



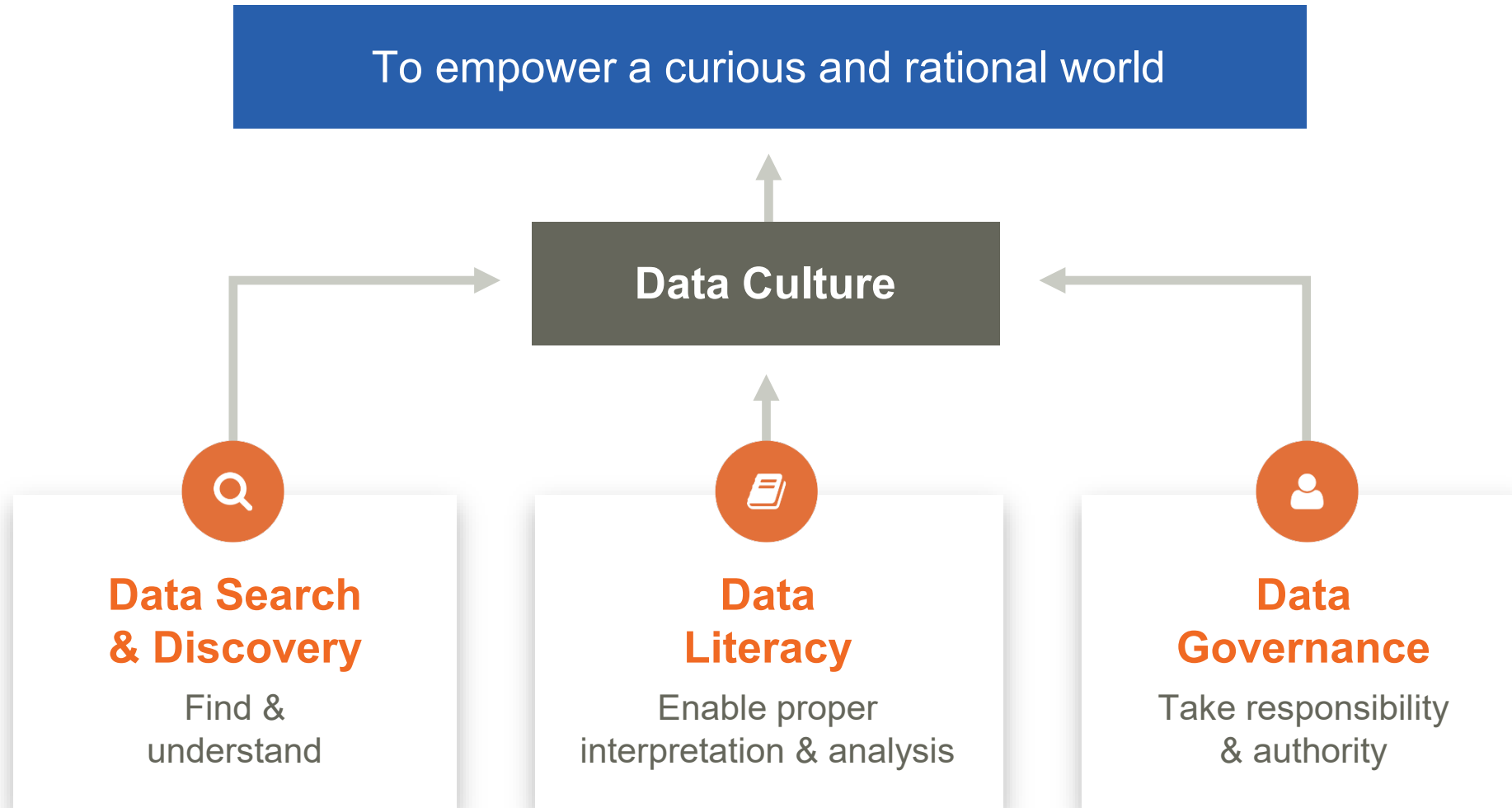
Data Catalogs

Data Intelligence + Human Brilliance – Automation on the Data Intelligence Platform

Jason Lim

Sept 20, 2023


The Alation Vision



What is Data Intelligence?

