

The logo for Liferay Boot Camp Summer Edition is centered in the upper half of the slide. It features a stylized sun with rays emanating from it. To the left of the sun are two speech bubbles; the top one contains the letters 'SMC' and the bottom one contains a grid of dots. The text 'Liferay Boot Camp' is written in a large, bold, sans-serif font, with 'Summer Edition' in a smaller font below it. The entire logo is set against a background of orange and blue wavy shapes.

# Liferay Boot Camp Summer Edition

## How to Efficiently Query Your Data Lake Using SQL



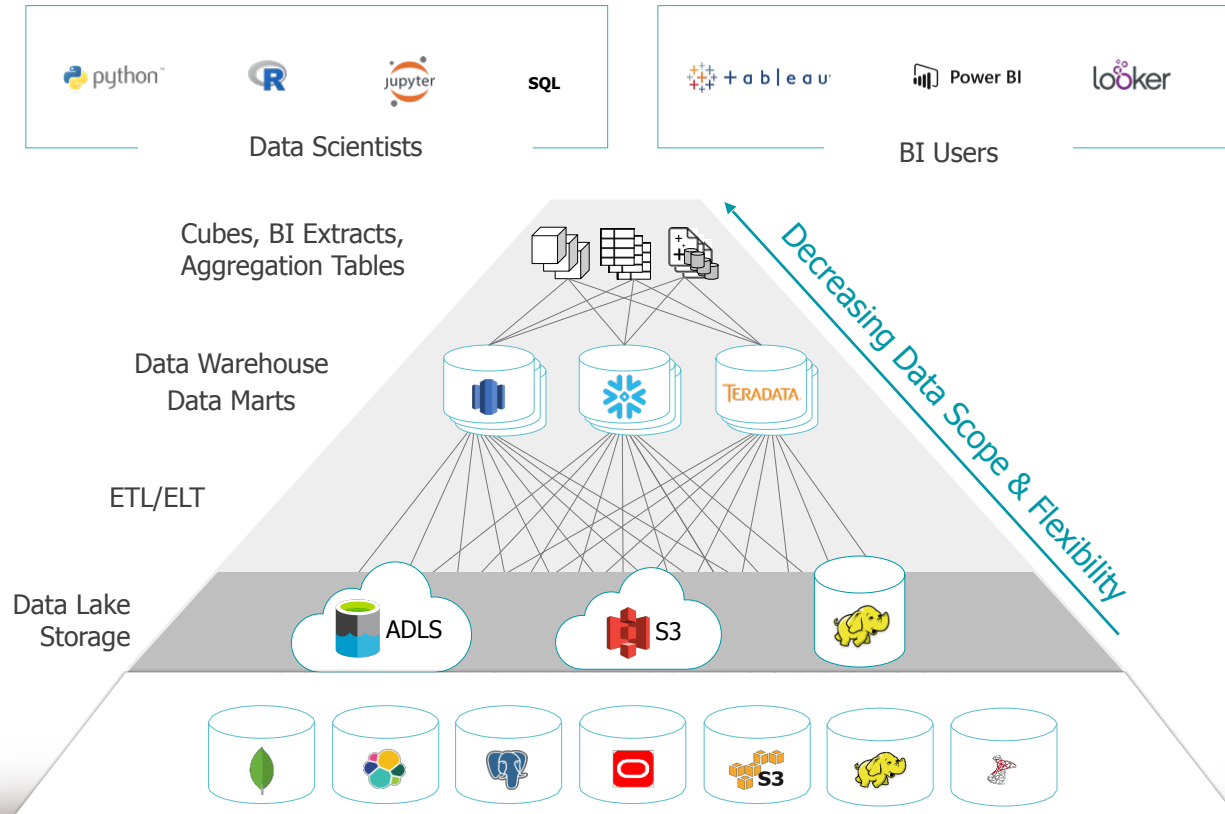
Stefano Pampaloni

CEO@Seacom



# Data Movement

# Data Movement



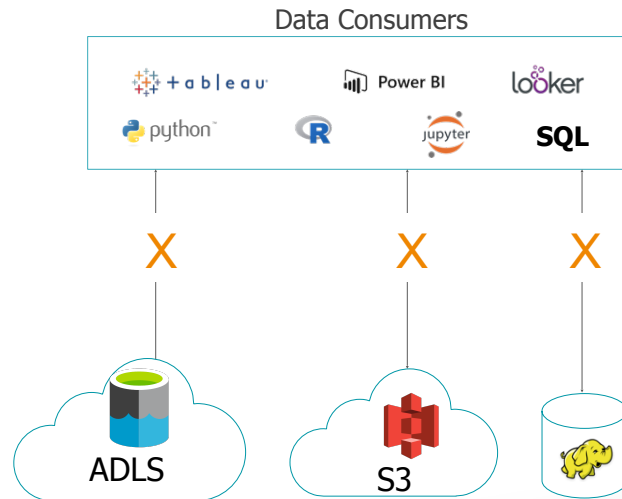
# The Data Lake is Exploding, Yet The Data Remains Inaccessible

