



# Accelerating Insight Creation with Data Mesh

April 2024

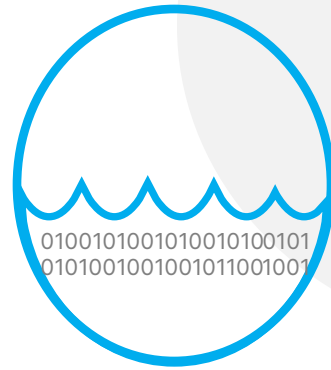
# Journey to a Democratized Mesh: Overview



Leading Financial  
Research and  
Data Solutions



1984



Centralized Data  
Lake  
and Lakehouse



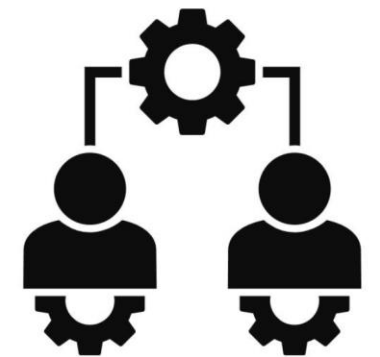
2018



Self-Service  
Platform



2021



Data Mesh  
Operating Model



2024

# About Morningstar

## Empowering Investor Success

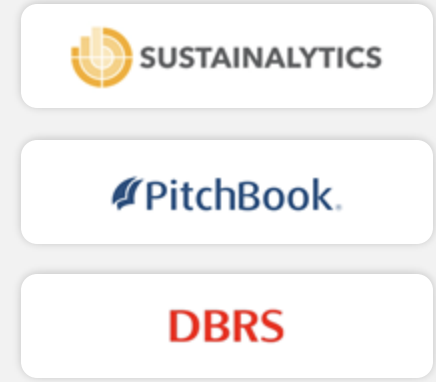


## 40 Years and Counting

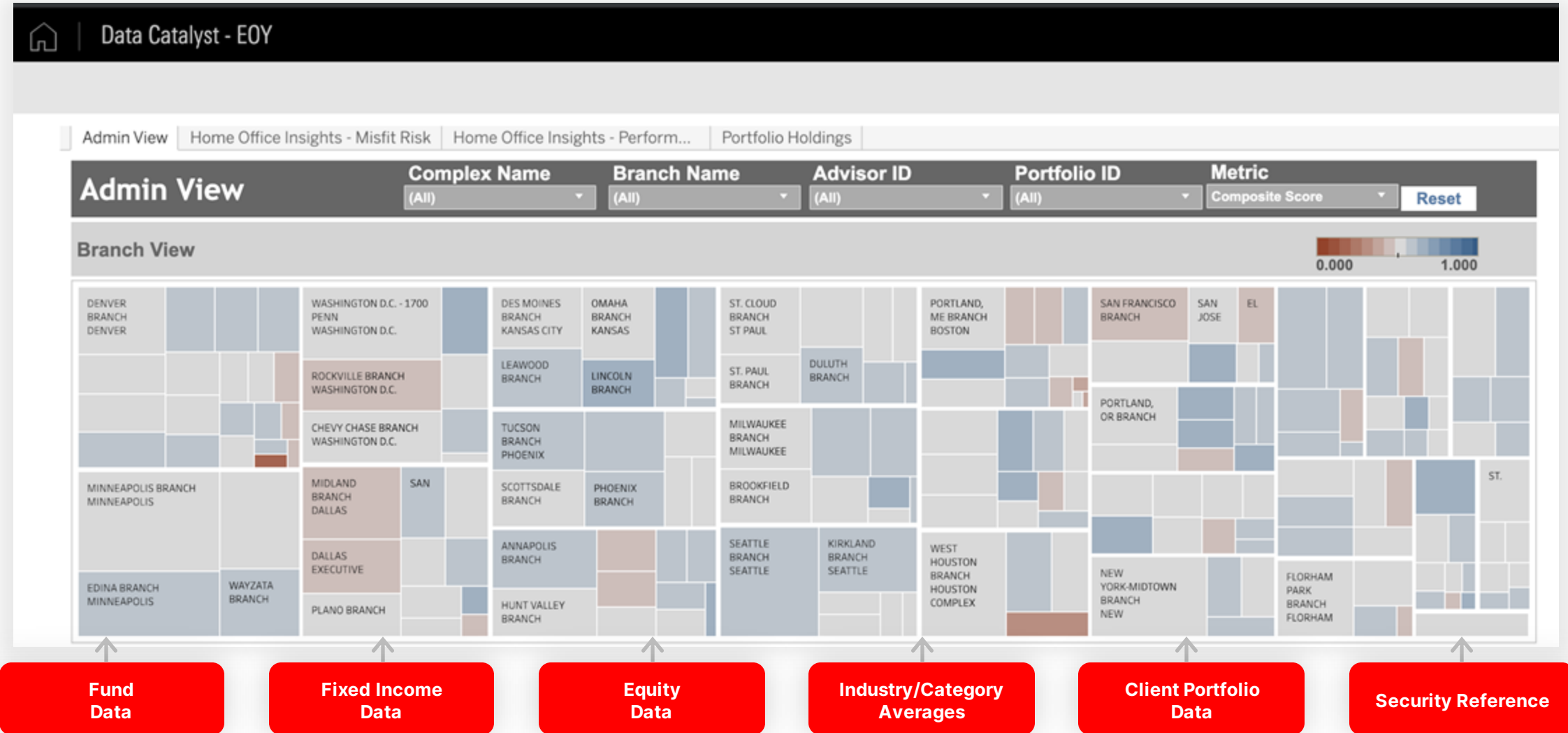


Morningstar Style Box

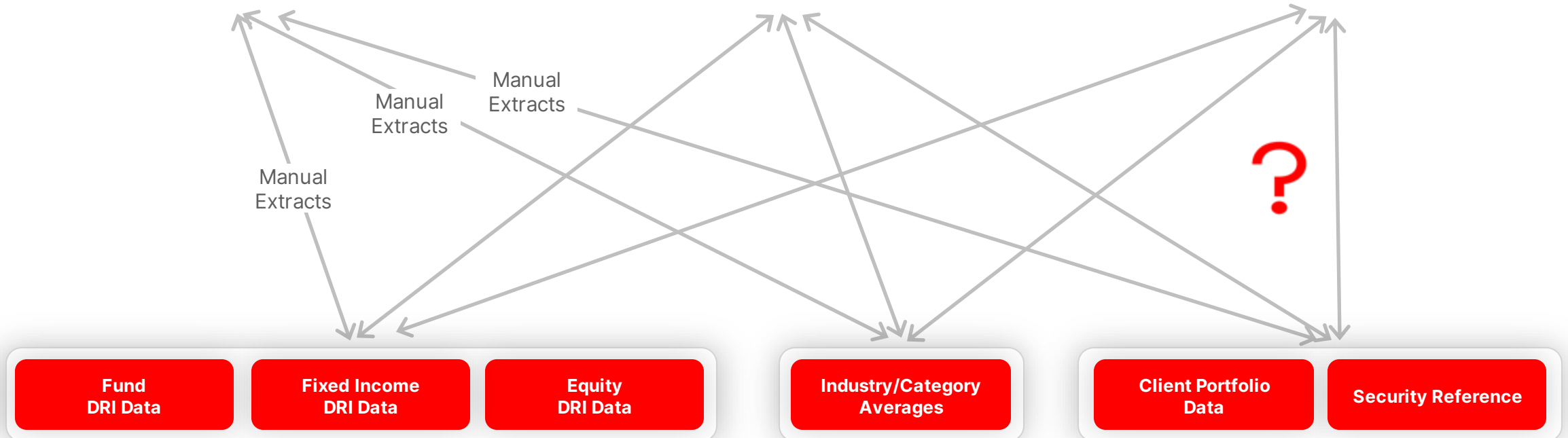
|       |       |        |        |
|-------|-------|--------|--------|
|       |       |        | Large  |
|       |       |        | Medium |
|       |       |        | Small  |
| Value | Blend | Growth |        |



