

# Care to Share?

A New Framework for Identity & Data Sharing

# About Me

- 10 Years Experience in Data, Technology, and Finance
  - 2x Founder
  - Growth Startups: KoalaFi & Clearco
- Profitability Focus to Scale
  - Increased KoalaFi's EBIT by 20M
  - Reduced Clearco Credit Losses by 85%
  - Launched Data Products to increase revenue by 15%+



Alexander Lawson

# Data Sharing, Data Silos, & The Modern Data Platform

# What is data sharing?

Data is organized, version controlled, categorized, and accessible cross functionally

1. An associated that doesn't know SQL can make data driven decisions
2. An analyst can access and manipulate information easily to make data driven recommendations
3. Systems can access cross system data and adjust based on that data

Do you have?

Data Swamp

Data Silos

Excel Based Reporting

Data Scientists

Broken Pipelines

Data



Have you been told you need?

Data Warehouses

Data Cloud

Data Mesh

Data Lake

Data Engineers



# You are not alone

Data Mismanagement costs the US economy 3.1T Annually

## 80%

Of companies have difficulty managing data governance

Data governance and access management is a real problem.

Is our data secure?

- Security Conscious RBAC
- Obfuscation Layers
- Data Governance

## 82%

Of companies have difficulty controlling data costs

Long running queries, growing costs, data accuracy & trust.

Is our data up to date & accurate?

- Data Quality Audit
- Preparation Team & System
- Data Catalogue

## 32%

Of companies have difficulty with data sprawl

Multiple data warehouses, a data lake, and production access systems.

Is our all our data connected?

- Data Culture
- Self Service Tooling
- ETL Standards

# A Framework for the Modern Data Platform

# Building a Data Strategy

1. Data Culture Enabling Data Driven Decisions Making
2. Data Categorization and Management System
3. RBAC, Data Governance, & Data Obfuscation
4. Data Quality & Transformation Layers
5. Self Service and Enablement





# Data Culture Enabling Data Driven Decisions Making



1. Management and Executive Alignment
2. Data Governance Practices
  - a. Data Reporting
  - b. Data Categorization
  - c. Data Accessibility
3. Data Discovery Landscape
  - a. Data Search
  - b. Data Quality
  - c. Data Literacy & Education

# Data Categorization and Management System

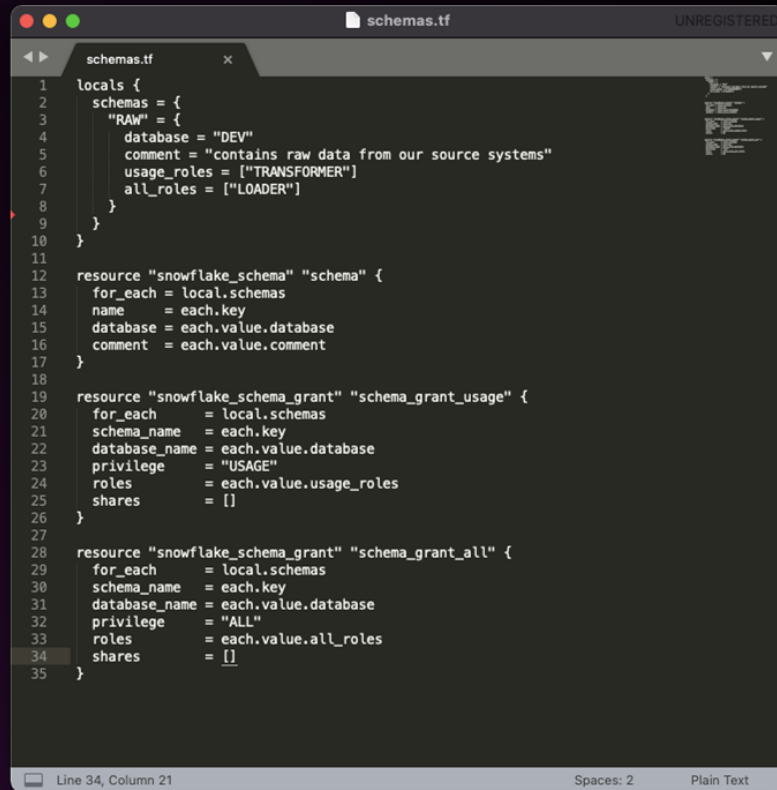
1. Search for data & conduct data investigations
2. What are these data sets and /or attributes?
3. What metadata can we associate with these data sets and / or attributes?
4. Distribute Trial Knowledge
5. Reduce Data Errors & Gaps

