Get unlimited access to the best of Medium for less than $5/week. Become a member

Alchemesh: Data mesh framework — The genesis

Alexander Gutton · Following
Published in Alchemesh · 5 min read · May 25, 2021

Alchemesh

Data product hub

An data becomes increasingly vital to decision-making, many companies are rethinking their organization to embrace data. After a week of reading, I've been inspired by the concepts of data mesh and the idea of treating data as a product. Inspired by the principles of data mesh, Alchemesh tell me here, at the start of a new journey: Building a Data Mesh framework.

Data mesh is a decentralized sociotechnical approach to share, access, and manage analytical data in complex and large-scale environments — within or across organizations. Data mesh is a paradigm that helps organizations realize amazing robust governance and a product-centric approach.

However, implementing a Data Mesh presents numerous challenges and requires platform support.

Data Mesh: Beyond Technology

Contrary to popular belief, Data Mesh is not just about reorganizing teams. It's a blend of organizational changes, decentralized governance, and sophisticated and extensible platform. Data Mesh represents a profound transformation in the interactions between people, the technical architecture and solutions in the organization, based on principles:

- **Domain ownership:** Decentralize analytical data ownership to business domains closer to the data source or data consumers, and manage the data lifecycle independently based on those domains. This approach aligns business, technology and data, enabling scalability, agility, accuracy, and resiliency by reducing bottlenecks and ensuring localized change management.
- **Data as a product:** Domain-oriented data is shared as a product directly with data users, adhering to interoperable standards for discoverability, understandability, accessibility, security, and governance, ensuring interoperability, compatibility, increase value, and security. Each autonomous data product provides explicit, easy to use data sharing contracts and is managed independently, introducing the concept of a "data quantum" that encapsulates all necessary components for data sharing, aiming to provide data silos, foster a data-driven culture, and unlock business insights.
- **Self-serve data platform:** Platform-based infrastructure that enables data by managing the full lifecycle of data products and creating a reliable system of interconnected products, simplifying data discovery, access, and use. It aims to reduce the cost of decentralized data ownership, shorten data management complexity, engage a broader range of developers, and automate governance for security and compliance.
- **Federated computational governance:** A federated governance model with domain representatives, data platform members, and experts to balance local autonomy with global interoperability. This approach ensures data policy enforcement. It aims to derive value from interoperable data products, mitigate decentralization risks, integrate governance requirements, and reduce manual synchronization overhead.

Supporting the Transition to a Data Mesh

Implementing a Data Mesh is a complex and evolving process. Companies must adapt their organizational structures and technical capabilities. As new technologies emerge and organizations mature in their Data Mesh implementation, concepts and practices must evolve.

Data Mesh is far from a static solution. It must continually adapt to new cultures and technological advancements. Companies adopting this approach must be prepared to continually review and adjust their practices and tools.

Many challenges to address

When you start to data drive in Data Mesh implementation, you start to realize the every growing number of challenges that you will have to deal with, like:

- **Data Contracts:** These become crucial for formalizing dependencies between teams and their products. Data contracts clarify expectations and responsibilities, ensuring effective communication and collaboration.
- **Polysemes:** These elements enable different data products to interpret shared data differently, accommodating diversity and data consistency across the organization.
- **Data Products:** At the core of Data Mesh, these data products must be properly documented, maintained, and owned by teams. This includes defining metadata, quality standards, and mechanisms for updates and versioning.

Challenges of Autonomy

While essential, team autonomy inevitably leads to divergences in the technologies used and the best practices adopted. Some might be tempted to centralize decisions via a single platform. Technical stack (e.g. a JVE project with an Airflow instance). However, this might simply shift the bottleneck from data teams to platform teams, hindering the overall productivity by allowing one technical fix data product and providing a platform that fosters this dynamic.

This technological diversity can be seen as an asset if well managed, allowing each team to choose tools that best meet their specific needs. However, this can lead to increased complexity. However, it is essential to establish standards and best practices to ensure consistency and interoperability of the implemented solutions.

