



# EDM Webinar

## How Cloud is making it easier to integrate external data sources

*A conversation with*



**Matt Holzapfel**  
Head of Corporate Strategy  
Tamr



**Chris Napoli**  
Industry Principal, Financial Services, Head of Wealth & Asset Management  
Snowflake



# Moderated by **Mike Meriton** Co-Founder & COO, EDM Council

- Joined EDM Council full-time 2015 to lead Industry Engagement
- EDM Council Co-Founder & First Chairman (2005-2007)
- EDM Council Finance Board Chair (2007-2015)
- Former CEO GoldenSource (2002-2015)
- Former Executive for D&B Software and Oracle
- FinTech Innovation Lab – Executive Mentor (2011 – Present)



Moderator



**Mike Meriton**

Co-Founder & COO  
EDM Council



**Matt Holzapfel**

Head of Corporate Strategy  
Tamr



**Chris Napoli**

Industry Principal, Financial  
Services, Head of Wealth  
& Asset Management  
Snowflake

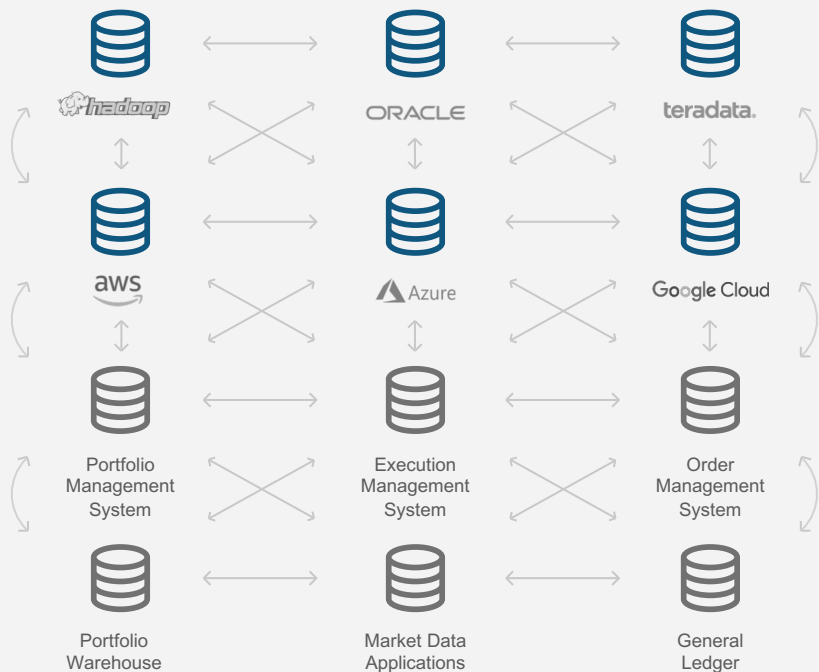




# How Cloud is making it easier to integrate external data sources

# Data and Technology Challenges

## Technology and Data Silos



## Technology and Data Challenges



Stale or incomplete data



Data inconsistencies & versioning issues



Poor data governance



Extended time to insight due to elongated data pipelines



Limited scale



Lack of resiliency

# Business Challenges



**Limited new data to fuel models, backtest, and perform research**



**Lack of scale to support compute-intensive algorithms**



**High costs and inefficiencies to meet regulatory obligations**



**Inability to integrate internal and external data for business functions**



**Inability to accurately value & service in a timely and automated way**



**Operational and regulatory risks with resilience & connectivity gaps**



# POLL #1 -

How many data cloud providers does your organization have?

