

GNARLY
Data_Waves

PRESENTED BY  **dremio**

EPISODE 17

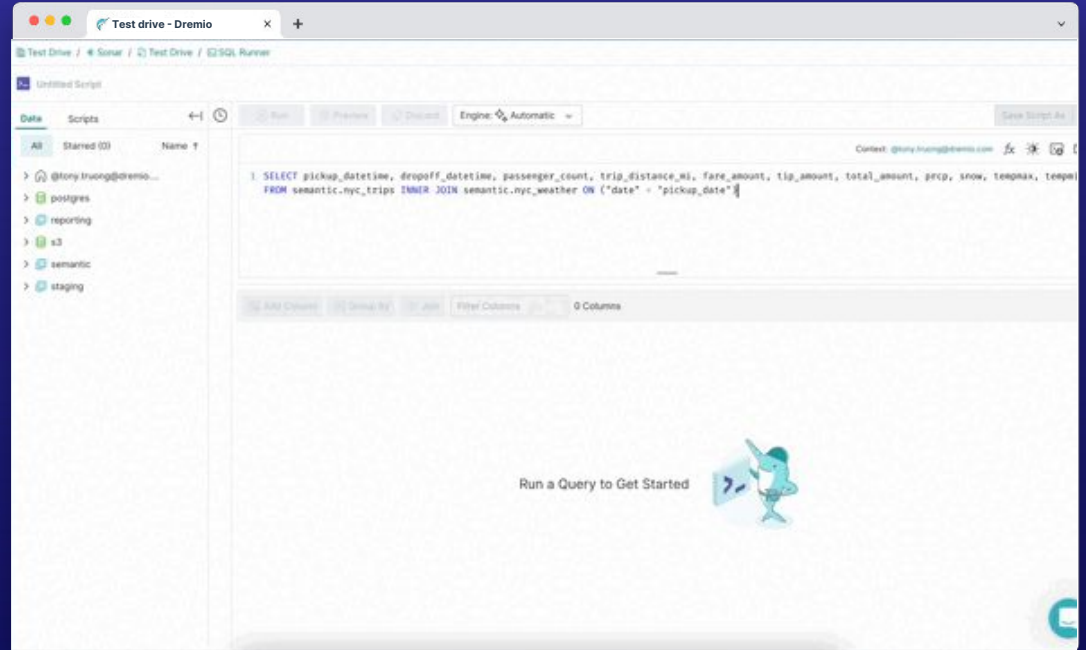
Unified Access for Your Data Mesh

Self-Service Data with Dremio's Semantic Layer

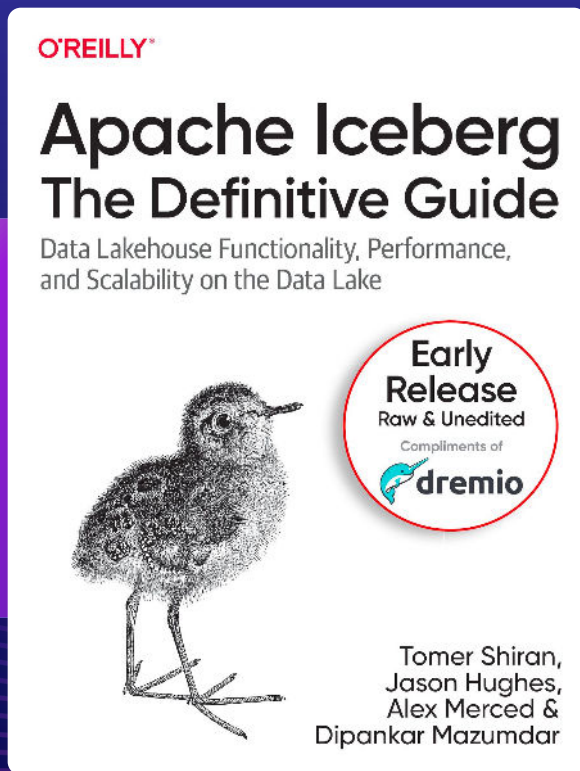
Experience the data lakehouse with Dremio Test Drive

- Sub-second query on 1 billion rows of data joining Amazon S3 with a Postgres database
- Connect to Tableau or Power BI and build a dashboard with this dataset
- Everything hosted by Dremio - 100% free for you

Start Test Drive



Apache Iceberg: The Definitive Guide



Upcoming shows

Register now

EPISODE 16

Automatic Data Optimization with Dremio Arctic



May 9th, 2023 at 8AM PST | 11AM EST | 4PM GMT

EPISODE 17

Unified Access for Your Data Mesh: Self-Service Data with Dremio's Semantic Layer



