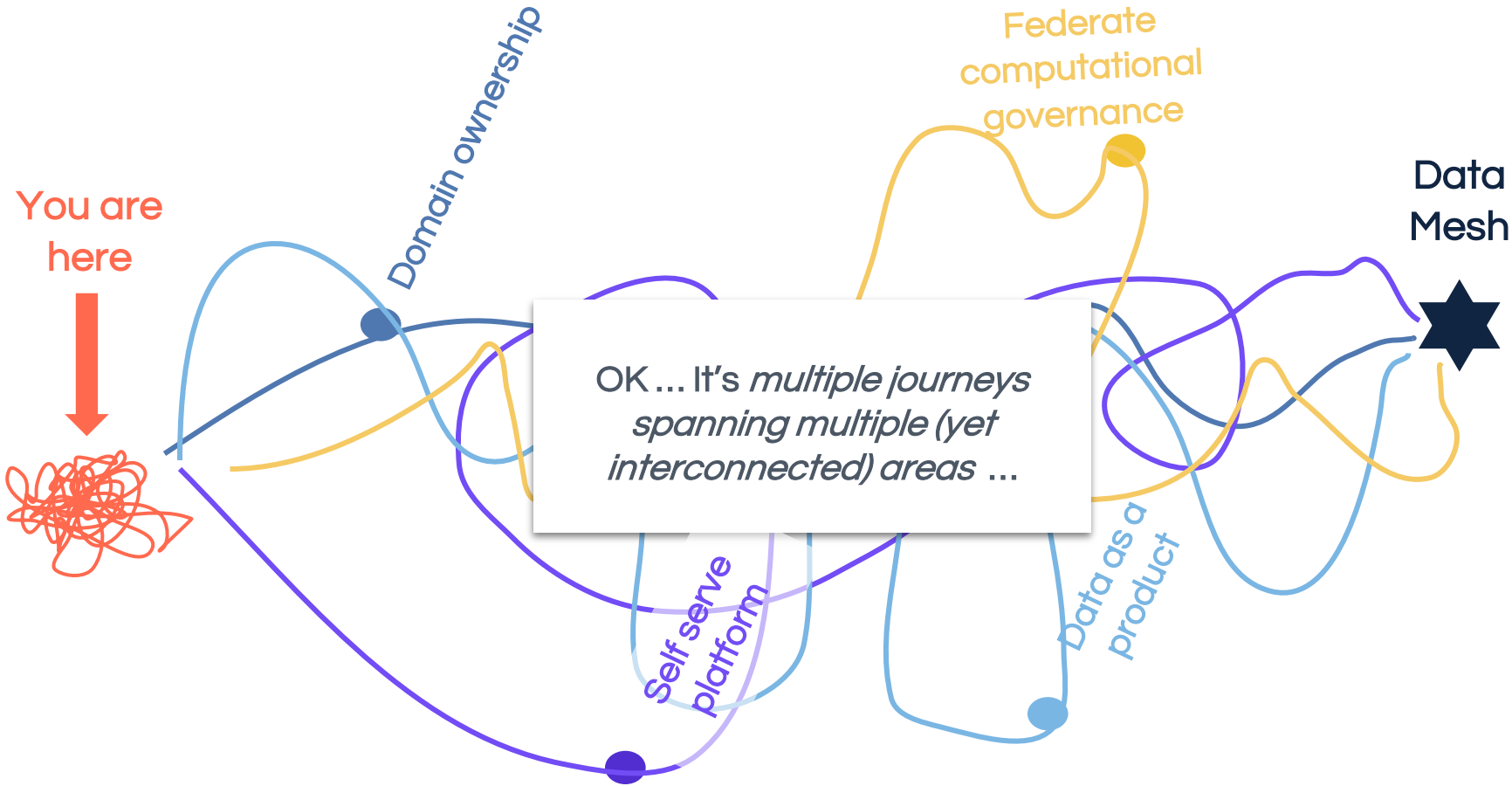
Federate computational governance

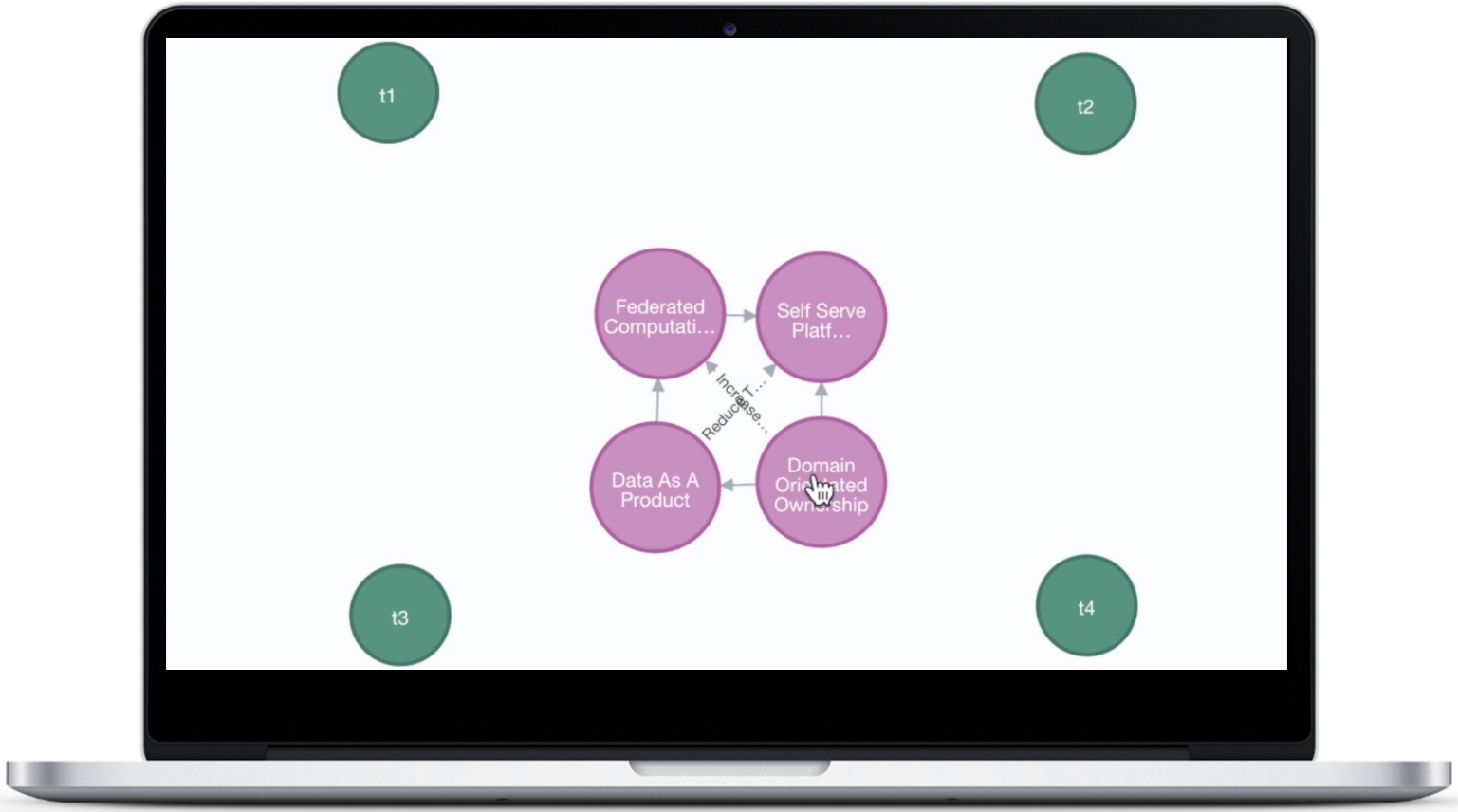
Data Mesh

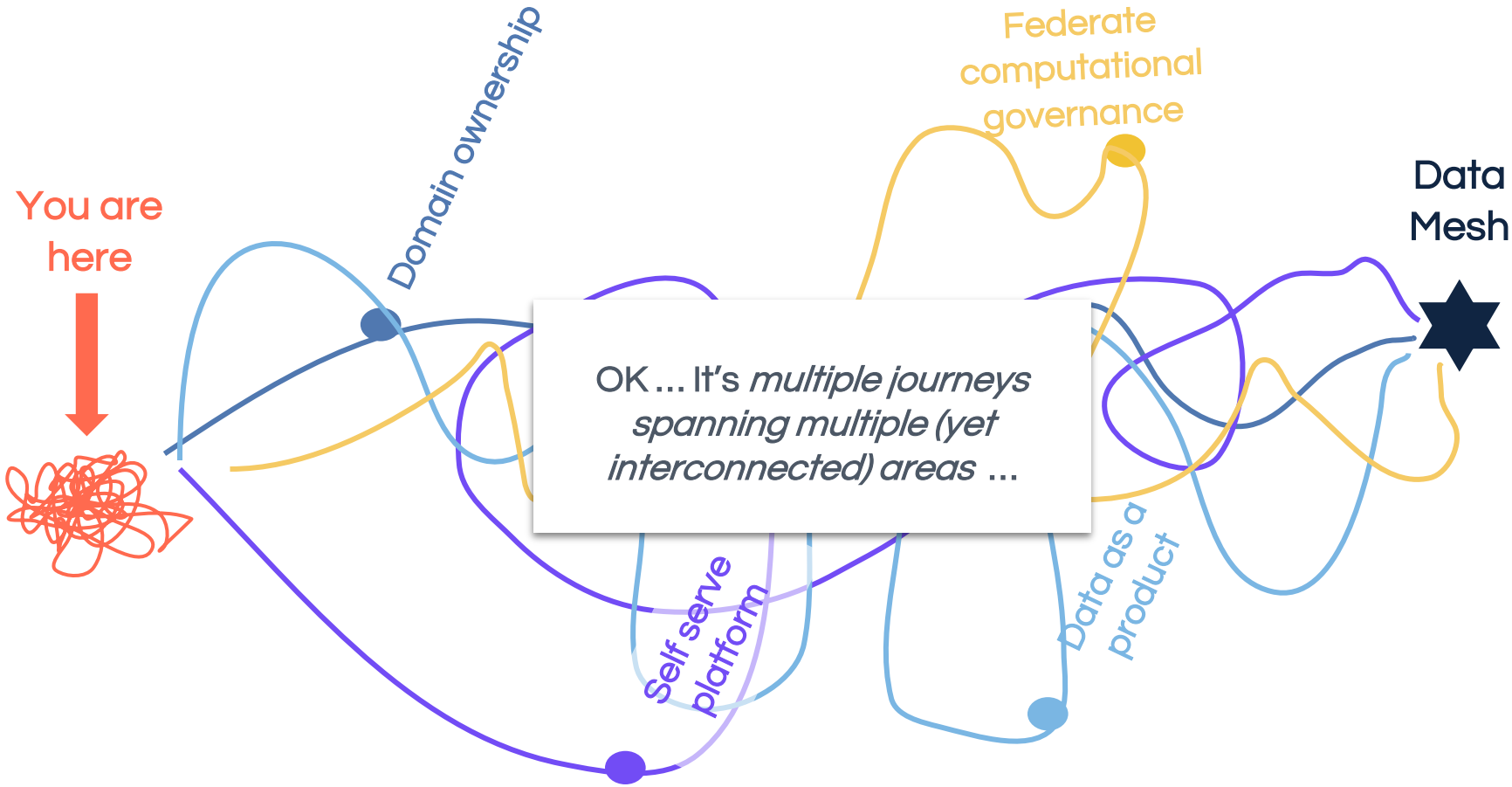
OK … It's *multiple journeys spanning multiple (yet interconnected) areas* …

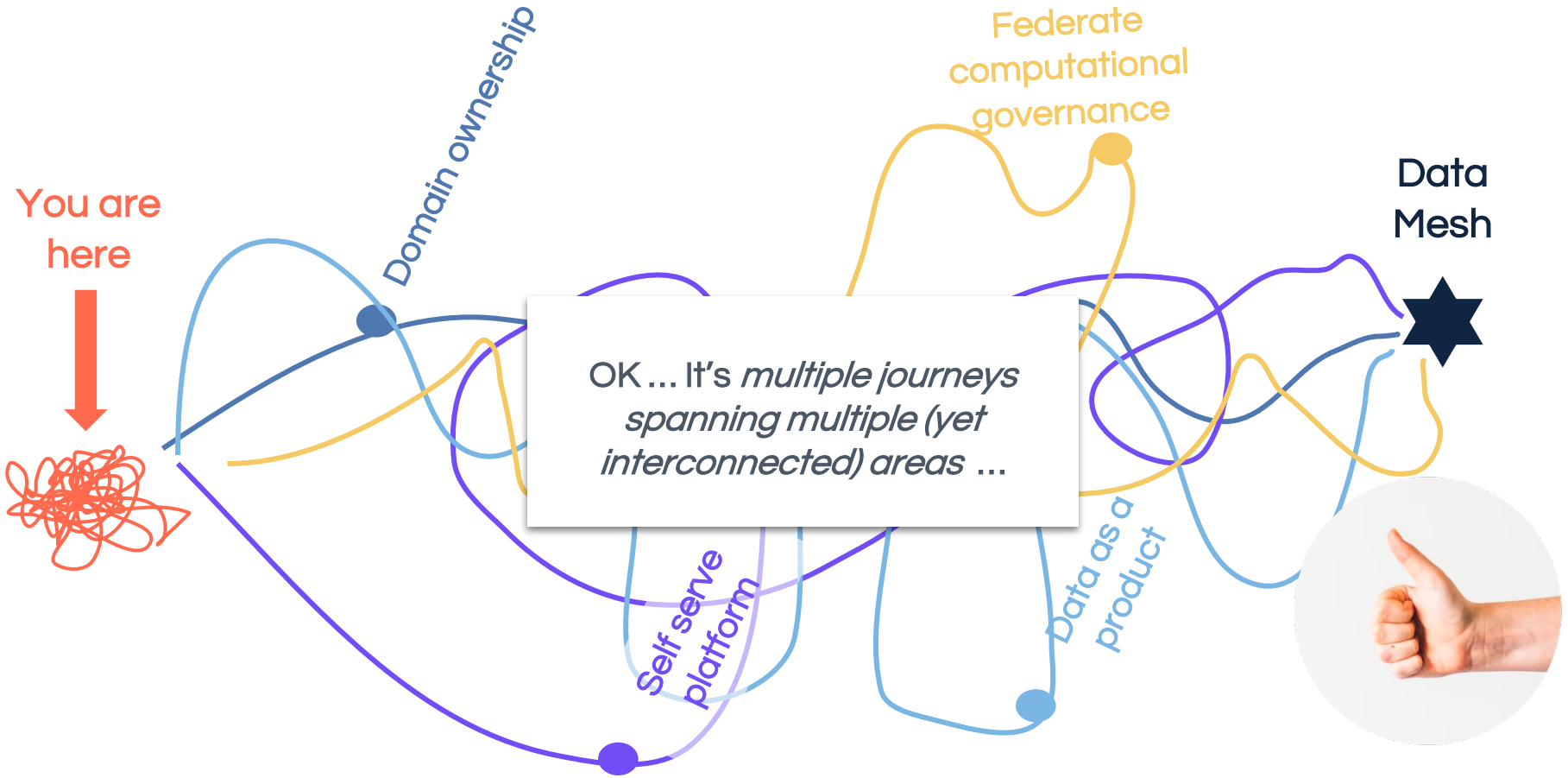
Self serve platform

Data as a product

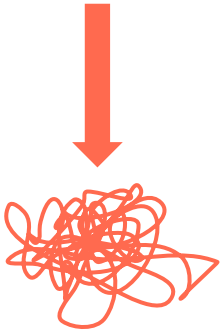

You are here
Domain ownership
Federate computational governance
Data Mesh
OK … It's *multiple journeys spanning multiple (yet interconnected) areas* …
Self serve platform
Data as a product


t1
t2
Federated Computati…
Self Serve Platf…
Increase T…
Reduce…
Data As A Product
Domain Oriented Ownership
t3
t4


You are here
Domain ownership
Federate computational governance
Data Mesh
OK … It's *multiple journeys spanning multiple (yet interconnected) areas* …
Self serve platform
Data as a product


You are here
Domain ownership
Federate computational governance
Data Mesh
OK … It's *multiple journeys spanning multiple (yet interconnected) areas* …
Self serve platform
Data as a product

You are here

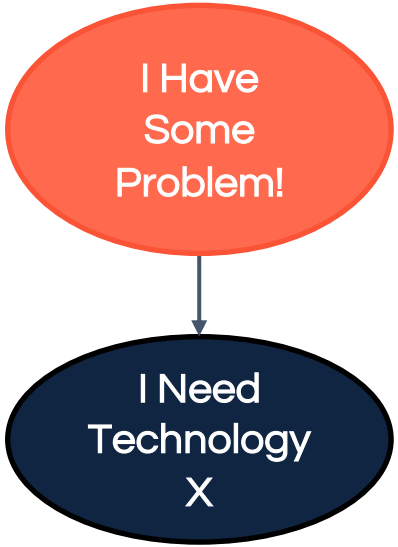


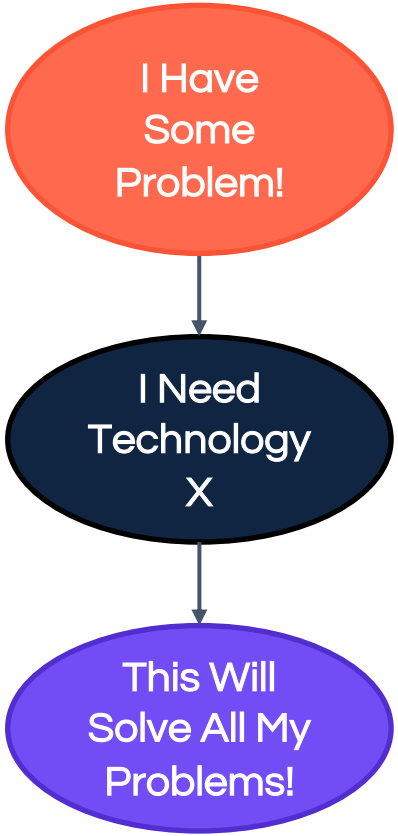
Assumes you know that the answer is “Data Mesh”