An organization's ability to deliver the

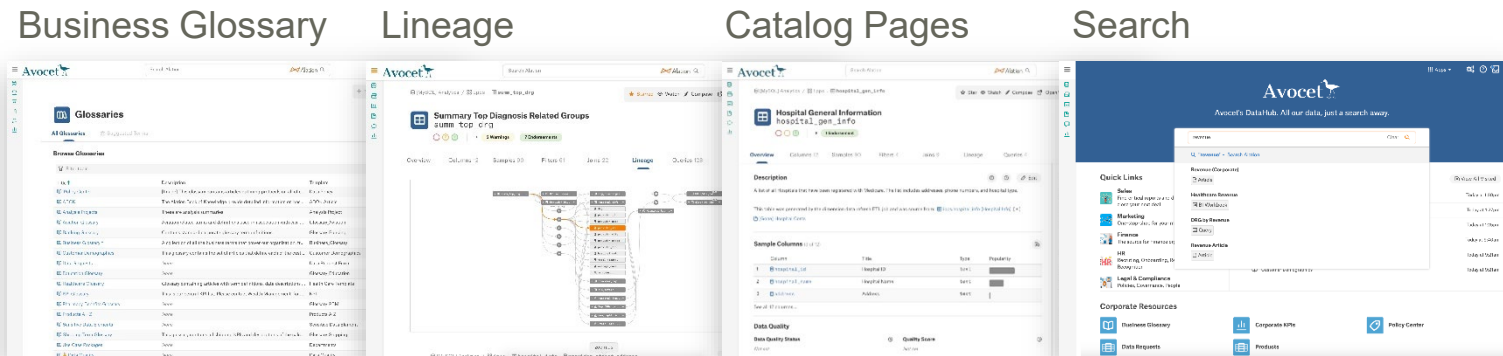
- right **data** to the
- right **person** at the
- right **time** for the
- right **purpose** in the
- right **medium/context**.



A **data intelligence** system learns and manages the **who, what, when, where, why** and **how** of data.

What is a Data Catalog?

- A repository of metadata on information sources across an organization
 - Search & discovery
 - Data governance & curation
 - Collaboration & analysis
- Catalogs a broad range of information assets
 - Data sets, tables, articles, reports, queries, visualizations, conversations
- Includes common functionality such as:



Answers these core questions:

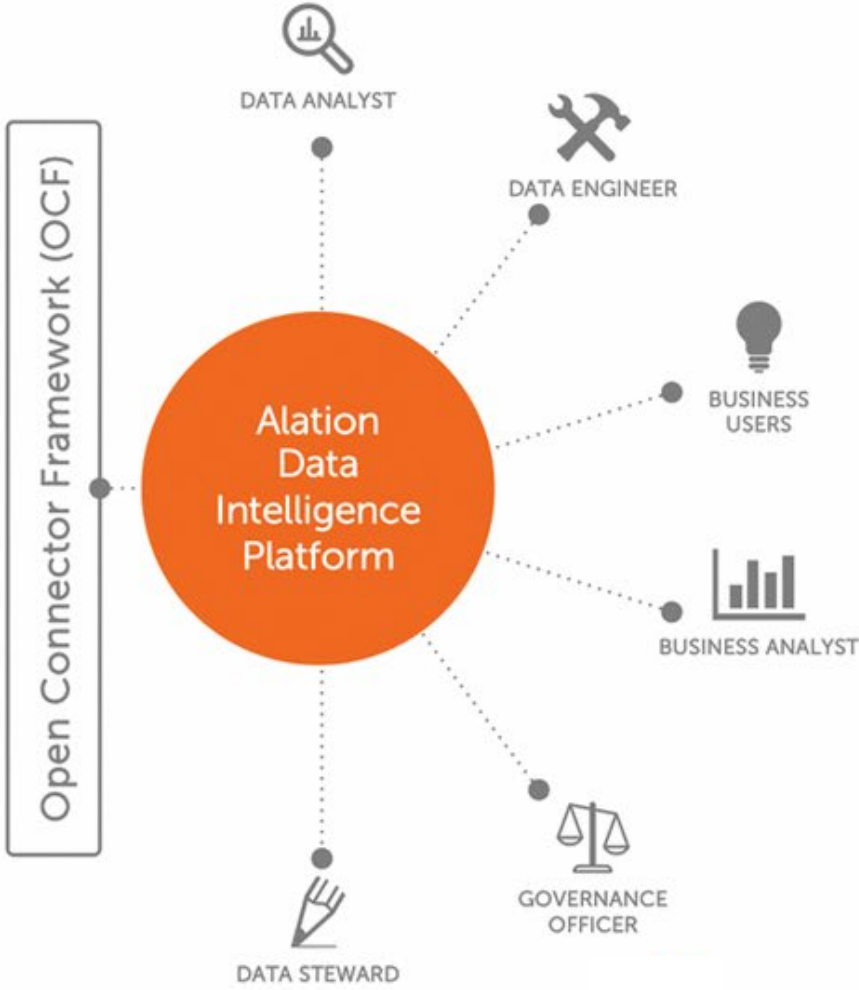
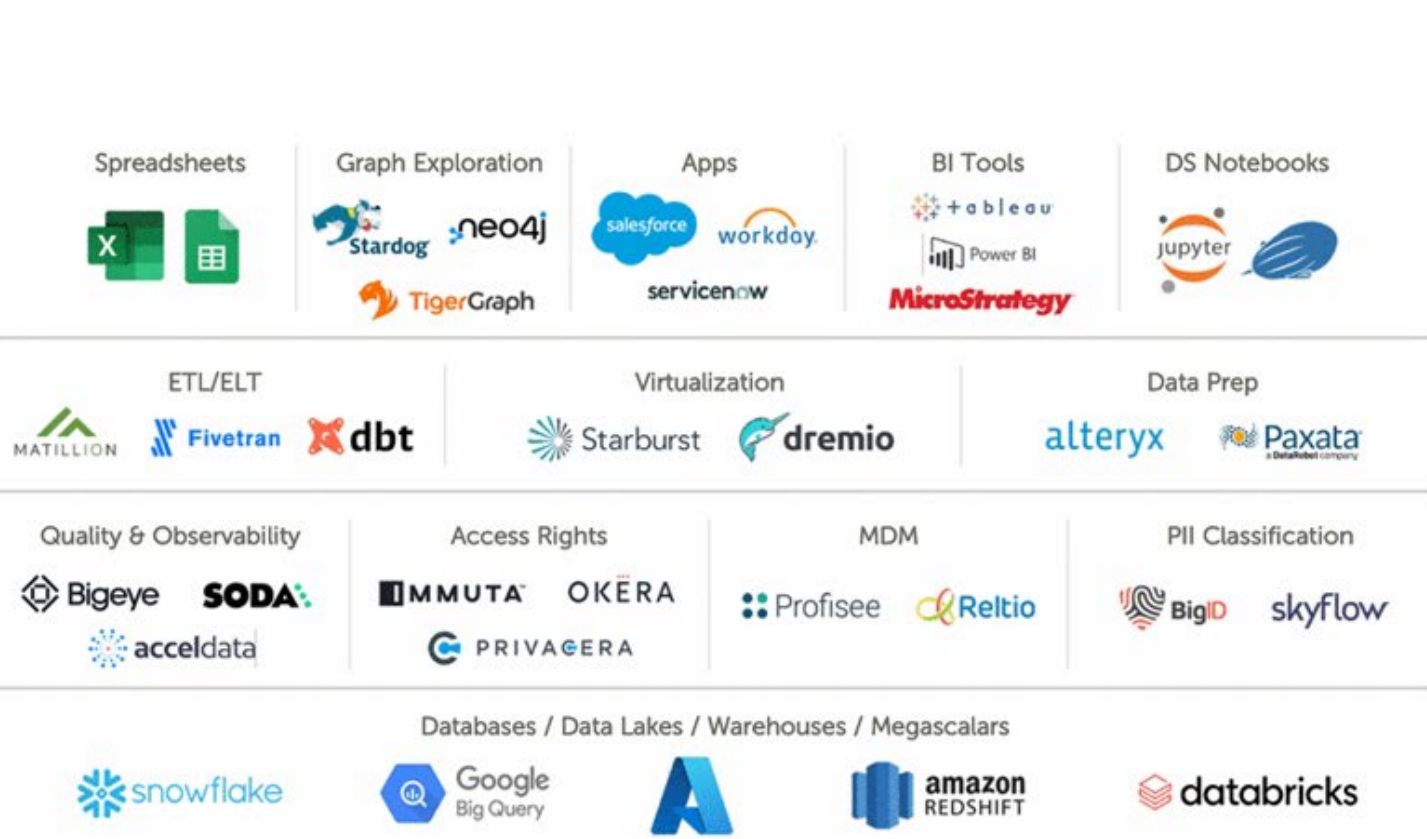
How to find information?