BNPL Details				
Field Label	Description	LookML Name	Type	SQL
IDs BNPL ID	BNPL table unique id	bnpl_details.id	Number	<code>\${TABLE}."ID"</code>
IDs Bill ID	Individual bill id used to determine every unique bill submitted.	bnpl_details.bill_id	String	<code>\${TABLE}."BILL_ID"</code>
Bnpl Type	type of BNPL product	bnpl_details.bnpl_type	String	<code>\${TABLE}."BNPL_TYPE"</code>
Created Date Date	Date-time at which bill is submitted.	bnpl_details.created_date	Date date	<code>(convert_timezone('UTC', 'America/Toronto', \${TABLE}."CREATED_AT":timestamp_ntz))</code>
Created Date Month	Date-time at which bill is submitted.	bnpl_details.created_month	Date month	<code>(convert_timezone('UTC', 'America/Toronto', \${TABLE}."CREATED_AT":timestamp_ntz))</code>
Created Date Quarter	Date-time at which bill is submitted.	bnpl_details.created_quarter	Date quarter	<code>(convert_timezone('UTC', 'America/Toronto', \${TABLE}."CREATED_AT":timestamp_ntz))</code>

# RBAC, Governance Application, & Data Obfuscation

Protect your data, Maximize User Potential

1. Centralized Data Governance Policies
2. PII Obfuscations or Data Removal
3. Code Driven Data Access Protocols
  - a. Team, Role, Zone, and Department Level



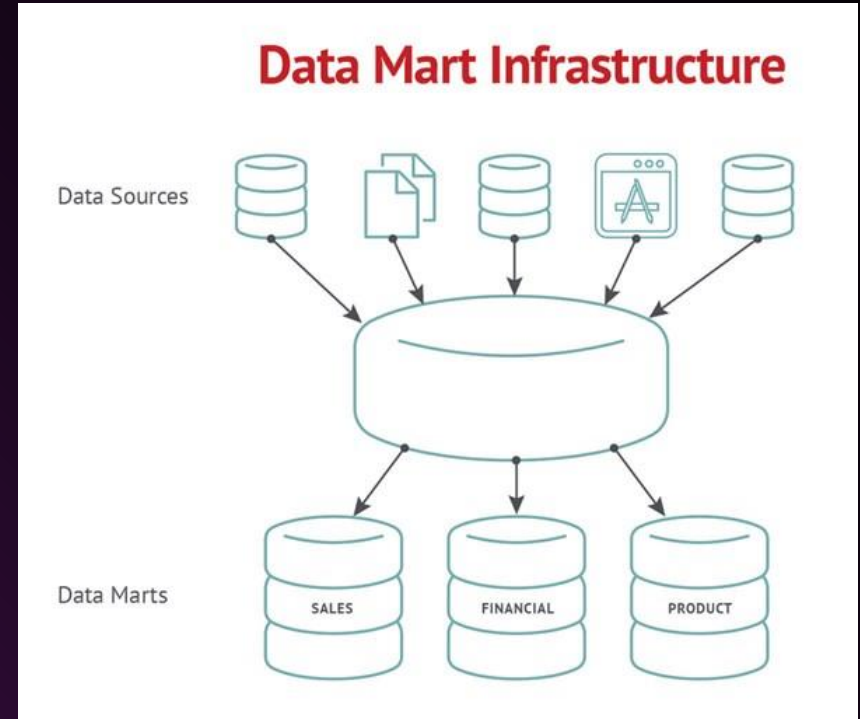
```
schemas.tf
1 locals {
2   schemas = {
3     "RAW" = {
4       database = "DEV"
5       comment = "contains raw data from our source systems"
6       usage_roles = ["TRANSFORMER"]
7       all_roles = ["LOADER"]
8     }
9   }
10 }
11
12 resource "snowflake_schema" "schema" {
13   for_each = local.schemas
14   name     = each.key
15   database = each.value.database
16   comment  = each.value.comment
17 }
18
19 resource "snowflake_schema_grant" "schema_grant_usage" {
20   for_each = local.schemas
21   schema_name = each.key
22   database_name = each.value.database
23   privilege   = "USAGE"
24   roles       = each.value.usage_roles
25   shares      = []
26 }
27
28 resource "snowflake_schema_grant" "schema_grant_all" {
29   for_each = local.schemas
30   schema_name = each.key
31   database_name = each.value.database
32   privilege   = "ALL"
33   roles       = each.value.all_roles
34   shares      = []
35 }
```