May 16, 2023 at 8AM PST | 11AM EST | 4PM GMT

EPISODE 18

Easily Migrate Hadoop Workloads to AWS with Dremio



May 23, 2023 at 8AM PST | 11AM EST | 4PM GMT

EPISODE 19

How MSK Accelerates Cancer Research with Dremio's Data Lakehouse



May 31, 2023 at 8AM PST | 11AM EST | 4PM GMT



AWS Dev Day: **Chicago**

Experience Dremio Cloud and
Tableau on Amazon S3

May 17th, 10am CST-12pm | Lunch to follow



AWS Summit **Toronto**

June 14th, 2023
Metro Toronto Convention Centre



AWS Dev Day: **New York**

Experience Dremio Cloud

July 25th, 10:00 am to 1:00 pm
Endeavor 1 | Courtyard - 461 W 34th St.
New York, NY 10001



AWS Summit **New York**

July 26th, 2023
Javits Center



DATAfestival #Munich

June 13th – 14th 2023

DATA. PEOPLE. EVERYWHERE.

Tickets 2023



Coalesce by dbt

Oct 16-20, 2023
Hilton Bayfront San Diego



BIGDATA & AI | by Corp
P A R I S

TIME TO ACCELERATE

September 25 & 26, 2023

Paris Convention Center

Unified Access for Your Data Mesh

Self-Service Data with Dremio's Semantic Layer



Alex Merced

Developer Advocate, Dremio



Data Utopia

- Data is consistent
- Data is fresh
- Data is correct
- Data is discoverable
- Data is accessible by the right people

Data Reality

- Data is inconsistent (so many copies being used with opaque lineage)
- Data is stale (too much data for a centralized team to make SLAs)
- Data is correct (too much data to monitor quality within SLAs)
- Data is discoverable (data in many places with little documentation)
- Data is accessible by the right people (due to many sources access rules are opaque and distributed)

Solution

Data Mesh

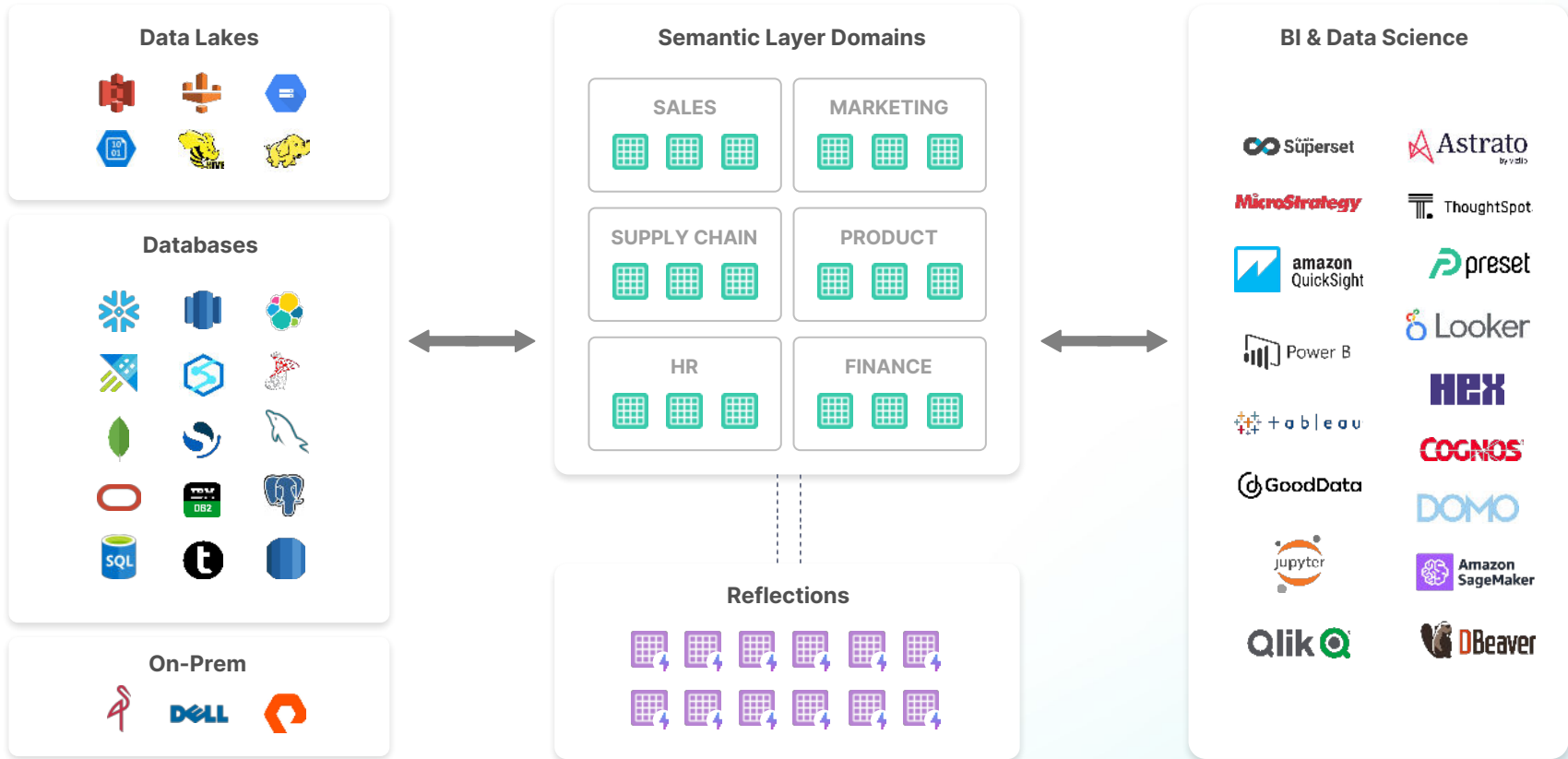
- a. An architecture that relies on decentralized creation of data into data products using shared governance and standards.
 - i. Meet SLAs as different teams focus on different parts of the data making freshness and correctness easier to achieve