# Integrating Diverse Datasets: Morningstar Data Catalyst

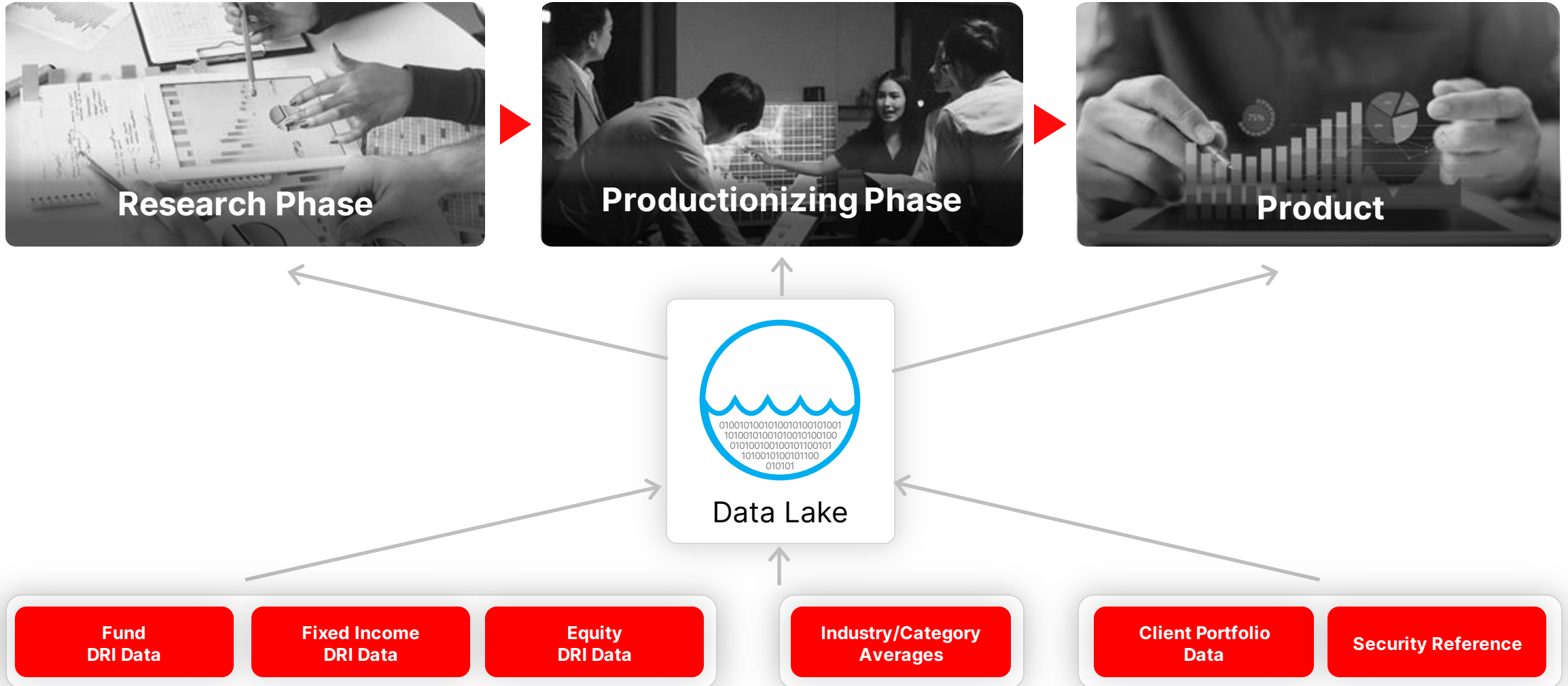


# Fragmented Data Sources -> Handoffs



# Started our data journey by centralizing data

Morningstar's data lake



# Making Data Consumer Friendly

Morningstar's Lakehouse

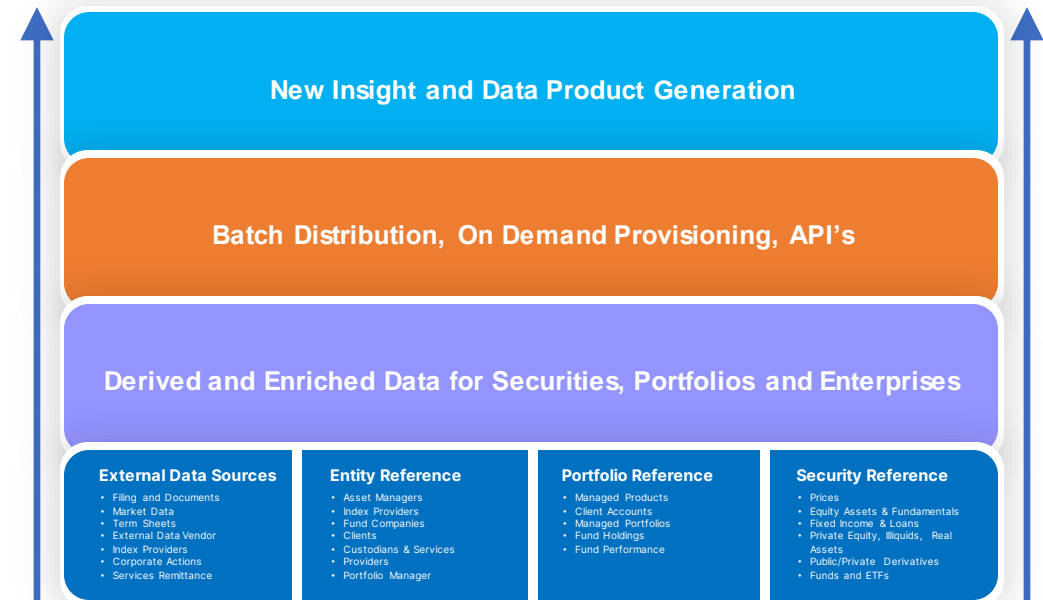
## Redshift

addresses query speed



## Modeling

addresses data ambiguity



# Realization of Problem - Technology Alone is not the Solution



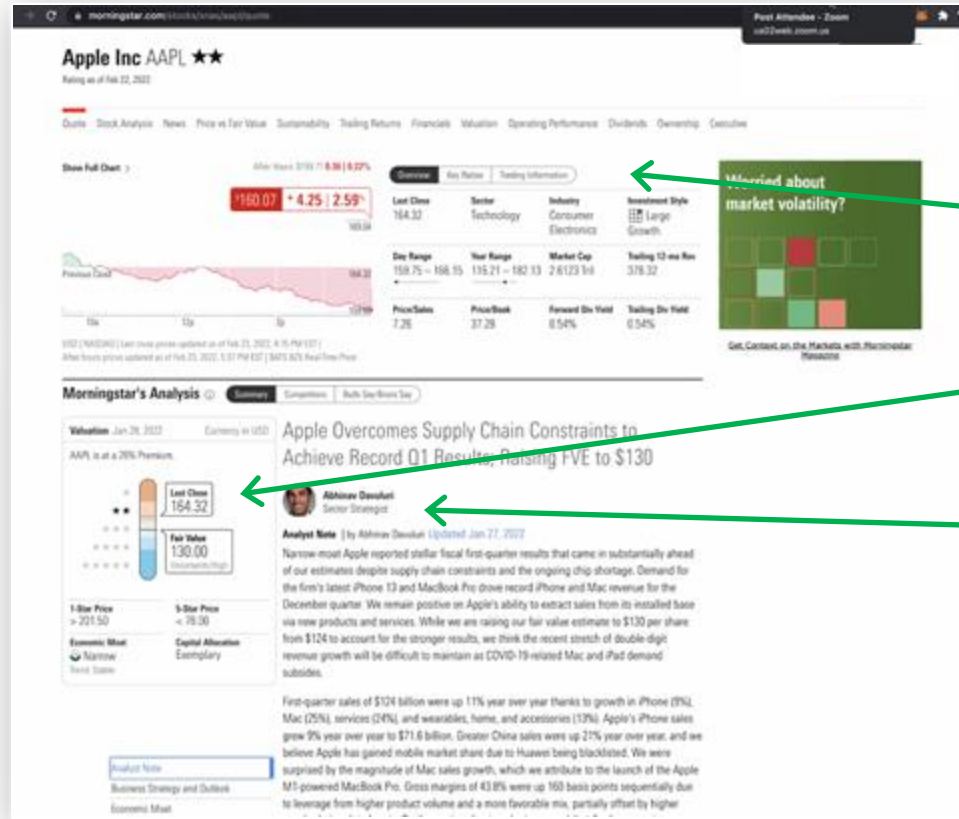
## Technology

Supported by:  
People and Research



## People & Research

Supported by:  
Technology



Technology

Data

Research

People



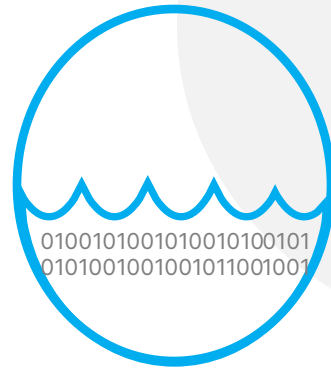
# Journey to a Democratized Mesh: Self-Service