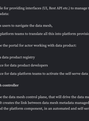
Our Vision: A Framework for Data Mesh

Given these insights and lacking open-source or proprietary alternatives about transitioning from the Modern Data Stack to Data Mesh principles, I decided to develop a framework for Data Mesh governance. The goal is not to offer an off-the-shelf product but to provide a flexible and modular tool. The framework includes:

- **Standardize Interfaces:** Offer a common working framework for data domains, data products, output ports, data contracts, etc., thus facilitating evaluation and understanding.
- **Support Platform Teams:** Assist in creating self-serve data platforms through component standardization while enabling implementation flexibility.
- **Provide Modular Components:** Supply "Lego-like" platform components, allowing teams to choose from commonly used components or have your team enact to translate the data mesh resources in platform.

This framework is designed to be modular and adaptable, enabling companies to act according to their specific needs. Whether it's implementing a data contract, creating a template for defining business definitions, the framework aims to provide a solid foundation for implementing and managing a Data Mesh.

Alchemesh: Layers



The Alchemesh framework will be composed of three layers:

Alchemesh core

Responsible for providing interfaces (UI, Rest API etc.) to manage the data mesh metadata.

- enables users to navigate the data mesh,
- allows platform teams to translate all this into platform provisioning.

This will be the portal for actor working with data product.

- Act as a data product registry
- Interface for data product developers
- Interface for data platform teams to activate the self-serve data platform

Alchemesh controller

This will be the data mesh control plane, that will drive the data mesh platform. It creates the link between data mesh metadata managed by the core and the platform component, in an automated and self-service manner.

Alchemesh platform components

Set of 'LEGO' Packaging Platform Components for half-brevity. The platform components are divided into several categories:

- **Infrastructure Platform Component:** Defines the platform foundation to ensure data mesh support, such as provider gateways, VMs, registries, Kubernetes clusters, and ...
- **Output Port Platform Component:** Automates storage components on the infrastructure to expose data produced by data products, ensuring interoperability and access management.
- **Input Port Platform Component:** Automates components to consume data from operational systems and makes it available to the data product infrastructure, allowing the associated code to be formed to and produce the value of the data product.
- **Code Platform Component:** Automates the business logic on the infrastructure, enabling the use of incoming data to produce the desired outcomes.

Open source

Each of these three layers can be used independently and partially! With the probability of replacing each of the solutions with custom, depending on the use to be made of each:

- The console part can be used as a metadata layer for the data mesh, then connected external via interfaces (Rest, GraphQL, Events) by the company's platform teams to integrate with their orchestration systems (GUIs, Gitlab controller, Kubernetes controller etc.) to create the link between the mesh metadata and the platform.
- The controller must be able to drive the platform components offered by Alchemesh as well as those produced by the organization using the solution.
- The platform components do not need to be specialized to meet the requirements of Alchemesh or even part Data Mesh. Indeed, the goal is to provide a set of "Lego" blocks to conduct the specialized, IaC system, i.e. an infrastructure component that allows me to create a GKE cluster via Terraform, it must be usable to create a GKE cluster in a traditional Terraform enterprise environment without having to use the console or controller, and the same goes for an output port to manage storage and data for an output port to manage storage and access rights on Hashicorp.

Conclusion

Data Mesh represents a profound transformation in data management, requiring a shift in mindset and organizational capabilities. With this framework, we aim to provide a flexible and modular solution that supports this transition by offering standardized tools and interfaces while maintaining team autonomy. We must, at our level, to participate in the emulation and reflection around Data Mesh to try to identify the best thinking to help leverage the benefits of Data Mesh, successfully navigating the challenges of its implementation.

We are still in the early development of this framework based on our experiences in Data Mesh. Implementing a product also gives us the Framework to grow our reflection, starting from core concepts (e.g. data domains, data products, data contracts, etc.) to including them with features that are critical for data mesh, such as data quality, observability, feedback loops, etc.). This series of articles will allow us to share our reflections and decisions we had to predict with the community!

In the next article, we will focus on the north star architecture we currently have to develop this framework and then present to you the resource modeling (data products, technical stacks, etc.) we have for our MVP.