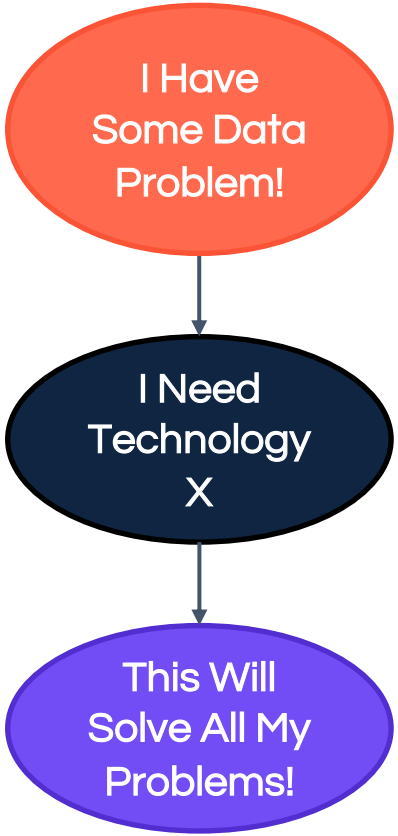
Data Mesh

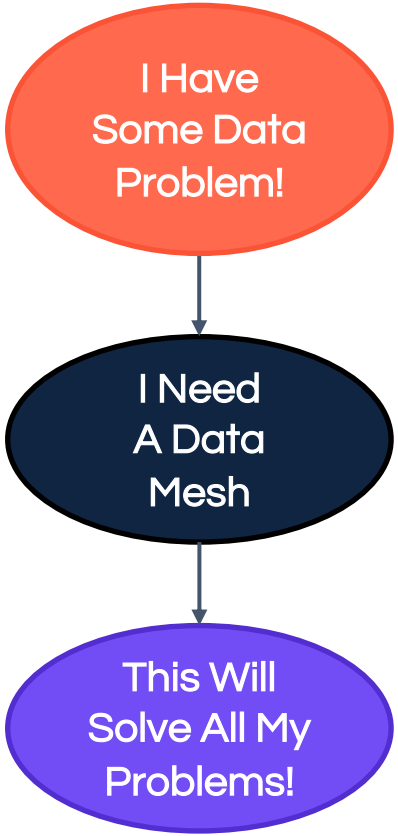
But you want to be …

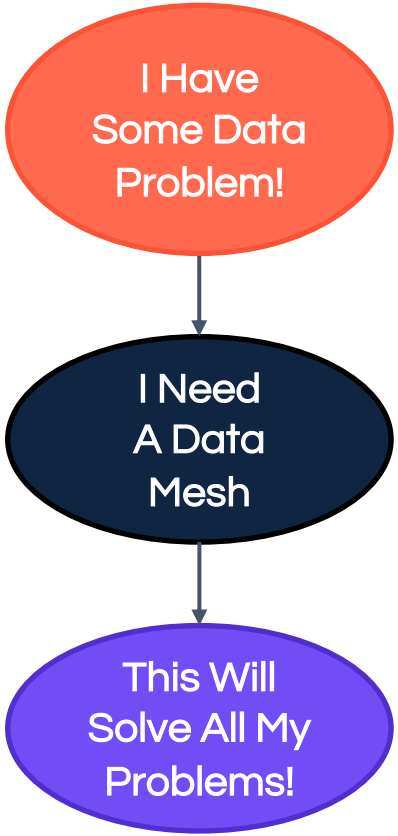
I Have
Some
Problem!


I Have Some Problem!
I Need Technology X

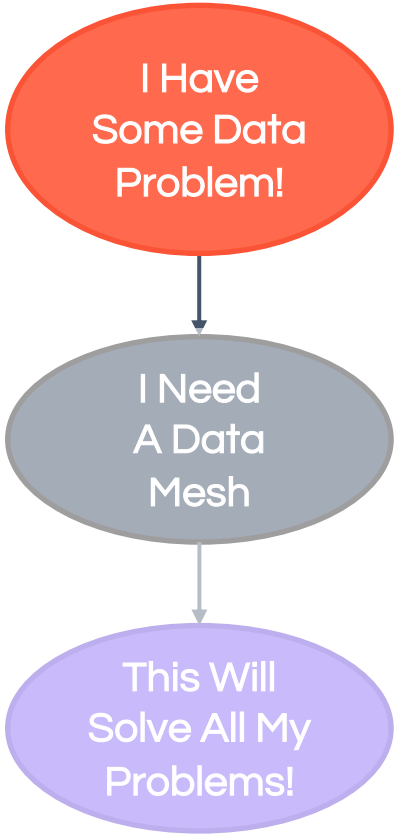

I Have Some Problem!
I Need Technology X
This Will Solve All My Problems!


I Have Some Data Problem!
I Need Technology X
This Will Solve All My Problems!

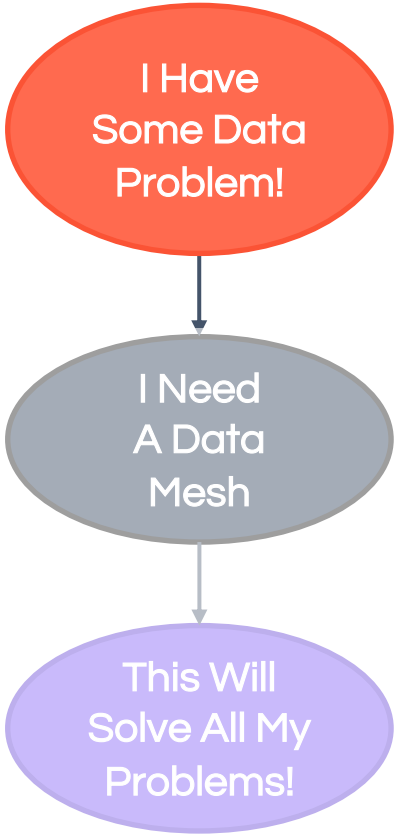

I Have
Some Data
Problem!
I Need
A Data
Mesh
This Will
Solve All My
Problems!

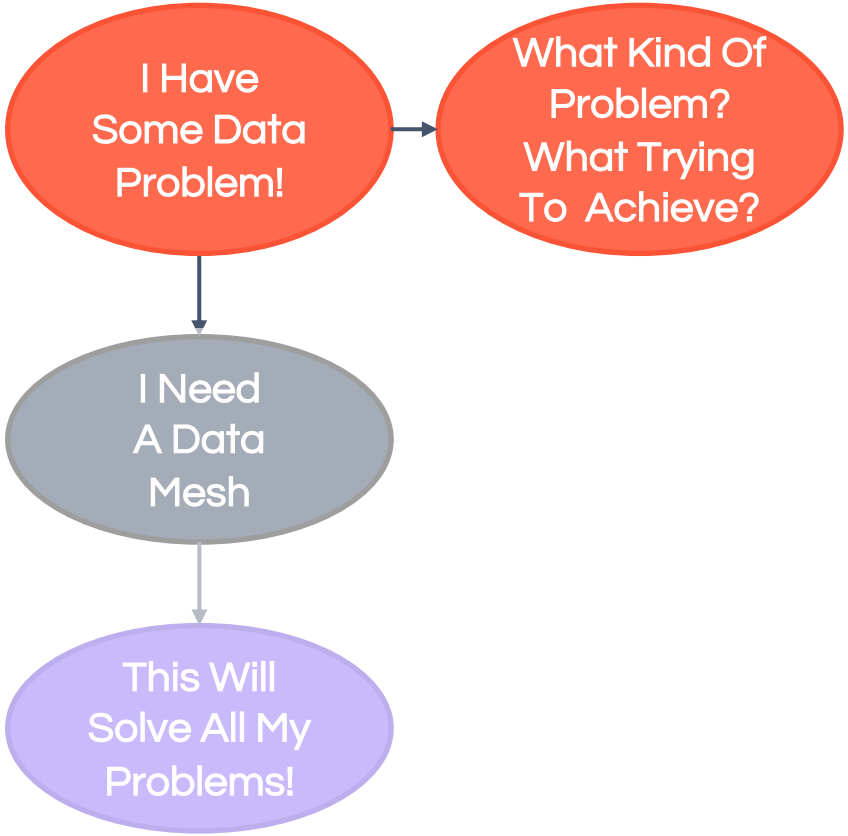


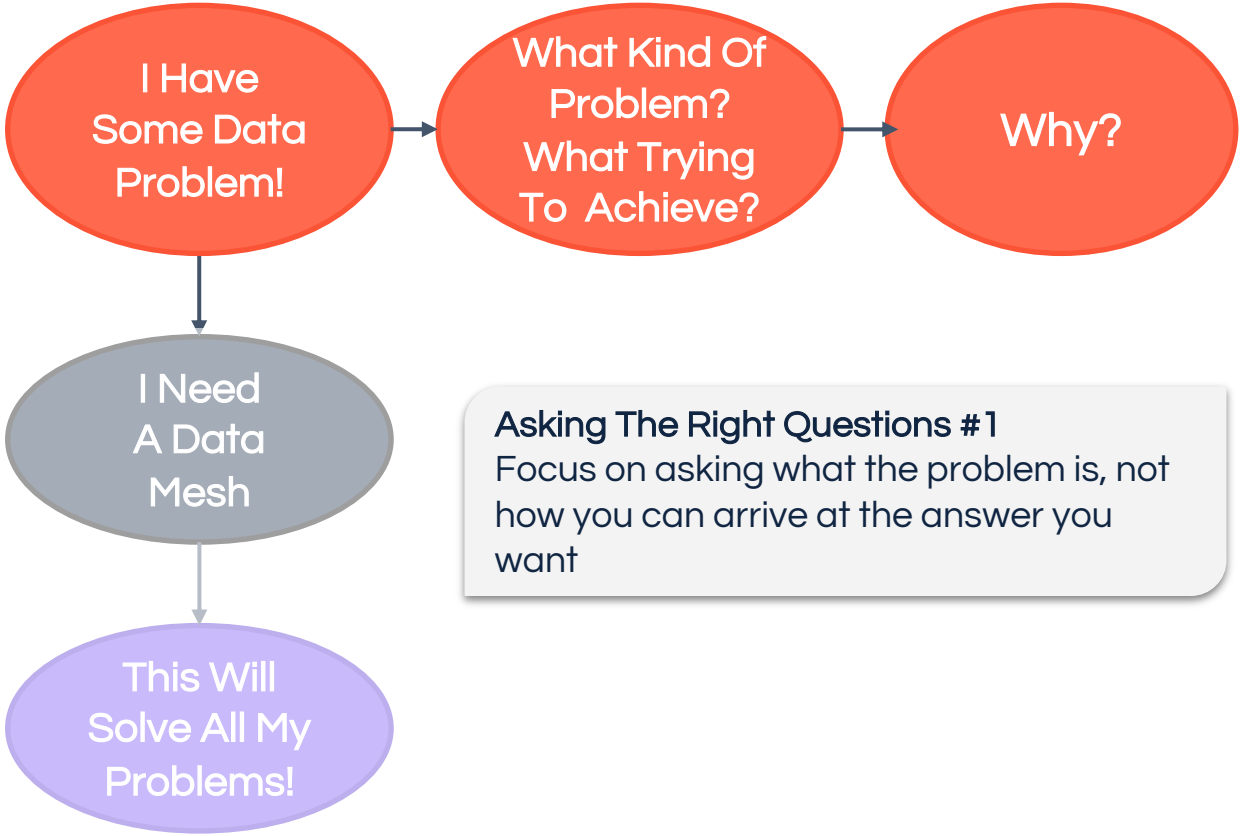
Reality Check #4
Data Mesh is not a piece of tech / platform you can buy & install


I Have
Some Data
Problem!
I Need
A Data
Mesh
This Will
Solve All My
Problems!

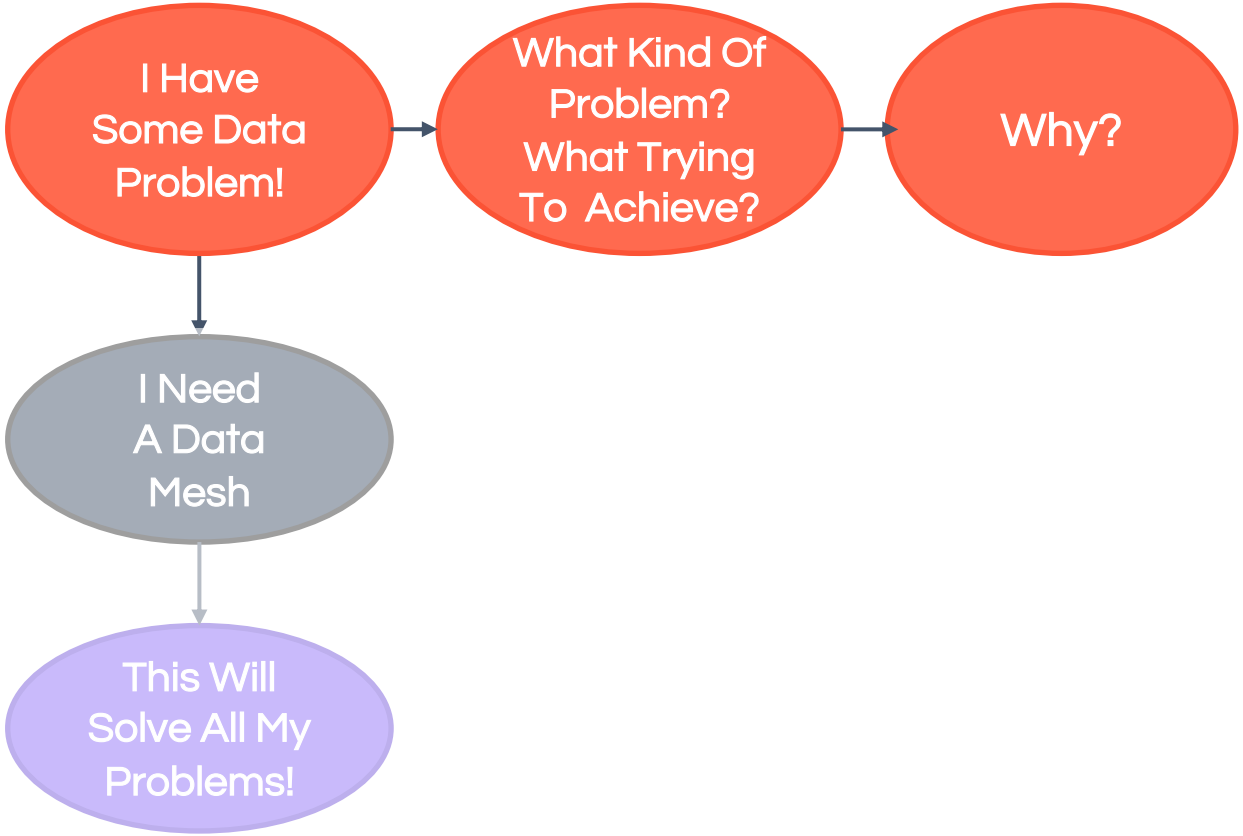
Ask the right questions

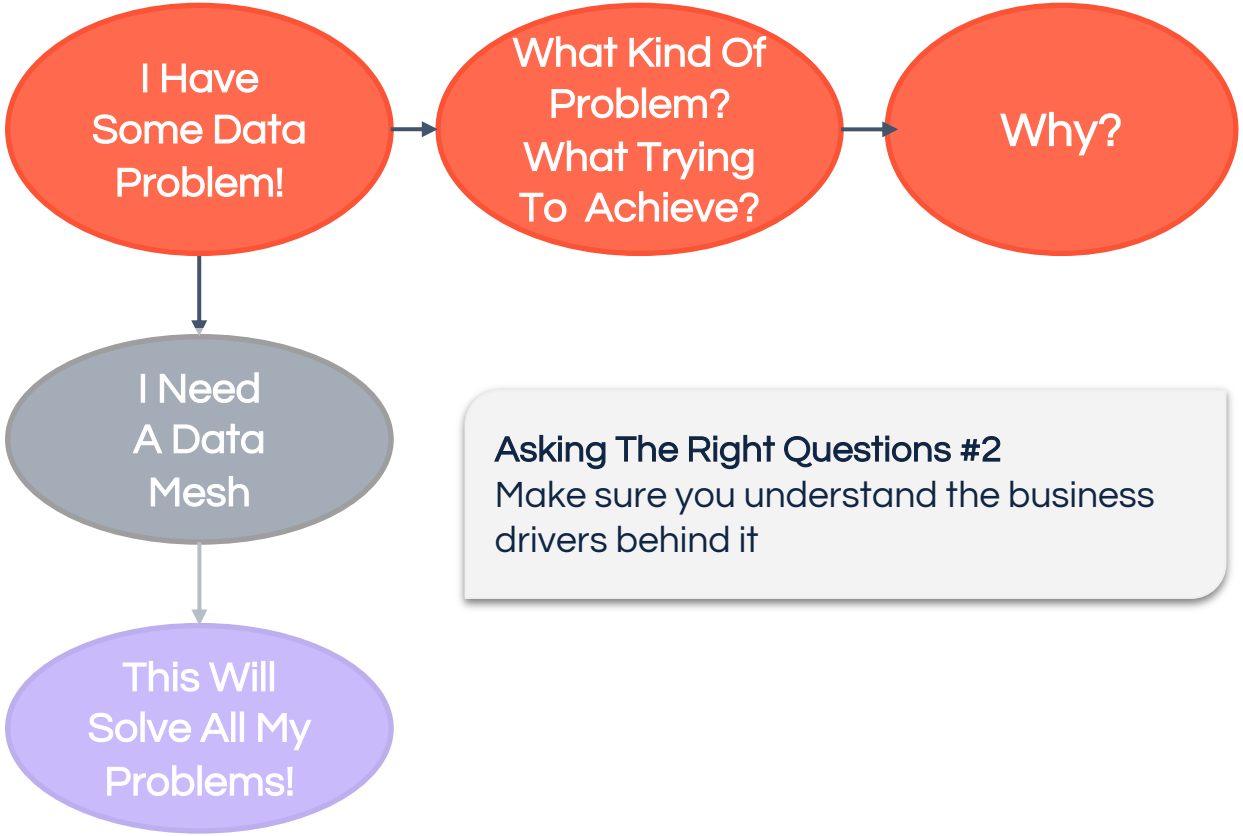

I Have
Some Data
Problem!
I Need
A Data
Mesh
This Will
Solve All My
Problems!


I Have Some Data Problem!
What Kind Of Problem? What Trying To Achieve?
I Need A Data Mesh
This Will Solve All My Problems!