Leading Financial  
Research and  
Data Solutions



1984



Centralized Data  
Lake  
and Lakehouse



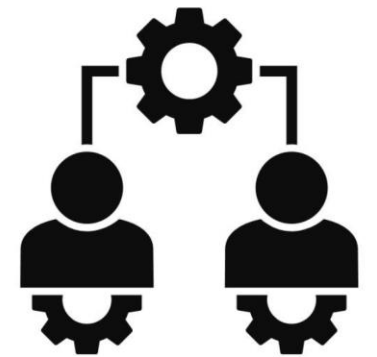
2018



**Self-Service  
Platform**



2021



Data Mesh  
Operating Model



2024

# Problem - How Do We Scale Through Our People?

|                           | Data Ingestion  | Data Quality   | Data Modeling  |
|---------------------------|---|--|--|
| <b>People Processes</b> ▶ | <u>Content specialists</u> and <u>developers</u> create ETL pipelines.. | <u>Business analyst</u> , <u>domain SME's</u> and <u>developers</u> set validation rules...  | <u>Content specialists</u> and <u>DBAs</u> define modeling...          |
| <b>How?</b> ▶             | ... back and forth communication and handoffs                           |  |  |
| <b>Scaling Issues</b> ▶   | 1000s of lines of ETL code to write                                     | Validation rule volume is unruly to manage <ul style="list-style-type: none"><li>• Sheer volume is an issue</li><li>• But validation rules can change as market does</li><li>• 10's of 1000's of tests</li></ul> | Keeping methodology and code in sync is problematic<br><br>1000+ views |

# Benefits - Data Ingestion

Self-service beyond the specialists



```
AWS Glue > Developer Guide
root
|-- role: string
|-- seats: int
|-- org_name: string
|-- links: array
|   |-- element: struct
|   |   |-- note: string
|   |   |-- url: string
|   |-- type: string
|-- sort_name: string
|-- area_id: string
|-- images: array
|   |-- element: struct
|   |   |-- url: string
|-- on_behalf_of_id: string
|-- other_names: array
|   |-- element: struct
|   |   |-- note: string
|   |   |-- name: string
|   |   |-- lang: string
|-- contact_details: array
|   |-- element: struct
|   |   |-- type: string
|   |   |-- value: string
|-- name: string
|-- birth_date: string
|-- organization_id: string
|-- gender: string
|-- classification: string
|-- death_date: string
|-- legislative_period_id: string
|-- identifiers: array
|   |-- element: struct
```

**AWS Glue  
Code + Airflow  
Complexity**



etleap Home Activities Connections Search + CREATE Demo User

← BACK Create Pipeline

| event_type | referrer                                      | ipaddress    | joindate   | cookie               | event                     |
|------------|---|--------------|------------|----------------------|---------------------------|
| Click      | https://www.doggyswag.com/about               | 67.234.12.43 | 2016-03-19 | NLDWBHYZLCL9ZX1UC6VB | 2019-05-07T18:42:55-07:00 |
| Click      | https://www.doggyswag.com/catalog/rubberballs | 67.234.12.43 | 2019-03-16 | GTJWE050Q3QPSSFXXYDX | 2019-05-07T16:24:30-07:00 |
| PageLoad   | null  | 67.234.12.43 | 2015-12-12 | 9D3DCUEN9XF8HECD37FA | 2019-05-07T05:52:42-07:00 |
| PageLoad   | null  | 213.34.56.96 | 2017-01-13 | M0TG7RKOON8DC684CS6F | 2019-05-07T16:43:31-07:00 |
| PageLoad   | null  | 39.43.233.2  | 2016-03-10 | SKRMQHK6T1JFDHDCWE9V | 2019-05-07T07:31:58-07:00 |
| Click      | https://www.doggyswag.com/catalog/push toys   | 213.34.56.96 | 2015-04-30 | ZXNFKPZYKPMHPSV8KGI  | 2019-05-07T20:35:39-07:00 |
| PageLoad   | null  | 67.234.12.43 | 2015-12-09 | CE7MMXONS3LXZ2M9B98  | 2019-05-07T17:37:53-07:00 |
| PageLoad   | null  | 54.193.34.3  | 2015-02-19 | ELD3DCULTHFFKL7BMAJW | 2019-05-07T00:41:20-07:00 |
| Click      | https://www.doggyswag.com/account             | 12.13.200.32 | 2016-01-08 | HMP2QSGGAZXXHR2KNTDQ | 2019-05-07T12:03:39-07:00 |
| PageLoad   | null  | 54.193.34.3  | 2016-01-05 | WGAW8DWQZBKNOF39ZAME | 2019-05-07T07:08:47-07:00 |
| PageLoad   | null  | 39.43.233.2  | 2017-12-29 | 6R1KQXJAF68YB8CFQJRX | 2019-05-07T19:23:23-07:00 |
| PageLoad   | null  | 39.43.233.2  | 2018-12-02 | LHL1R88IAP8FXJ3YH3UH | 2019-05-07T17:49:57-07:00 |

Script

- 1 Split [data] repeatedly on newLine
- 2 Split [data] once on space character
- 3 Rename column split to event\_type
- 4 Flatten JSON object in split1 with all fields
- 5 Interpret joindate as date with format MM/dd/yyyy

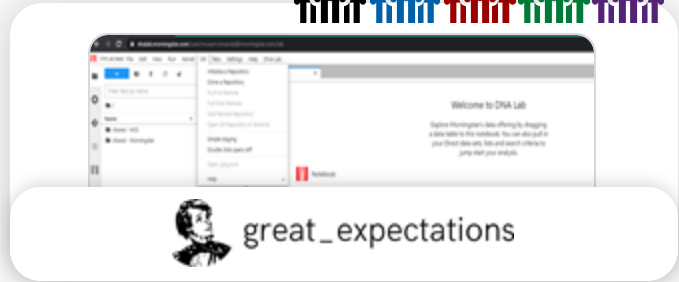
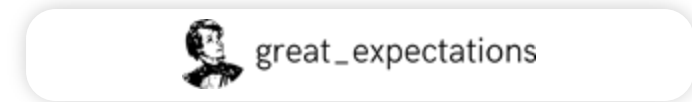
+ ADD SCRIPT STEP

