Sigma Computing

EPIC Data Lab Retreat 2023

About

SaaS startup for cloud business data analytics

Founded in 2014

- Sutter Hill Ventures + Rob Woollen & Jason Frantz

Series C, ~500 employees

About

SaaS startup for cloud business data analytics

Founded in 2014

- Sutter Hill Ventures + Rob Woollen & Jason Frantz

Series C, ~500 employees

Product

Sigma Workbooks

Spreadsheet for cloud data warehouses (CDWs)

Familiar, *Scalable*, *Collaborative*, *Expressive*

About

SaaS startup for cloud business data analytics

Founded in 2014

Sutter Hill Ventures + Rob Woollen & Jason Frantz

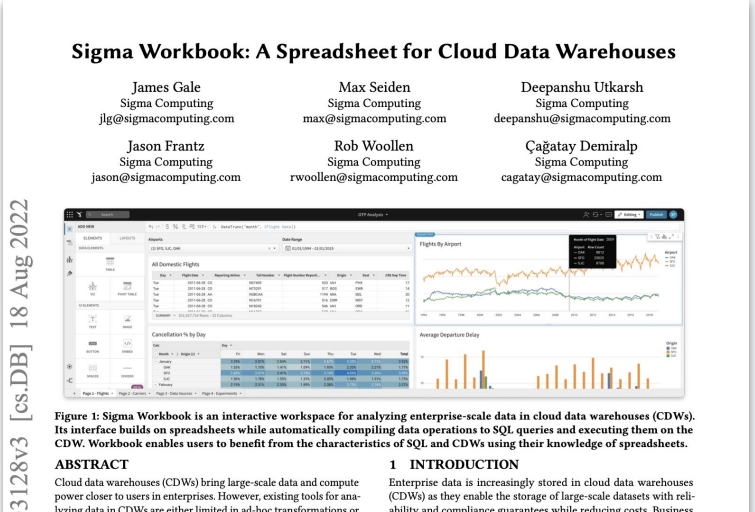
Series C, ~500 employees

Product

Sigma Workbooks

Spreadsheet for cloud data warehouses (CDWs)

Familiar, *Scalable*, *Collaborative*, *Expressive*



Sigma Workbook: A Spreadsheet for Cloud Data Warehouses

James Gale
Sigma Computing
jlg@sigmacomputing.com

Max Seiden
Sigma Computing
max@sigmacomputing.com

Deepanshu Utkarsh
Sigma Computing
deepanshu@sigmacomputing.com

Jason Frantz
Sigma Computing
jason@sigmacomputing.com

Rob Woollen
Sigma Computing
rwoollen@sigmacomputing.com

Çağatay Demiralp
Sigma Computing
cagatay@sigmacomputing.com

Figure 1: Sigma Workbook is an interactive workspace for analyzing enterprise-scale data in cloud data warehouses (CDWs). Its interface builds on spreadsheets while automatically compiling data operations to SQL queries and executing them on the CDW. Workbook enables users to benefit from the characteristics of SQL and CDWs using their knowledge of spreadsheets.

ABSTRACT

Cloud data warehouses (CDWs) bring large-scale data and compute power closer to users in enterprises. However, existing tools for ana-

1 INTRODUCTION

Enterprise data is increasingly stored in cloud data warehouses (CDWs) as they enable the storage of large-scale datasets with reli-

About

SaaS startup for cloud business data analytics

Founded in 2014

Sutter Hill Ventures + Rob Woollen & Jason Frantz

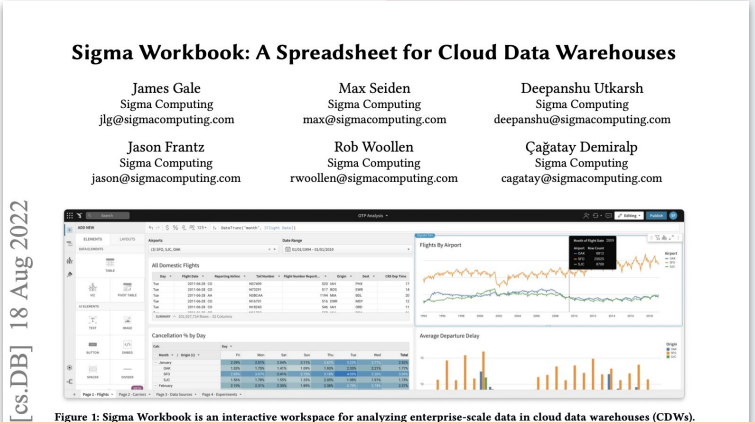
Series C, ~500 employees

Product

Sigma Workbooks

Spreadsheet for cloud data warehouses (CDWs)