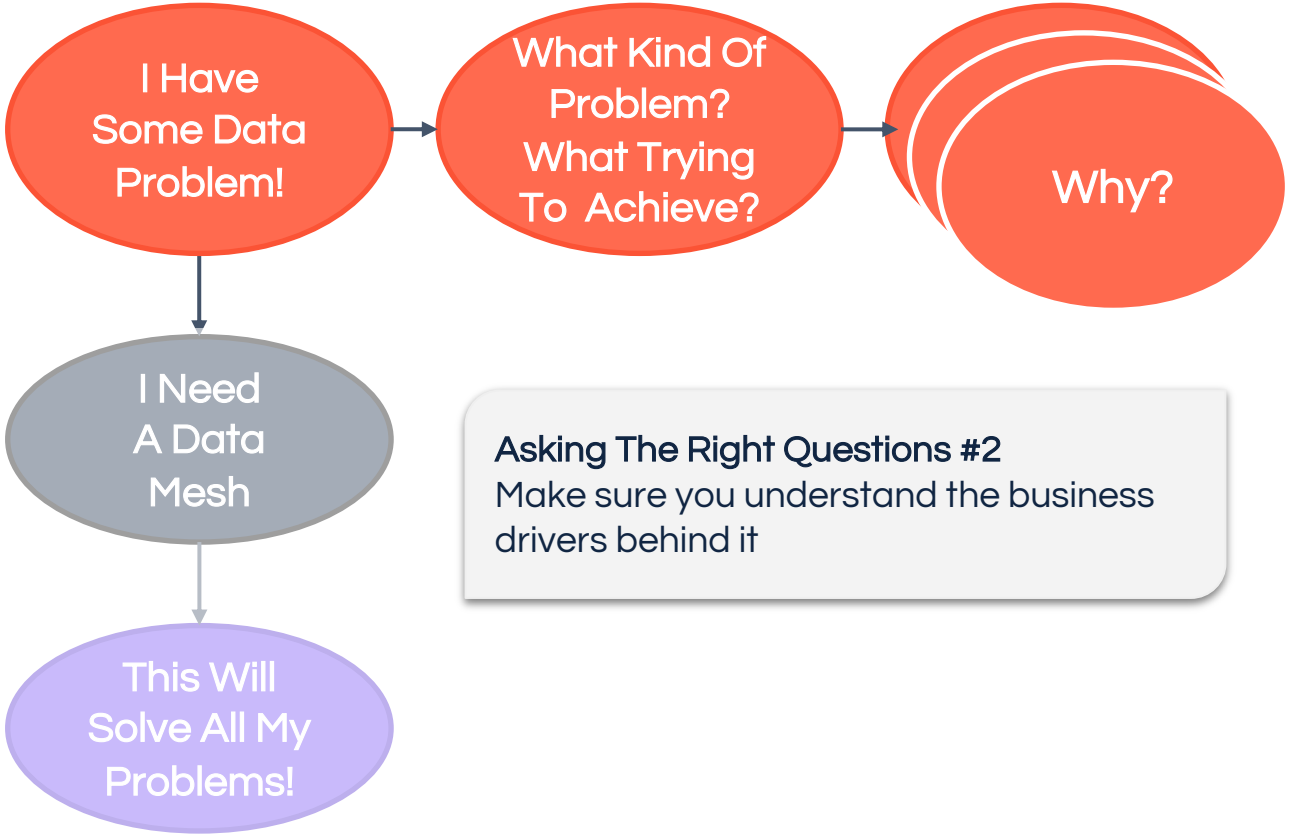


Asking The Right Questions #1
Focus on asking what the problem is, not how you can arrive at the answer you want

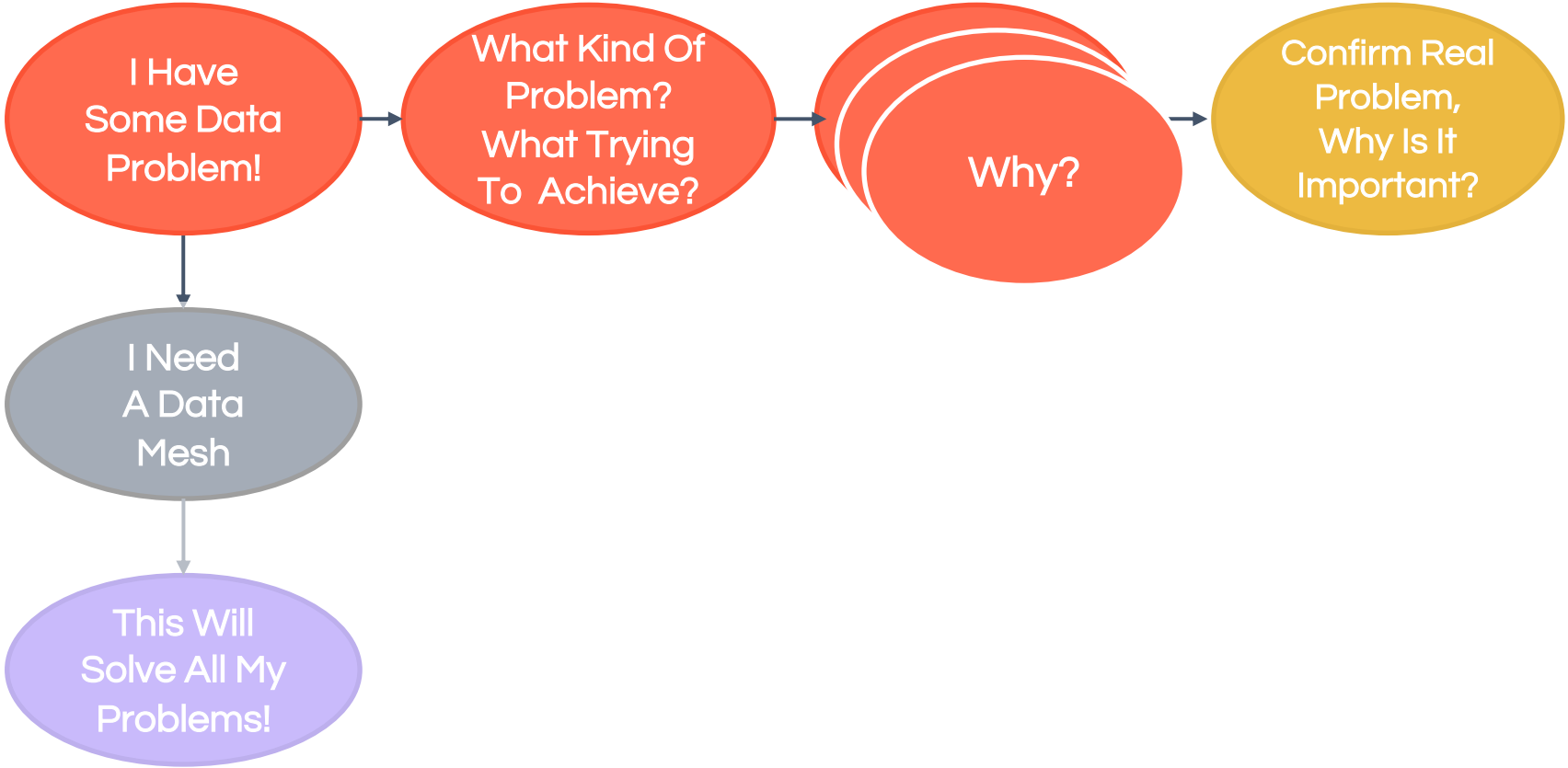

I Have Some Data Problem!
What Kind Of Problem? What Trying To Achieve?
Why?
I Need A Data Mesh
This Will Solve All My Problems!

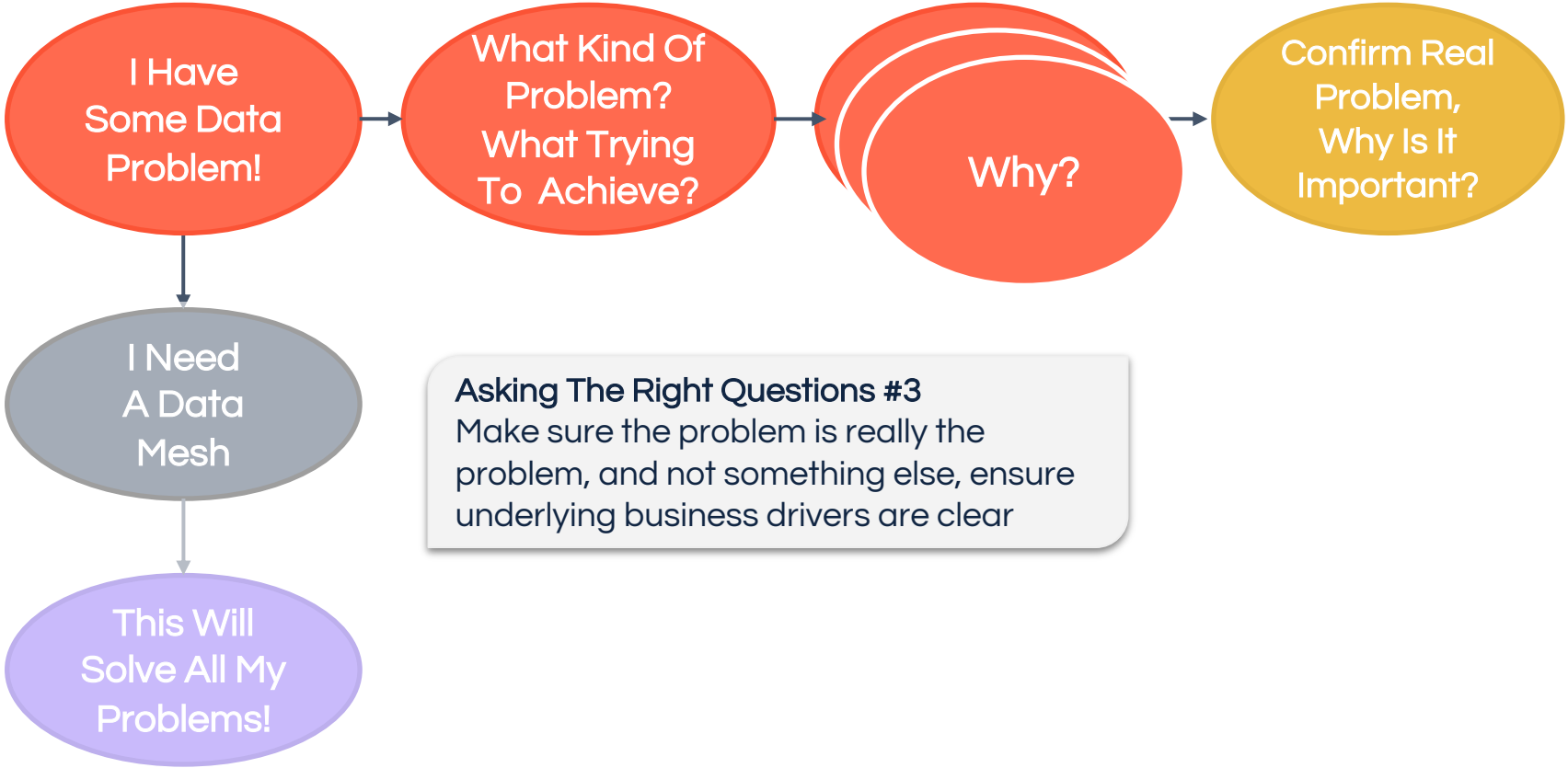


Asking The Right Questions #2
Make sure you understand the business drivers behind it



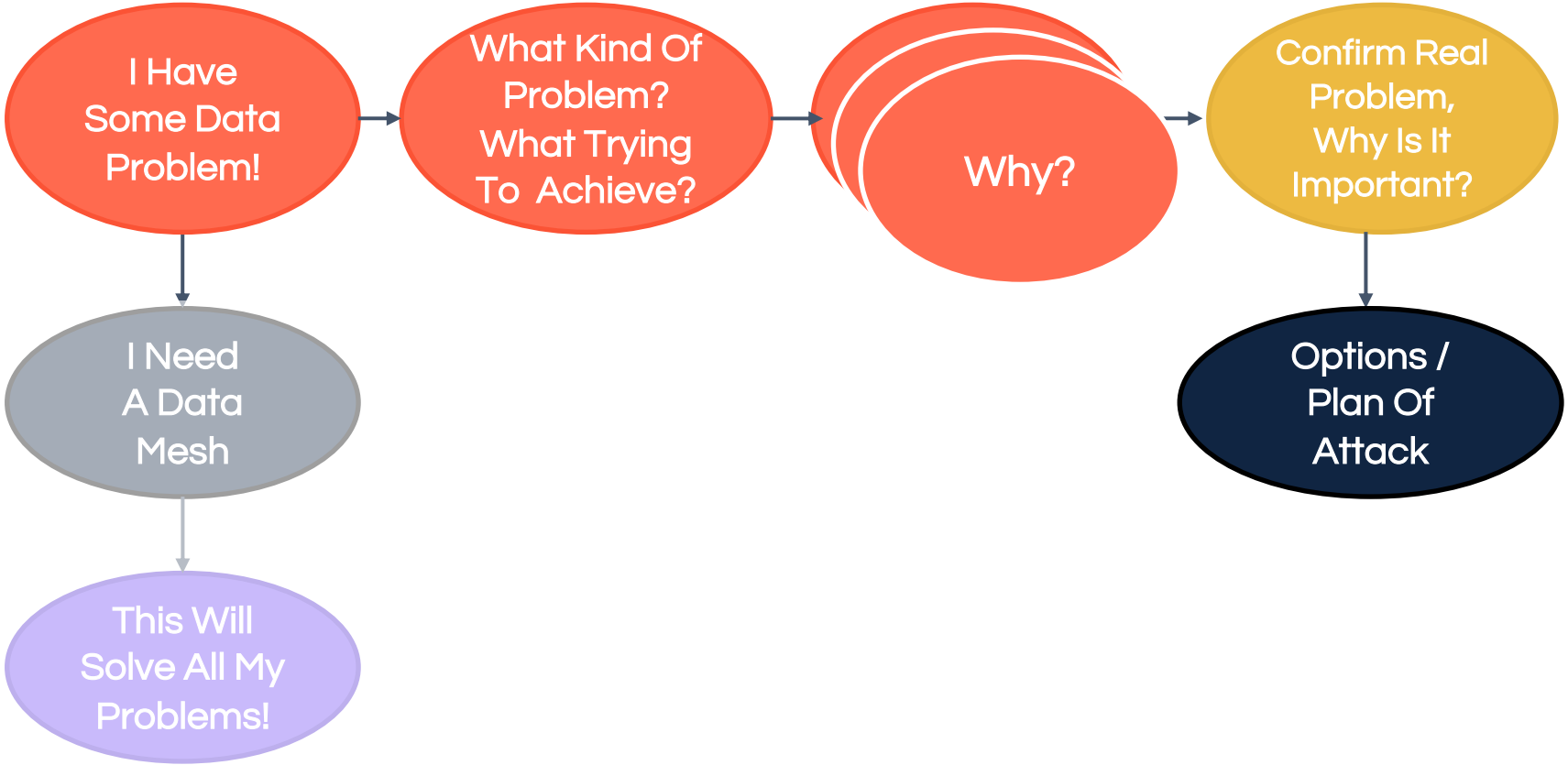
Asking The Right Questions #2
Make sure you understand the business drivers behind it


I Have Some Data Problem!
What Kind Of Problem? What Trying To Achieve?
Why?
Why?
Why?
Confirm Real Problem, Why Is It Important?
I Need A Data Mesh
This Will Solve All My Problems!