# External Data Sets: Empowering Access & Collaboration

<p><b>COMMERCE</b></p>	<p><b>DEMOGRAPHIC</b></p>	<p><b>ECONOMY</b></p>	<p><b>ENERGY</b></p>
<p><b>ESG</b></p>	<p><b>FINANCIAL</b></p>	<p><b>GEOSPATIAL</b></p>	<p><b>GOVERNMENT</b></p>
<p><b>HEALTH &amp; LIFE SCIENCES</b></p>	<p><b>IDENTITY</b></p>	<p><b>MARKETING</b></p>	<p><b>MEDIA</b></p>
<p><b>SECURITY</b></p>	<p><b>TRANSPORTATION</b></p>	<p><b>TRAVEL</b></p>	<p><b>WEATHER</b></p>





# Security & Governance

## Know Your Data

Understand, classify, track data & data changes, and monitor usage

Object Tagging

Access History

Classification

Object Dependencies



## Protect Your Data

Secure sensitive data with policy-based access controls

Row Access Policies

Conditional Masking

Dynamic Data Masking

Anonymization

External Tokenization

## Integrate Your Data

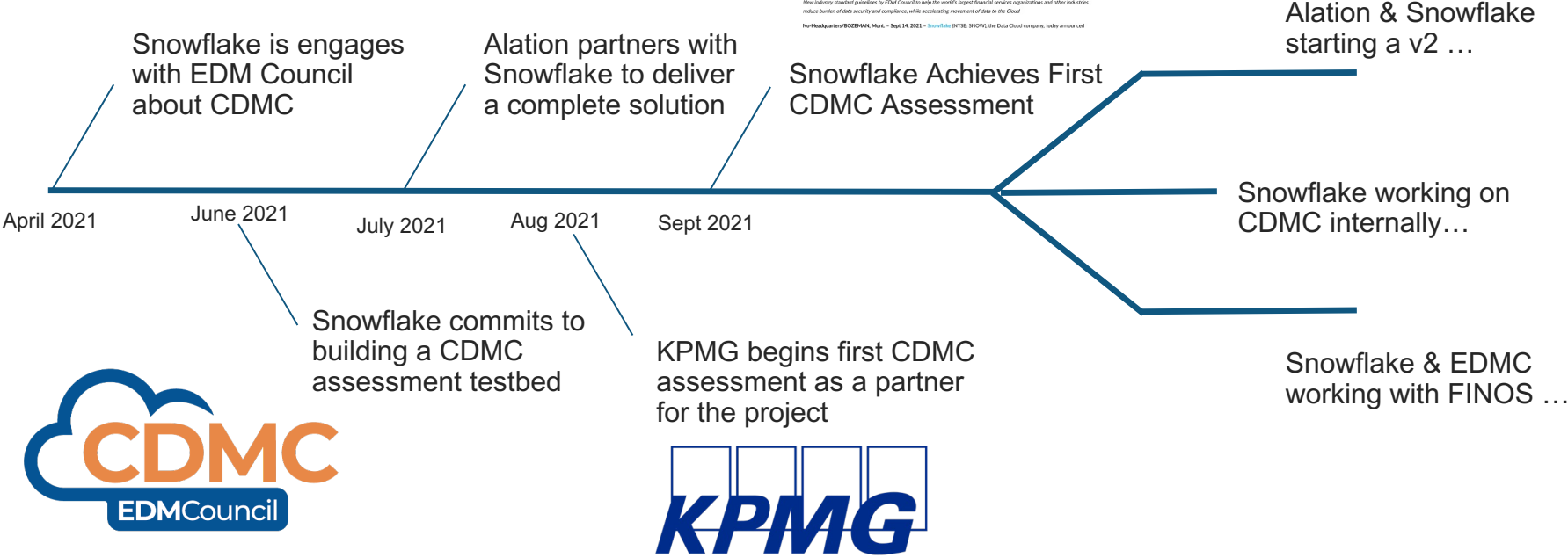
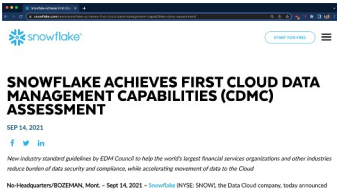
Securely collaborate and share data across teams and regions