Familiar, Scalable, Collaborative, Expressive



Sigma Workbook: A Spreadsheet for Cloud Data Warehouses

James Gale
Sigma Computing
jlg@sigmacomputing.com

Max Seiden
Sigma Computing
max@sigmacomputing.com

Deepanshu Utkarsh
Sigma Computing
deepanshu@sigmacomputing.com

Jason Frantz
Sigma Computing
jason@sigmacomputing.com

Rob Woollen
Sigma Computing
rwoollen@sigmacomputing.com

Çağatay Demiralp
Sigma Computing
cagatay@sigmacomputing.com

Figure 1: Sigma Workbook is an interactive workspace for analyzing enterprise-scale data in cloud data warehouses (CDWs).

https://www.sigmacomputing.com/free-trial

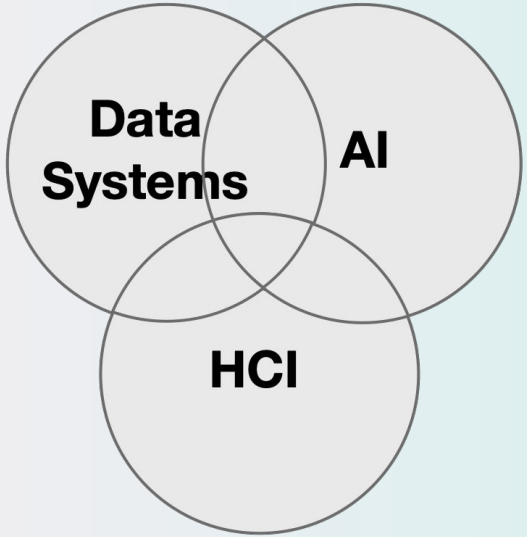
Sigma Computing Research

Multidisciplinary [intelligent] data systems research group

sigma

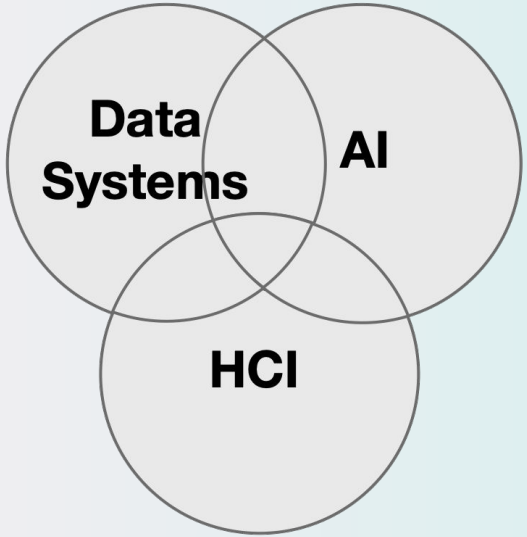
Sigma Computing Research

Multidisciplinary [intelligent] data systems research group



Sigma Computing Research

Multidisciplinary [intelligent] data systems research group



Data Systems: Includes, e.g, databases, information retrieval, programming languages.

AI: Artificial intelligence, including machine learning (ML) and natural language processing (NLP).

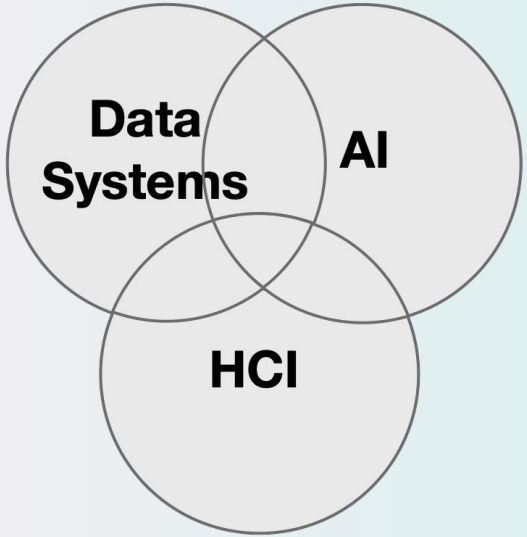
HCI: Human-computer interaction, including data visualization and visual analytics.

sigma

Sigma Computing Research

Multidisciplinary [intelligent] data systems research group

Operates like *a startup within a startup*



Data Systems: Includes, e.g, databases, information retrieval, programming languages.

AI: Artificial intelligence, including machine learning (ML) and natural language processing (NLP).