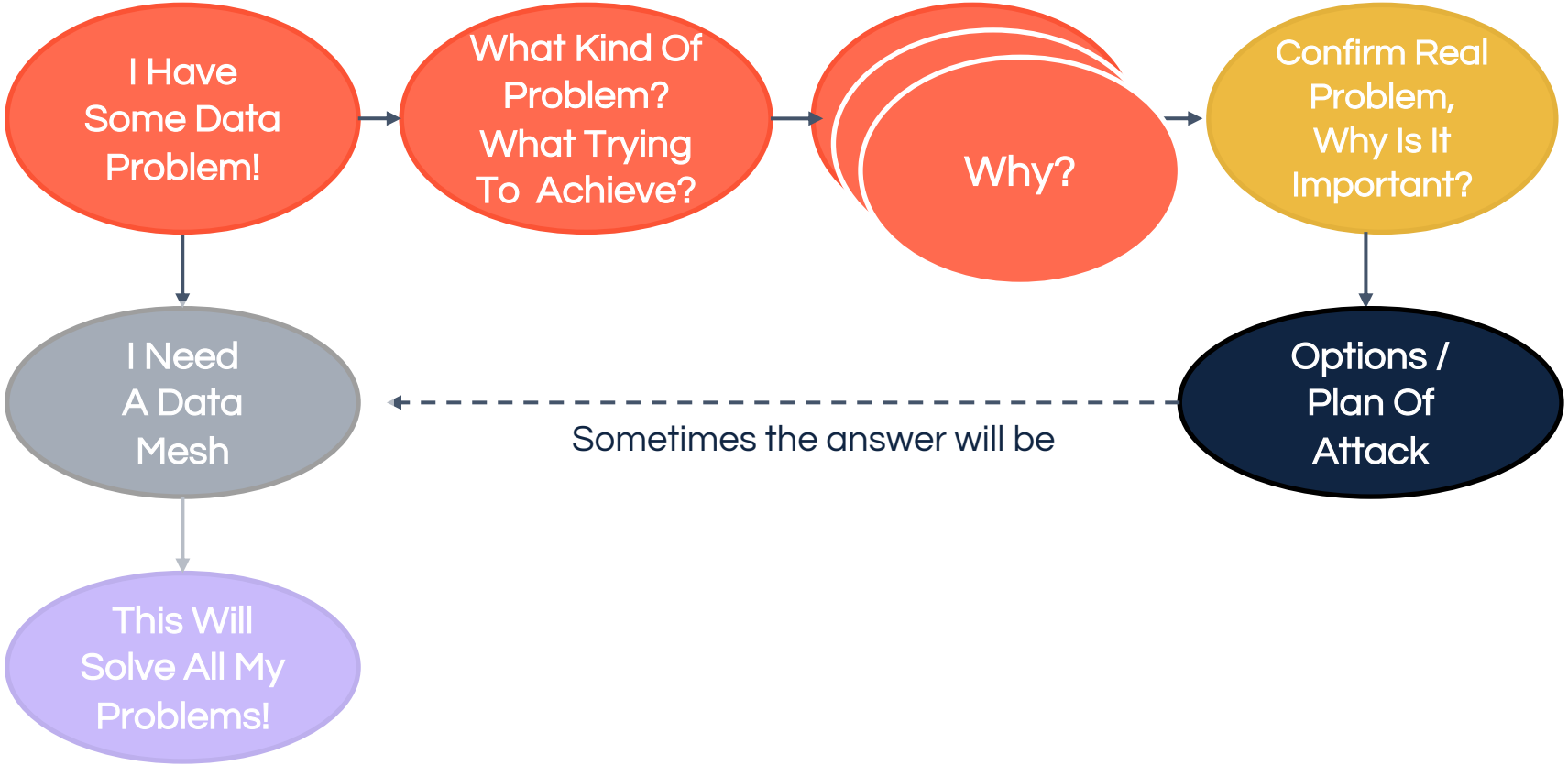


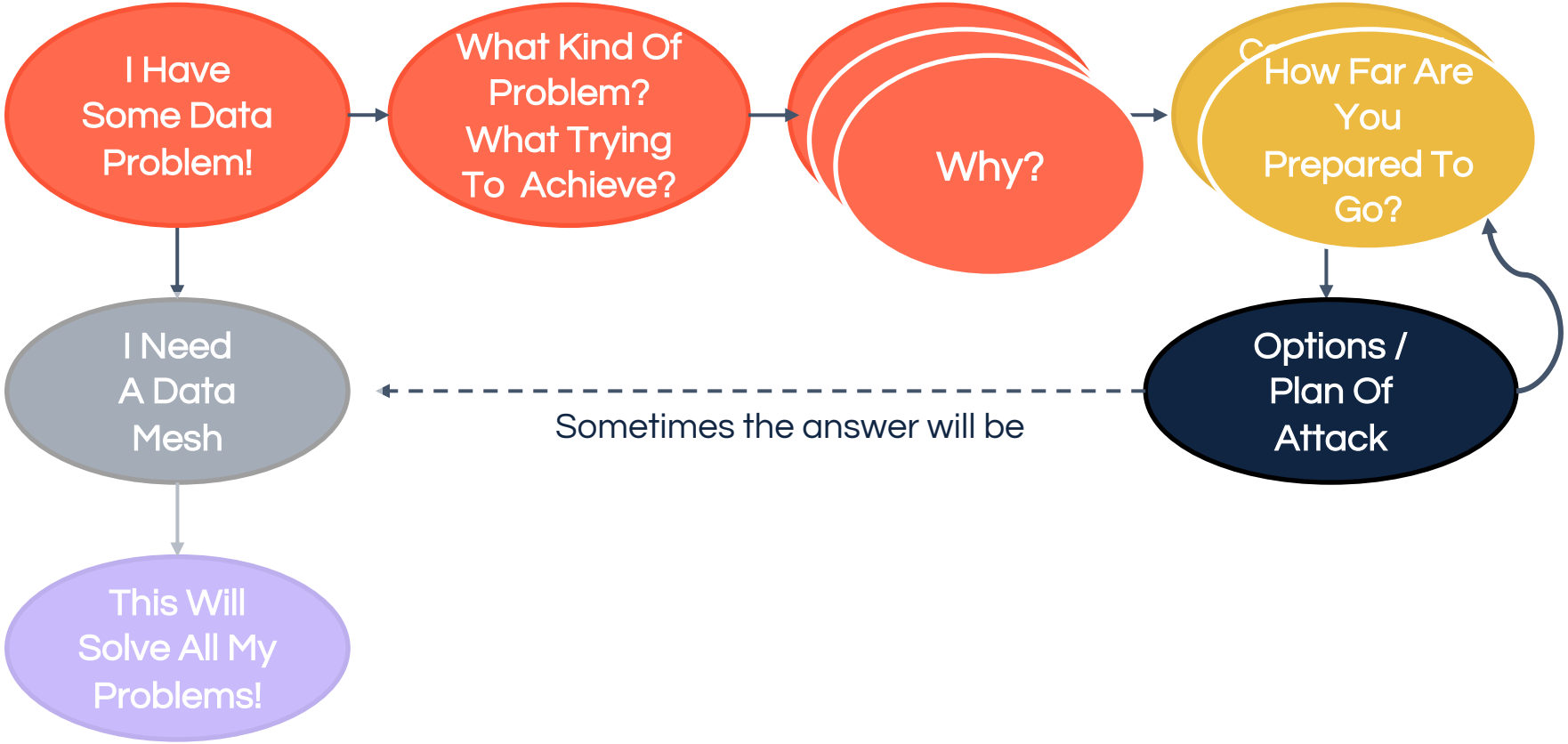
Asking The Right Questions #3

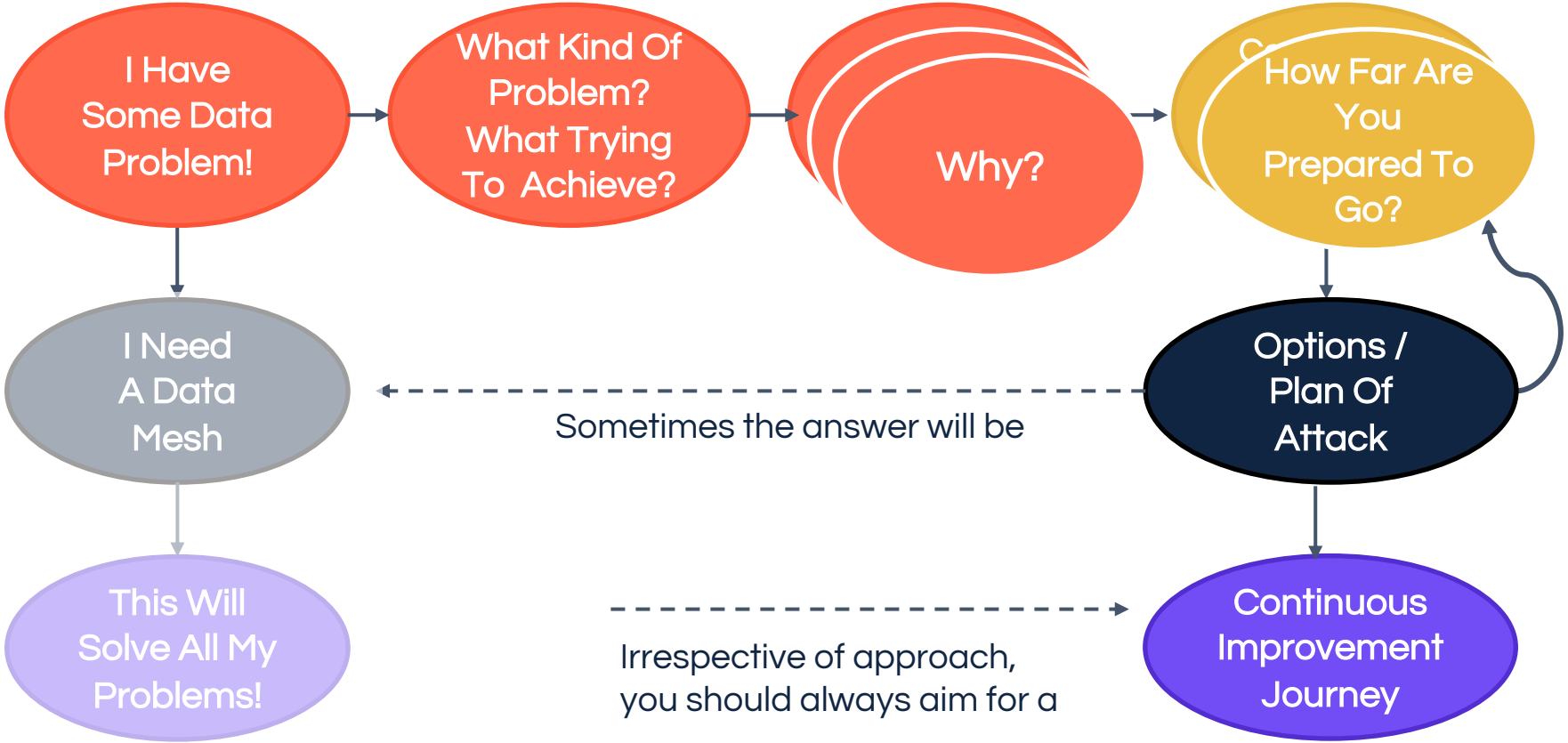
Make sure the problem is really the problem, and not something else, ensure underlying business drivers are clear

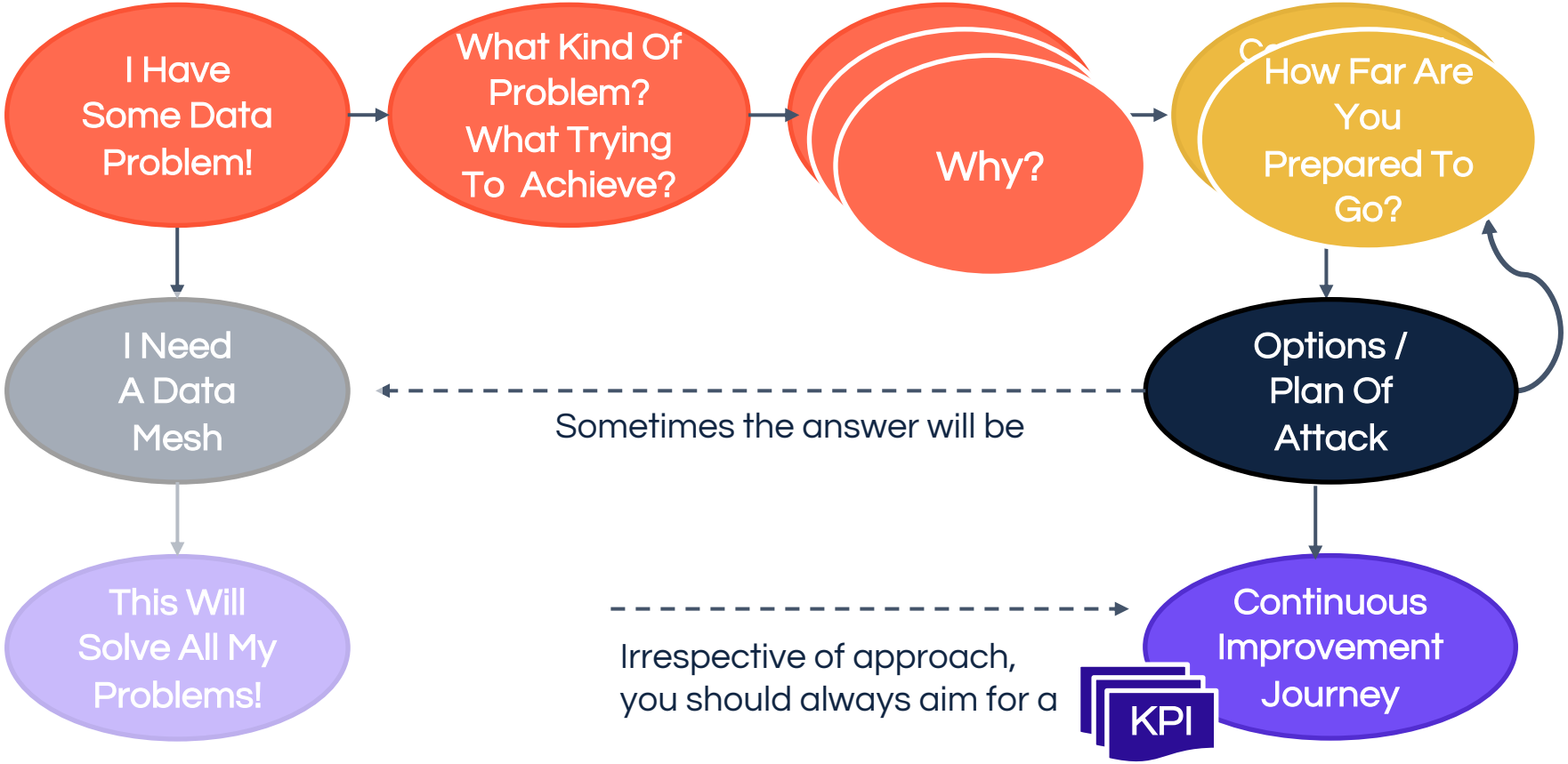

I Have Some Data Problem!
What Kind Of Problem? What Trying To Achieve?
Why?
Confirm Real Problem, Why Is It Important?
I Need A Data Mesh
Options / Plan Of Attack
This Will Solve All My Problems!

Choose the right approach for the job


I Have Some Data Problem!
What Kind Of Problem? What Trying To Achieve?
Why?
Confirm Real Problem, Why Is It Important?
I Need A Data Mesh
Options / Plan Of Attack
Sometimes the answer will be
This Will Solve All My Problems!


I Have Some Data Problem!
What Kind Of Problem? What Trying To Achieve?
Why?
How Far Are You Prepared To Go?
I Need A Data Mesh
Options / Plan Of Attack
Sometimes the answer will be
This Will Solve All My Problems!

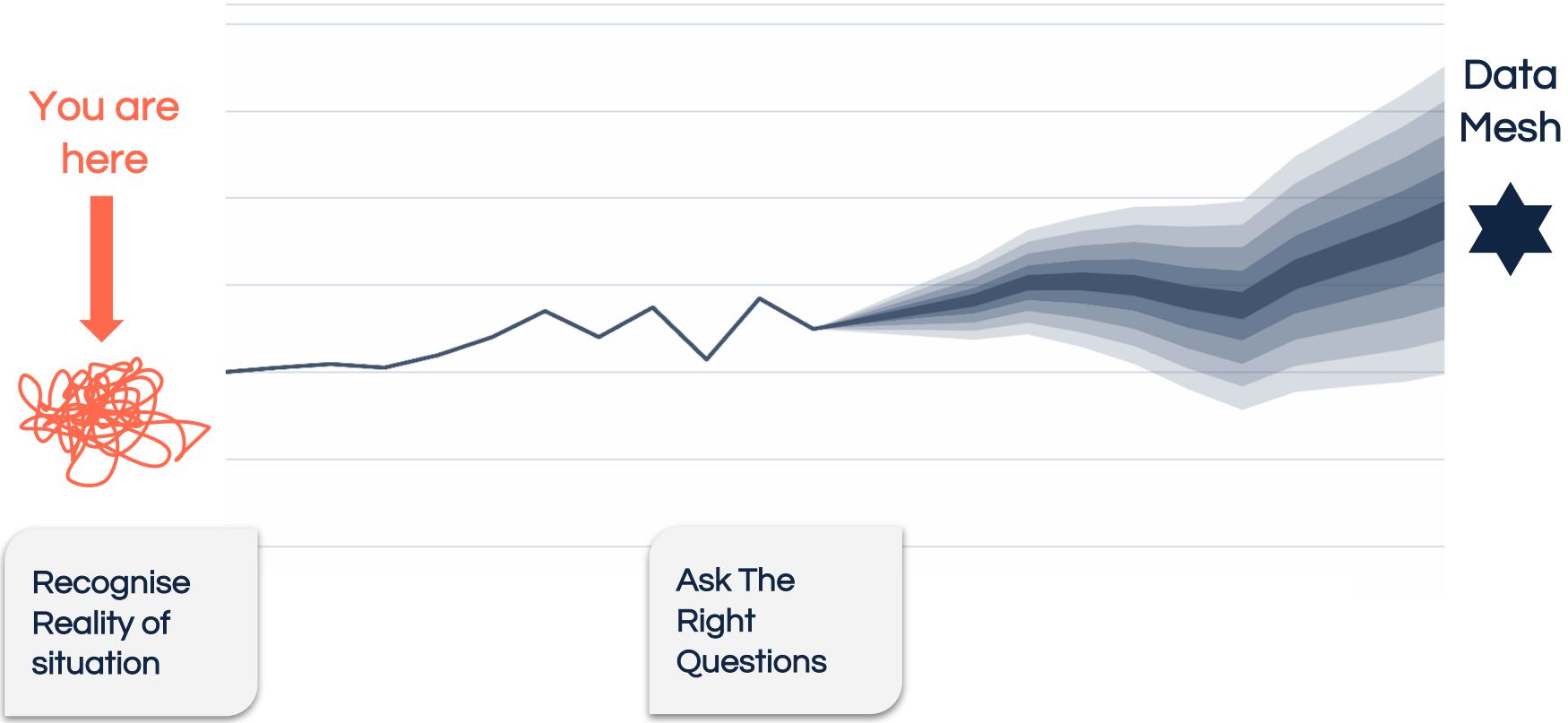

I Have Some Data Problem!
What Kind Of Problem? What Trying To Achieve?
Why?
How Far Are You Prepared To Go?
I Need A Data Mesh
Options / Plan Of Attack
Sometimes the answer will be
This Will Solve All My Problems!
Irrespective of approach, you should always aim for a
Continuous Improvement Journey

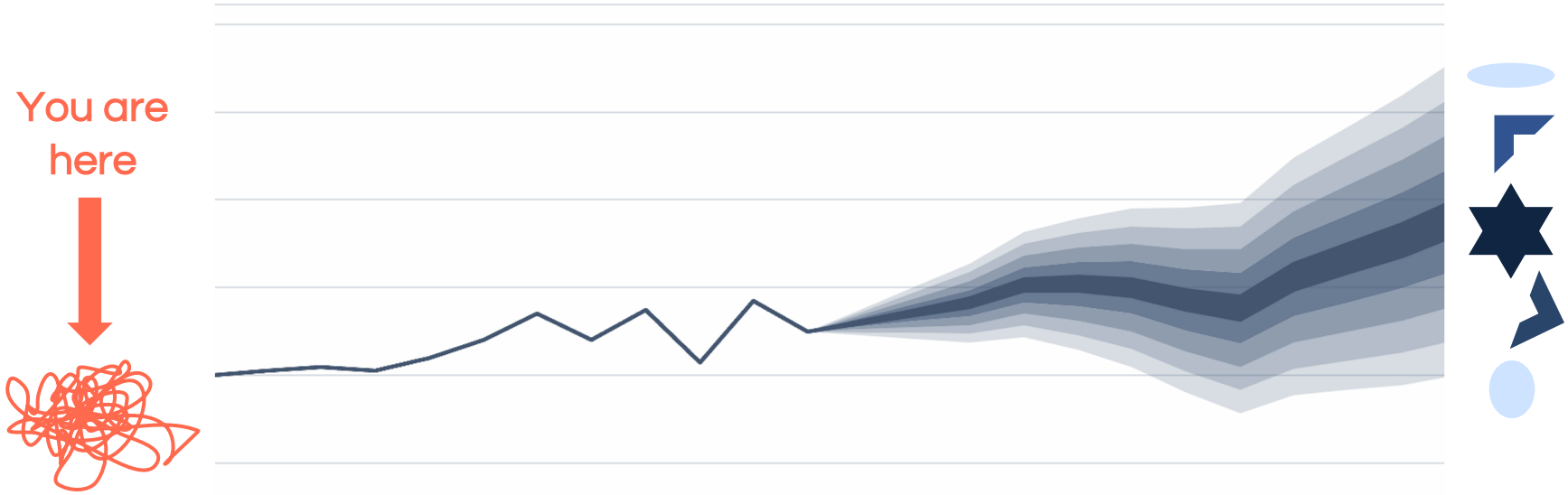

I Have Some Data Problem!
What Kind Of Problem? What Trying To Achieve?
Why?
How Far Are You Prepared To Go?
I Need A Data Mesh
Options / Plan Of Attack
Sometimes the answer will be
This Will Solve All My Problems!
Irrespective of approach, you should always aim for a
KPI
Continuous Improvement Journey