Data Lakes are becoming the **primary place** that data lands

But



Consuming the data is **too slow & too difficult**





**SQL**

**Data Scientists**



**BI Users**

## Data Catalog

Interactive data discovery

## Data Acceleration

OLAP and ad hoc queries at interactive speed

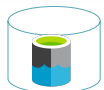


## Data Curation

Wrangle, prepare, enrich, join any source

## SQL Execution Engine

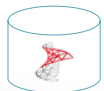
Apache Arrow-based columnar engine



Hadoop

Cloud Object

Analytical D/W

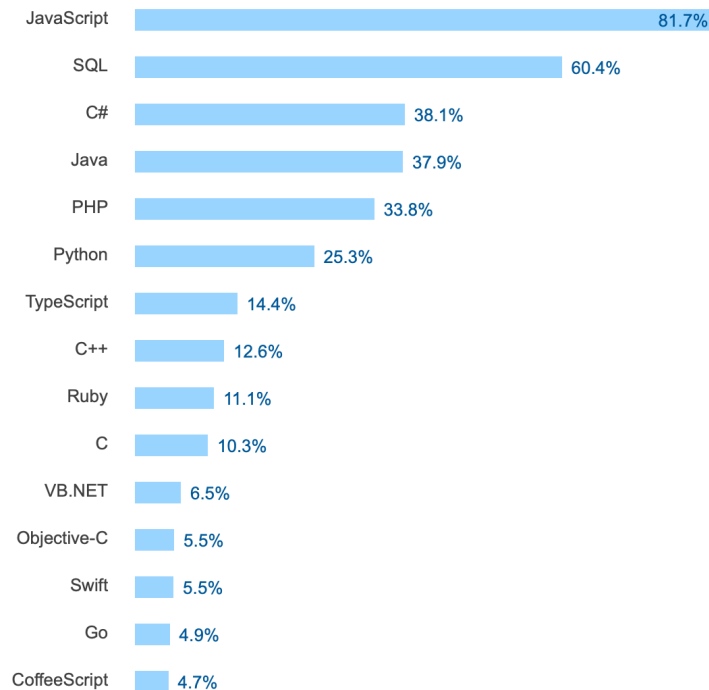


Relational

Operational Database

Non-Relational

# SQL as a common language

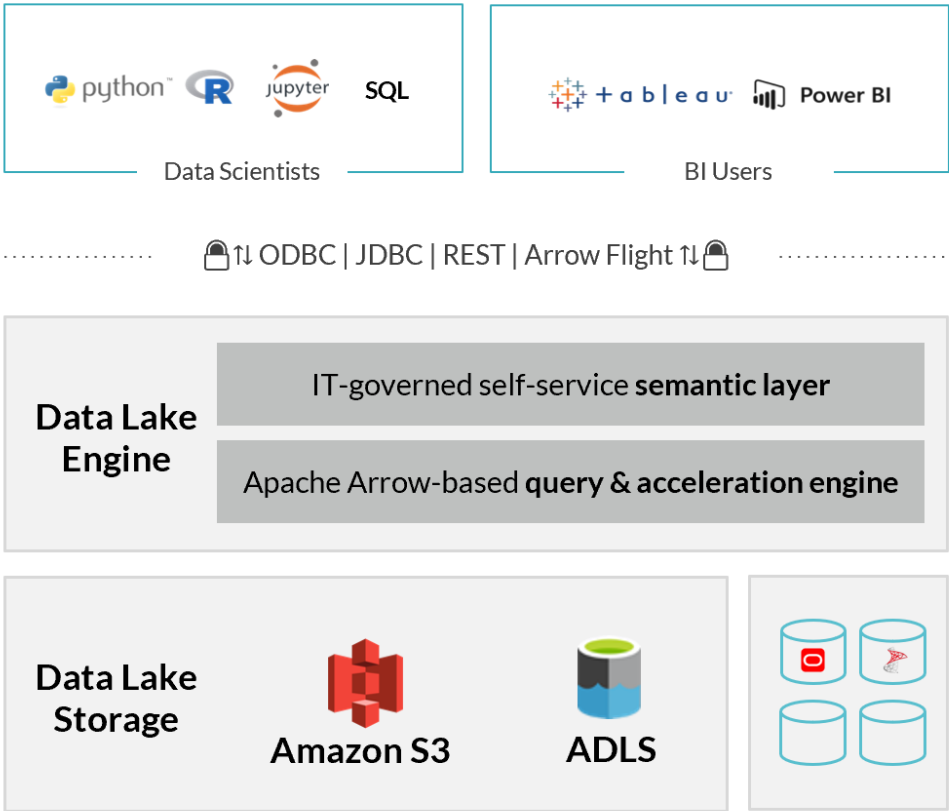


# The Data Lake Engine



Query data lake storage directly,  
minimize data movement, and maintain  
governance and control