HCI: Human-computer interaction, including data visualization and visual analytics.

sigma

Sigma Computing Research

Multidisciplinary [intelligent] data systems research group

Operates like *a startup within a startup*

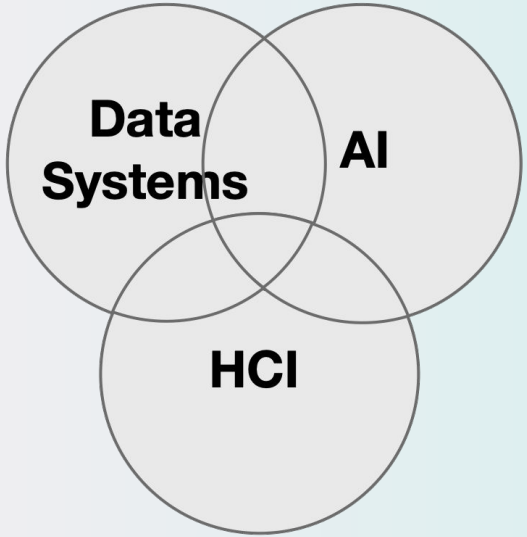
Team

Alex Bäuerle 

Çağatay Demiralp 

Georg Menzl 

Jim Gale 



Data Systems: Includes, e.g, databases, information retrieval, programming languages.

AI: Artificial intelligence, including machine learning (ML) and natural language processing (NLP).

HCI: Human-computer interaction, including data visualization and visual analytics.

sigma

Sigma Computing Research

Multidisciplinary [intelligent] data systems research group

Operates like *a startup within a startup*

Team

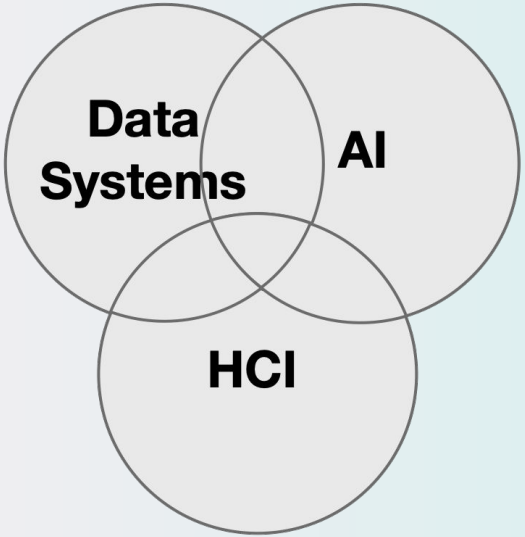
Alex Bäuerle 

Çağatay Demiralp 

Georg Menzl 

Jim Gale 

Collaborative



Data Systems: Includes, e.g, databases, information retrieval, programming languages.

AI: Artificial intelligence, including machine learning (ML) and natural language processing (NLP).

HCI: Human-computer interaction, including data visualization and visual analytics.

Sigma Computing Research

Multidisciplinary [intelligent] data systems research group

Operates like *a startup within a startup*

Team

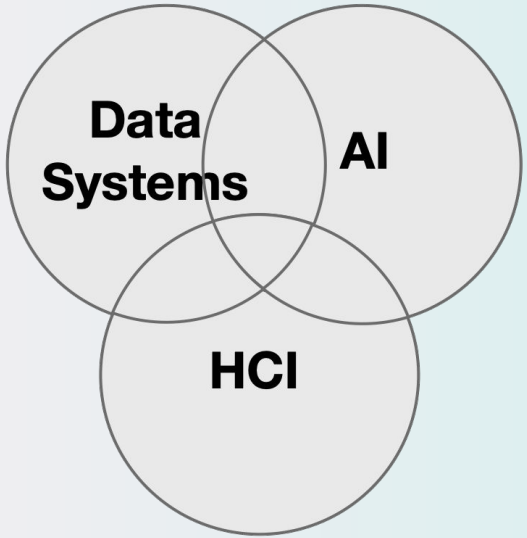
Alex Bäuerle 

Çağatay Demiralp 

Georg Menzl 

Jim Gale 

Collaborative



Data Systems: Includes, e.g, databases, information retrieval, programming languages.

AI: Artificial intelligence, including machine learning (ML) and natural language processing (NLP).