Solution

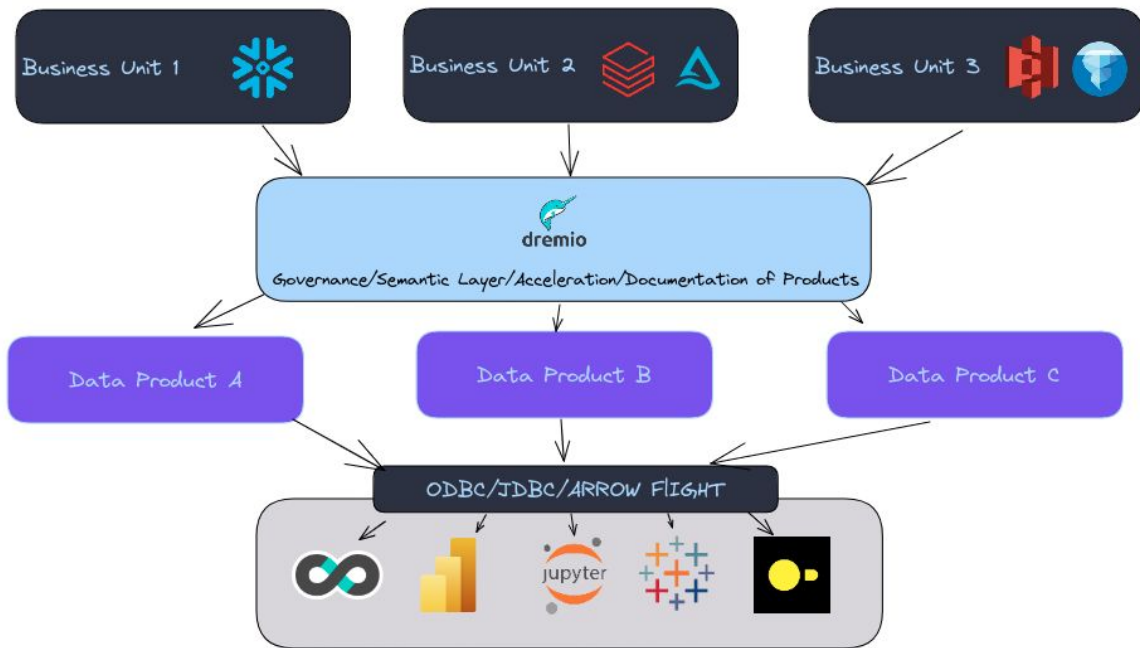
Dremio's Semantic Layer

- a. A layer for organizing, documenting, and governing your data across different sources with visibility into your data's lineage, further enhanced by Dremio's no copy architecture.
 - i. One central place for all end users to discover data across all your data products
 - ii. One central place to manage access all your data by decentralized teams
 - iii. Documentation and Lineage provide context to end users to better use data