# The Data Lake Engine

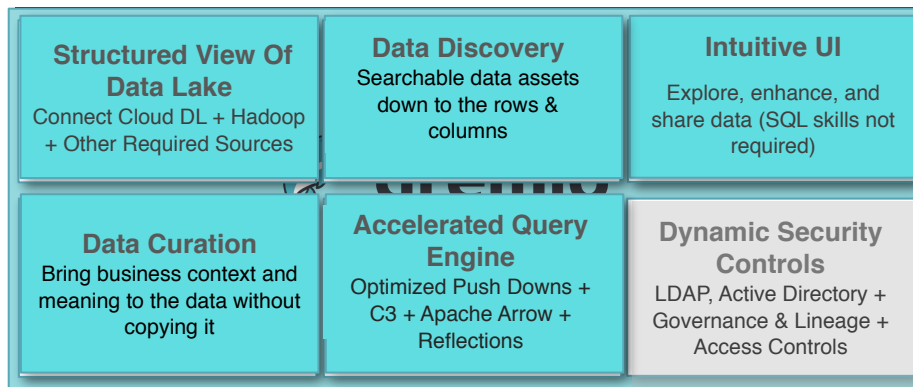


# Dremio's Data Lake Engine

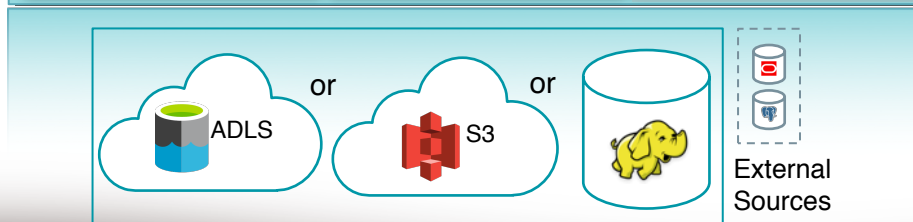
## Data Consumers



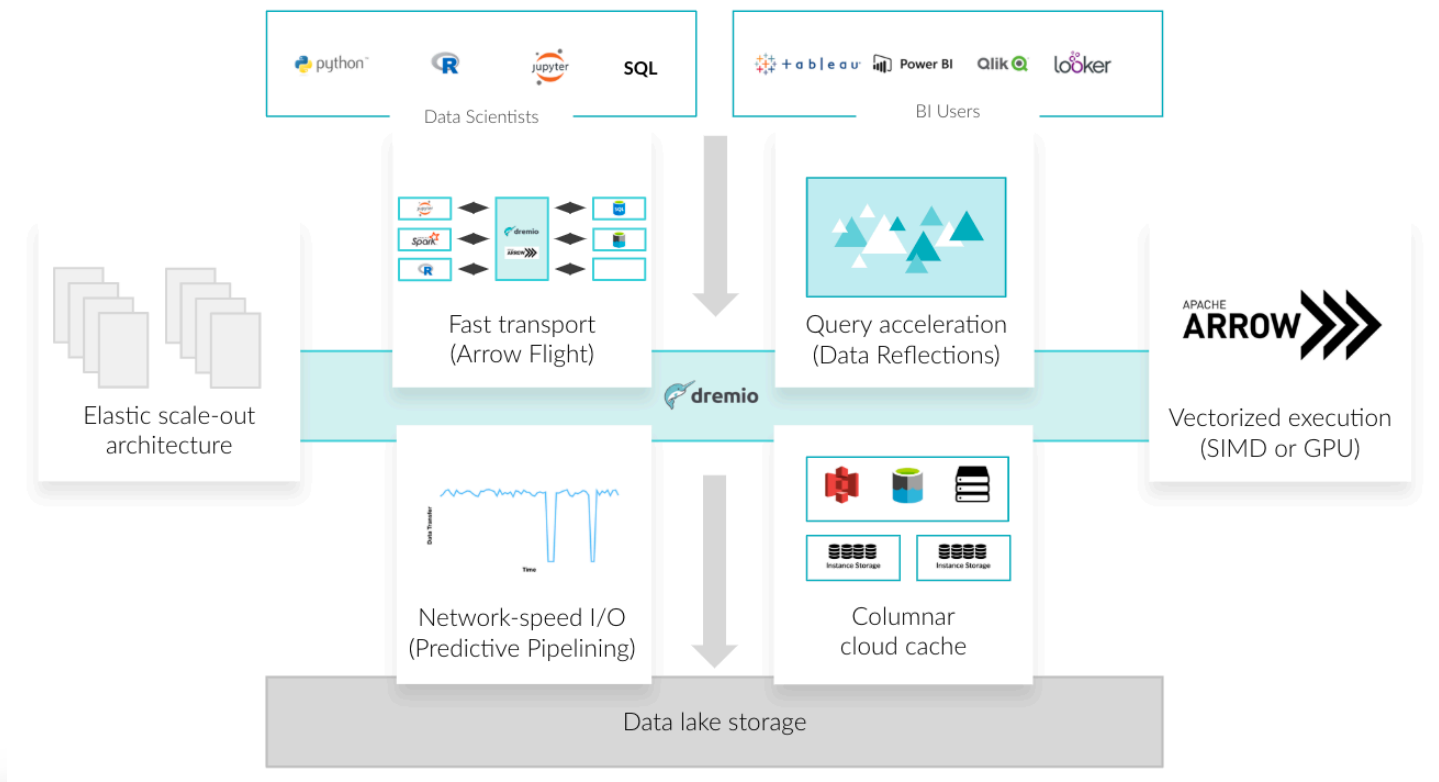
## Data-as-a-Service



## Data Sources

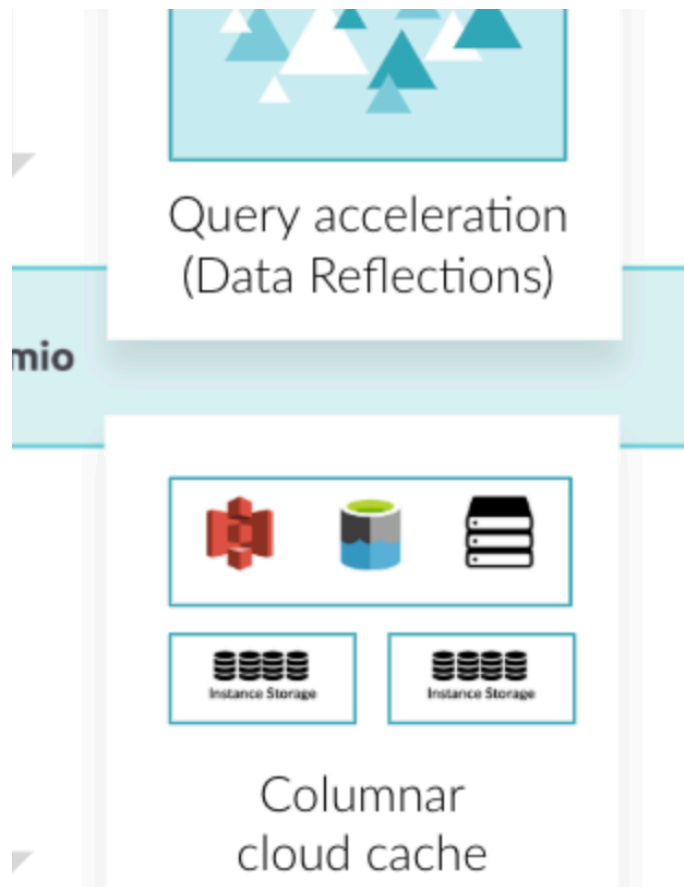


# Lightning-fast queries, directly on data lake storage



## Data Reflections – the on switch for extreme speed

With a few clicks, Dremio lets you create a Data Reflection, a physically