To wrap our product a little, here are some sketches of the Alchemesh controller:

Data Mesh Hub

Data product overview

Data product hub

Data platform component hub

Written by Alexander Gutton
Data Engineer | Data Vis Advocate
Lead Data Mesh - Alchemesh

Follow 1K

More from Alexander Gutton and Alchemesh

Alexander Gutton · 2 days ago
A Reference Operator for Apache Airflow
Data quality manage your data processing at scale with Airflow Alchemesh Operator

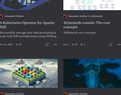
Mar 16, 2021 · 5 min read

Alexander Gutton in Alchemesh · 2 days ago
Alchemesh: Concepts The core concepts
Alchemesh core concepts

Jun 1, 2021 · 5 min read

Alexander Gutton in Alchemesh · 2 days ago
Alchemesh: Data mesh implementation on GCP (Part1)
Alchemesh: Data mesh implementation using Hashicorp for provisioning and Dagster ...

Aug 2, 2021 · 5 min read

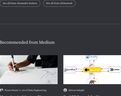


Alexander Gutton · 2 days ago
Jira/Gitlab Beyond the Hype
Jira/Gitlab Beyond the Hype Alchemesh Jira/Gitlab Beyond the Hype Alchemesh ...

Jun 1, 2021 · 5 min read

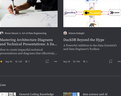
See all from Alexander Gutton · See all from Alchemesh

Recommended from Medium



Data Engineer in Data Engineering
Mastering Architecture Diagrams: Tools, Techniques, and Tips of the Trade
Mastering Architecture Diagrams: Tools, Techniques, and Tips of the Trade ... presentation and diagrams that effectively ...

Mar 31 · 10 min read



Ashish Kumar
Jira/Gitlab Beyond the Hype
Jira/Gitlab Beyond the Hype Alchemesh Jira/Gitlab Beyond the Hype Alchemesh ...

Jun 1 · 5 min read

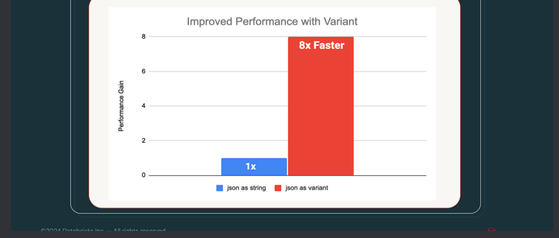
Lists

Stories to Read with Medium
Curated by Medium Staff · 10 stories · 1 follower

Product Management
Product Management Fundamentals
Curated by Adrienne Tan · 18 stories · 1 follower

Generative AI Recommendation Systems
Curated by Vaibhav Gupta · 19 stories · 1 follower

Data Science
Data Science Tools
Curated by Jeff Hale · 19 stories · 1 follower

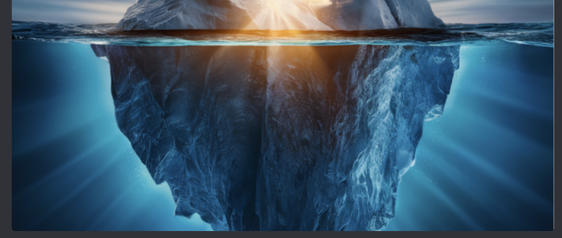


Archana Goyal

What's Next for Apache Spark 4.0: A Comprehensive Overview with...

Apache Spark has established itself as a leading platform for big data processing, a...

Aug 25 379 2



Shawn Gordon

What The Heck is Apache Polaris?

Introduction

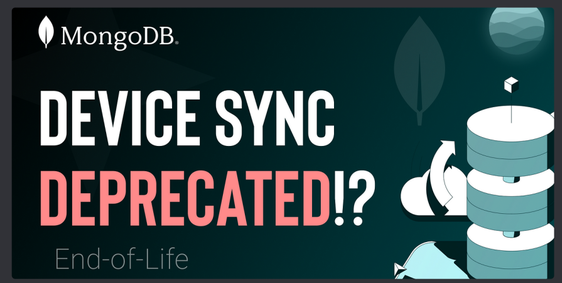
Sep 12 7

Jagadesh Jamjala

Essential Data Engineering Tools in 2024

Are you keeping up with the latest tools in data space?

May 26 129 1



Stevdza-San

This is BAD! MongoDB is Shutting Down their Services? 😭

About

Sep 12 759 38

See more recommendations