

DATA+AI
SUMMIT 2022

The Modern Metadata Platform

What, Why, and How?

ORGANIZED BY  databricks



Mars Lan
CTO, Metaphor Data

Agenda

- What is the Modern Metadata Platform (MMP)?
- Why do you need a MMP?
- How to build a great MMP?



Metaphor Data CTO & co-founder

DataHub co-creator

Ex-LinkedIn metadata team TL

What is MMP?

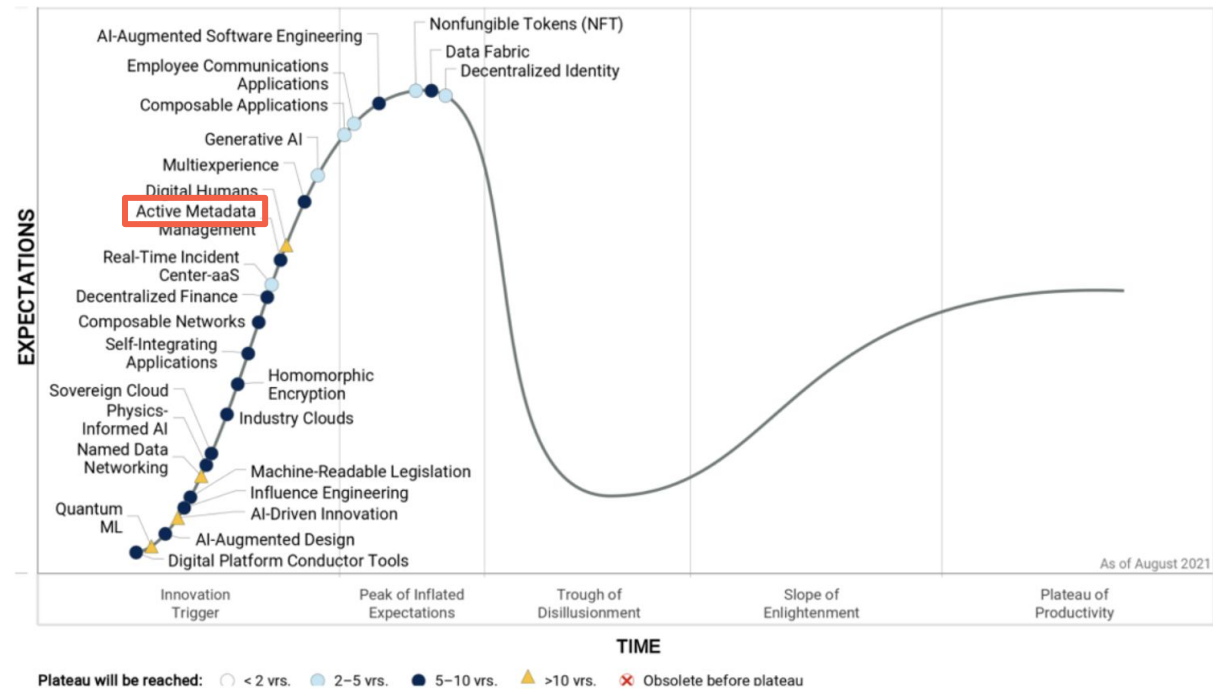
Industry Trends

Gartner: "Active Metadata Management"

An emerging set of capabilities across multiple data management markets resulting from continuous metadata management innovation.

Make metadata actionable

Figure 1. Hype Cycle for Emerging Technologies, 2021



Source: Gartner (August 2021)