optimized data structure that can accelerate a variety of query patterns. Create as many or as few as you want; Dremio invisibly and automatically incorporates Reflections in query plans and keeps their data up to date.



# Data Reflections improve productivity & reduce data copies

User picks best optimization



User-specific Cubes, BI Extracts, Aggregation Tables



Domain-specific Data Marts

Source Data

Dremio picks best optimization



Shared Experience Model



Data Reflections

Source Data

TODAY

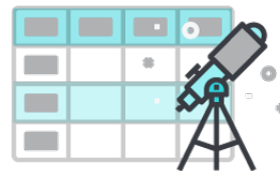
DREMIO

## Data Reflections get you there faster by going a shorter distance



### Like an index

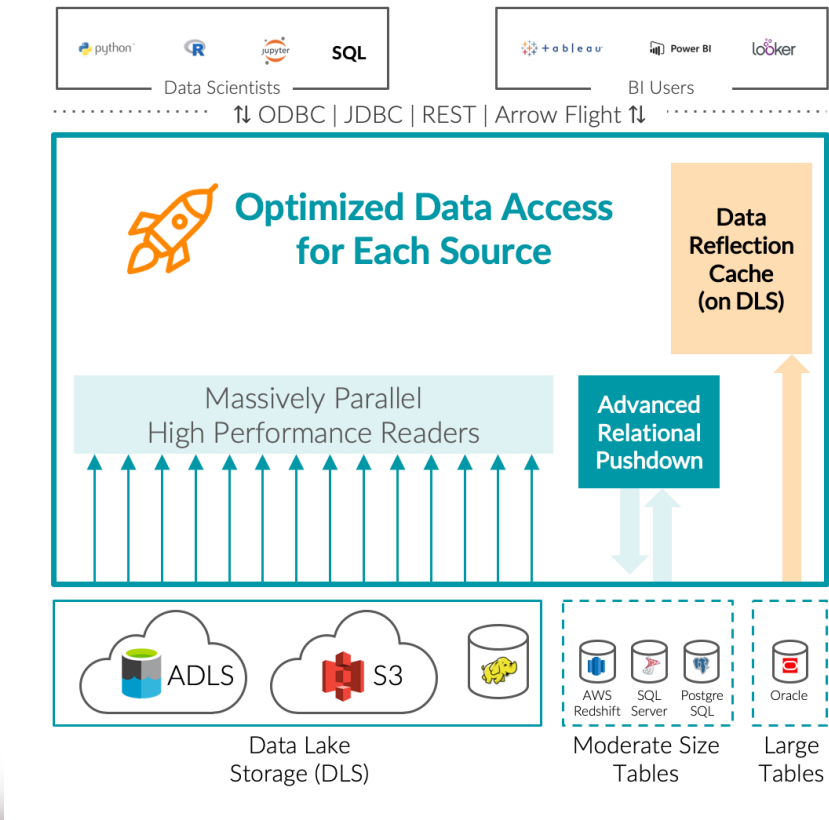
- ✓ Users don't change their behavior, dashboards, etc
- ✓ Dramatically improves query efficiency
- ✓ Many options for different optimizations
- ✓ Queries can be resolved without push-down
- ✓ System-managed maintenance, security



