

# Solutions to Sample Tasks

Tolgonai Kozhoeva

March 23, 2022

# The list of tasks:

- What is data Mesh?
- When can Data Mesh be useful?
- Visualize Airbyte's architecture and technology and describe how each sync mode technically works.
- What do you see as the most promising technology in the field of modern data platforms?
- What ten questions would you ask to be able to work on the project as quickly as possible?

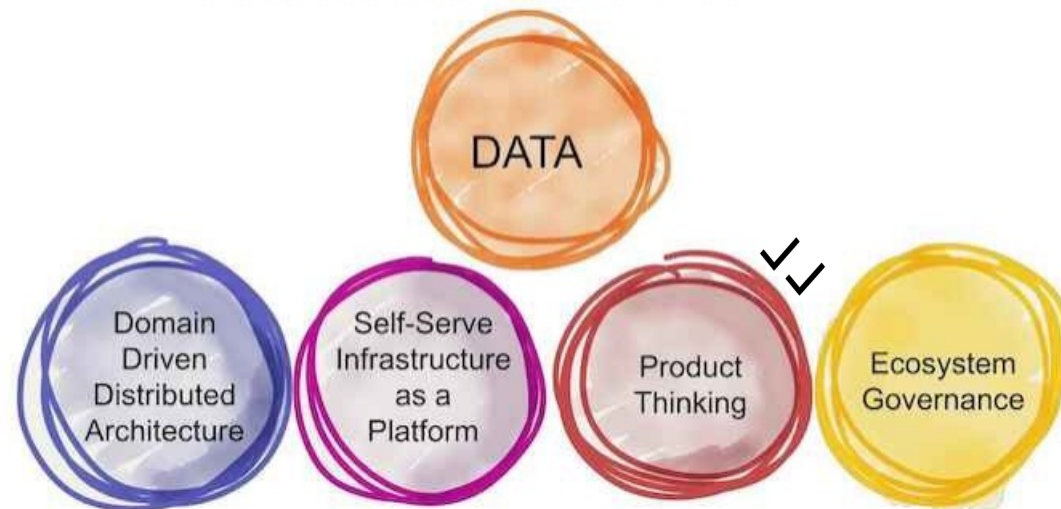
# What is driving the need for data mesh?



Source: [martinflower.com](https://martinflower.com)

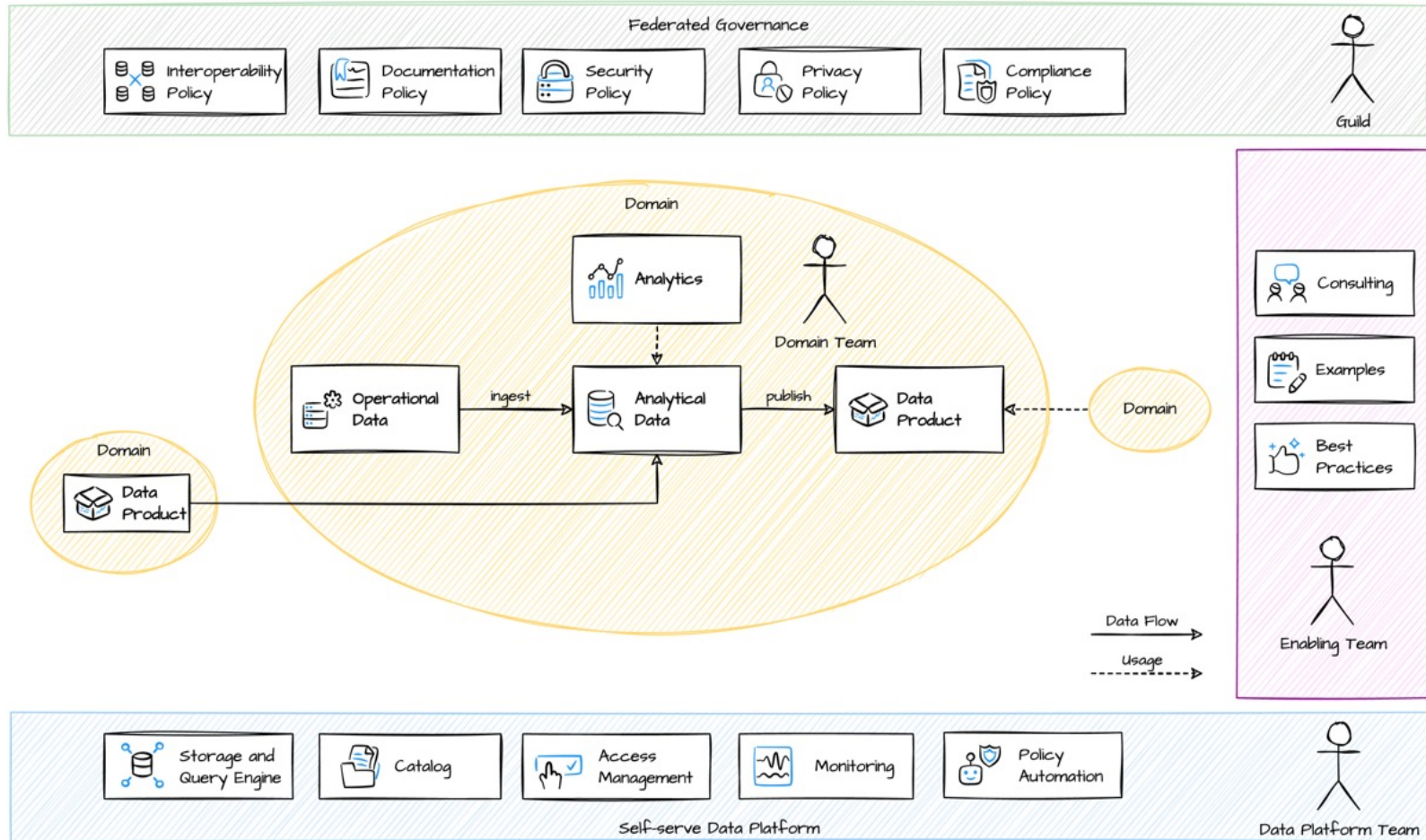
# What is Data Mesh?