747576

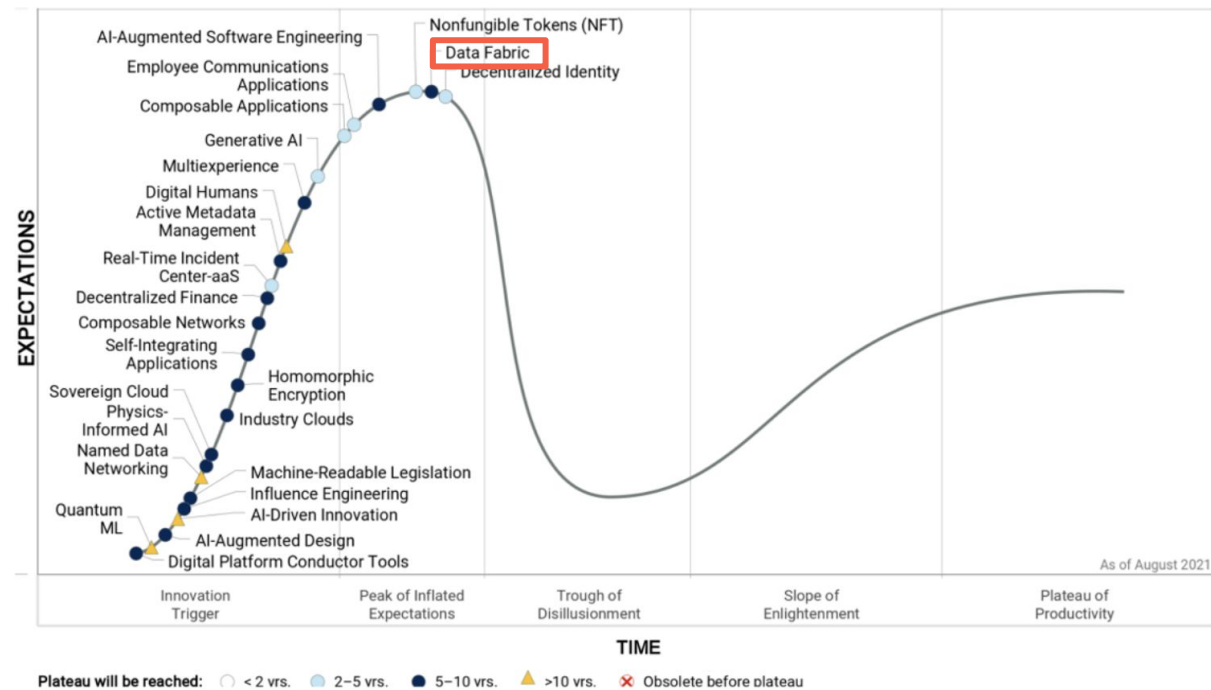
Industry Trends

Forrester: "Enterprise Data Fabric"

Focuses on automating the process integration, transformation, preparation, curation, security, governance, and orchestration to enable analytics and insights quickly for business success.

Enable self-service through metadata

Figure 1. Hype Cycle for Emerging Technologies, 2021



Source: Gartner (August 2021)

747576

Metadata, Metadata, Everywhere

- Use cases

- Data discovery
- Data lineage
- Data observability
- Data governance
- Data privacy
- Data/ML Ops