Secure Data Sharing

Data Exchange

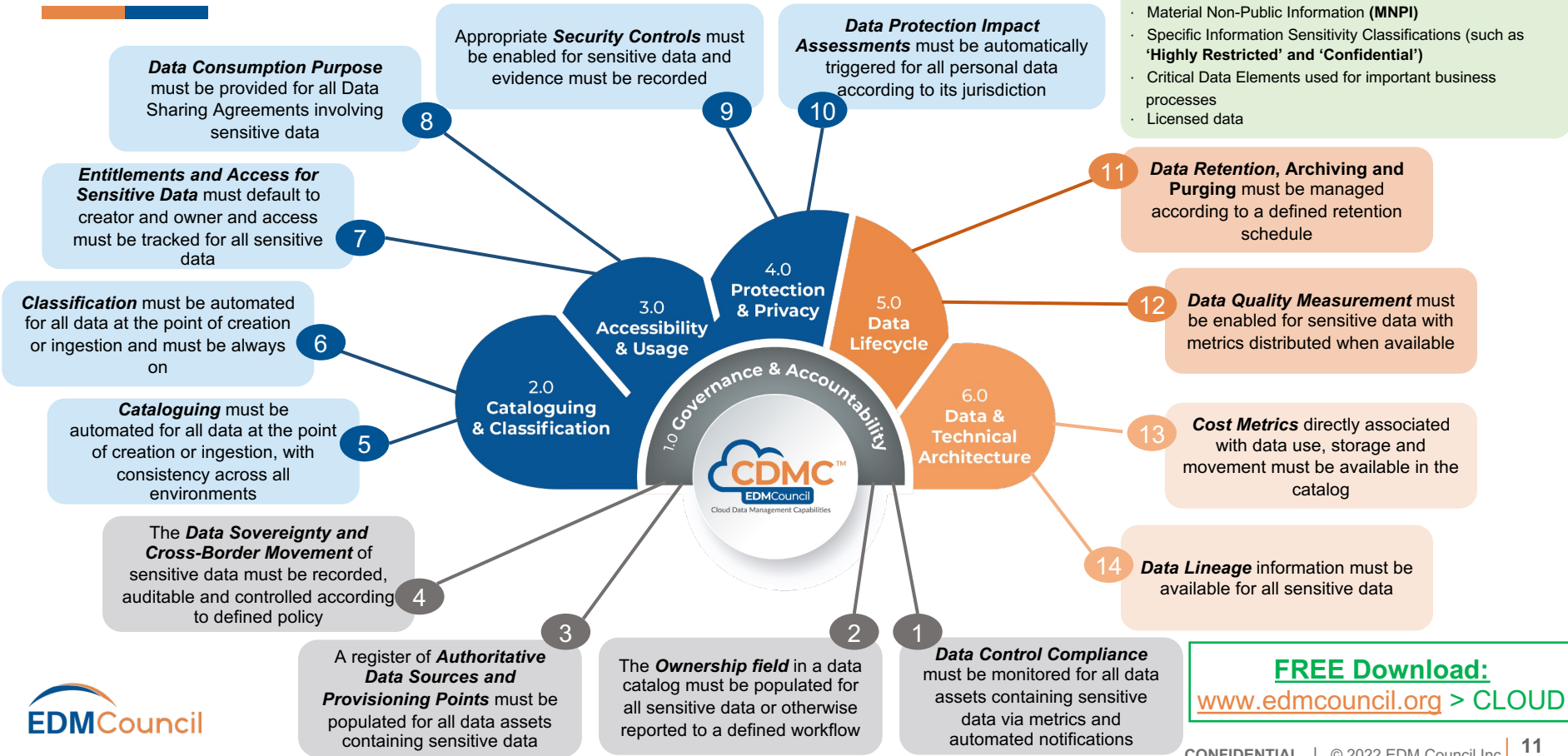


# Snowflake's CDMC Journey (so far)



# Cloud Data Management Capabilities (CDMC)

## 14 Key Controls for Managing Data Risk



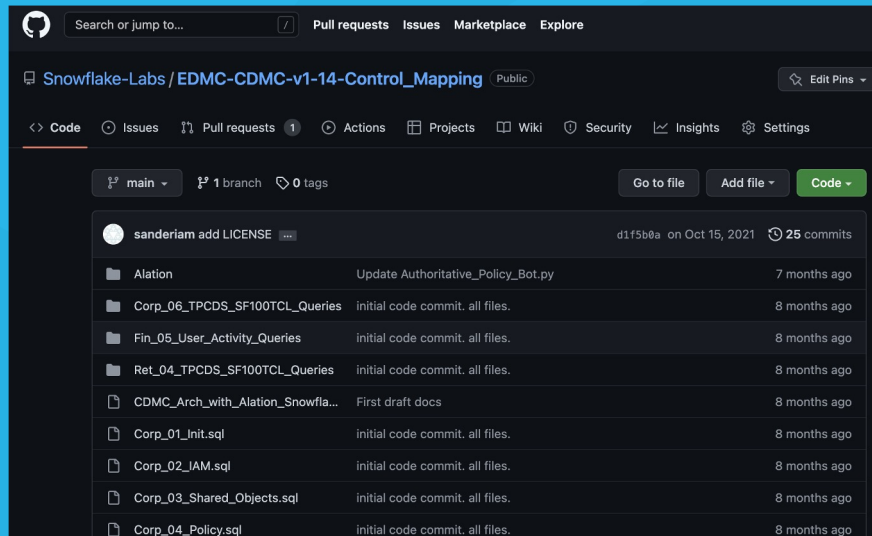
**Sensitive Data** includes classifications such as:

- Personal Information (PI) / Sensitive Personal Data
- Personally Identifiable Information (PII)
- Personal Health Information (PHI)
- Company or Client Identifiable Information
- Material Non-Public Information (MNPI)
- Specific Information Sensitivity Classifications (such as 'Highly Restricted' and 'Confidential')
- Critical Data Elements used for important business processes
- Licensed data

**FREE Download:**  
[www.edmcouncil.org](http://www.edmcouncil.org) > CLOUD

# Assets Available Now

- All the materials used to pass the CDMC assessment are on github
  - [https://github.com/Snowflake-Labs/EDMC-CDMC-v1-14-Control\\_Mapping](https://github.com/Snowflake-Labs/EDMC-CDMC-v1-14-Control_Mapping)
- The Snowflake **Financial Services Data Summit** to get access to keynotes, breakout sessions, industry content and more from leading experts in financial services
- “The journey to Processing PII in the Data Cloud”
  - [BUILD 2021 session on YouTube](#)
  - [Snowflake Blog Post](#)





## POLL #2 -

What data sets do you find most useful to incorporate into your analysis?

# External Data: An Opportunity & Challenge

~ **1/2**

Companies reported using **external data** in their analytics activities.<sup>1</sup>

**92%**

professionals said their firms needed to **increase use of external data**<sup>2</sup>