NEXT



# Benefits - Data Quality

Increased input, visibility, integration



JupyterLab

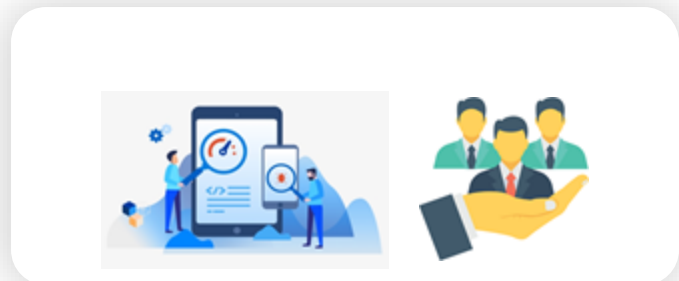


Fix

Update Test Suite



OR



Data SME / Client

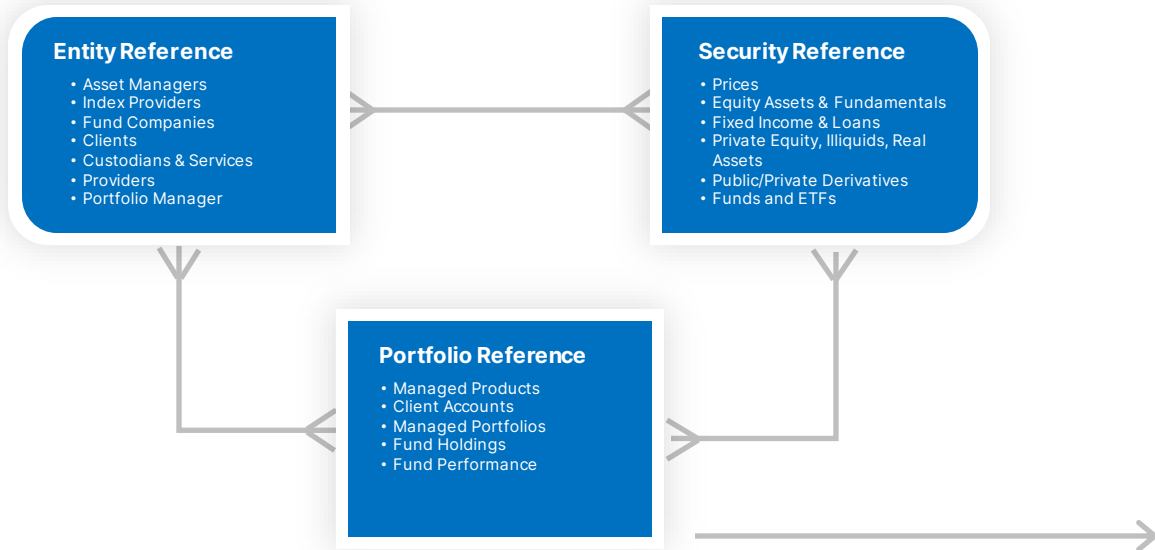


Reporting Dashboard

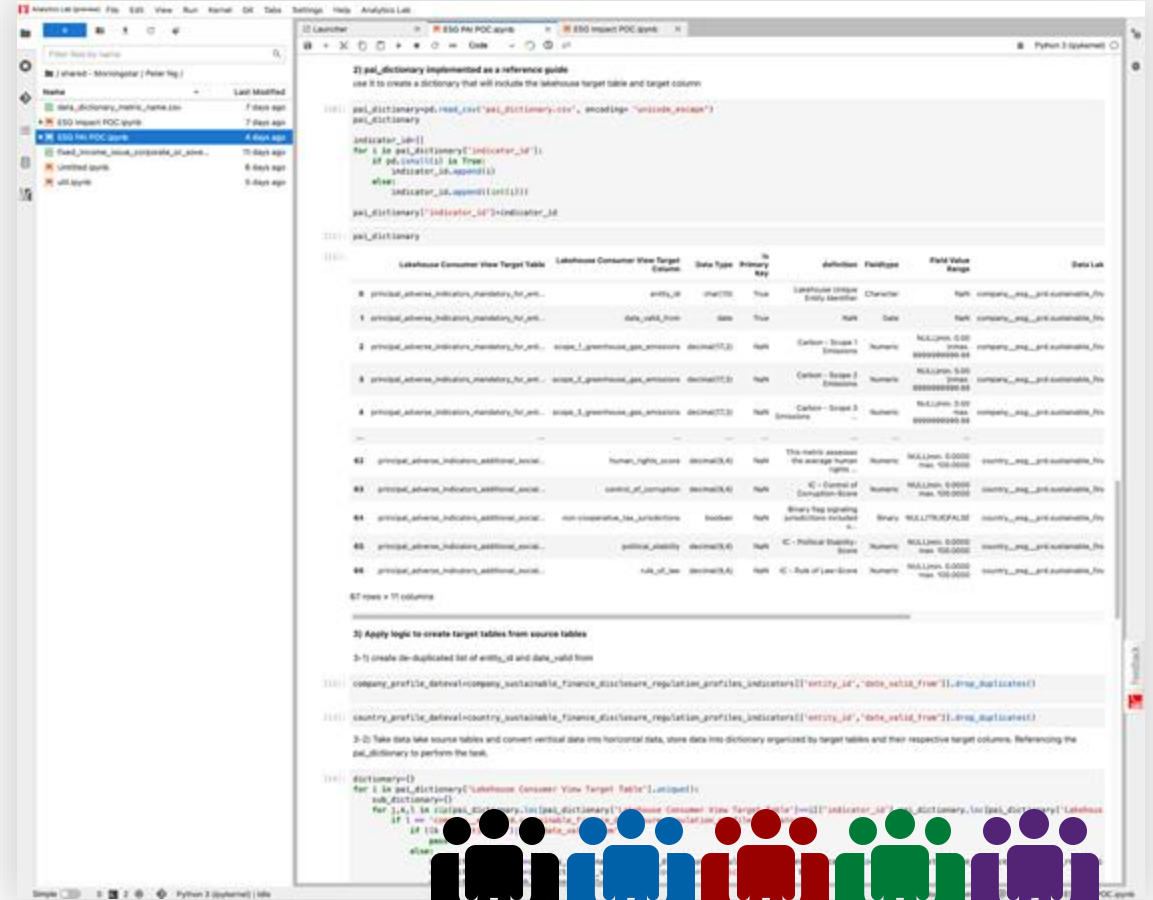


# Benefits - Data Modeling

Data SMEs informing data relationships



Tie model development with methodology documentation



# Making Data Consumer Friendly

Morningstar's Lakehouse

## Redshift

addresses query speed



**Data Lake**  
**etleap**

## Modeling

addresses data ambiguity

### Entity Reference

- Asset Managers
- Index Providers
- Fund Companies
- Clients
- Custodians & Services
- Providers
- Portfolio Manager

### Portfolio Reference

- Managed Products
- Client Accounts
- Managed Portfolios
- Fund Holdings
- Fund Performance

### Security Reference

- Prices
- Equity Assets & Fundamentals
- Fixed Income & Loans
- Private Equity, Illiquids, Real Assets
- Public/Private Derivatives
- Funds and ETFs

# Making Data Consumer Friendly

Morningstar's Lakehouse

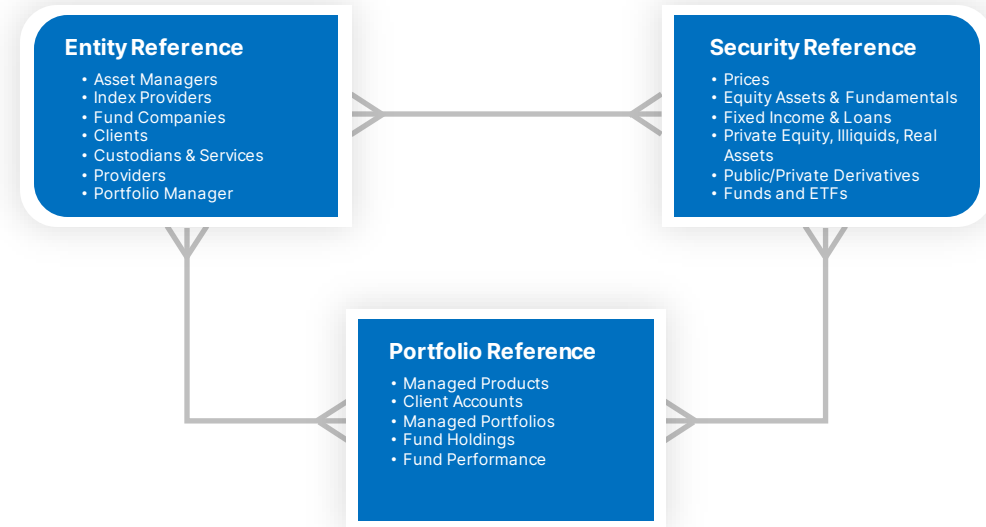
## Redshift

addresses query speed



## Modeling

addresses data ambiguity

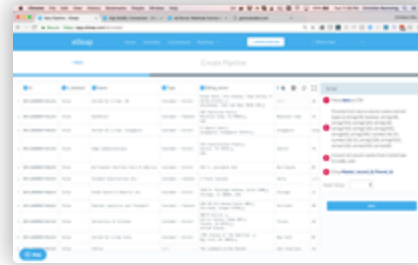


# Benefits - Scaling Through Self-Service

Solution



## Data Ingestion



## Data Quality



## Data Modeling



Results



ETL onboarding from weeks to minutes

Total production pipelines: **2900 and counting**

Dozens of teams producing ingestion

Current tests: **24k+**  
Current executions: **15mil**, with **2.2%** failure rate

