













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)





Today's panel





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



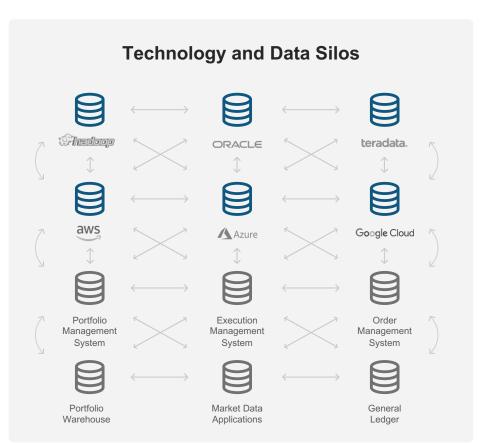






How Cloud is making it easier to integrate external data sources

Data and Technology Challenges



Technology and Data 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



Inability to integrate internal and external data for business functions



Lack of scale to support compute-intensive algorithms



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



High costs and inefficiencies to meet regulatory obligations



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



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 policybased access controls

Row Access Policies	Conditional Masking
Dynamic Data Masking	Anonymization
External Tokenization	

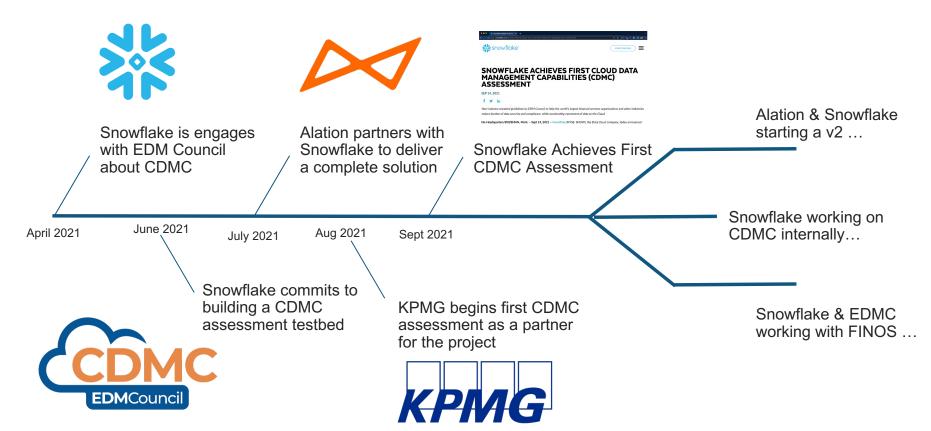
Integrate Your Data

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) Personal Information (PI) / Sensitive Personal Data Personally Identifiable Information (PII) 14 Key Controls for Managing Data Risk Personal Health Information (PHI) Company or Client Identifiable Information Material Non-Public Information (MNPI) Data Protection Impact Appropriate Security Controls must Specific Information Sensitivity Classifications (such as Assessments must be automatically be enabled for sensitive data and 'Highly Restricted' and 'Confidential') Data Consumption Purpose triggered for all personal data evidence must be recorded Critical Data Elements used for important business must be provided for all Data according to its jurisdiction processes Sharing Agreements involving 8 Licensed data sensitive data Data Retention, Archiving and Entitlements and Access for Purging must be managed Sensitive Data must default to according to a defined retention creator and owner and access schedule must be tracked for all sensitive data 4.0 **Protection** Classification must be automated Data Quality Measurement must 3.0 & Privacy be enabled for sensitive data with for all data at the point of creation Accessibility Data Lifecycle metrics distributed when available or ingestion and must be always & Usage Covernance & Account on 2.0 Cataloguing Cataloguing must be Data & & Classification Cost Metrics directly associated automated for all data at the point 5 with data use, storage and of creation or ingestion, with movement must be available in the consistency across all catalog environments The Data Sovereignty and Cross-Border Movement of Data Lineage information must be sensitive data must be recorded. available for all sensitive data auditable and controlled according to defined policy Data Control Compliance A register of Authoritative The Ownership field in a data **FREE Download:** must be monitored for all data **Data Sources and** catalog must be populated for assets containing sensitive www.edmcouncil.org > CLOUD Provisioning Points must be all sensitive data or otherwise

reported to a defined workflow

EDMCouncil

populated for all data assets

containing sensitive data

11

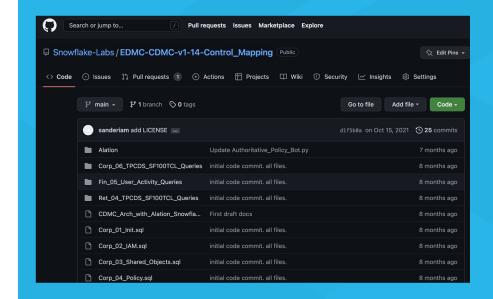
data via metrics and

automated notifications

Sensitive Data includes classifications such as:

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
- 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.¹

92%

professionals said their firms needed to **increase use of external data**²

54%

said their companies plan to **increase spending on external data**²



No primary keys between different sources



Data quality gets worse for longtail sources



Reliant on third-party matching services

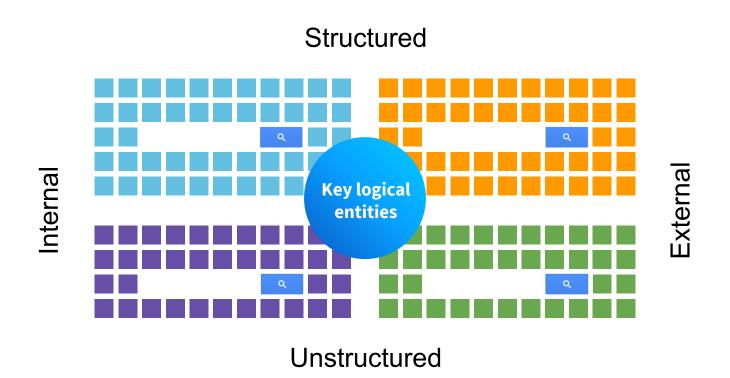


Inconsistent internal and external quality standards

^{..} Gartner, "Gartner survey shows organizations are slow to advance in data and analytics," press release, February 5, 2018

^{2.} Forrester, "<u>Digital is driving the next generation of data marketplaces</u>," December 2017.

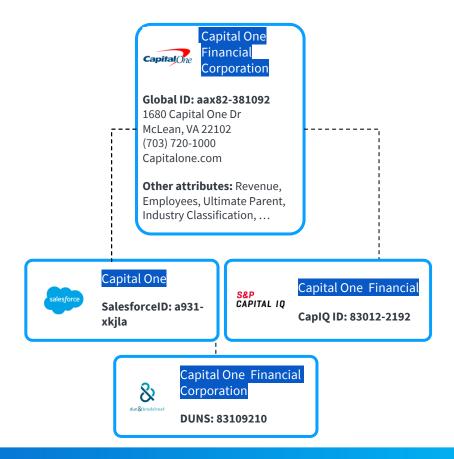
Goal: Complete data organized by logical entity



All Rights Reserved – Tamr, Inc.

15

Goal: Complete data organized by logical entity



All Rights Reserved – Tamr, Inc.



POLL #3 -

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

Strong business case for external data...







e.g. B2B prospecting, TAM analysis

Firmographic data

Portfolio Monitoring

e.g. Opportunity tracking, buyer trends

Forecasting

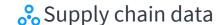
e.g. Segment forecasting, real estate pricing

e.g. Exposure analysis, supplier risk

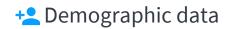
Risk Management













All Rights Reserved - Tamr, Inc.

18

...and need to automate

Minimize Total Cost of Ownership of Data Mastering



Improve Productivity for Functional Teams by Having the Right Data

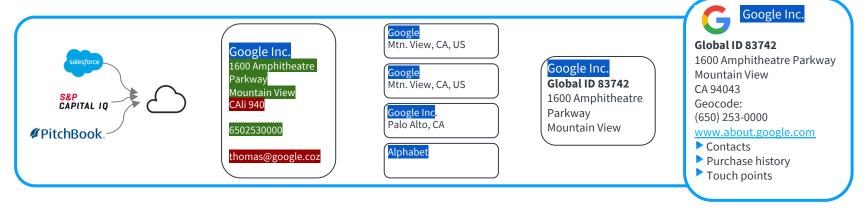


Build Applications & Analytics on Top of Integrated Data Asset



All Rights Reserved – Tamr, Inc.

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

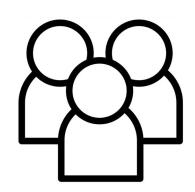
Create & maintain a single version of each entity, consolidating the best information

Blackstone

Before: Many silos, long time-to-insight



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

Blackstone

After: Centralized, curated data for fast insights







Easy, self-serve access to insight-ready data

Data centralized in Snowflake

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



More data-driven business culture

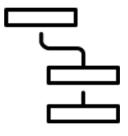
Lessons Learned



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 <u>tamr.com</u> and <u>snowflake.com</u> to learn more





Head of Corporate Strategy





Chris Napoli chris.napoli@snowflake.com

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