HCI: Human-computer interaction, including data visualization and visual analytics

Coming soon! https://research.sigmacomputing.com

Research Challenges & Opportunities

Usability

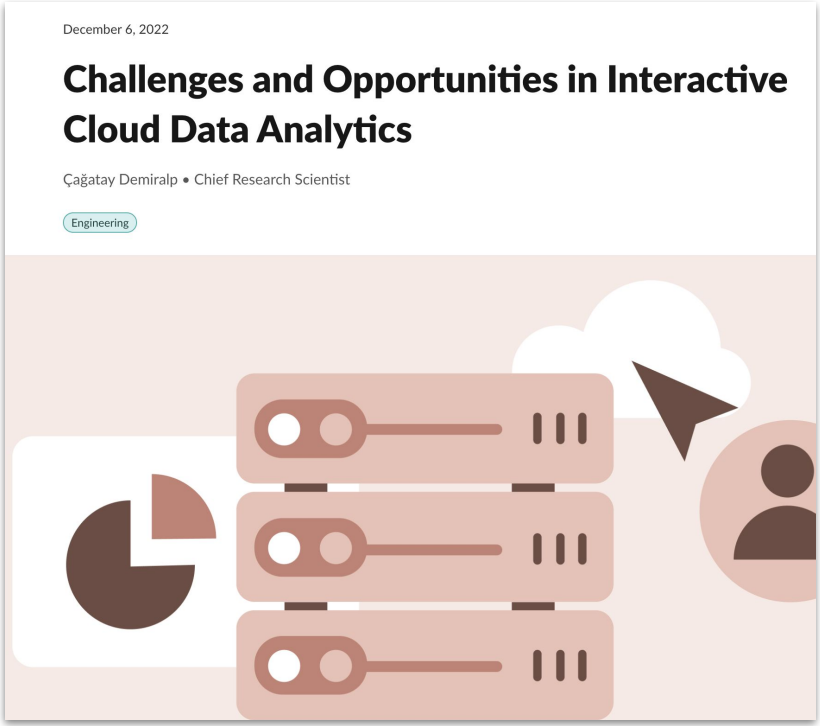
Performance

Discovery

Integration

Administration

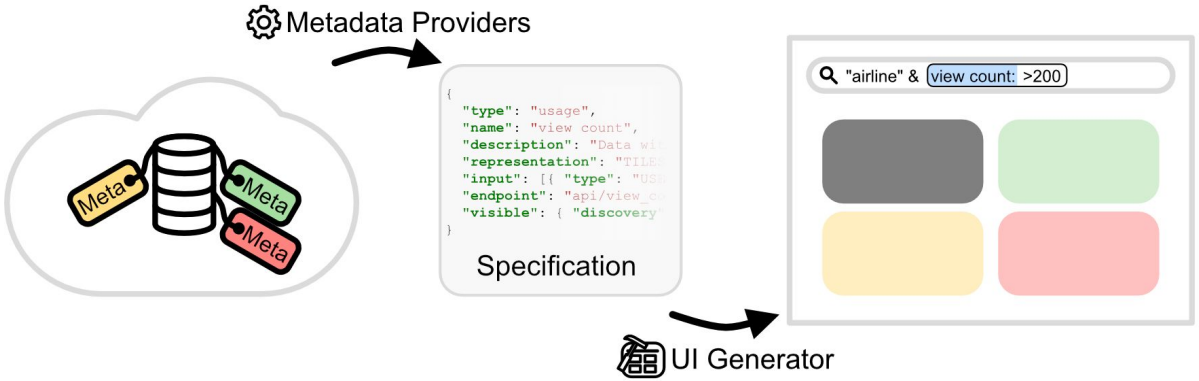
Infrastructure



sigma

Humboldt: Metadata-Driven Extensible Data Discovery

Joint work with Mike Stonebraker, MIT



sigma

Large number of tables + derived artifacts

sigma

“Either you know where the data is or you ask an expert who knows.”

sigma

“Either you know where the data is or you ask an expert who knows.”

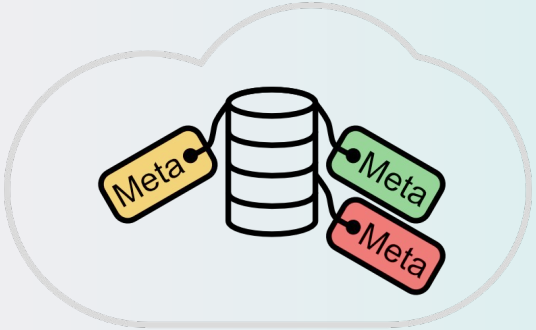
“Anyone who can use a spreadsheet can do data analysis.”