Models: **600+** by core team, POC'ing Finance team

BA's: **Hundreds** scattered across the organization



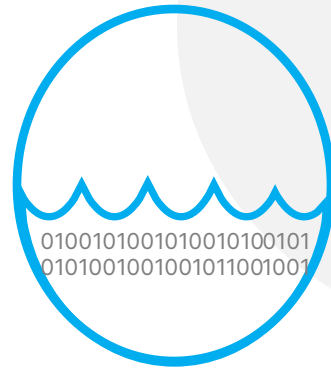
# Journey to a Democratized Mesh: Self-Service



Leading Financial  
Research and  
Data Solutions



1984



Centralized Data  
Lake  
and Lakehouse



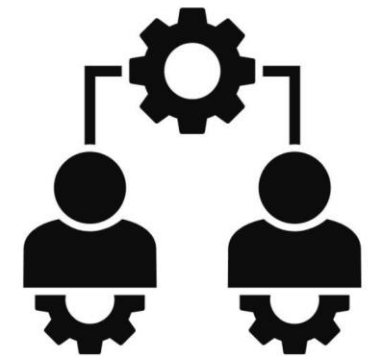
2018



Self-Service  
Platform



2021



**Data Mesh  
Operating Model**



2024

# Hueler + Morningstar Analytics

## Morningstar Acquires Hueler Analytics' Stable Value Data and Index

The addition of stable-value product data to Morningstar data increases visibility for assets in defined-contribution (DC) retirement plans



NEWS PROVIDED BY  
**Morningstar, Inc.** →  
03 Feb, 2020, 09:00 ET

SHARE THIS ARTICLE

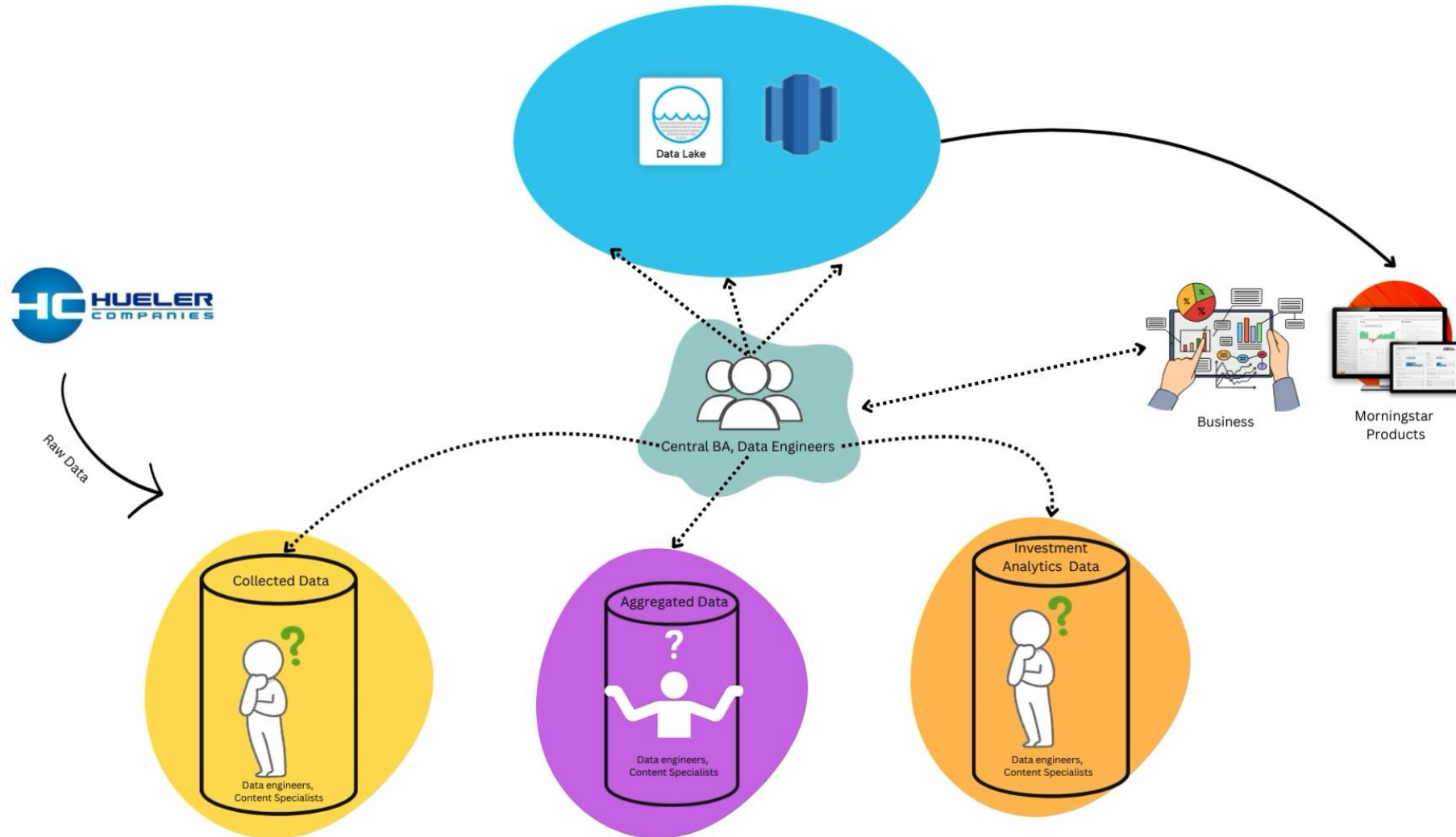


CHICAGO, Feb. 3, 2020 /PRNewswire/ -- **Morningstar, Inc.** (Nasdaq: **MORN**), a leading provider of independent investment research, today announced its acquisition of **Hueler Analytics'** Stable Value Comparative Universe Data and Stable Value Index. Terms of the transaction were not disclosed.