## DATA MESH PRINCIPLES



Full accreditation for this photo and subsequent information belongs to [Zhamak Dehghani](#)

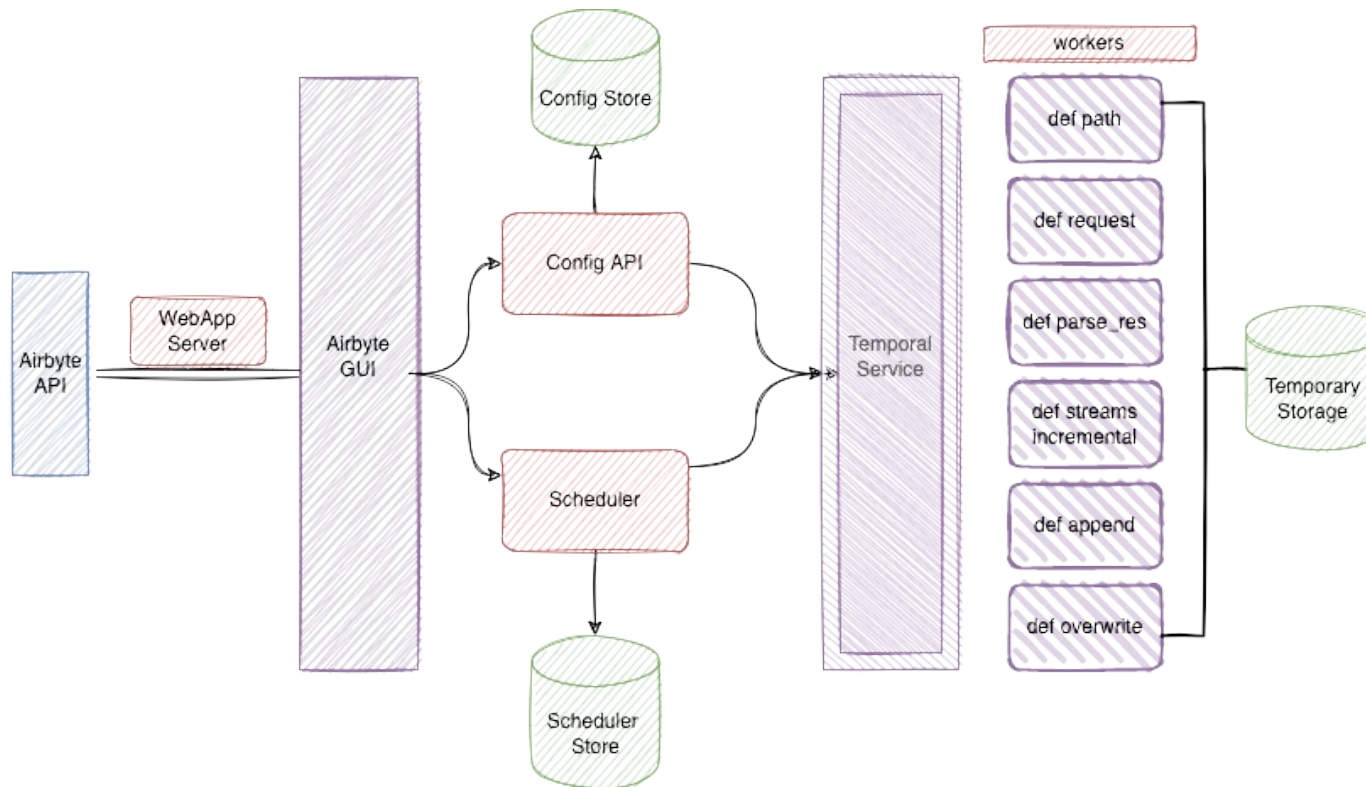
# Data Mesh Architecture



# When can Data Mesh be useful?

- Rapid scaling of a company ✓
- Poliferation of data sources ✓
- Ubiquitous usage of data ✓
- Challenges with data availability and accessibility at scale ✓
- Domain-driven development ✓
- Started working with Microservices ✓
- Cloud migration is a viable option ✓

# Airbyte's architecture overview



# Sync modes

**Connection settings**ENABLED ☒

Source connector

Destination connector

Sync frequency

GitHub - Issues →

BigQuery →

12 hours ▾

Namespace Prefix - Where to replicate your data to in your destination

/configured-setting

Select the data you want to sync [Update latest source schema](#)

☒

	Data type	Cleaned name	Primary key	Cursor field	Sync settings
> <input checked="" type="checkbox"/> Table name		table_name			Full refresh - overwrite ▾
> <input checked="" type="checkbox"/> Table name		table_name		timestamp_id ▾	Incremental - append ▾
✓ <input checked="" type="checkbox"/> Table name		table_name	Primary_key_id ▲	Column Name ▲	Incremental - deduped + story ▾
Column name		column_name	<input type="checkbox"/>	<input type="checkbox"/>	
Column name	Number Nullable	column_name	<input type="checkbox"/>	<input checked="" type="checkbox"/>	
Column name	Object Nullable	column_name	<input checked="" type="checkbox"/>	<input type="checkbox"/>	