You are here

Data Mesh

Recognise Reality of situation

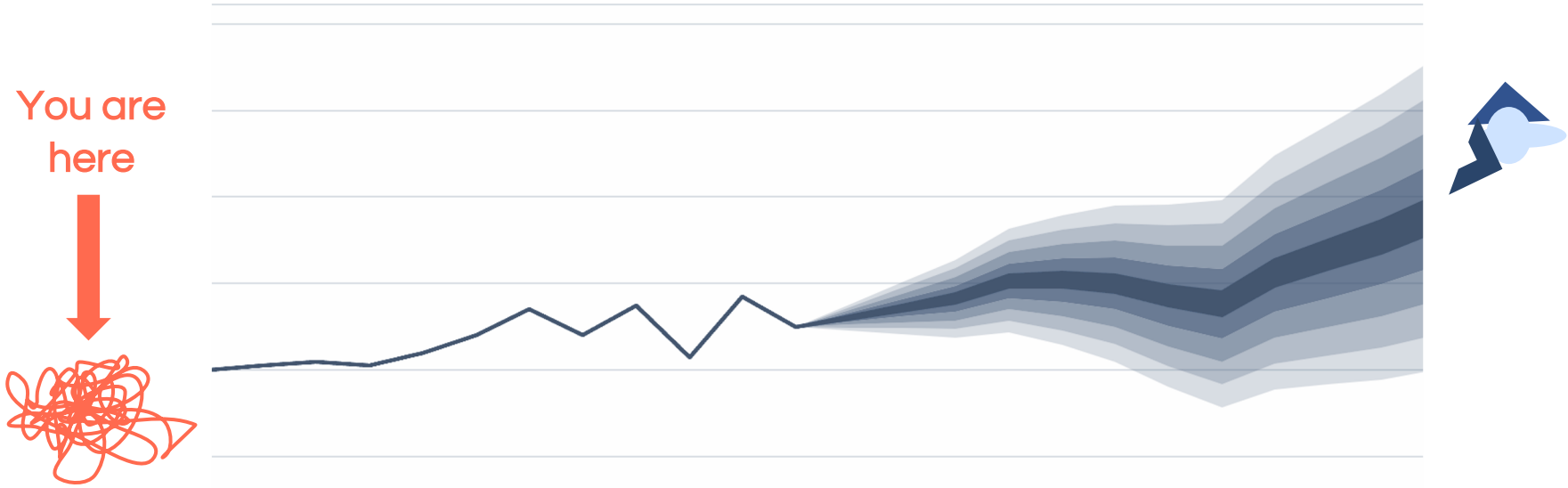

You are here
Data Mesh
Recognise Reality of situation
Ask The Right Questions



Recognise Reality of situation

Ask The Right Questions

Chose The Right Approach For Job

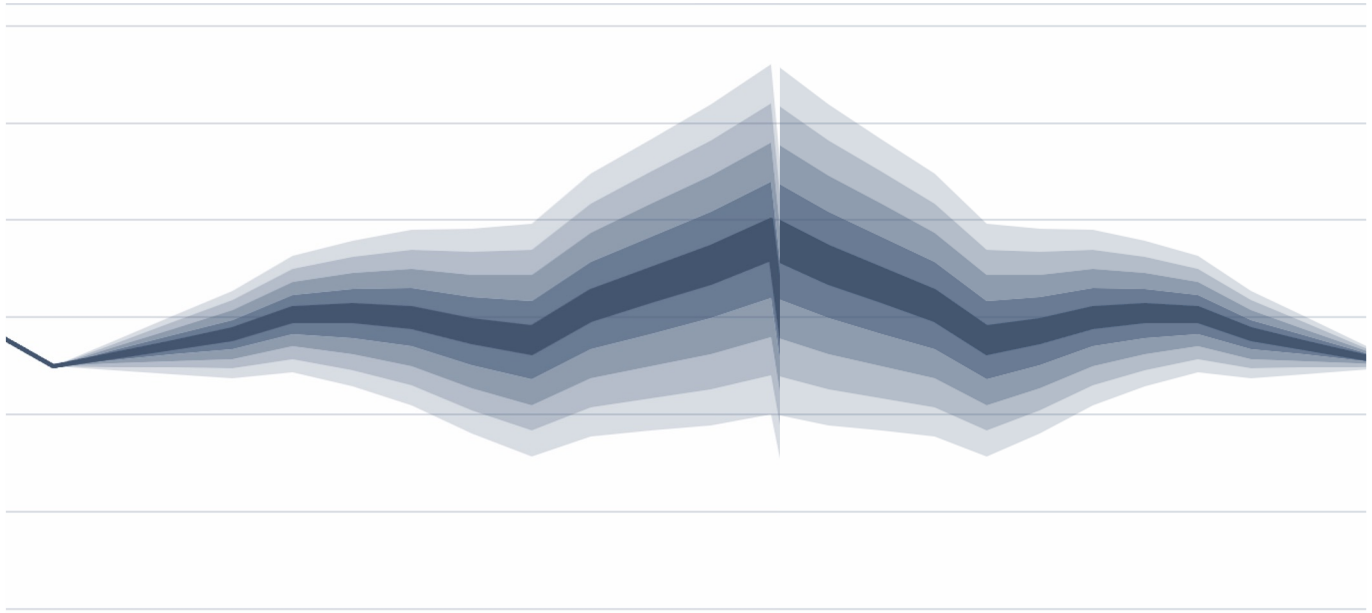


Recognise Reality of situation

Ask The Right Questions

Chose The Right Approach For Job

You are here

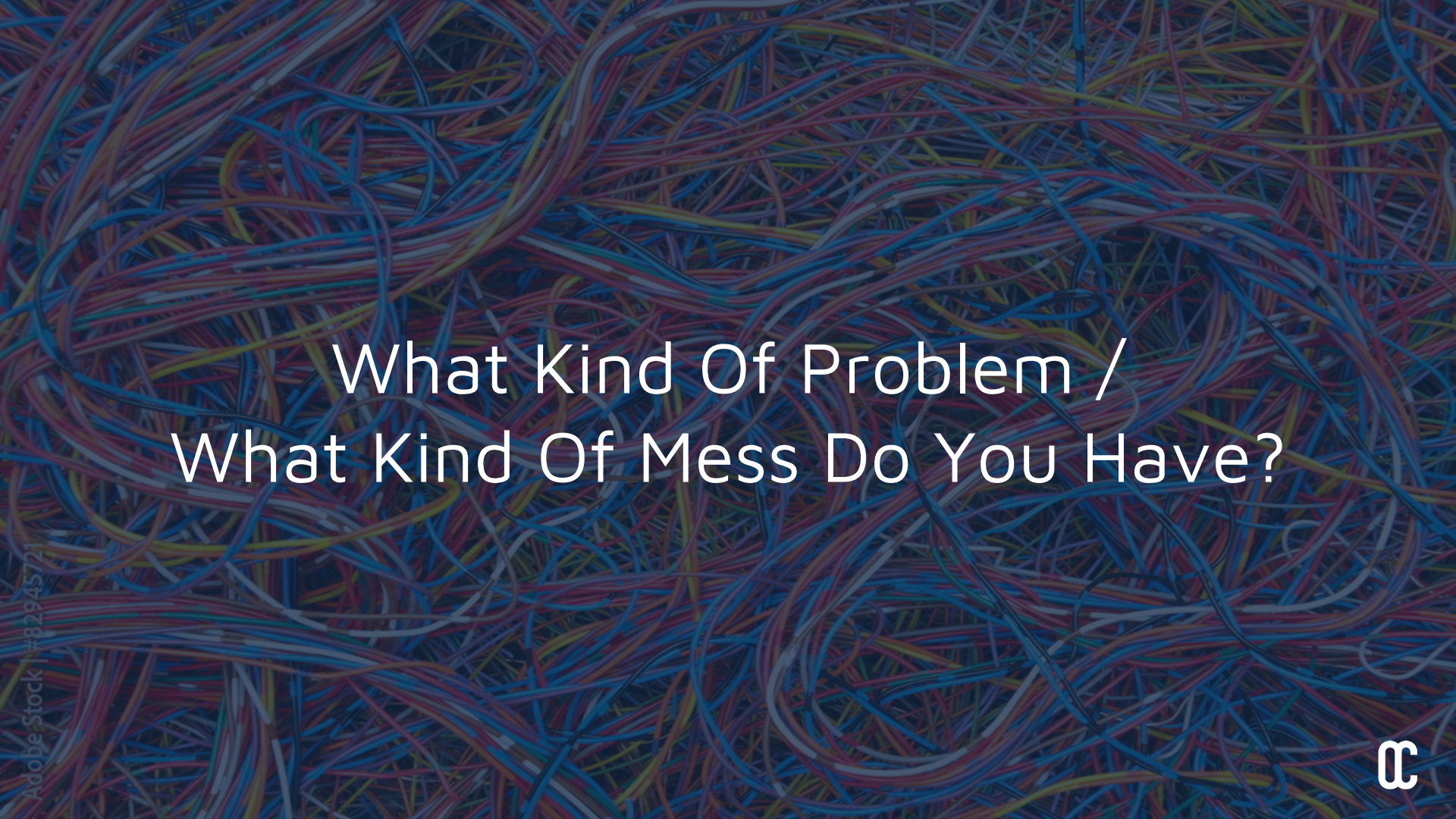


Considerations On Path To Mesh: Key Takeaways

- Recognise the reality of the situation you're in
- Ask the right questions
- Choose the right approach for the job
- Data Mesh is not the only answer!
- Data Mesh comes in many shapes & sizes

3

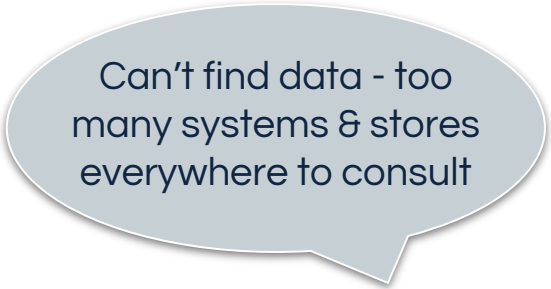
Tracing a journey from mess → mesh


What Kind Of Problem /
What Kind Of Mess Do You Have?
Adobe Stock | #82945721

Types of Problems

As a Data Scientist / Analyst

Types of Problems



As a Data Scientist / Analyst

Types of Problems

Can't find data - too many systems & stores everywhere to consult

As a Data Scientist / Analyst

Can't easily access or make sense of data in data lake (data swamp)

Types of Problems

Can't find data - too many systems & stores everywhere to consult

As a Data Scientist / Analyst

Can't easily access or make sense of data in data lake (data swamp)

As a CTO

Types of Problems

Can't find data - too many systems & stores everywhere to consult

As a Data Scientist / Analyst

Can't easily access or make sense of data in data lake (data swamp)

[Insight driven] projects taking too long to complete, business is losing out on opportunities

As a CTO

Types of Problems

Can't find data - too many systems & stores everywhere to consult

As a Data Scientist / Analyst

Can't easily access or make sense of data in data lake (data swamp)

[Insight driven] projects taking too long to complete, business is losing out on opportunities

As a CTO

Overloaded data teams, not enough people to do the work

Types of Problems

Can't find data - too many systems & stores everywhere to consult

As a Data Scientist / Analyst

Can't easily access or make sense of data in data lake (data swamp)

[Insight driven] projects taking too long to complete, business is losing out on opportunities

Need to get a better handle on where sensitive data is being used - worried we are exposing things like PII unnecessarily

As a CTO

Overloaded data teams, not enough people to do the work

How can I solve my mess?

Can't find data - too many systems & stores everywhere to consult

As a Data Scientist / Analyst

Can't easily access or make sense of data in data lake (data swamp)

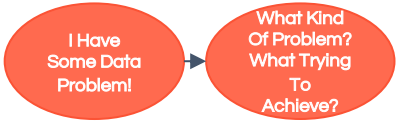
How can I solve my mess?

Can't find data - too many systems & stores everywhere to consult

As a Data Scientist / Analyst

Can't easily access or make sense of data in data lake (data swamp)

How can I solve my mess?



Can't find data - too many systems & stores everywhere to consult

As a Data Scientist / Analyst

Can't easily access or make sense of data in data lake (data swamp)

How can I solve my mess?

Can't find data - too many systems & stores everywhere to consult

As a Data Scientist / Analyst

Can't easily access or make sense of data in data lake (data swamp)

Why Is It Important To Address?

- If I can more easily find the data I need, it will be easier & faster to build my models

How can I solve my mess?

Can't find data - too many systems & stores everywhere to consult

As a Data Scientist / Analyst

Can't easily access or make sense of data in data lake (data swamp)

Why Is It Important To Address?

- If I can more easily find the data I need, it will be easier & faster to build my models

Why Is that Important ?

- Models are key part to how we deliver insight / value to users
- We are growing and can't afford to slow down in this area

How can I solve my mess?

Can't find data - too many systems & stores everywhere to consult

As a Data Scientist / Analyst

Can't easily access or make sense of data in data lake (data swamp)

Why Is It Important To Address?

- If I can more easily find the data I need, it will be easier & faster to build my models

Why Is that Important ?

- Models are key part to how we deliver insight / value to users
- We are growing and can't afford to slow down in this area

Why Is that Important ?

- My boss says so …

How can I solve my mess?

Can't find data - too many systems & stores everywhere to consult

As a Data Scientist / Analyst

Can't easily access or make sense of data in data lake (data swamp)

Why Is It Important To Address?

- If I can more easily find the data I need, it will be easier & faster to build my models

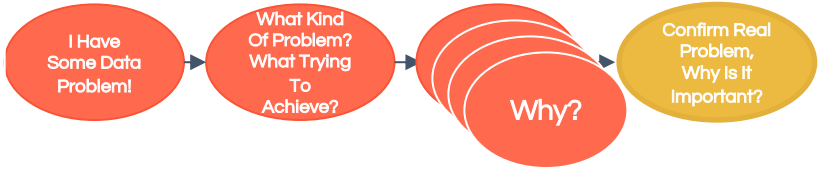
Why Is that Important ?

- Models are key part to how we deliver insight / value to users
- We are growing and can't afford to slow down in this area

Why Is that Important ?

- Our competition is already doing this, if we don't adapt, we die!

How can I solve my mess?



As Data scientists & analysts its difficult to find, acquire and make sense of data to build models and run reports easily - too complex, too many moving parts, too slow.

With business growing & competition increasing, continually being able to derive value from these models and feedback into product offerings will become a crucial part of the companies survival!

How can I solve my mess?

...

Data Catalogue?

Data Lakehouse?

Data Mesh?

How can I solve my mess?

...

Data Catalogue?

Data Lakehouse?

Data Mesh?

How can I solve my mess?

...

Data Catalogue?

Data Lakehouse?

Data Mesh?

How can I solve my mess?

...

Data Catalogue?

Data Lakehouse?

Data Mesh?

How can I solve my mess?

...

Data Catalogue?

Data Lakehouse?

Data Mesh?

Considering Data Mesh?
Do we want/need to be able :

- ❑ Deal well with change and business complexity
- ❑ Remain agile especially when growing quickly
- ❑ Get better value from your data investment

How can I solve my mess?

Options / Plan Of Attack

...

Data Catalogue

Data Lakehouse

Data Mesh

Considering Data Mesh?
Do we want/need to be able :

- Deal well with change and business complexity
- Remain agile especially when growing quickly
- Get better value from your data investment

How can I solve my mess?

How Far Are You Prepared To Go?

...

Data Catalogue

Data Lakehouse

Data Mesh???

"Everybody wants to go to heaven, but nobody wants to die."

Quote Unknown (possibly Alfred King)

How can I solve my mess?

How Far Are You Prepared To Go?

...

Data Catalogue

Data Lakehouse

Data Mesh???

Are You Prepared To Change where necessary:

- ❑ Team Structures
- ❑ Operating Model
- ❑ Governance Approach
- ❑ Architectural Approach

How can I solve my mess?

How Far Are You Prepared To Go?

...

Data Catalogue

Data Lakehouse

Data Mesh

Are You Prepared To Change where necessary:

- Team Structures
- Operating Model
- Governance Approach
- Architectural Approach

How can I solve my mess?

How Far Are You Prepared To Go?

...

Data Catalogue

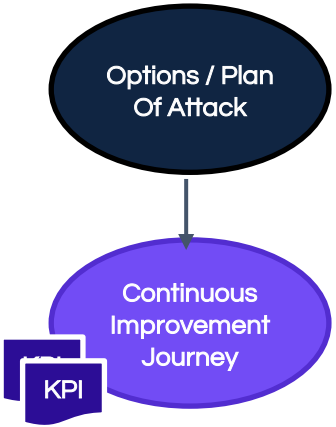
Data Lakehouse

Data Mesh

Are You Prepared To Change where necessary:

- Team Structures
- Operating Model
- Governance Approach
- Architectural Approach

How can I solve my mess?