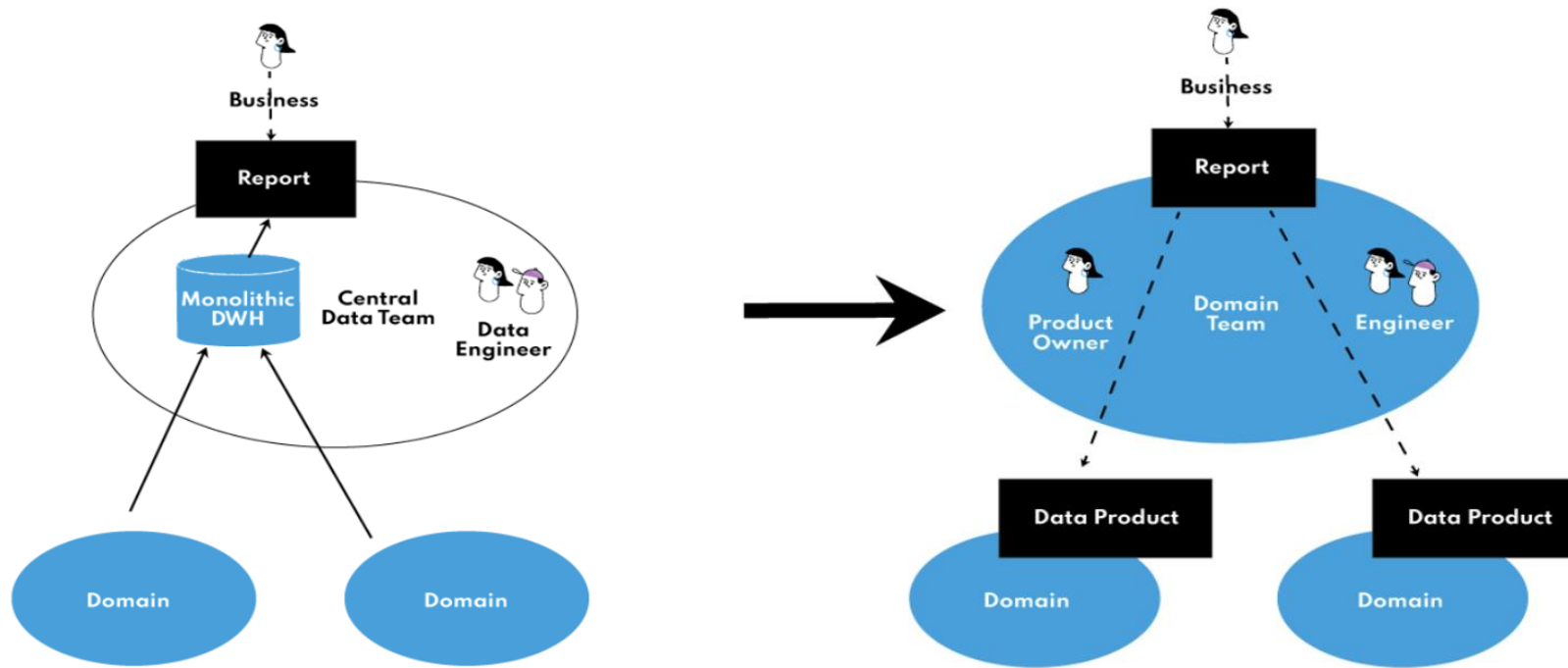
Founded in 1987 in Minneapolis, Minn., Hueler is the premiere independent data and research firm providing reporting and systems designed for the annuity and stable-value marketplace. Hueler Analytics' distribution encompasses advisors, investment managers, product providers, plan fiduciaries, and consultants. Hueler Analytics' Stable Value Comparative Universe Data provides broad market coverage of stable-value investments, including stable-value pooled funds, insurance



# A siloed operating model

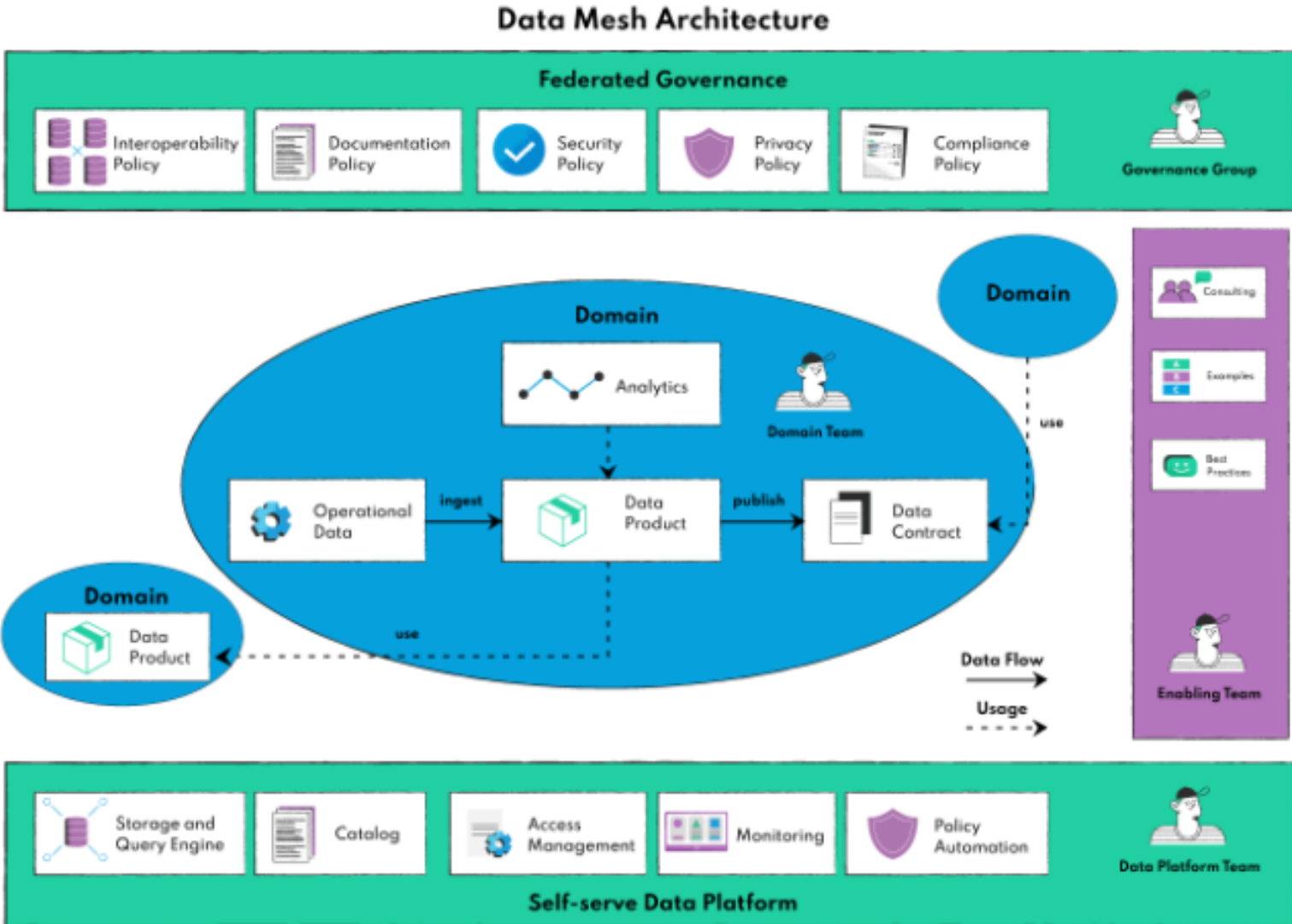


# Realization: headwinds in our operating model



“Data mesh is a decentralized sociotechnical approach to share, access, and manage analytical data in complex and large-scale environments—within or across organizations”  
- Dehghani, Zhamak

# Solution - Data Mesh Architecture



---

# Benefits

- **Democratizing Data / Promoting a Culture of Data Literacy**
- **Scalable, Agile and Autonomous**
- **Enhanced Data Quality**
- **Foster Innovation**

# Responsibilities of data product owner / domain team

| Traditional Agile Product Owner   | Data Product Owner   |
|---|--|
| What are the <b>workflows</b> I need to build for my users?                               | What are the <b>data views</b> I need to build for my users?                                       |
| This user wants access to my <b>software</b> product – let me ensure they have an account | This user wants access to my <b>data</b> product – let me ensure they are entitled properly        |
| I will sign-off after every release my <b>workflows are still functioning</b> properly    | I will sign-off after every release my <b>data exists in all products</b> properly                 |
| I need to open a P1 because <b>no users can login to my platform</b>                      | I need to open a P1 because <b>my latest data is not available for users in Lakehouse</b> to query |