Save changes

Cancel



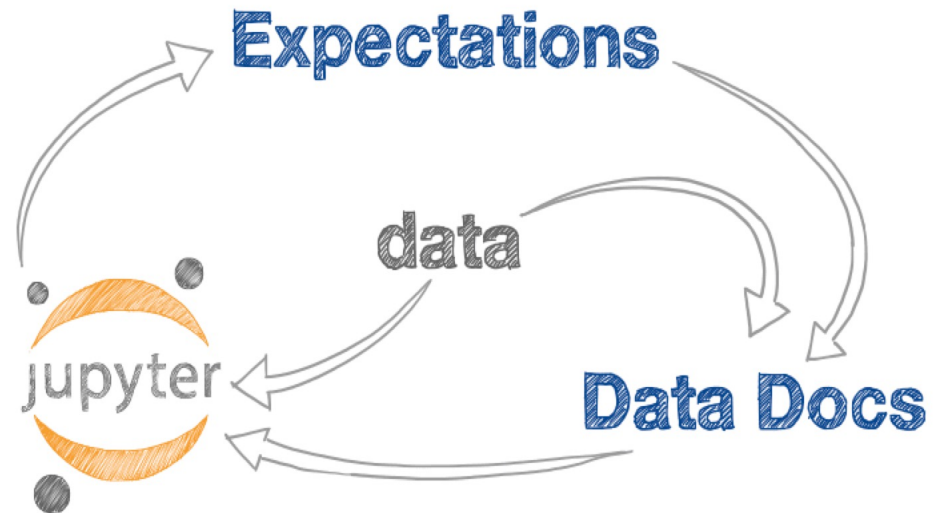
# What do you see as the most promising technology in the field of modern data platforms?

- Great Expectations
  - It is a Python library that allows you to validate, document, and profile your data to make sure the data are as you expected.

- Why is it promising?

It enables to:

- automatically create a set of tests for your data
- build your expectations using Jupyter Notebook
- create a comprehensive and human-readable documentation
- integrate well with DAG execution tools



Source: [greatexpectations.io](https://greatexpectations.io)

# What ten questions would you ask to be able to work on the project as quickly as possible?

1. What are the primary objectives, goals, KPIs and scale of a project?
2. Who are the key stakeholders and their needs/expectations?
3. What are the project management framework and decision making processes?
4. What is the timeline of the project?
5. What are the tech and data stacks?
  - Are there any technical or tooling requirements?
6. What is the scope of the project and overall budget?
7. What are my and other team member's roles and responsibilities? What is expected from me?
8. What are the priorities, and if potential risks identified, what are those?
9. How communication between team members and relevant stakeholders are organized (format, frequency, duration, medium)?
10. What are the reporting and technical documentation requirements and scheduling?