Eg HR domain

- 15% reduction in rejection of candidates at 1st hire phase
- 10% improvement in retention rate for key talent
- Cost per hire remains under $X.XX

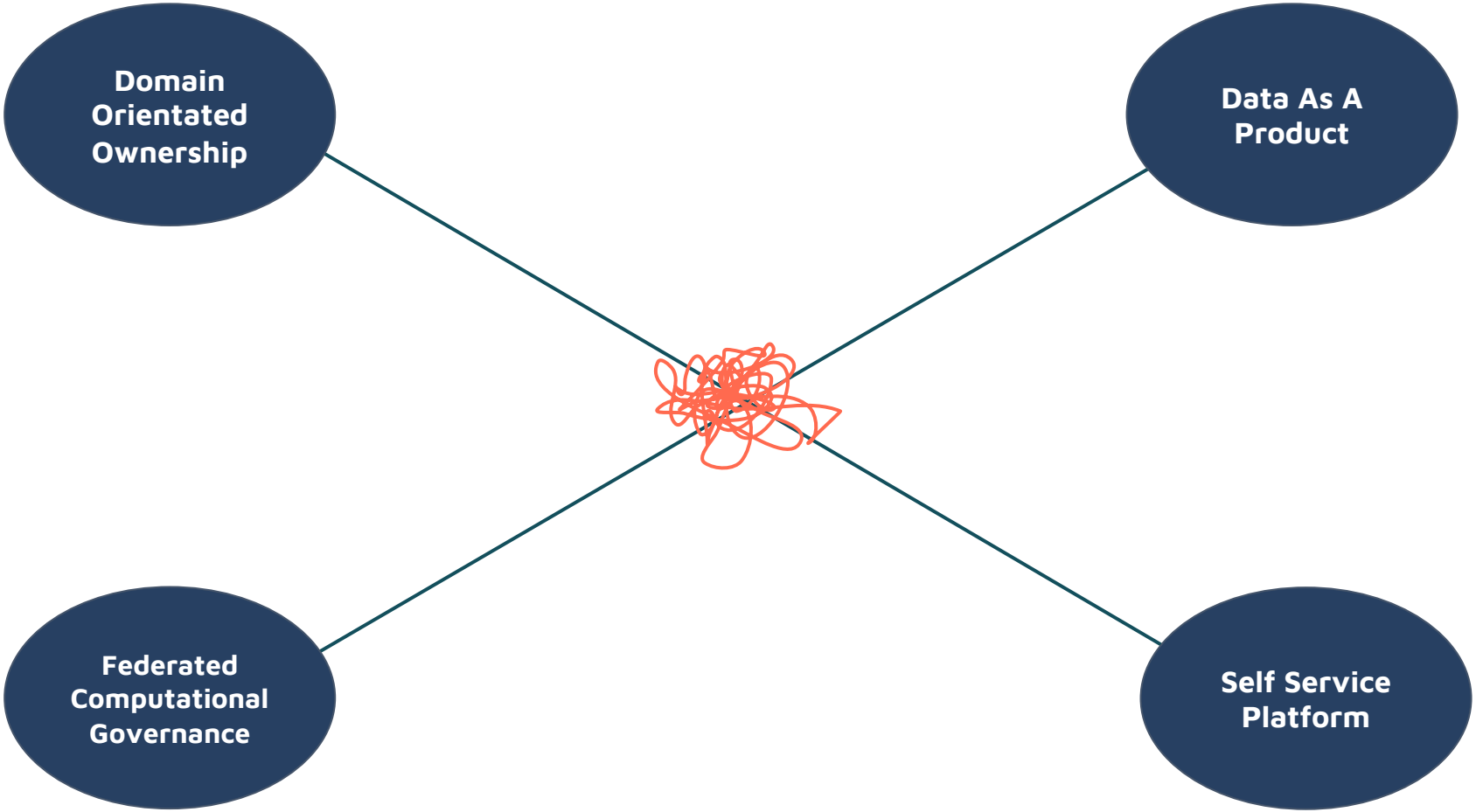
Tracing A Journey: Key Takeaways

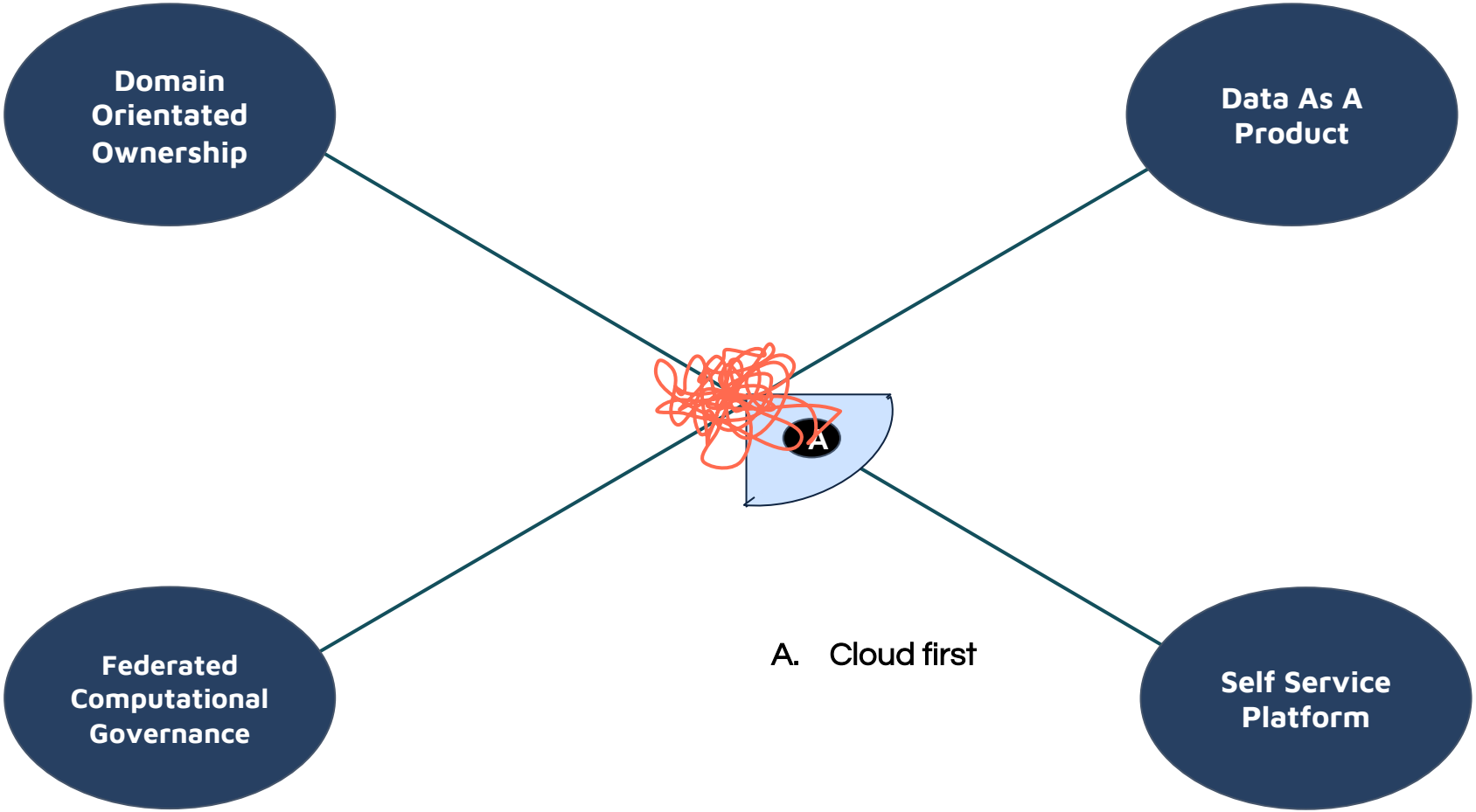
- Data Mesh is not for everyone
- But even if it is, how far are you prepared to go?
- Data Mesh can require big changes to people & org
  - This can scupper many folks / organisations which are not ready or aware

4

Evolutionary Considerations

With some top tips along the way!


Domain Orientated Ownership
Data As A Product
Federated Computational Governance
Self Service Platform



A. Cloud first

#1 Lay a solid foundation: start with cloud

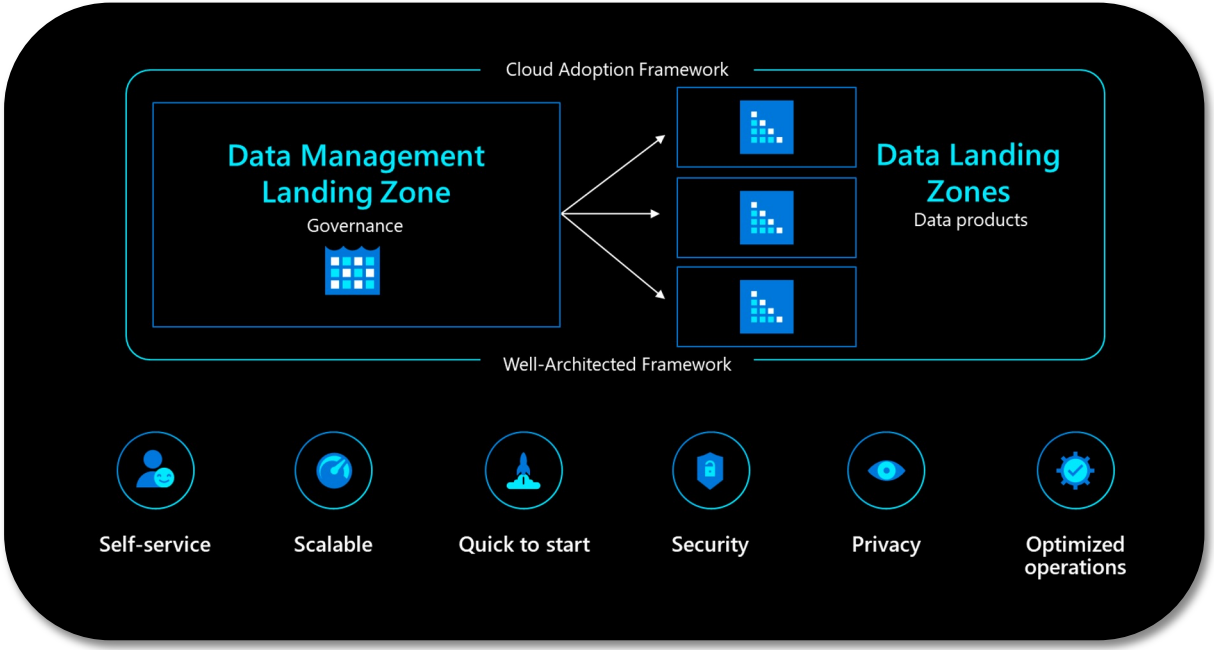
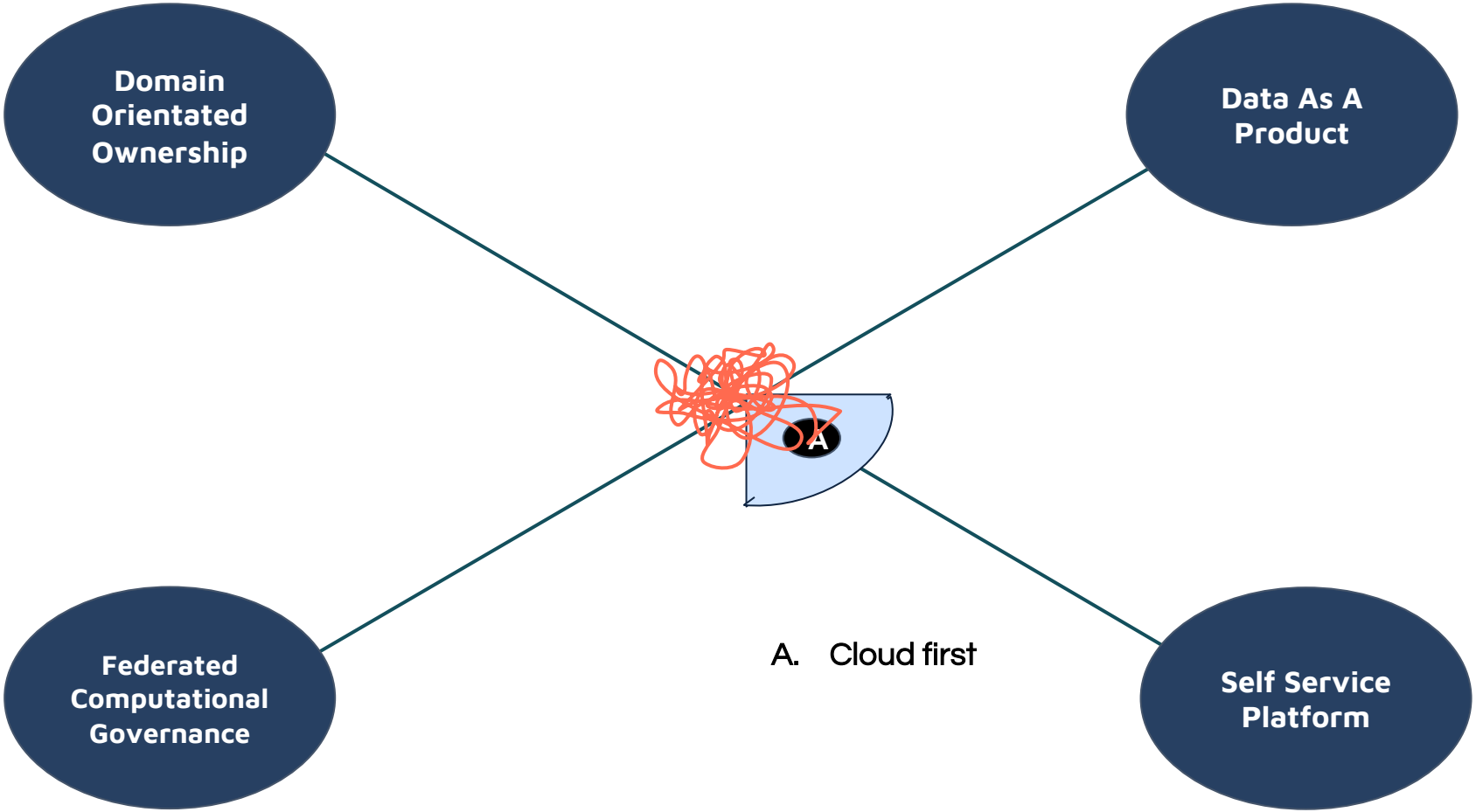
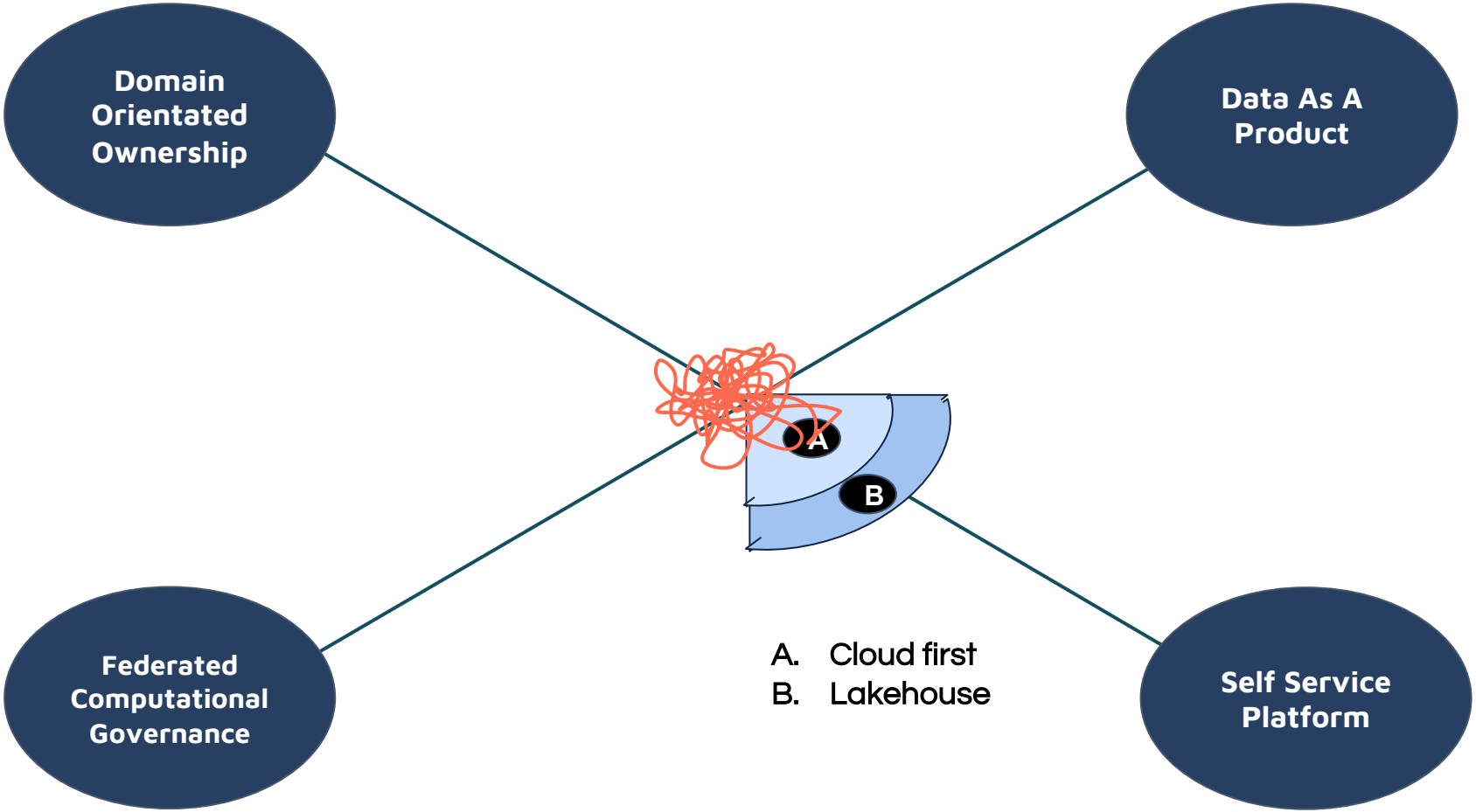


Image Credit: Microsoft Cloud Adoption Framework - Cloud Scale Analytics
https://learn.microsoft.com/en-us/azure/cloud-adoption-framework/scenarios/cloud-scale-analytics/



A. Cloud first



A. Cloud first
B. Lakehouse

#2 Be pragmatic about decentralization

#2 Be pragmatic about decentralization

- Better ownership
- Less contention
- Easier to scale

Centralised

Decentralised

#2 Be pragmatic about decentralization

- Large cloud bills ???
- Performance
  (copying data everywhere) ???

- Better ownership
- Less contention
- Easier to scale

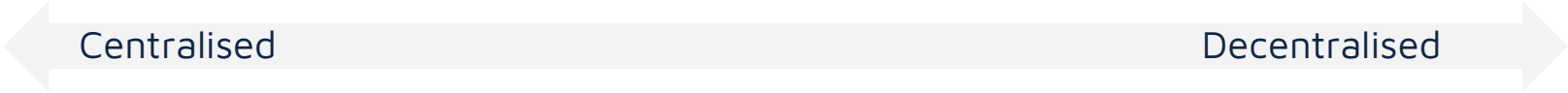
Centralised

Decentralised

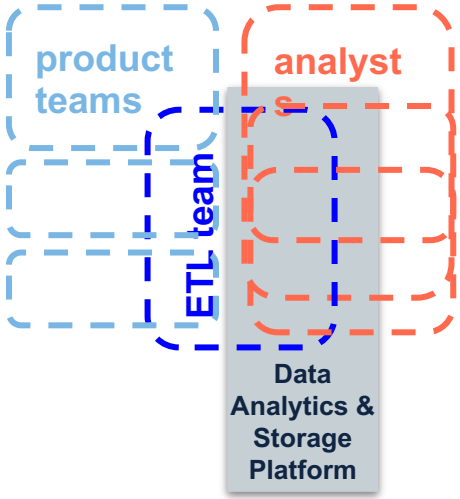
#2 Be pragmatic about decentralization

What is most important to you?

How far are you prepared to go?