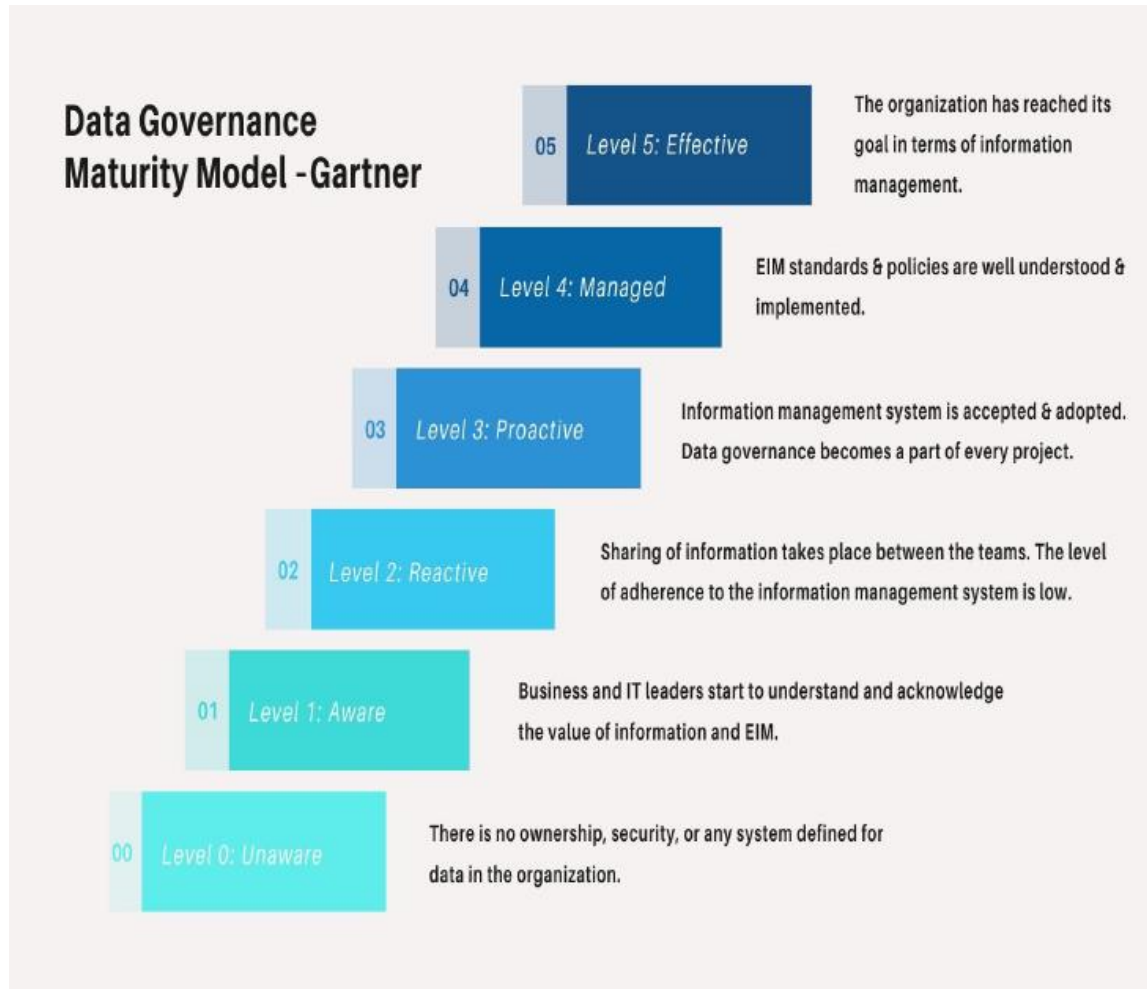
---

# Our approach

We cannot change what we are not aware of, and once we are aware, we cannot help but change  
– Sheryl Sandberg



# Step 1: Measuring Maturity



## An Example – Data Quality (Availability)

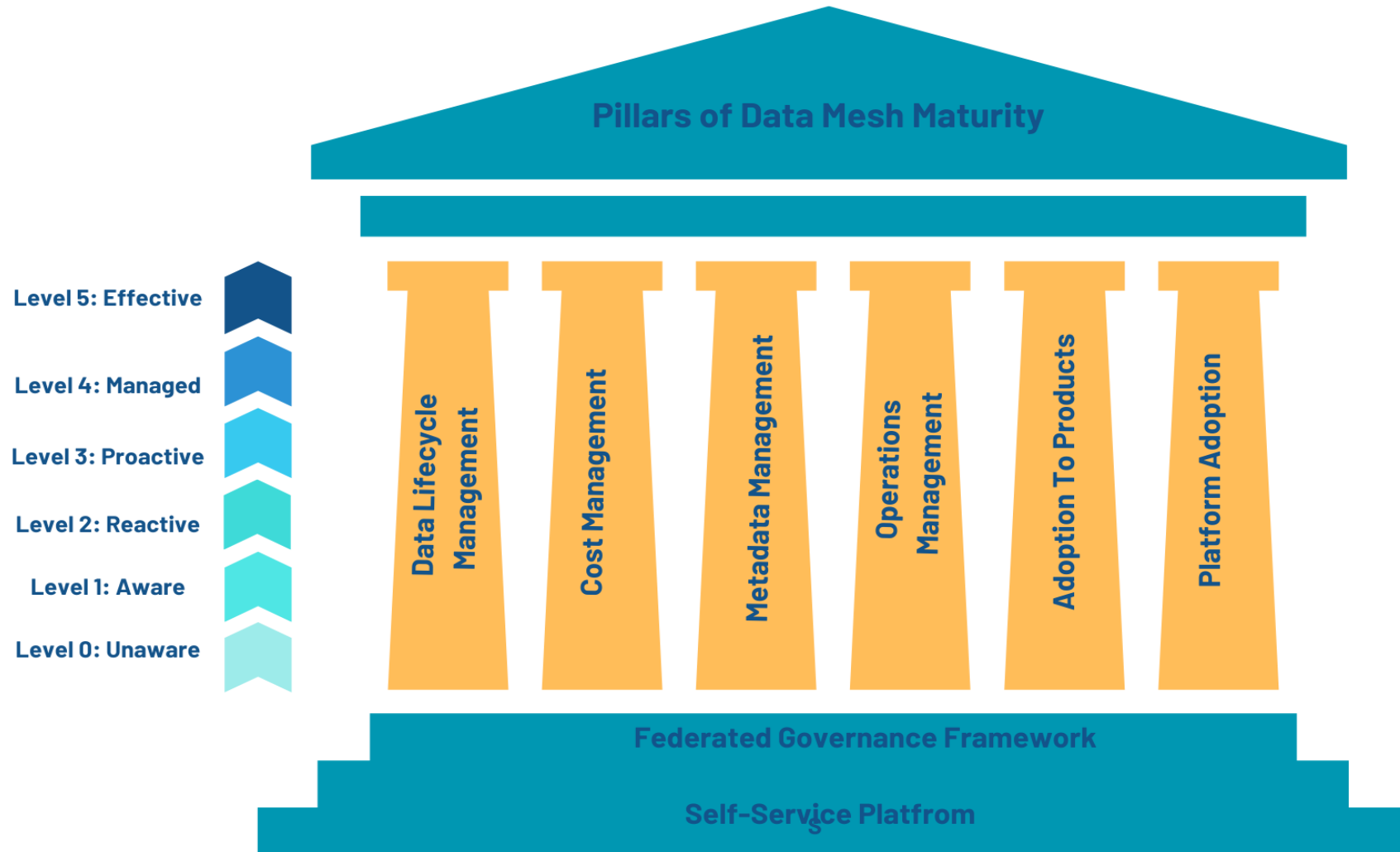
Level n+1: my data is available in all software products