Line 34, Column 21      Spaces: 2      Plain Text

# Data Quality & Transformation Layers

Make Data Accessible & Maintainable

- Data Aggregation, Normalization, Joining, and Filtering  
Streamlining Query Performance
- Data Quality Validations through Automated Data Testing
- Leverage CI/CD for Data Transformation



















# Self Service and Enablement


1. Multi-Tenant BI Solution
  - a. Low No Code BI Tool like Tableau, PowerBI, or Looker
  - b. SQL Access to the mart and warehouse level
2. Access to Data Pipeline Tools for Easy Stand Up of New Data Sources
  - a. Build directly into the data catalogue
  - b. Upload data to their Sandbox
3. Systematic Access to new data sources
  - a. Automated Learning as new data enters

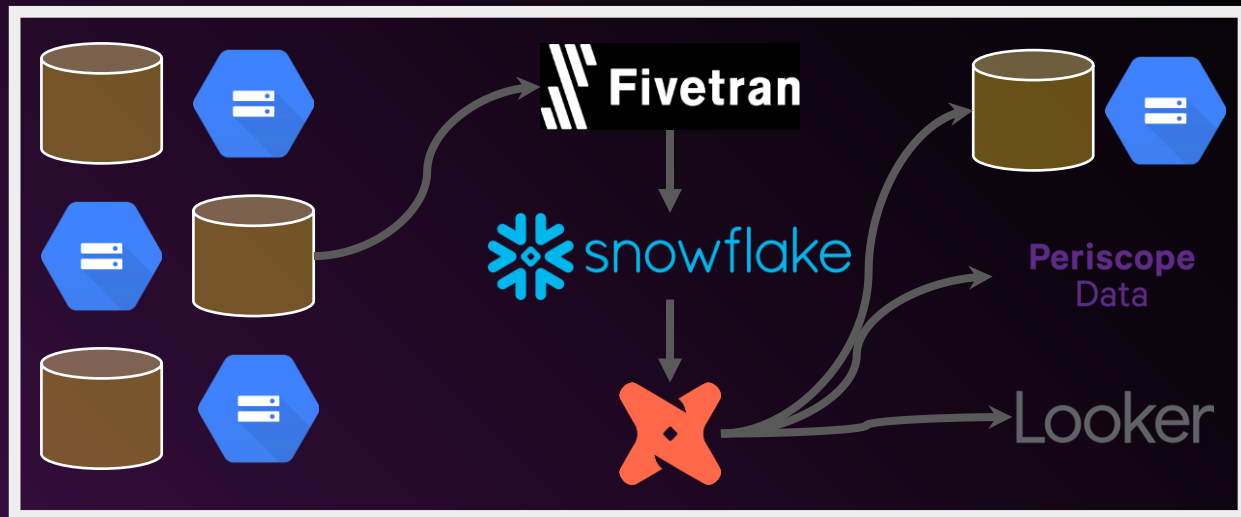
# What does the modern data platform look like?

Clearco uses proprietary machine learning models, trained across a growing database of 35,000+ businesses, to predict a company's future performance.

 Sources of system of record data

Storefront Data	   
Transaction Data	
Real Time Access To Multiply Bank Accounts	
Accounting Systems	   
Marketing Platform Data	    
Other 3rd Party & Internal Data	

 Data Platform



How did you overcome your data silos?

What are your tips for data sharing?

**Q&A**