Can it be used?

Should it be used?

How should it be used?

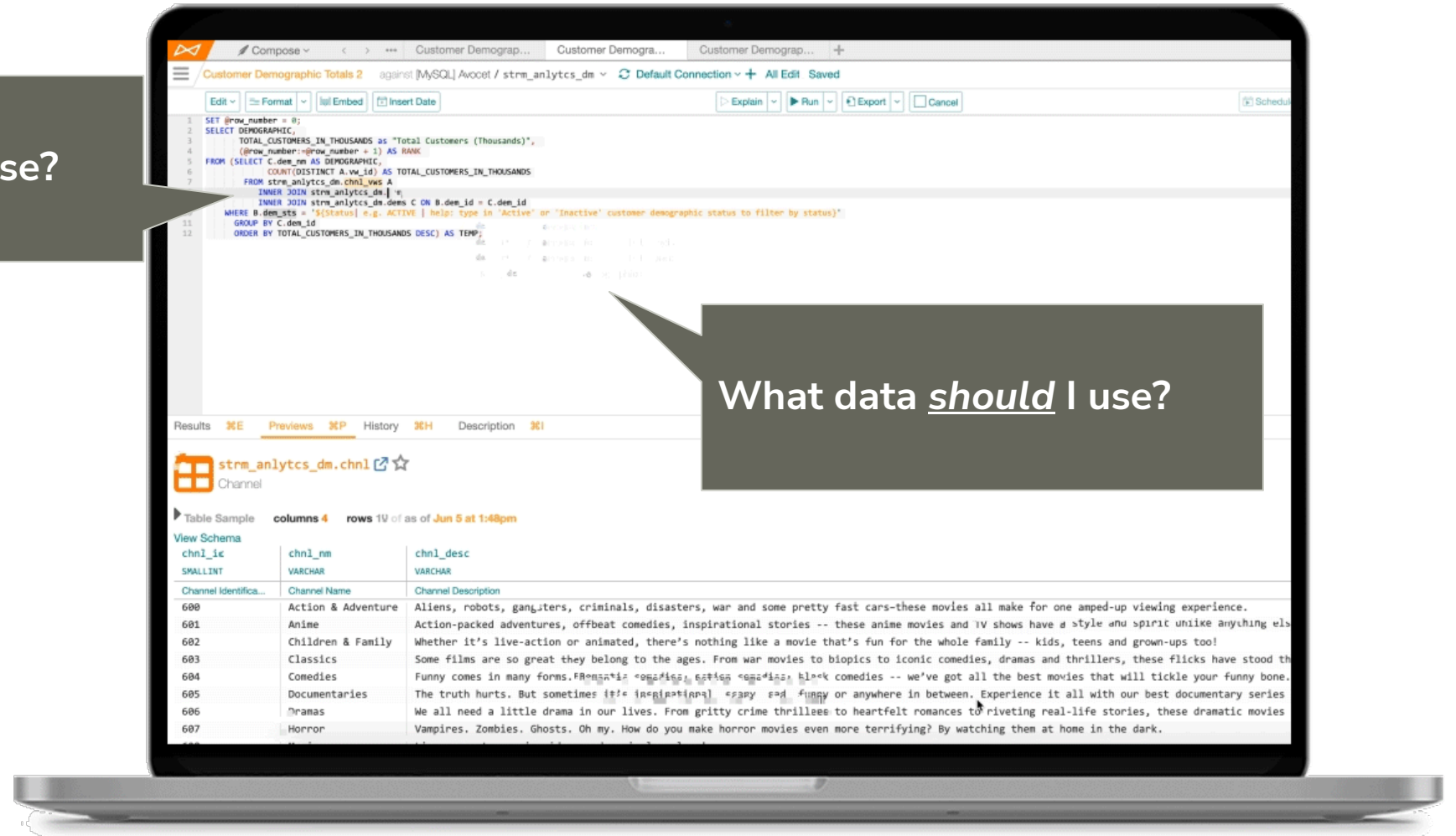
Universal Connectivity to Data Sources



Data Analysts and Scientists

What data can I use?

What data should I use?



Data Engineers and DataOps

Disc measure

Overview Fields 2 Sources 1 **Lineage**

Lineage Chart Viewing Default View Use Comp

Filters

Tableau Server 1

CNTY+ (testenv) Disc measure Dashboard 2

PostgreSQL-Alatio...

Where did the data come from?

Table
physical_name

1 Warning 3 Endorsements 1 Deprecation

Overview Columns 45 Samples 45 Filters 45 Joins 45 Lineage Queries 45 Health 1

Description