Level n: my data is available in Lakehouse

Level 1: I am verifying the accuracy of my data in the lake

Level 0: my data is being published to data lake

# Our Maturity Model

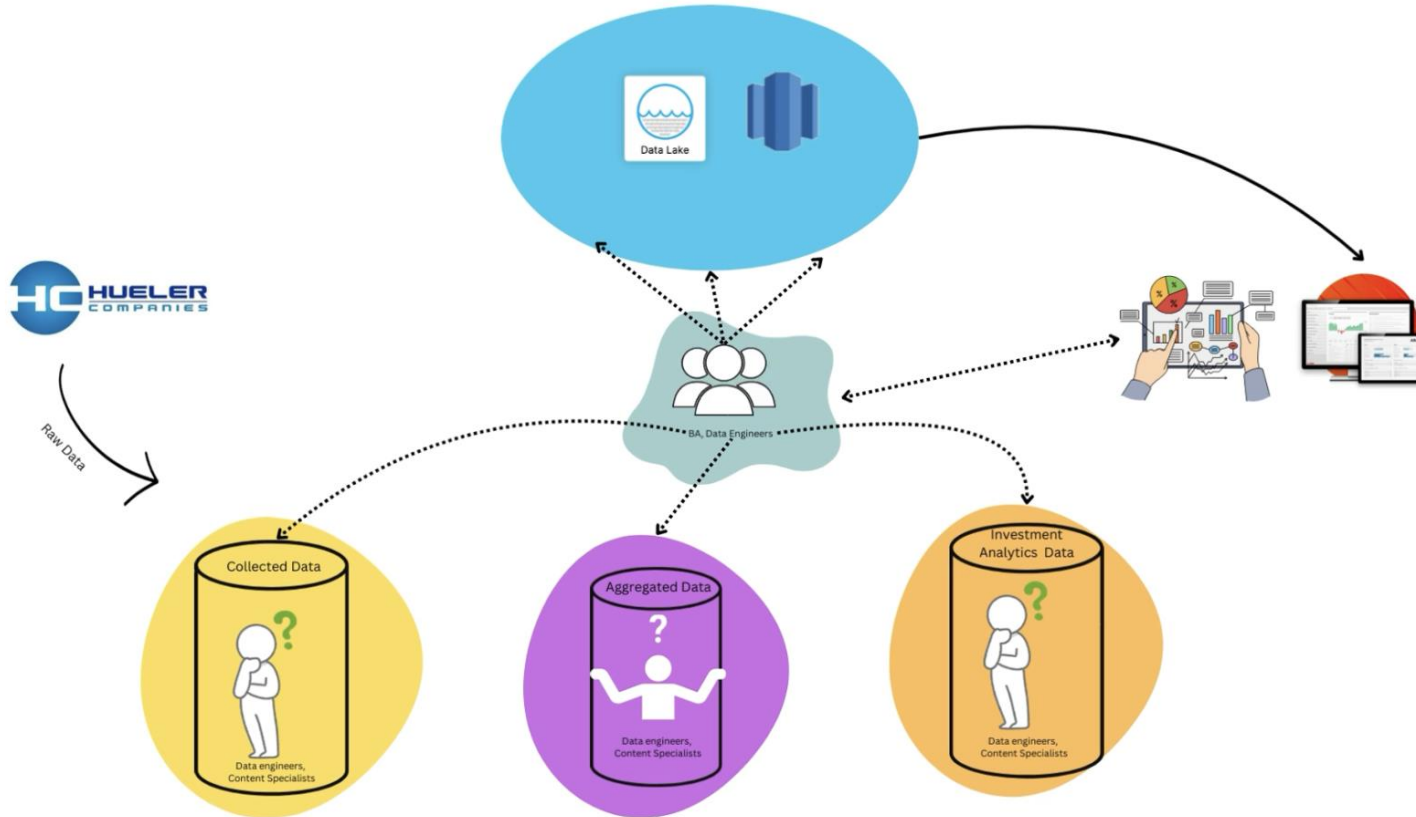


# Step 2: Increasing Efficiency

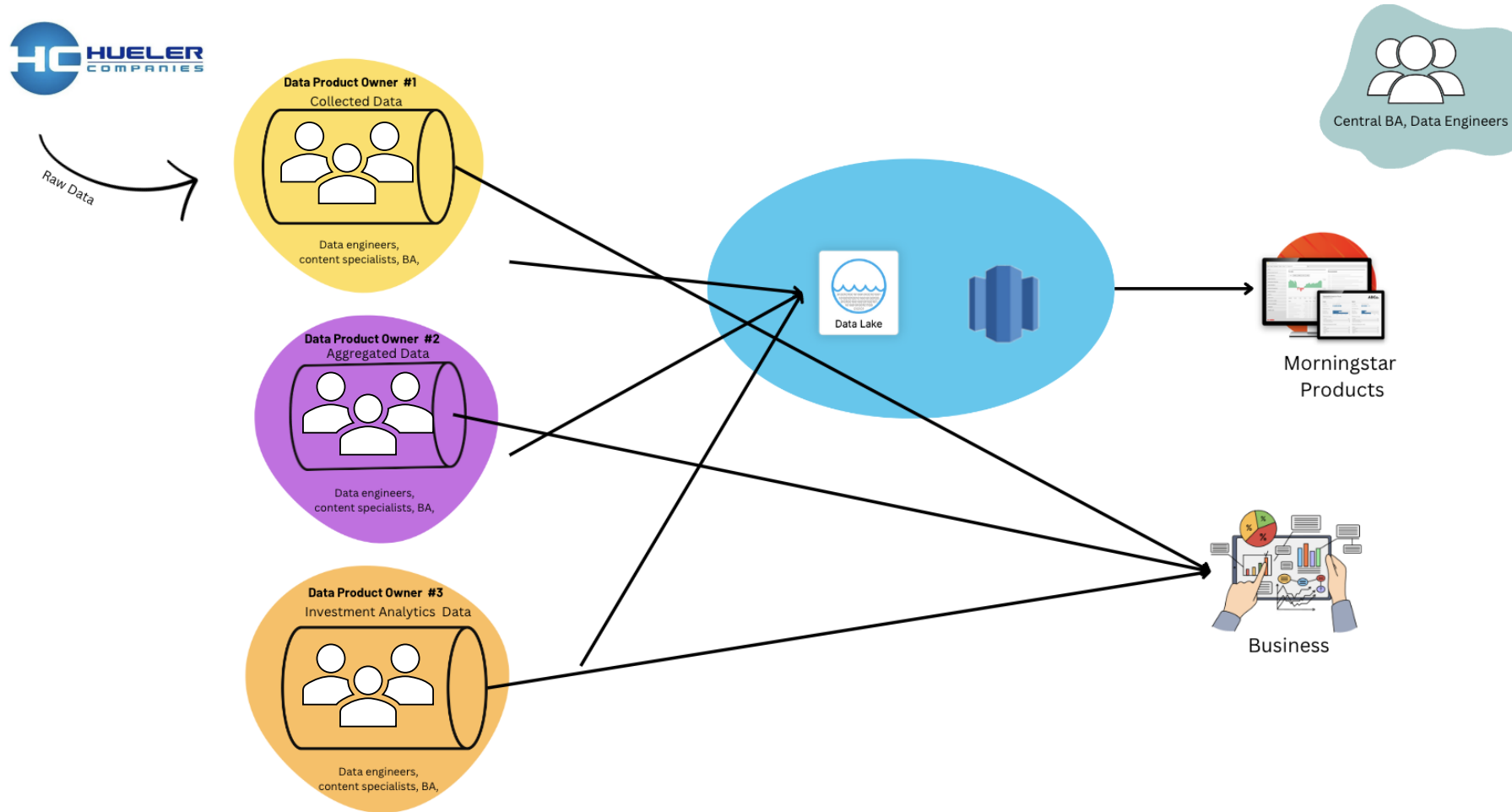


END GOAL:  
measure  
domains/product  
maturity across  
against  
cost/value of  
procuring the  
data

# Hueler and Morningstar- a siloed operating model



# Moving to a collaborative operating model



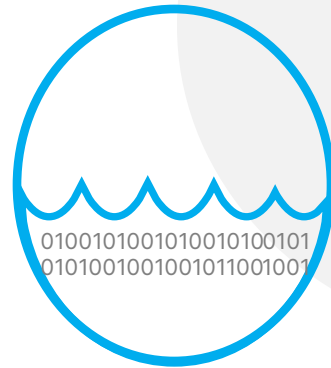
# Democratized Data Mesh Journey



Leading Financial  
Research and  
Data Solutions



1984



Centralized Data  
Lake  
and Lakehouse



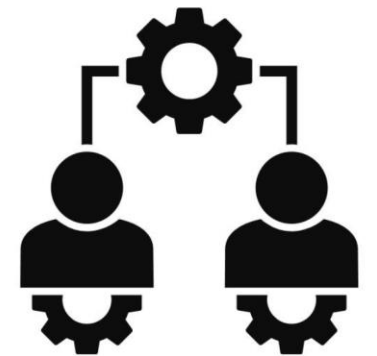
2018



Self-Service  
Platform



2021



Data Mesh  
Operating Model



2024

---

# Thank you!

We are looking to collaborate and exchange ideas:



Jeff Hirsch



Anusha Dwivedula