### Like a materialized view

- ✓ Full power of SQL – join, filter, transform, etc
- ✓ Row-level fidelity, or summarized data
- ✓ Pre-computed aggregations, joins
- ✓ Pre-sort, pre-partition
- ✓ Columnarized, compressed on file system

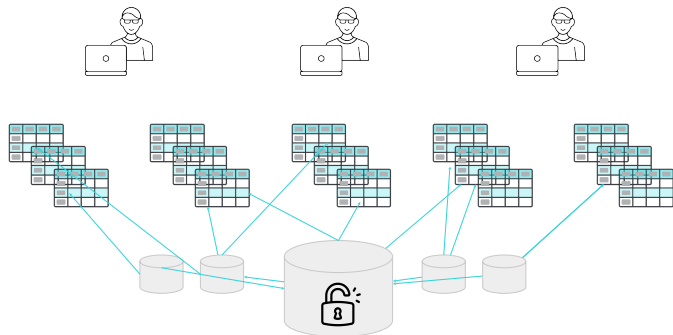
# The Data Lake Engine





# Security & Governance

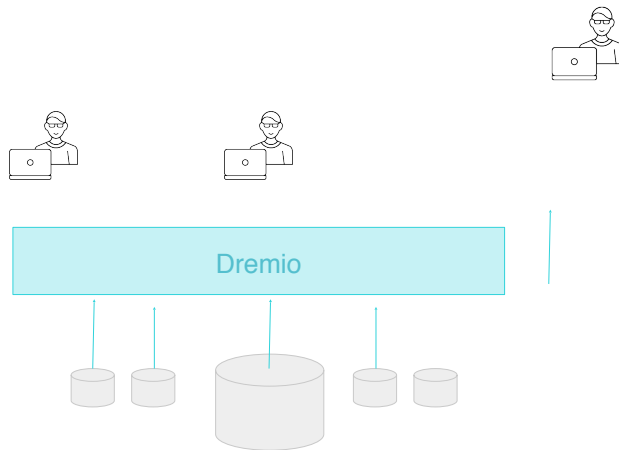
# Control your risk around Shadow Analytics



## Physical security controls

Users are taking matters into their own hands:

- ✗ Copying data into spreadsheets
- ✗ Downloading into 3<sup>rd</sup> party storage services
- ✗ Taking outside your enterprise access controls

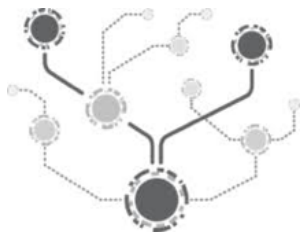


## Attract more bees with honey

Keep users coming back to a governed, secure environment

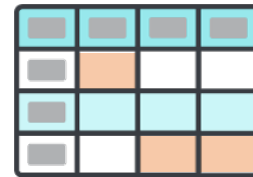
- ✓ Self-service: removes reasons to “go-around”
- ✓ Simple, intuitive access for all approved datasets
- ✓ Great experience with any tool, high performance

# Secure and govern analytics with confidence




## Automatic Data Provenance & Lineage

- ✓ Across all enterprise sources, virtual datasets
- ✓ Easily analyze patterns of use across users, tools
- ✓ Evaluate “what if” scenarios for downstream analysis
- ✓ Query time determination of access control behavior