Data Discovery

Finding data first step to analysis

Metadata readily available

Manual process of visualizing data

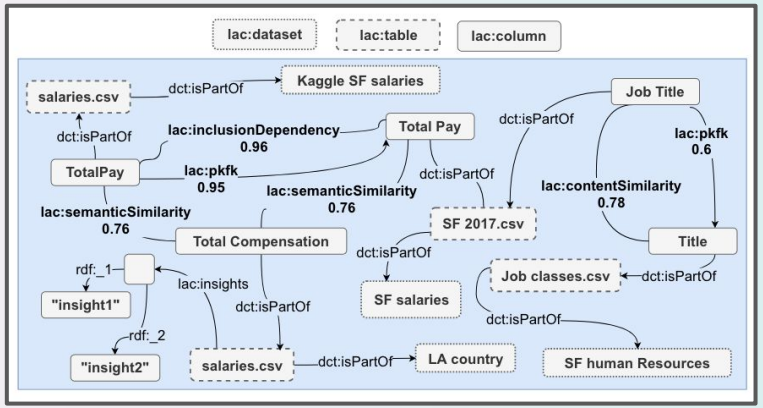


Data Relationships

Existing techniques to create data relationships.

Aurum[1], Seeping Semantics[2], KGLac[3], WarpGate[4], …

People don’t want to explore raw knowledge graphs or embeddings.



[1] Fernandez, Raul Castro, et al. "Aurum: A data discovery system." ICDE (2018).
[2] Fernandez, Raul Castro, et al. "Seeping semantics: Linking datasets using word embeddings for data discovery." ICDE (2018).
[3] Helal, Ahmed, et al. "A demonstration of kglac: A data discovery and enrichment platform for data science." VLDB (2021).
[4] Cong, Tianji, et al. "WarpGate: A Semantic Join Discovery System for Cloud Data Warehouse." CIDR (2022).

sigma

Metadata Providers

Methods to get data based on metadata

Best represented with different visualizations

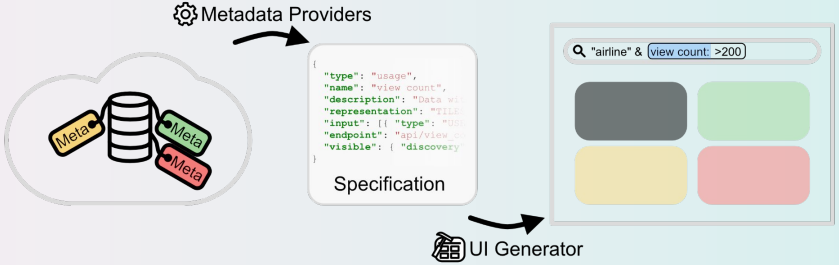
Require different input data

sigma

Humboldt

System framework to connect metadata with UI

Specification to generate UI automatically



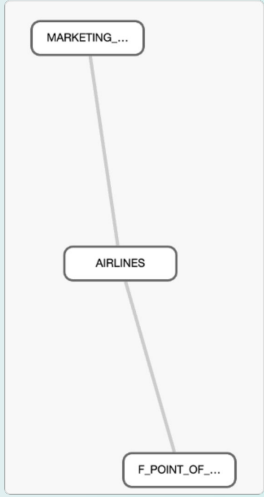
sigma

Specification-based UI

Specify parameters of the provider

Automatically generate a UI

```
{
  "type": "joinable",
  "name": "Name-Based",
  "description": "Informs about joinable
  tables by looking at column names.",
  "representation": "GRAPH",
  "input": [
    { "type": "TABLEID", "required": true }
  ],
  "endpoint": "api/name_joinability",
  "visible": {
    "discovery": true,
    "search": true,
  },
}
```



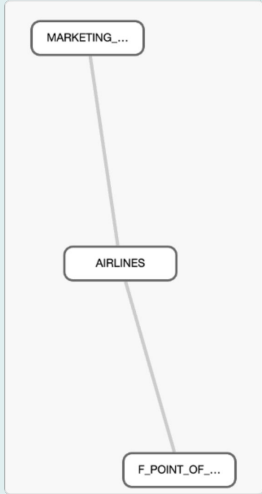
sigma

Specification-based UI

Specify parameters of the provider

Automatically generate a UI

```
{
  "type": "joinable",
  "name": "Name-Based",
  "description": "Informs about joinable
  tables by looking at column names.",
  "representation": "GRAPH",
  "input": [
    { "type": "TABLEID", "required": true }
  ],
  "endpoint": "api/name_joinability",
  "visible": {
    "discovery": true,
    "search": true,
  },
}
```