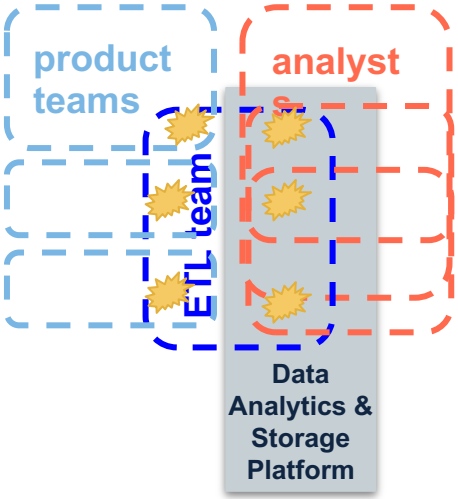
#2 Be pragmatic about decentralization



Centralised

Decentralised

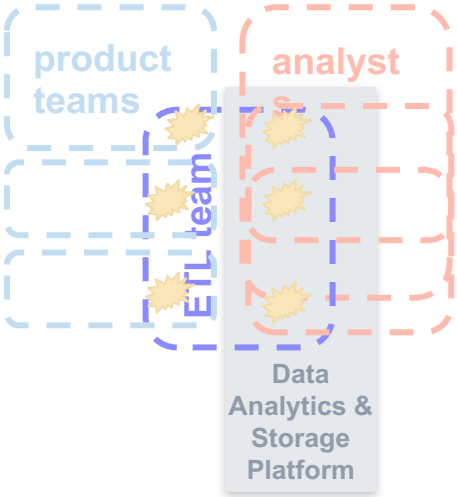
#2 Be pragmatic about decentralization

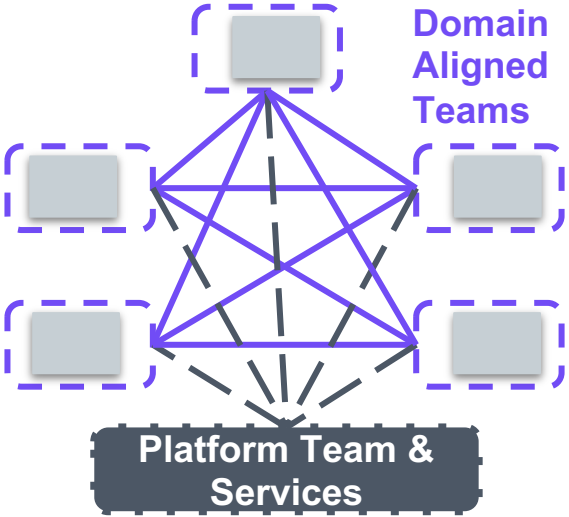


Centralised

Decentralised

#2 Be pragmatic about decentralization

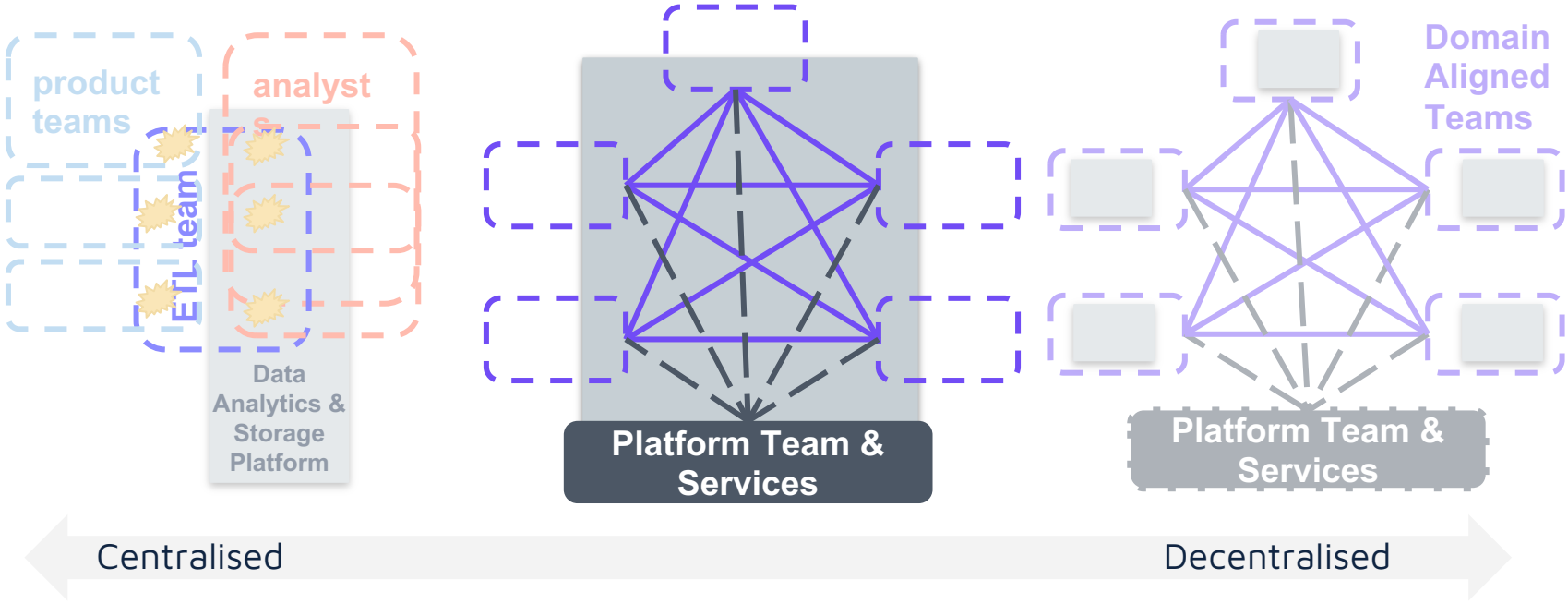




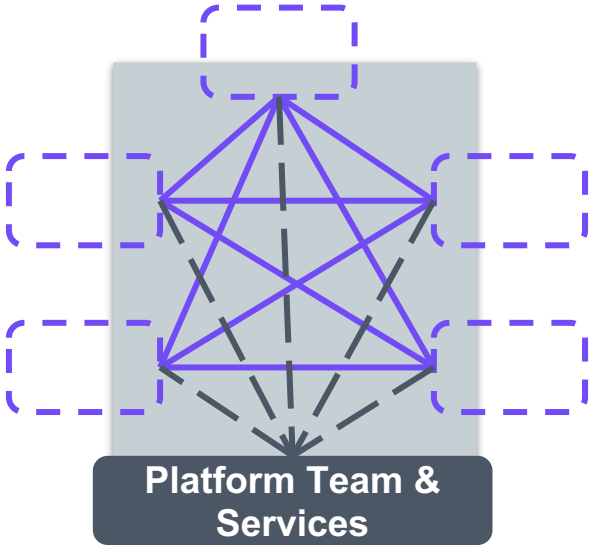
Centralised

Decentralised

#2 Be pragmatic about decentralization



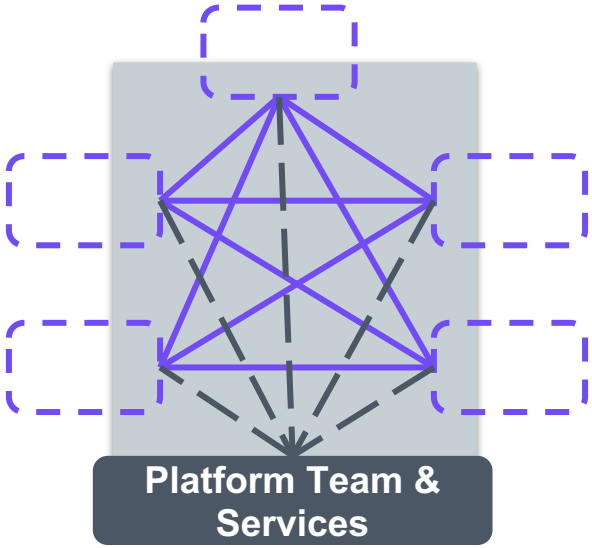
#2 Be pragmatic about decentralization



Centralised

Decentralised

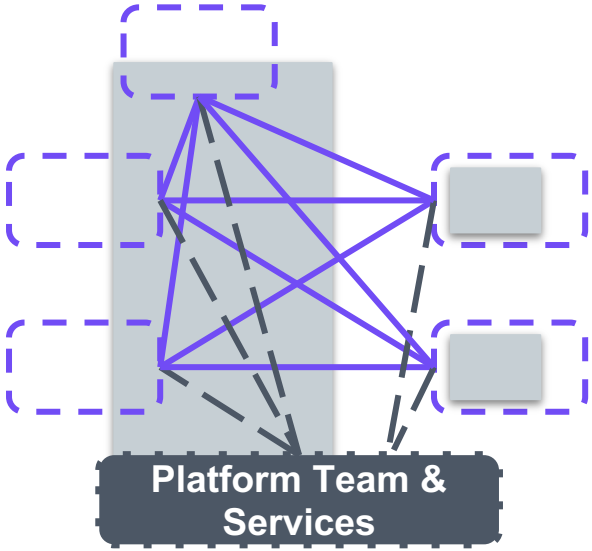
#2 Be pragmatic about decentralization



Centralised

Decentralised

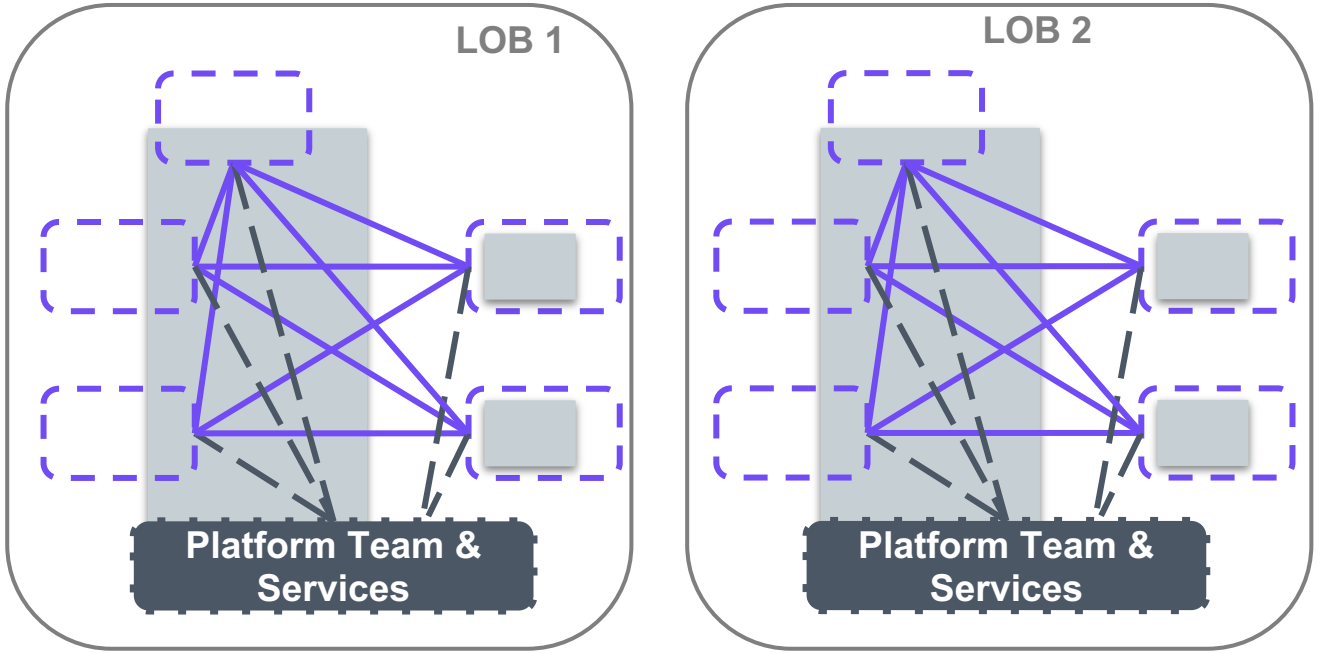
#2 Be pragmatic about decentralization



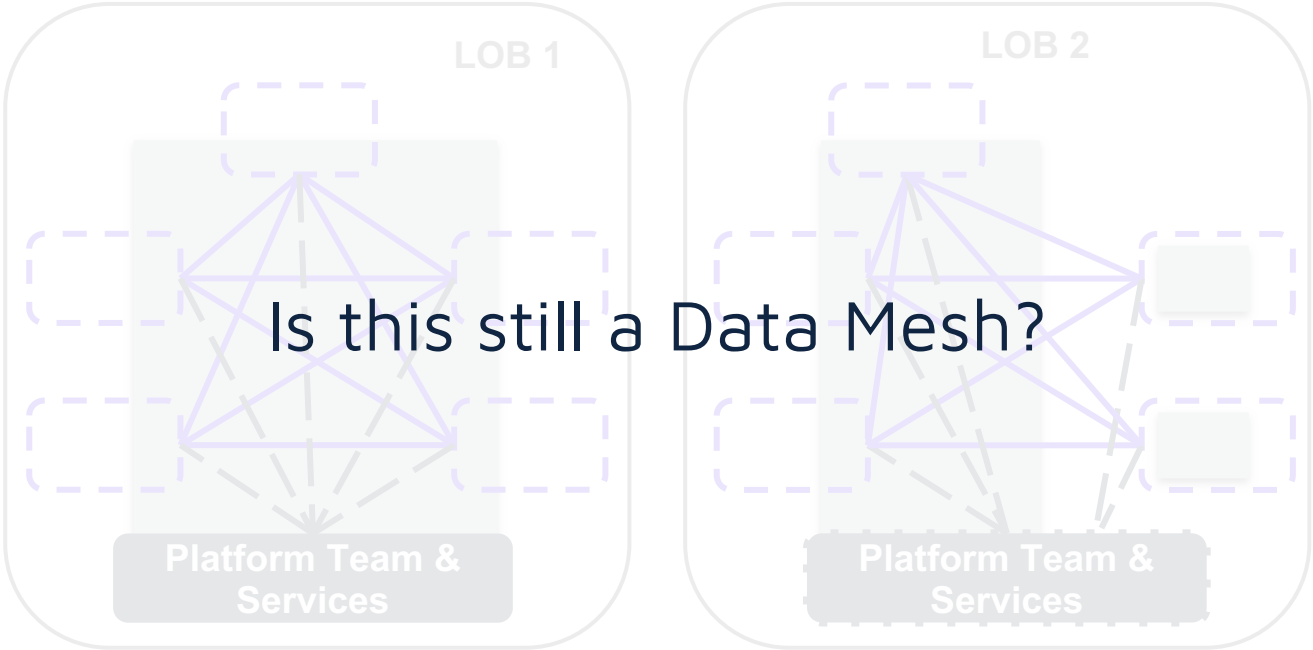
Centralised

Decentralised

#2 Be pragmatic about decentralization

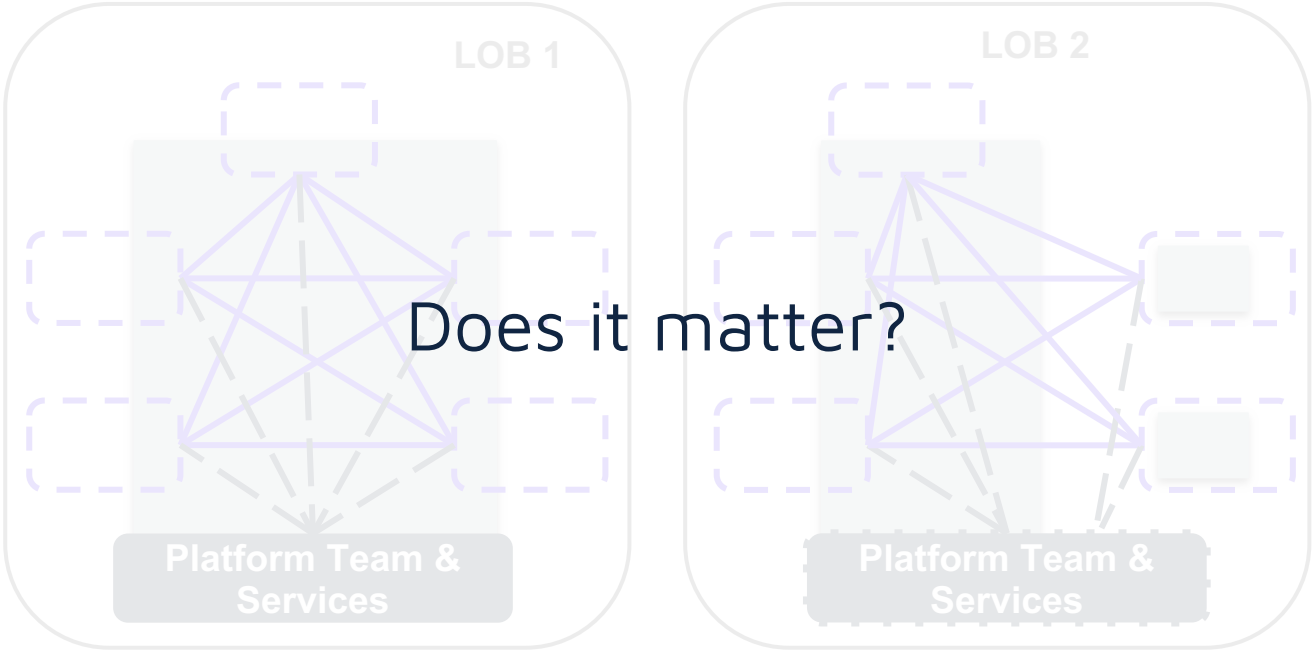


#2 Be pragmatic about decentralization



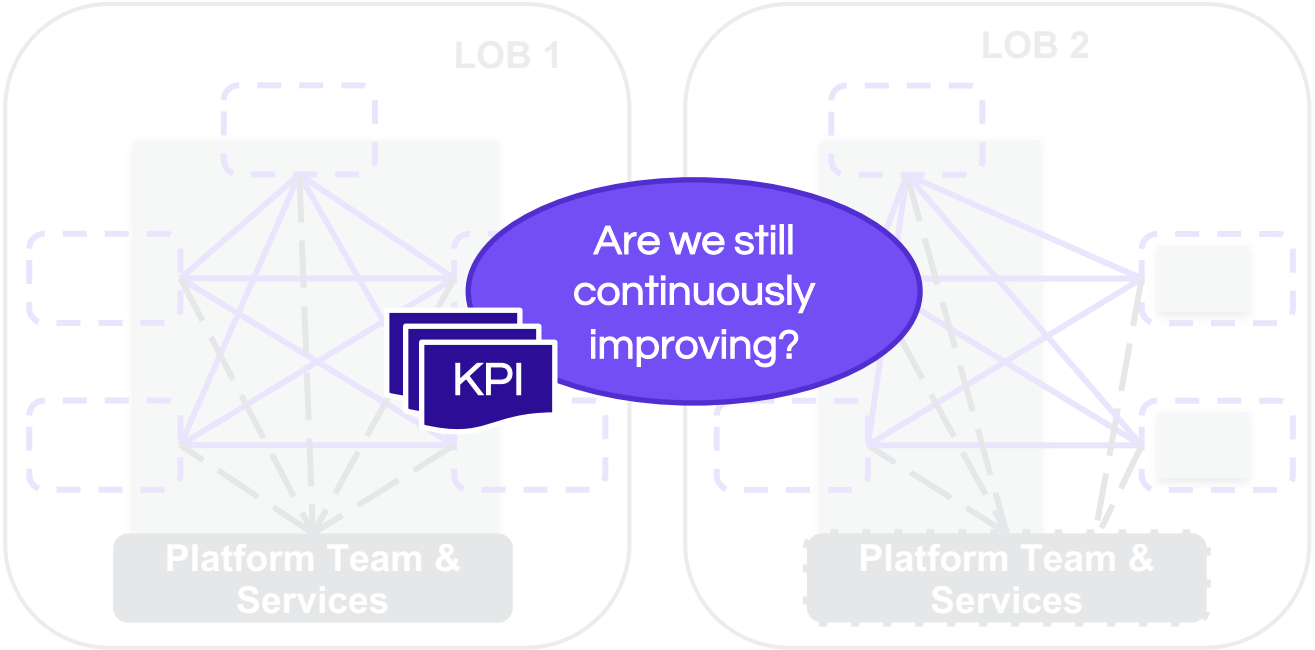
Is this still a Data Mesh?