- New paradigms

- Data OS
- Data Mesh

Metadata Management

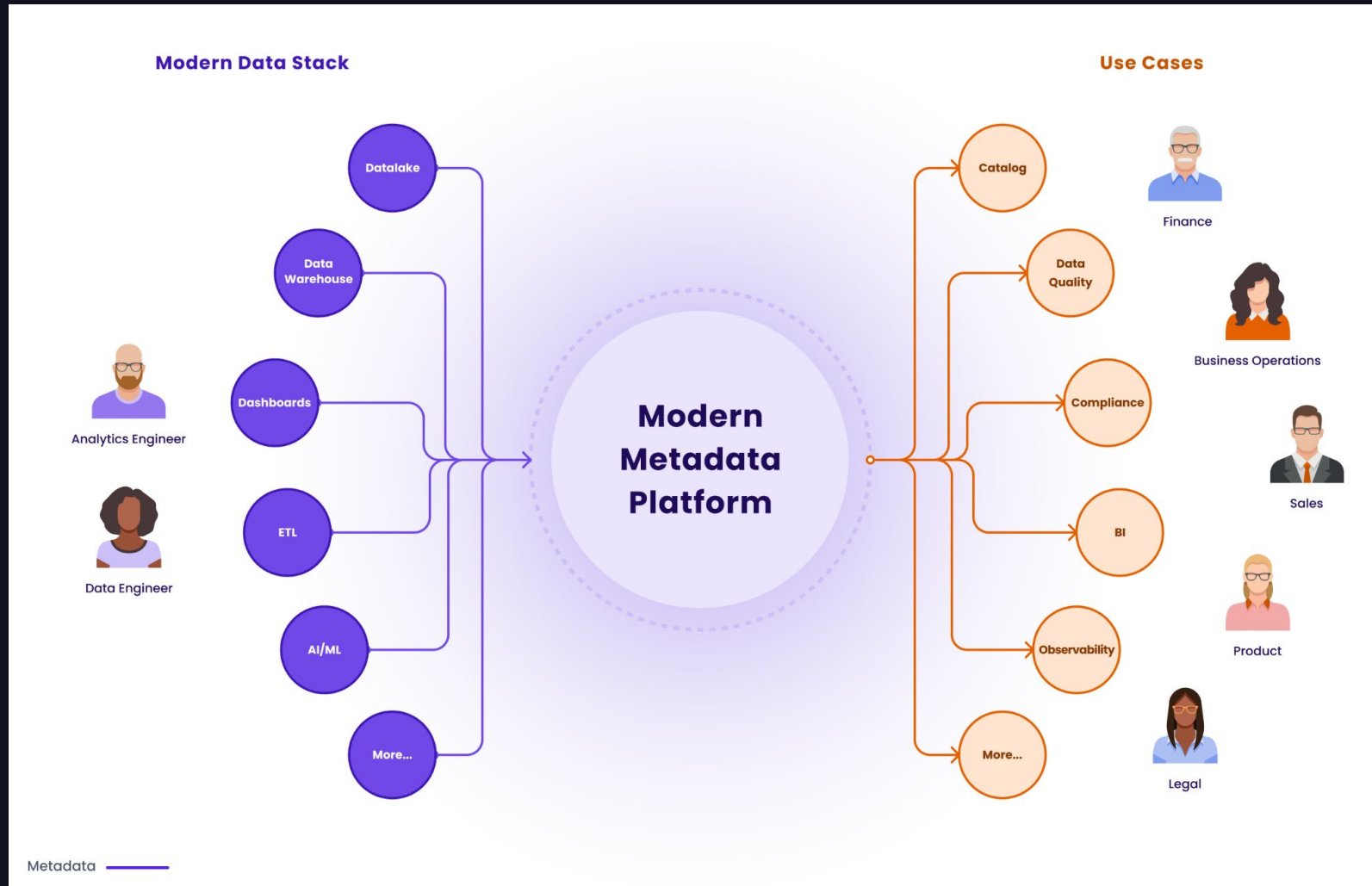
- Before MDS

- Only concerns data eng
- Operational use cases
- Bundled with ETL
- Homogenous stack ⇒ Well integrated metadata