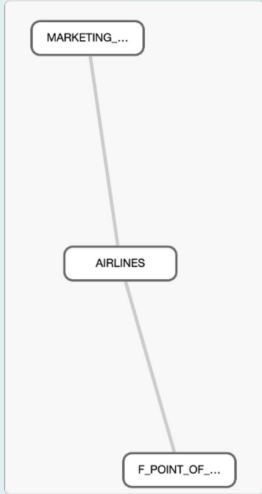
sigma

Specification-based UI

Specify parameters of the provider

Automatically generate a UI

```
{
  "type": "joinable",
  "name": "Name-Based",
  "description": "Informs about joinable
  tables by looking at column names.",
  "representation": "GRAPH",
  "input": [
    { "type": "TABLEID", "required": true }
  ],
  "endpoint": "api/name_joinability",
  "visible": {
    "discovery": true,
    "search": true,
  },
}
```



sigma

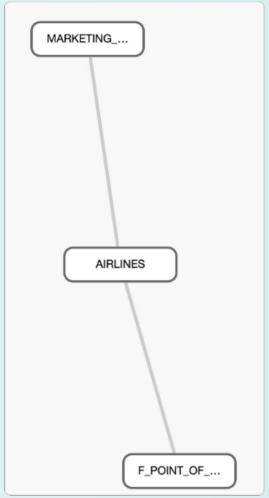
Specification-based UI

Specify parameters of the provider

Automatically generate a UI

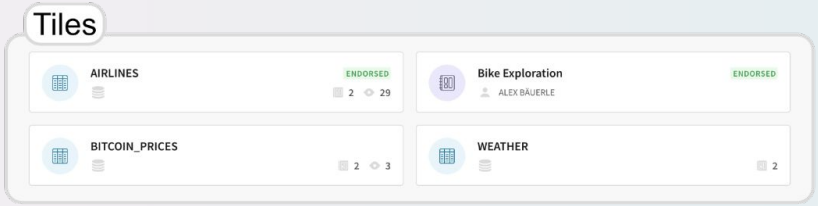
No manual UI updates when changing or adding providers

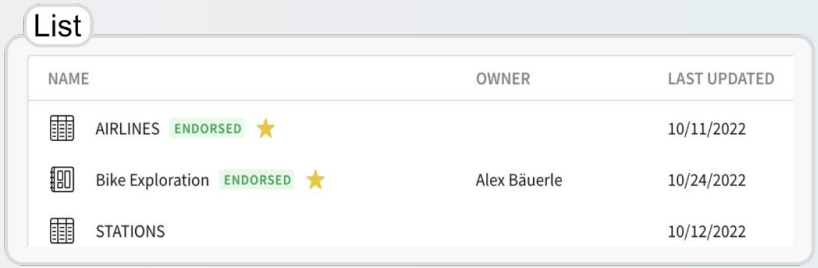
```
{
  "type": "joinable",
  "name": "Name-Based",
  "description": "Informs about joinable
  tables by looking at column names.",
  "representation": "GRAPH",
  "input": [
    { "type": "TABLEID", "required": true }
  ],
  "endpoint": "api/name_joinability",
  "visible": {
    "discovery": true,
    "search": true,
  },
}
```

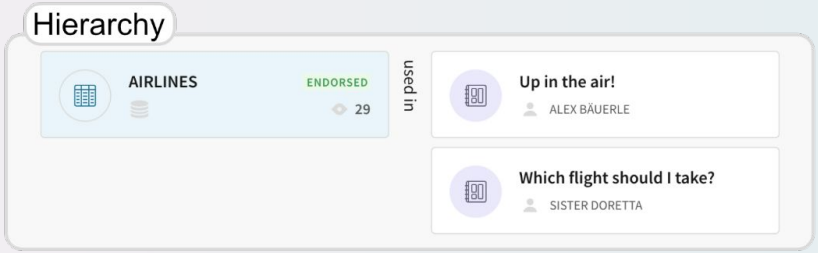


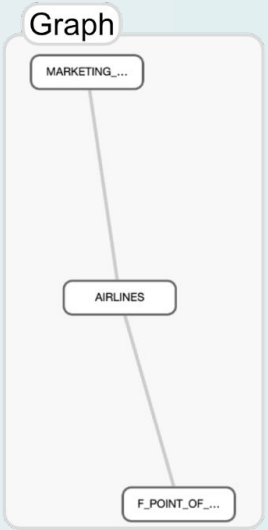
sigma

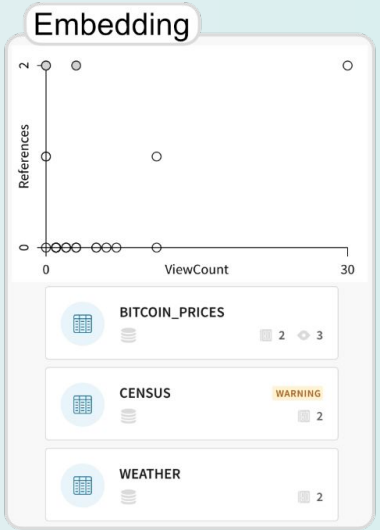
sigma


Tiles
AIRLINES
ENDORSED
2
29
Bike Exploration
ENDORSED
ALEX BÄUERLE
BITCOIN_PRICES
2
3
WEATHER
2