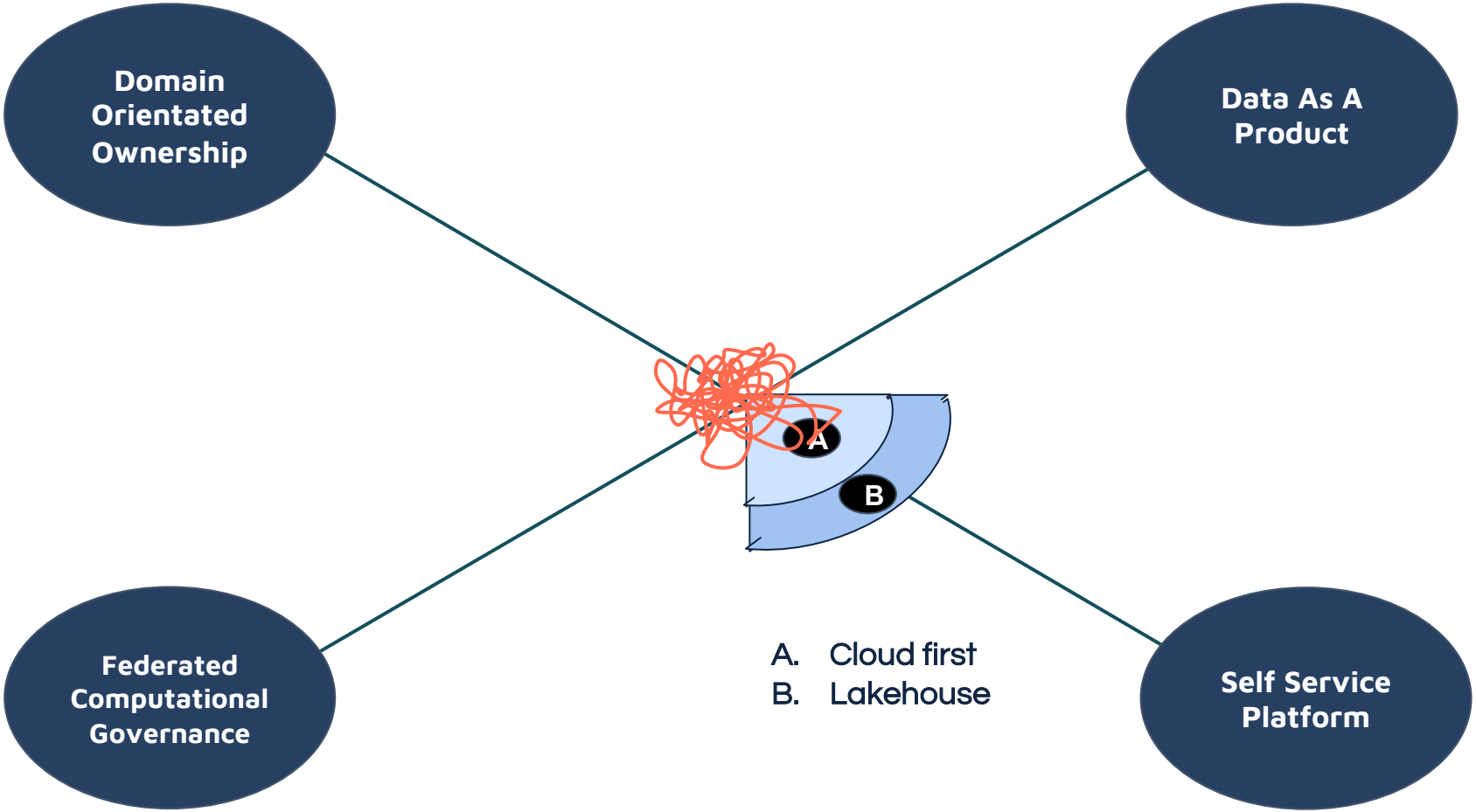
#2 Be pragmatic about decentralization



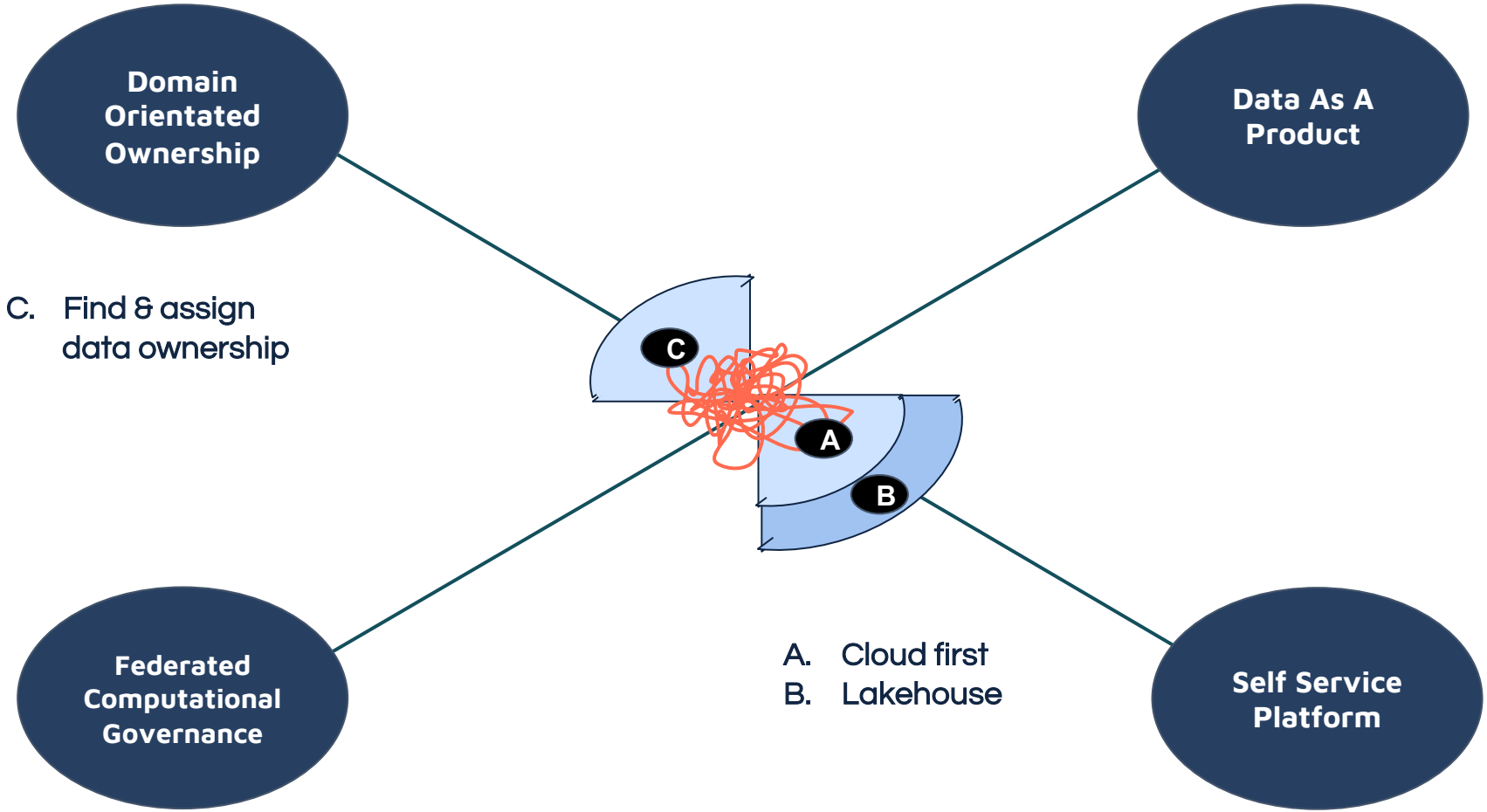
Does it matter?

#2 Be pragmatic about decentralization





A. Cloud first
B. Lakehouse

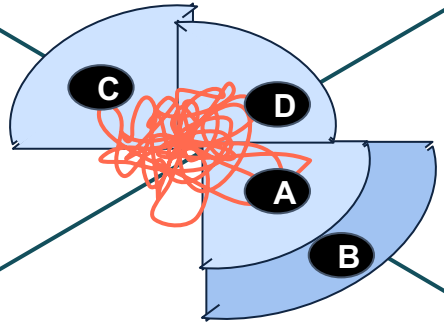

Domain Orientated Ownership
Data As A Product
C. Find & assign data ownership
C
A
B
Federated Computational Governance
A. Cloud first
B. Lakehouse
Self Service Platform

Domain Orientated Ownership

Data As A Product

C. Find & assign data ownership

D. DataSet catalogue



A. Cloud first
B. Lakehouse

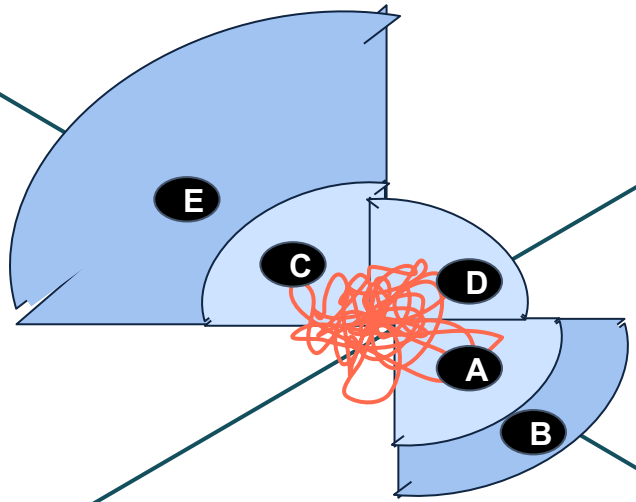
Federated Computational Governance

Self Service Platform

Domain Orientated Ownership

Data As A Product

C. Find & assign data ownership
E. Digital / cloud transformation



D. DataSet catalogue

A. Cloud first
B. Lakehouse

Federated Computational Governance

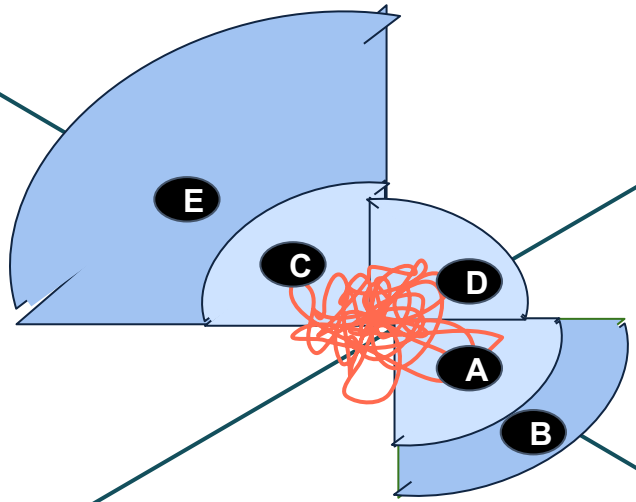
Self Service Platform

#3 Start small, learn, adapt, rinse repeat

- The Data Mesh Killer: Requiring a full re-org into domain teams up front
- Piggyback off a digital / cloud transformation
- Event Storming helpful to discover data products

Domain Orientated Ownership

Data As A Product



C. Find & assign data ownership
E. Digital / cloud transformation

D. DataSet catalogue

A. Cloud first
B. Lakehouse

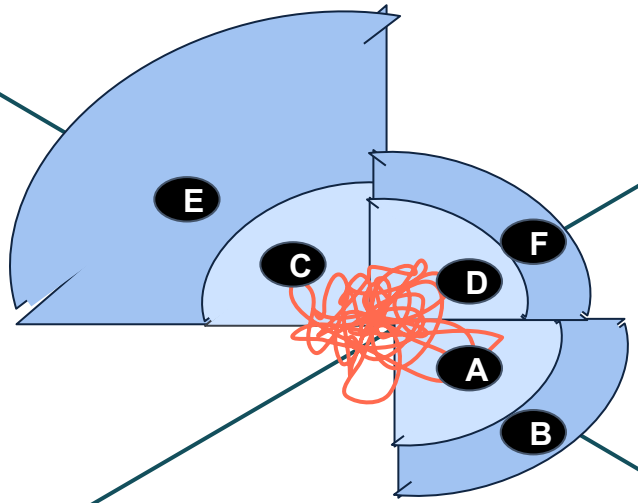
Federated Computational Governance

Self Service Platform

Domain Orientated Ownership

Data As A Product

C. Find & assign data ownership
E. Digital / cloud transformation



D. DataSet catalogue
F. Basic data catalogue

A. Cloud first
B. Lakehouse

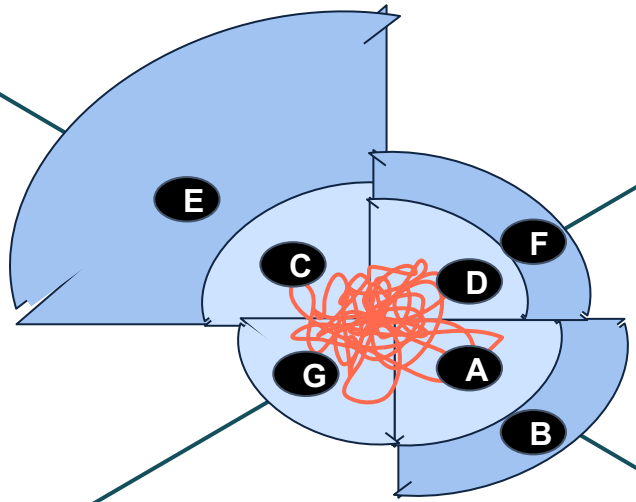
Federated Computational Governance

Self Service Platform

Domain Orientated Ownership

Data As A Product

C. Find & assign data ownership
E. Digital / cloud transformation



D. DataSet catalogue
F. Basic data catalogue

Federated Computational Governance

G. Extend catalogue with classification data (PII / sensitive etc)

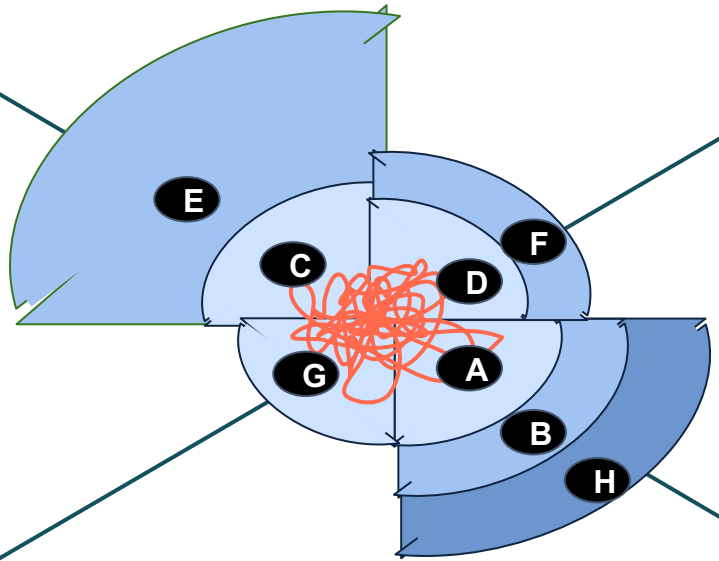
A. Cloud first
B. Lakehouse

Self Service Platform

Domain Orientated Ownership

Data As A Product

C. Find & assign data ownership
E. Digital / cloud transformation



D. DataSet catalogue
F. Basic data catalogue

Federated Computational Governance

G. Extend catalogue with classification data (PII / sensitive etc)

A. Cloud first
B. Lakehouse
H. Refactoring /integration tooling

Self Service Platform

#4 DIY Integration required

- Tools still immature
- You will need to a lot of work in this space

Evolutionary Considerations: Key Takeaways

- It's not all or nothing, can make progress along each principle
  - And still derive value - although may be not be fully realised

#1 Lay a solid foundation: start with cloud

#2 Be pragmatic about decentralization

#3 Start small, learn, adapt, rinse, repeat

#4 Custom "glue/integration" tooling dev still required

5

Conclusion

What does it take to move towards acquiring a half decent data mesh solution these days?

Especially when you feel you are starting with a data mess

Key Considerations On Journey

- Recognise the reality of the situation you're in
- Ask the right questions
- Choose the right approach for the job
- Data Mesh is not the only answer!
- Data Mesh comes in many shapes & sizes

Key Considerations On Journey

Data Mesh is not the goal in and of itself, but it does *provide a path and good principles* to help guide us towards achieving our strategic objectives

By tracing an example journey we saw ...

- Data Mesh is not for everyone
- But even if it is, how far are you prepared to go?
- Data Mesh requires big changes of people & orgs
    - This can scupper things if not ready or aware

Evolutionary Considerations

- It’s not all or nothing, can make progress along each principle
  - And still derive value - although may not be fully realised

#1 Lay a solid foundation: start with cloud

#2 Be pragmatic about decentralization

#3 Start small, learn, adapt, rinse, repeat

#4 Custom “glue/integration” tooling dev still required

Questions?

OpenCredo
A TRIFORK COMPANY

info@opencredo.com

www.opencredo.com

goto;

Don’t forget to
rate this session
in the **GOTO Guide app**