Governed, Domain-based Semantic Layer



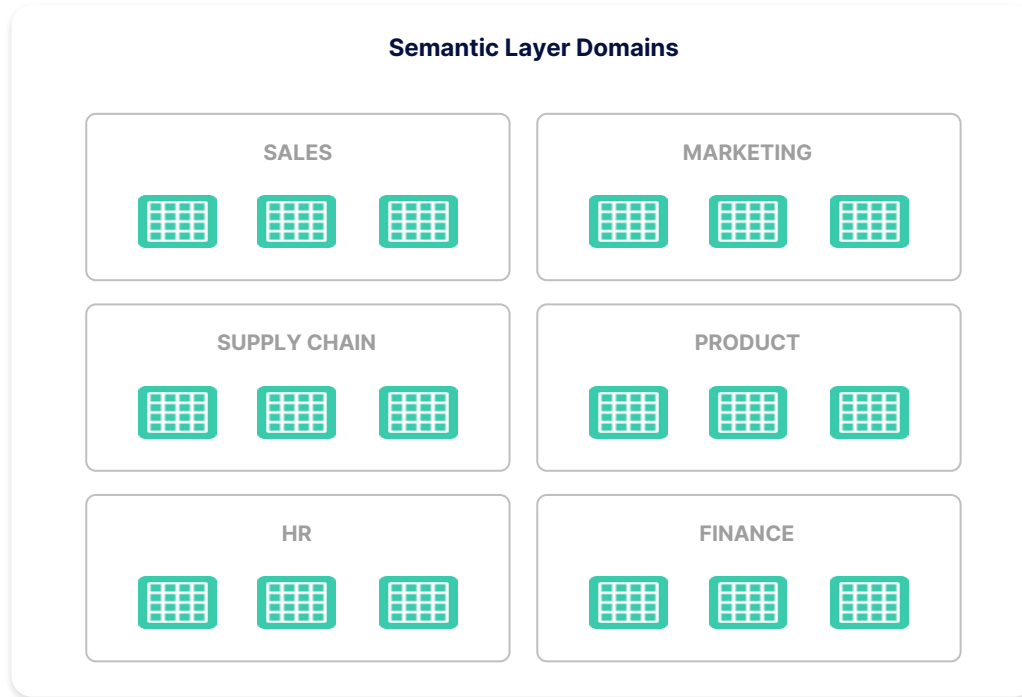
Understanding Data Mesh



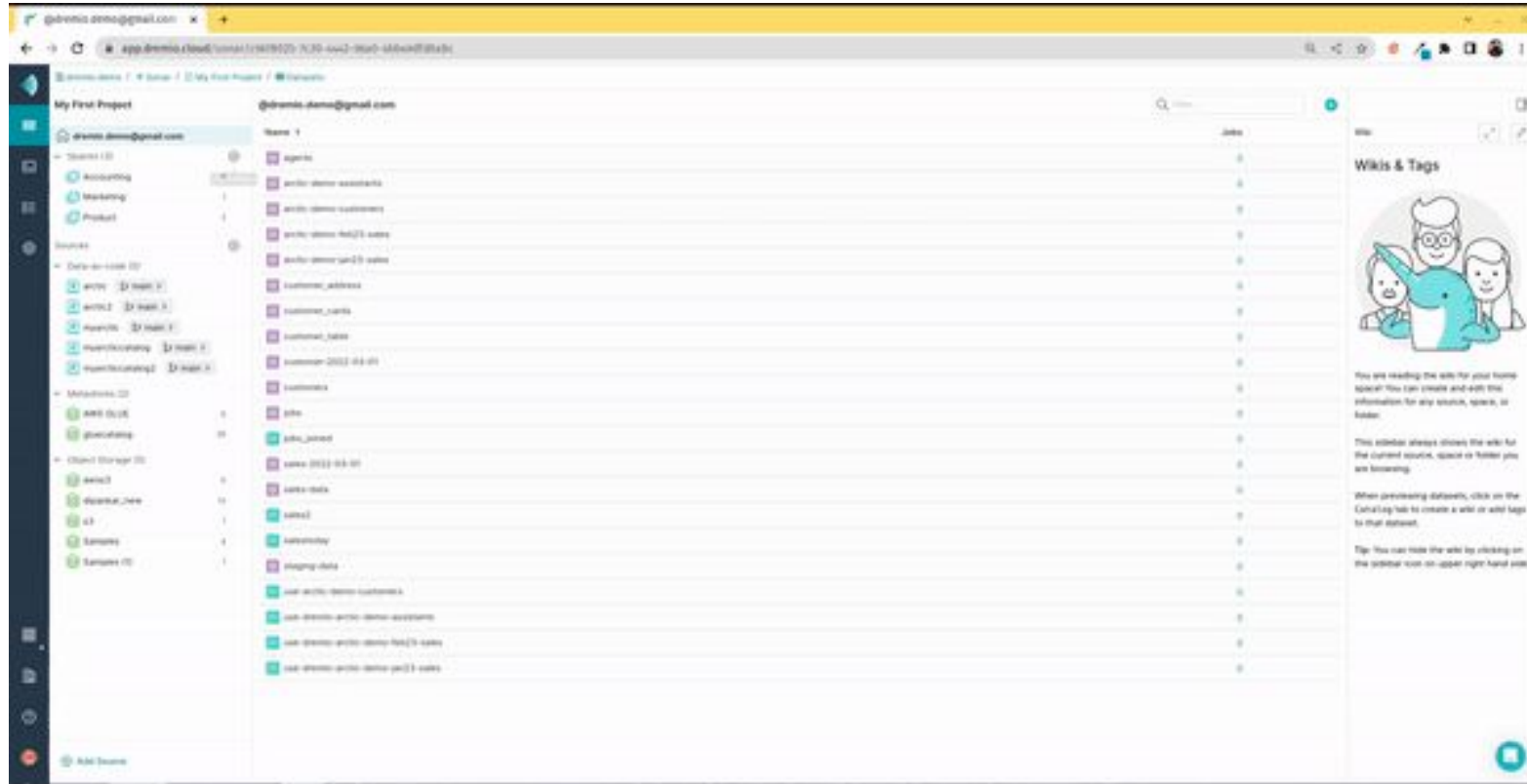
Four Principles of Data Mesh

- Domain Oriented Decentralization
- Data as Product
- Self-Service Infrastructure
- Computational Governance