**54%**

said their companies plan to **increase spending on external data**<sup>2</sup>



No primary keys between different sources



Data quality gets worse for longtail sources



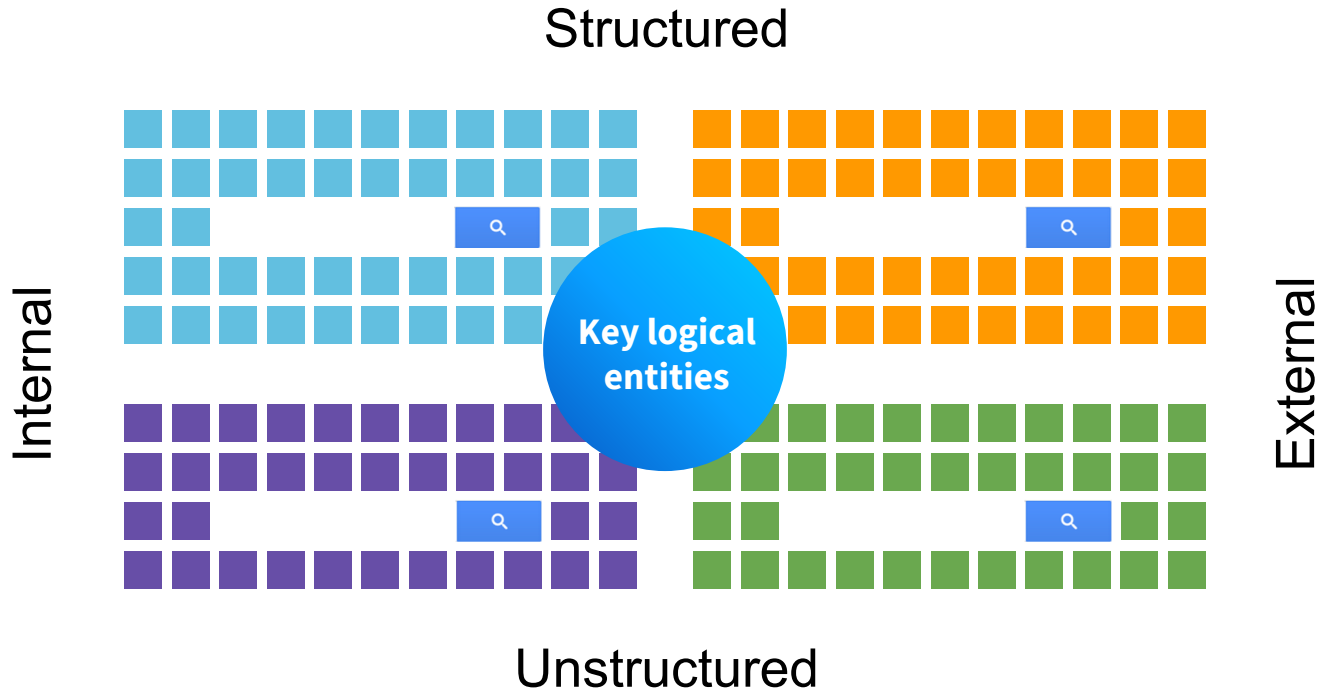
Reliant on third-party matching services



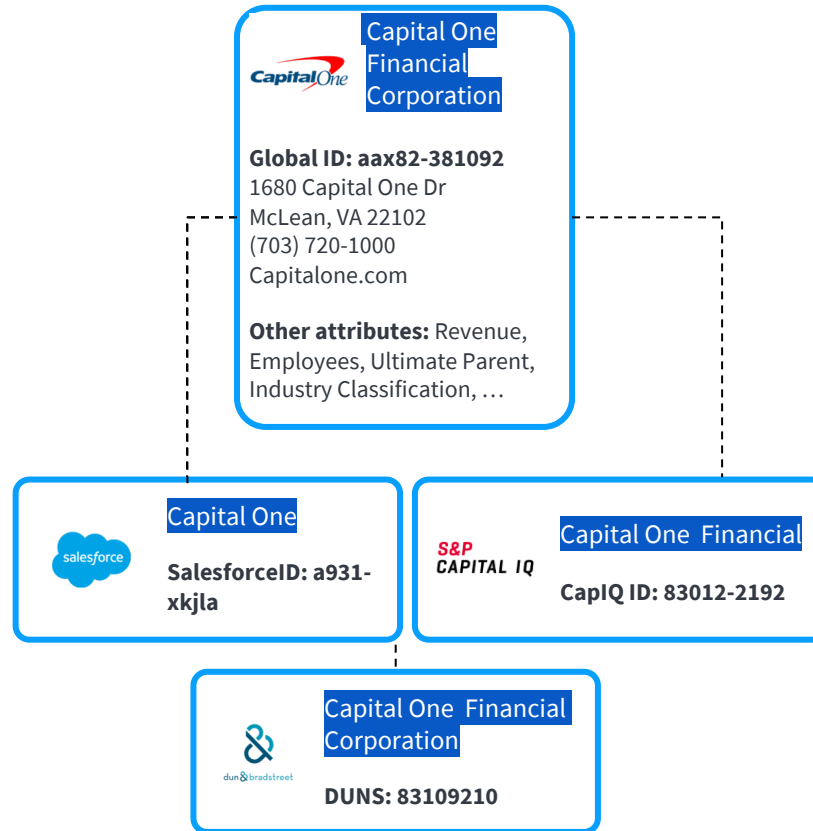
Inconsistent internal and external quality standards

1. Gartner, "[Gartner survey shows organizations are slow to advance in data and analytics](#)," press release, February 5, 2018
2. Forrester, "[Digital is driving the next generation of data marketplaces](#)," December 2017.