The new data... hospitals that receive Medicare Inpatient charges, paid under Medicare based on a rate per discharge... DRGs represent payments through... These charges. This data provides insight into doctor

Is this quality data?

Sample Columns

See all 45 columns

| Column | Title | Health | Type | Popularity ^ |
|--------------------|------------|----------|------|--------------|
| 1 client_plan_tier | Plan Tier | 🟢 | text | |
| 2 client_uid | Client UID | 🚩 1 | text | |
| 3 client_card_zip | Card Zip | ⚠️ 2 | text | |
| 4 _loaded_at | Loaded at | No Rules | text | |

How much do people use this data?

Business Users

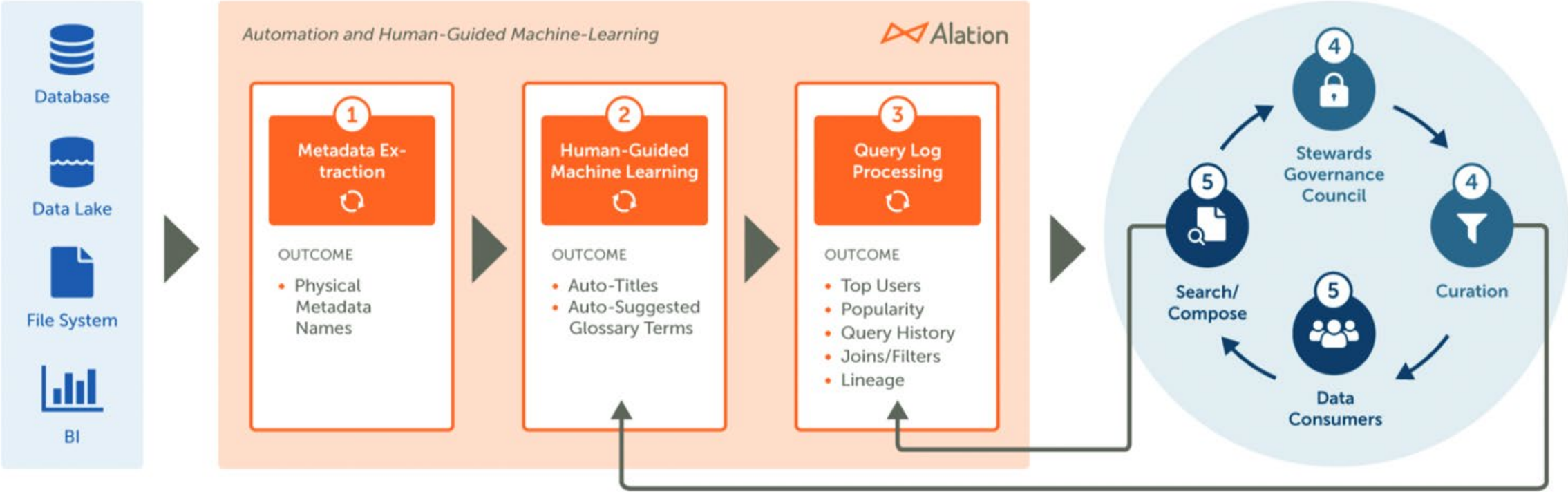
How do I get updated data quickly?