- After MDS

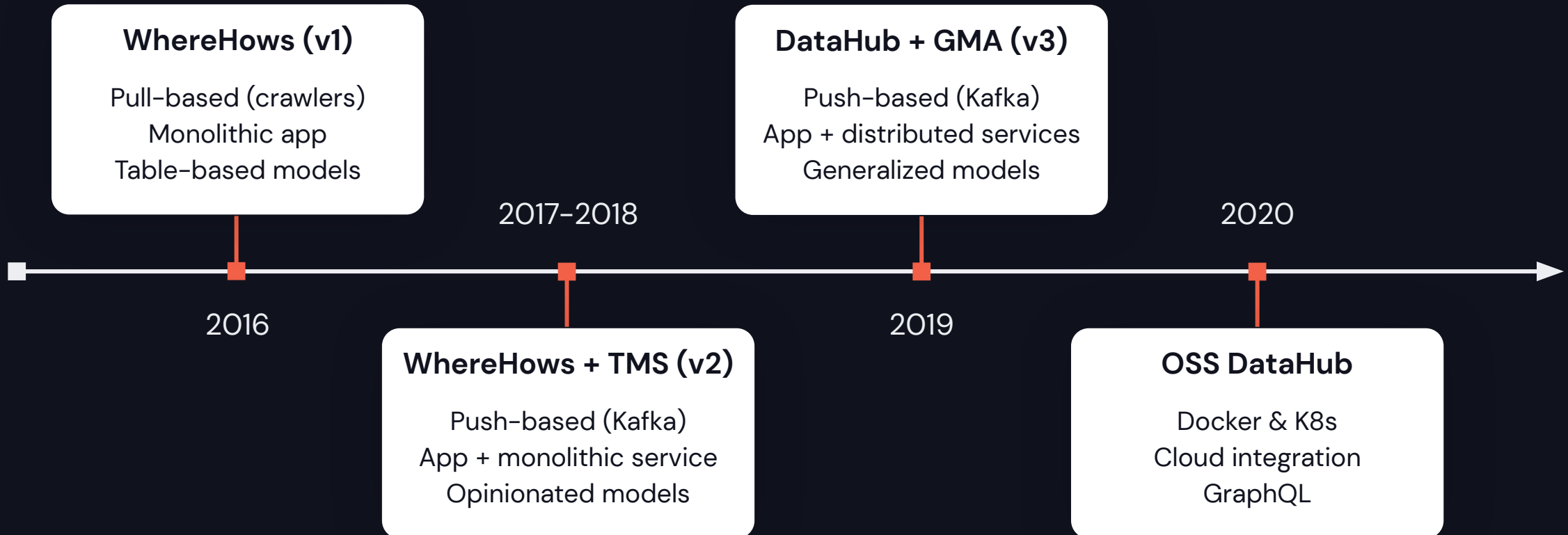
- Enabler for cross-team collaboration
- Operational + organizational use cases
- Stand-alone platform
- Heterogenous stack ⇒ Metadata silos

MMP – The Missing Link



LinkedIn's Metadata Journey

From monolithic app to metadata platform



Why Do You Need a MMP?

Why Do You Need a MMP?

Specifically, why a “modern” version?

- Complexity
 - Democratized, distributed, fragmented
 - Best-of-breed SaaS tools
 - Open-source solutions
 - Proprietary business definitions/concepts
- Scale
 - Every schema, column, dataset, dashboard, query, job run, access history, ...
 - Graph with 100s of millions of vertex & edges
 - Small data problem?

Why Do You Need a MMP?

“Your metadata can be as large and as complex as your data and deserve to be treated with the same respect.”

How to Build a Great MMP?

How to Build a Great MMP?

A great metadata platform \approx a great data platform

- Scalable
- Reliable
- Extensible
- Rich APIs
- Easy to integrate

Scalability

- Transactional DB is often the bottleneck
- Storing large amounts of data is easy
 - NoSQL, NewSQL
- Support all types of queries is hard
 - Complex joins
 - Graph traversal
 - Full-text search

Reliability

Reliable service \neq reliable data

- More machine = more reliable?
- Data reliability
 - Real-time sync between systems
 - Zero-downtime rebuild/bootstrapping
 - Full change history & audit trail

Extensibility

Extensible data model

- Loosely typed
 - Key-value pairs
 - Unlimited extensibility
 - Easy to write, hard to read
- Strongly typed
 - Schema evolution
 - All fields are optional (proto3)