# Goal: Complete data organized by logical entity



# Goal: Complete data organized by logical entity



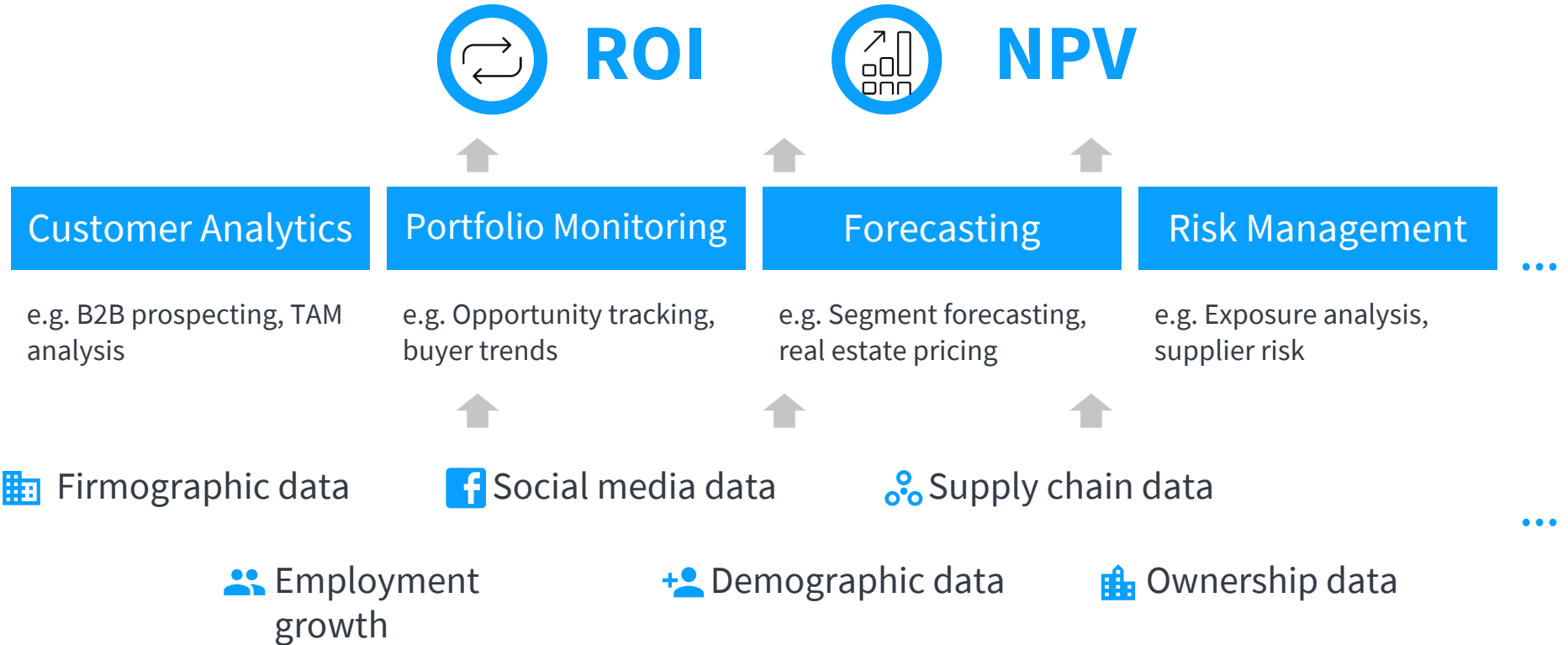




## POLL #3 -

How many sources of third-party data do you use?