Step 1 - Determine Domains

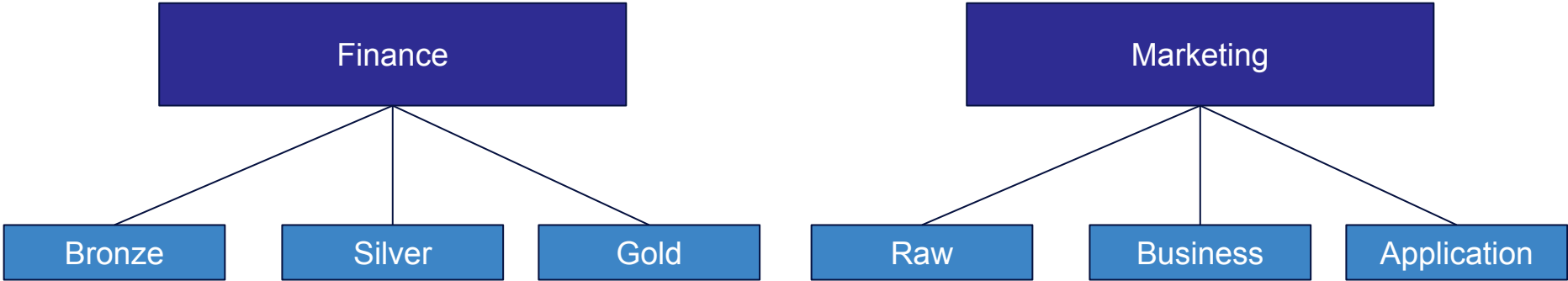


Step 2 - Create a Space for Each Domain



The screenshot displays the Dremio interface for a workspace named '@dremio.demos@gmail.com'. On the left, a navigation pane shows a tree structure of resources including 'Spaces (1)', 'Accounting', 'Marketing', 'Product', 'Databases (1)', 'Datawarehouse (1)', 'Data Lake (1)', 'Data Lake 2 (1)', 'Data Lake 3 (1)', 'Data Lake 4 (1)', 'Data Lake 5 (1)', 'Data Lake 6 (1)', 'Data Lake 7 (1)', 'Data Lake 8 (1)', 'Data Lake 9 (1)', 'Data Lake 10 (1)', 'Data Lake 11 (1)', 'Data Lake 12 (1)', 'Data Lake 13 (1)', 'Data Lake 14 (1)', 'Data Lake 15 (1)', 'Data Lake 16 (1)', 'Data Lake 17 (1)', 'Data Lake 18 (1)', 'Data Lake 19 (1)', 'Data Lake 20 (1)', 'Data Lake 21 (1)', 'Data Lake 22 (1)', 'Data Lake 23 (1)', 'Data Lake 24 (1)', 'Data Lake 25 (1)', 'Data Lake 26 (1)', 'Data Lake 27 (1)', 'Data Lake 28 (1)', 'Data Lake 29 (1)', 'Data Lake 30 (1)', 'Data Lake 31 (1)', 'Data Lake 32 (1)', 'Data Lake 33 (1)', 'Data Lake 34 (1)', 'Data Lake 35 (1)', 'Data Lake 36 (1)', 'Data Lake 37 (1)', 'Data Lake 38 (1)', 'Data Lake 39 (1)', 'Data Lake 40 (1)', 'Data Lake 41 (1)', 'Data Lake 42 (1)', 'Data Lake 43 (1)', 'Data Lake 44 (1)', 'Data Lake 45 (1)', 'Data Lake 46 (1)', 'Data Lake 47 (1)', 'Data Lake 48 (1)', 'Data Lake 49 (1)', 'Data Lake 50 (1)', 'Data Lake 51 (1)', 'Data Lake 52 (1)', 'Data Lake 53 (1)', 'Data Lake 54 (1)', 'Data Lake 55 (1)', 'Data Lake 56 (1)', 'Data Lake 57 (1)', 'Data Lake 58 (1)', 'Data Lake 59 (1)', 'Data Lake 60 (1)', 'Data Lake 61 (1)', 'Data Lake 62 (1)', 'Data Lake 63 (1)', 'Data Lake 64 (1)', 'Data Lake 65 (1)', 'Data Lake 66 (1)', 'Data Lake 67 (1)', 'Data Lake 68 (1)', 'Data Lake 69 (1)', 'Data Lake 70 (1)', 'Data Lake 71 (1)', 'Data Lake 72 (1)', 'Data Lake 73 (1)', 'Data Lake 74 (1)', 'Data Lake 75 (1)', 'Data Lake 76 (1)', 'Data Lake 77 (1)', 'Data Lake 78 (1)', 'Data Lake 79 (1)', 'Data Lake 80 (1)', 'Data Lake 81 (1)', 'Data Lake 82 (1)', 'Data Lake 83 (1)', 'Data Lake 84 (1)', 'Data Lake 85 (1)', 'Data Lake 86 (1)', 'Data Lake 87 (1)', 'Data Lake 88 (1)', 'Data Lake 89 (1)', 'Data Lake 90 (1)', 'Data Lake 91 (1)', 'Data Lake 92 (1)', 'Data Lake 93 (1)', 'Data Lake 94 (1)', 'Data Lake 95 (1)', 'Data Lake 96 (1)', 'Data Lake 97 (1)', 'Data Lake 98 (1)', 'Data Lake 99 (1)', 'Data Lake 100 (1)'. The main area shows a table with columns 'Name' and 'Jobs'. The table contains 100 rows of data, including entries like 'agency', 'article-demos-accountants', 'article-demos-customer', 'article-demos-field-sales', 'article-demos-part2-sales', 'customer_address', 'customer_cards', 'customer_tags', 'customer_2022-08-01', 'customers', 'info', 'info_json', 'sales 2022-08-01', 'sales-data', 'sales1', 'saleshistory', 'shipping-data', 'user-article-demos-customer1', 'user-demos-article-demos-accountants', 'user-demos-article-demos-field-sales', and 'user-demos-article-demos-part2-sales'. On the right, a sidebar titled 'Wikis & Tags' features a cartoon illustration of three people and a blue whale, along with text explaining how to use wikis and tags in the workspace.

Step 3 - Standardize Semantics



Step 4 - Give end users access to production data

First-class SQL UDFs

f_x

```
mask_ssn(  
  ssn VARCHAR,  
  co VARCHAR)  
RETURNS VARCHAR
```

```
ALTER TABLE customers  
ALTER COLUMN ssn  
SET MASKING POLICY  
mask_ssn(ssn, country)
```

f_x

```
restrict_location(  
  st VARCHAR,  
  co VARCHAR)  
RETURNS BOOLEAN
```

```
ALTER TABLE customers  
ADD ROW ACCESS POLICY  
restrict_location(  
  state, country)
```

| customer_id | ssn | state | country |
|-------------|-------------|-------|---------|
| | ***-**-6543 | CA | US |
| | ***-**-1212 | CA | US |
| | ***-**-9831 | CA | US |
| | ***-**-3289 | CA | US |