What data can I use in my spreadsheet?

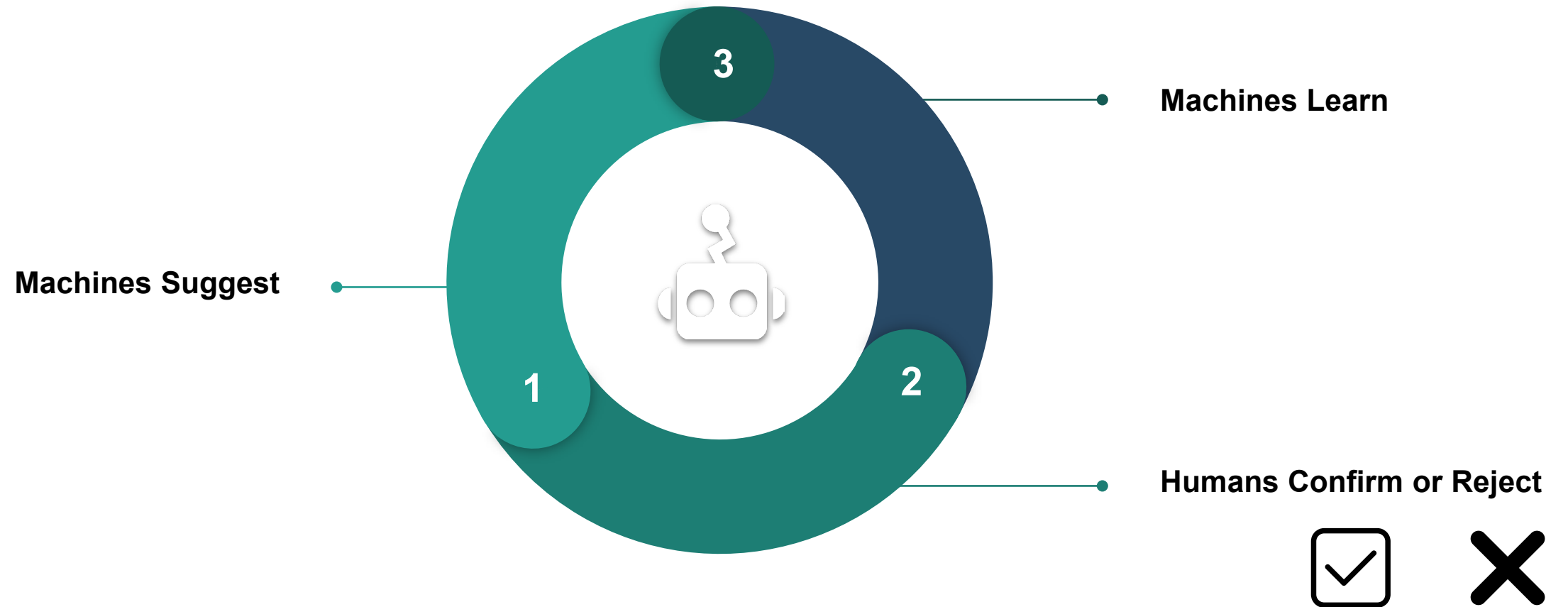
The screenshot shows a Google Sheet titled 'Bnk_Acct_Detail' with a 'Browse Catalog' dialog box open. The dialog has a search bar and a tree view of data sources. Under 'Postgres' > 'finance', the table 'Bnk_acct_detail' is selected. A 'Details' panel on the right shows the table name, trust check status (Endorsed by Rose Dore), description ('Bank account details'), owner ('Jane Smith'), and stewards ('Jane Smith, Ashwin Kashyap').