# Strong business case for external data...



# ...and need to automate

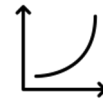
**Minimize Total Cost of  
Ownership of Data  
Mastering**



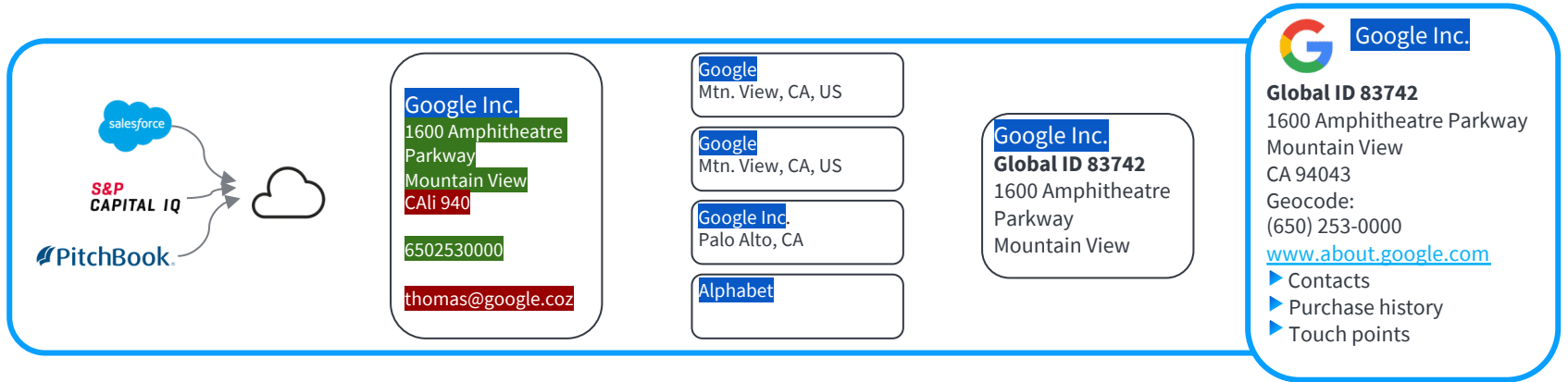
**Improve Productivity for  
Functional Teams by  
Having the Right Data**



**Build Applications &  
Analytics on Top of  
Integrated Data Asset**



# How to do it? Think modularly



## Data Aggregation

Establish 'raw zone' for landing disparate internal & external tables

## Data Cleaning

Validate & standardize raw data based on global reference data

## Record Matching

Cluster records within and across sources that represent the same entity

## Persistent ID

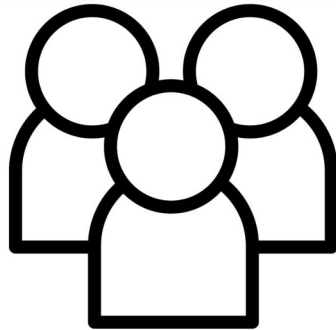
Assign each cluster a unique, persistent identifier to track the entity

## Mastered Record

Create & maintain a single version of each entity, consolidating the best information



**100+ on-prem instances**



**10 - 15 people needed to manually prepare entity data for analytics**



**Slow, manual processes for delivering insights**

**Many business decisions made without data**



**Data centralized in  
Snowflake**



**Data curated by entity (e.g.,  
legal entity, property)**



**Easy, self-serve access  
to insight-ready data**



**More data-driven  
business culture**

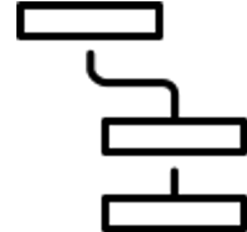
# Lessons Learned



**Humans + machine =  
better than humans or  
machine alone**



**Data cleaning matters a  
lot... even with  
“trusted” sources**



**Identify primary vs  
secondary sources**



**Q&A**



# See the integration in action!

Visit [tamr.com](https://tamr.com) and [snowflake.com](https://snowflake.com) to learn more



Matt Holzapfel  
[mh@tamr.com](mailto:mh@tamr.com)

Head of Corporate Strategy



Chris Napoli  
[chris.napoli@snowflake.com](mailto:chris.napoli@snowflake.com)

Industry Principal, Financial Services,  
Head of Wealth & Asset Management