## Row & Column-level Access Control


- ✓ Query time determination via RBAC
- ✓ Row & column level controls for all sources
- ✓ Dynamic masking based on LDAP/AD membership
- ✓ Extend the capabilities of all your sources













[Datasets](#) [Jobs](#)

[+ New Query](#)

 [Acceleration Tools](#) [Help](#) [Admin](#) [gnarly](#) 

 **customer360**  
Finance

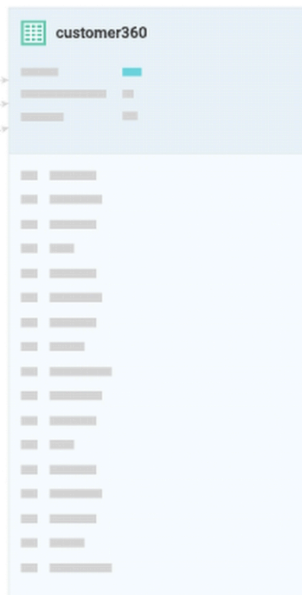
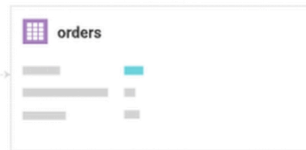
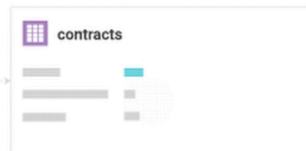
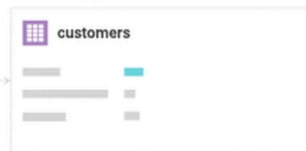
  JSON   Tableau   Save As...   Preview 

 Graph

Sources(3)




Parents(3)



# Masking & row/column-level access controls

```
1 SELECT
2     employee_id,
3     first_name,
4     last_name,
5     hire_date,
6     CASE
7         WHEN IS_MEMBER('hr_privileged') THEN ssn
8         ELSE REGEXP_REPLACE(ssn, '^\\d{5}', 'XXX-XX-')
9         END AS ssn_masked
10 FROM HR.employees
```

- ✓ Any source: S3, ADLS, HDFS, RDBMS, etc.
- ✓ IS\_MEMBER() integrated with LDAP
- ✓ Dynamic evaluation
- ✓ Masking, transformation, joins
- ✓ SQL-based

 **dremio**

Datasets

Jobs

>>



New Query


Help


Admin


kelly

New Query

  JSON

 Tableau

 Save As...

 Run

SQL Editor

Datasets

Functions

```
1 SELECT
2   employee_id,
3   first_name,
4   last_name,
5   hire_date,
6   CASE
7     WHEN IS_MEMBER('hr_privileged') THEN ssn
8     ELSE REGEXP_REPLACE(ssn, '^\\d{5}', 'XXX-XX-')
9   END AS ssn_masked
10 FROM HR.employees
```

Context: HR

Add Field

Group By

Join

Fields: 5 of 5

#	employee_id	first_name	last_name	hire_date	ssn_masked
100	Steven	King		2012-01-15	XXX-XX-6211
101	Neena	Kochhar		2014-11-05	XXX-XX-6695
102	Lex	De Haan		2016-01-11	XXX-XX-7273
103	Alexander	Hunold		2017-06-21	XXX-XX-8710
104	Bruce	Ernst		2014-06-06	XXX-XX-9809
105	David	Austin		2013-07-30	XXX-XX-1246
106	Valli	Pataballa		2013-03-01	XXX-XX-2986
107	Diana	Lorentz		2015-07-14	XXX-XX-3905
108	Nancy	Greenberg		2015-03-15	XXX-XX-4341
109	Daniel	Faviet		2013-12-08	XXX-XX-5214
110	John	Chen		2012-09-24	XXX-XX-6092
111	Ismael	Sciarra		2013-09-17	XXX-XX-7783
112	Jose Manuel	Urman		2014-12-07	XXX-XX-9207

# Domande



## CONTATTI

Stefano Pampaloni

Email [stefano.pampaloni@seacom.it](mailto:stefano.pampaloni@seacom.it)

Skype anteosnet

GitHub anteos



Thank you 😊