List
NAME
OWNER
LAST UPDATED
AIRLINES
ENDORSED
10/11/2022
Bike Exploration
ENDORSED
Alex Bäuerle
10/24/2022
STATIONS
10/12/2022


Hierarchy
AIRLINES
ENDORSED
29
used in
Up in the air!
ALEX BÄUERLE
Which flight should I take?
SISTER DORETTA


Graph
MARKETING_...
AIRLINES
F_POINT_OF_...


Embedding
References
2
0
0
ViewCount
30
BITCOIN_PRICES
2
3
CENSUS
WARNING
2
WEATHER
2


Categories
ENDORSED
NAME
OWNER
LAST UPDATED
LOCATION
AIRLINES
ENDORSED
10/11/2022
Shared with me
Bike Exploration
ENDORSED
Alex Bäuerle
10/24/2022
Shared with me
NONE

Search & Filters

Informed by same specification

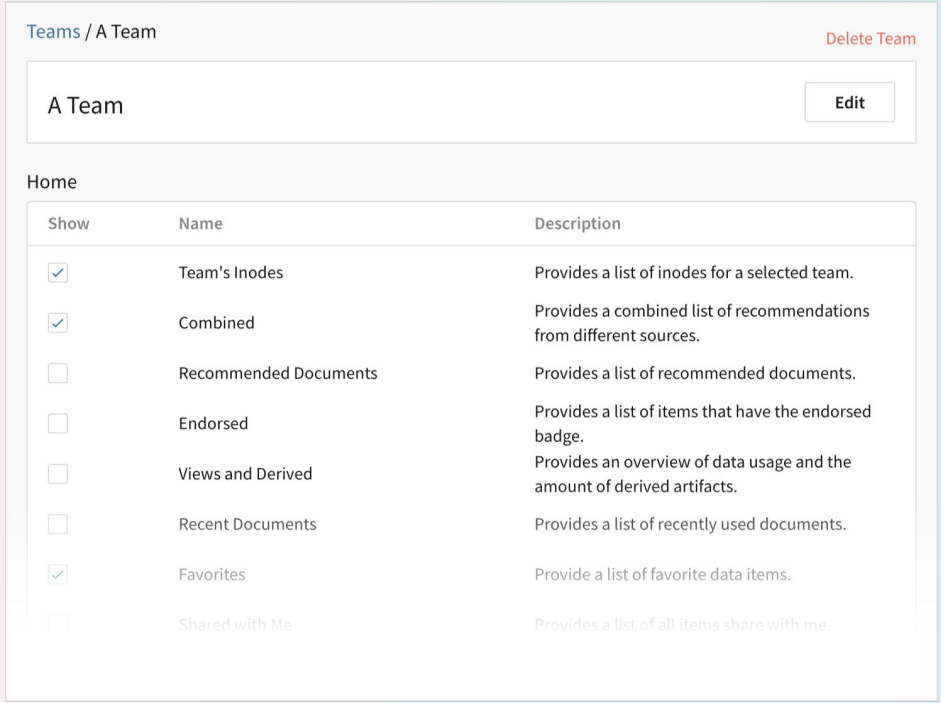
Can infer input variables and filter types