Rich APIs

REST alone is not enough

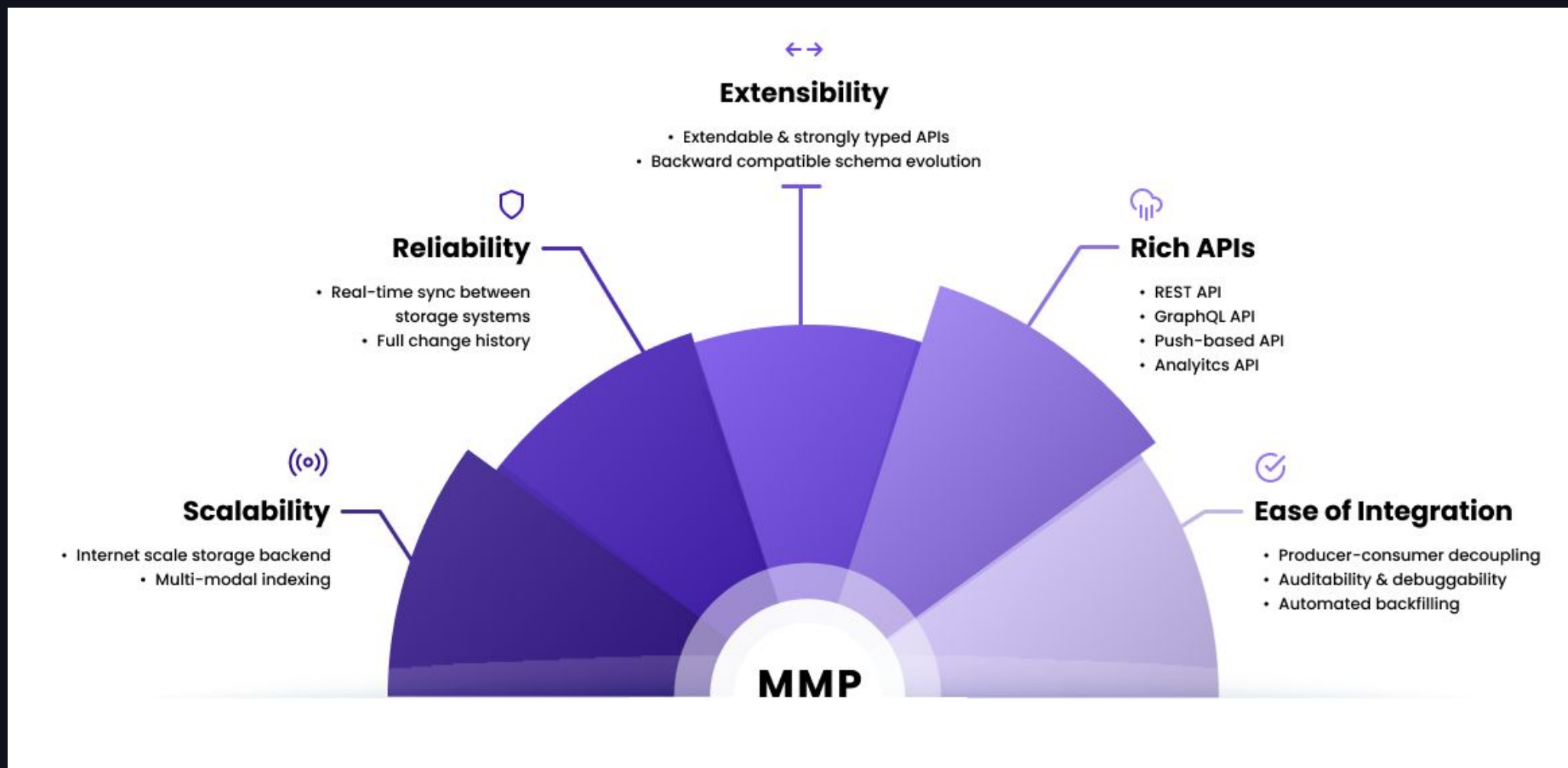
- REST API
 - Programmatic access
- GraphQL API
 - App integration, graph traversal
- Push-base API
 - Real-time, reactive
- Analytics API
 - OLAP

Ease of Integration

Metadata ETL

- API versioning
- Auditability & Debuggability
- Backfilling

How to Build a Great MMP



Summary

- What is MMP?
 - MDS + MMP = business & organizational challenges solved
- Why do you need a MMP?
 - Complexity & scale
- How to build a great MMP
 - Scalability, reliability, extensibility, rich APIs, ease of integration

DATA+AI
SUMMIT 2022

Thank you



Mars Lan

Twitter: @mars_lan

LinkedIn: [linkedin.com/in/marslan](https://www.linkedin.com/in/marslan)