Is the data in this spreadsheet compliant?

Process for Automation and Human-Guided ML



Human Guided Machine Learning



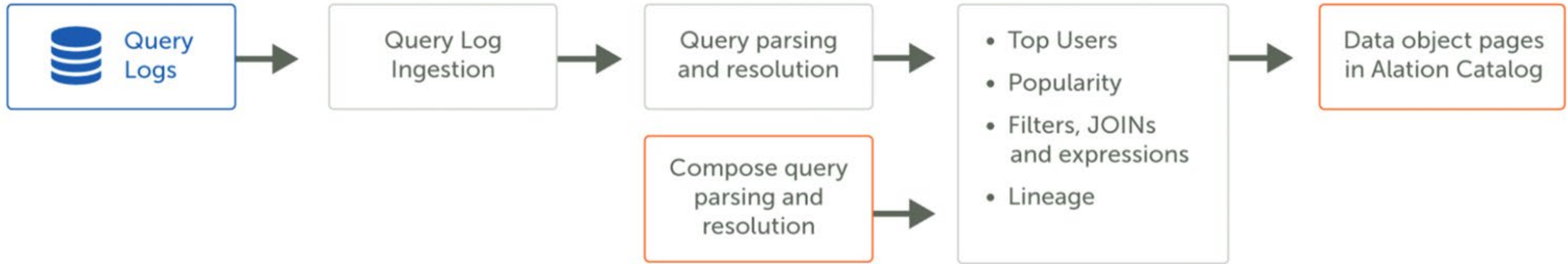
Lexicon Translates Technical to Business Terms

The screenshot shows the Lexicon application interface. At the top left is the Lexicon logo. Below it, the text "Suggested Expansions" is followed by a help icon. To the right are filter tabs for "All", "Confirmed", and "Unconfirmed", along with an "Add New" button. Below the filters is a search bar with a filter icon and the text "Filter table". The main content is a table with four columns: "Abbreviation", "Expands to", "Status", and "Abbreviation Frequency". The table contains several rows of data, all with a status of "Unconfirmed". A context menu is open over the row where the abbreviation is "id" and it expands to "Issue Date". The menu options are "Edit expansion" and "Add a different expansion".

| Abbreviation | Expands to | Status | Abbreviation Frequency ↓ |
|--------------|--------------|-------------|--------------------------|
| c | c | Unconfirmed | [Progress bar] |
| id | idea | Unconfirmed | [Progress bar] |
| id | ideation | Unconfirmed | [Progress bar] |
| id | Issue Date | Unconfirmed | [Progress bar] |
| sbqq | | Unconfirmed | [Progress bar] |
| leandata | | Unconfirmed | [Progress bar] |
| name | name | Unconfirmed | [Progress bar] |
| createddate | created date | Unconfirmed | [Progress bar] |

Making it easier for everyone to find and understand data

Query Log Processing





Demo

Resources

Alation, Inc.

<https://www.alation.com/>

What is a Machine Learning Data Catalog?

<https://www.alation.com/blog/machine-learning-data-catalog/>

Alation “Data Radicals” Podcast with Satyen Sangani, Alation CEO

<https://www.alation.com/podcast/>

revAlation 2023 User Conferences

<https://revalation.alation.com/>



Q&A