sigma

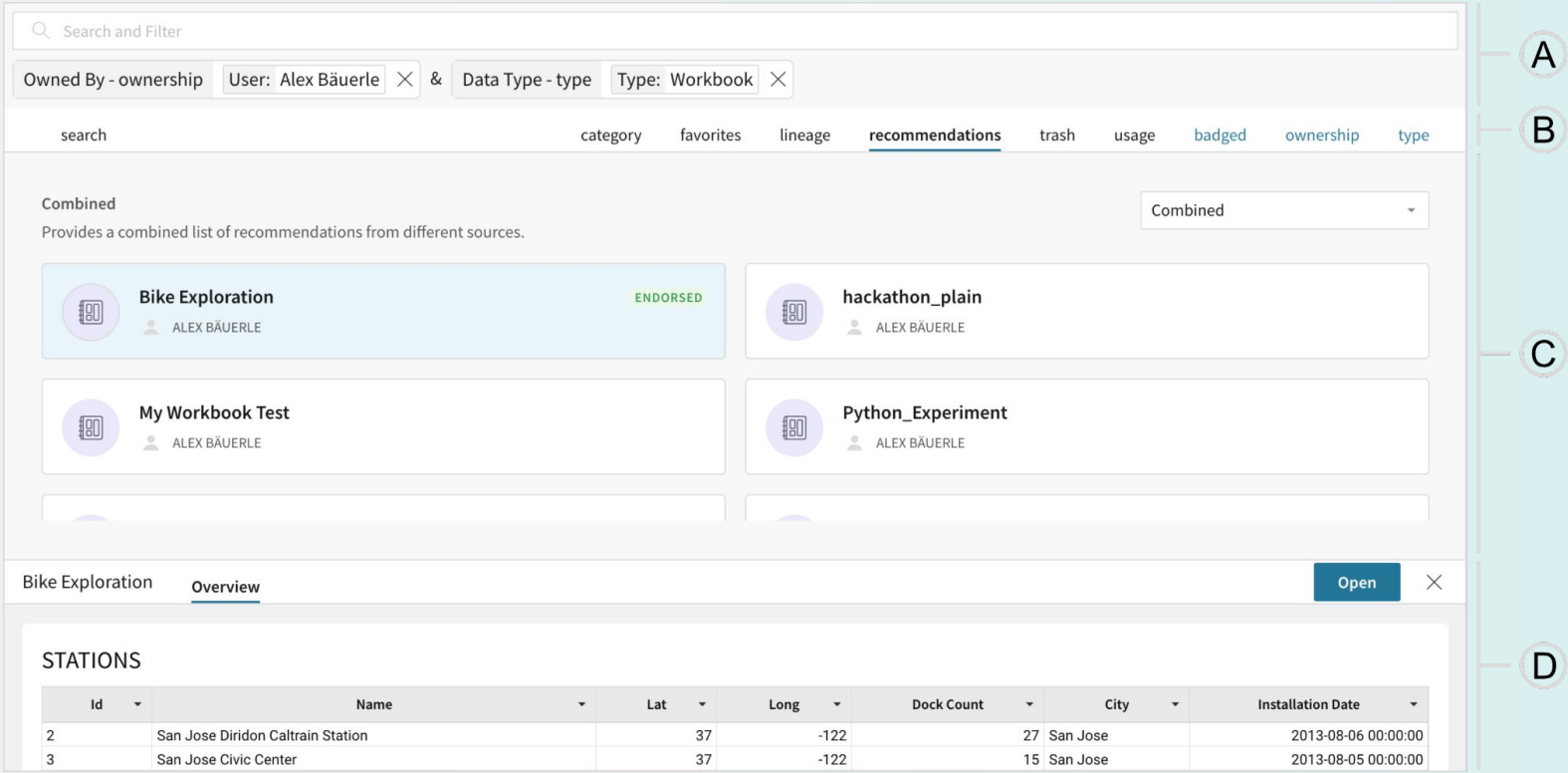
Configuration

Modify the specification

UI-based modifications



sigma


Search and Filter
Owned By - ownership
User: Alex Bäuerle
&
Data Type - type
Type: Workbook
search
category
favorites
lineage
recommendations
trash
usage
badged
ownership
type
Combined
Provides a combined list of recommendations from different sources.
Combined
Bike Exploration
ENDORSED
ALEX BÄUERLE
hackathon_plain
ALEX BÄUERLE
My Workbook Test
ALEX BÄUERLE
Python_Experiment
ALEX BÄUERLE
Bike Exploration
Overview
Open
STATIONS
Id
Name
Lat
Long
Dock Count
City
Installation Date
2
San Jose Diridon Caltrain Station
37
-122
27
San Jose
2013-08-06 00:00:00
3
San Jose Civic Center
37
-122
15
San Jose
2013-08-05 00:00:00
A
B
C
D

User types

Sigma Experts

Work is repeated.

Hard to find and identify the data they are looking for.

User types

Sigma Experts

Work is repeated.

Hard to find and identify the data they are looking for.

Discovery UI
Filters & Search

sigma

“I don’t even know what to do.”

User types

Sigma Experts

Work is repeated.

Hard to find and identify the data they are looking for.

Sigma Novices

Don’t know what data exists.

Don’t know what to do with data.

User types

Sigma Experts

Work is repeated.

Hard to find and identify the data they are looking for.

Sigma Novices

Don’t know what data exists.

Don’t know what to do with data.

Tailored views for recommended data

Humboldt: Metadata-Driven Extensible Data Dicstovery

Alex Bäuerle, Çağatay Demiralp, and Michael Stonebraker

Click to see the demo

sigma

Questions?