coalesce

Transforming Loan Warehousing

Scalable data pipelines that empower financial analysts

David Maguire, Data Engineer, dv01

Meet today's presenter from this company

David Maguire

Data Engineer

Tech Lead

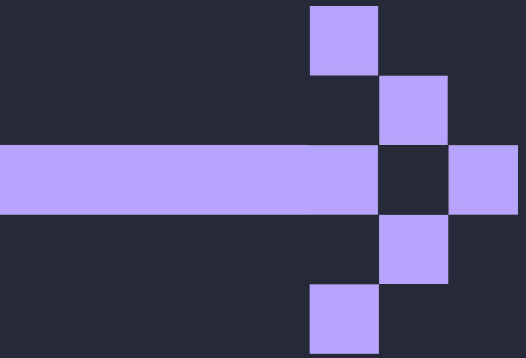
dmaguire@dv01.co

dv01

a Fitch Solutions Company

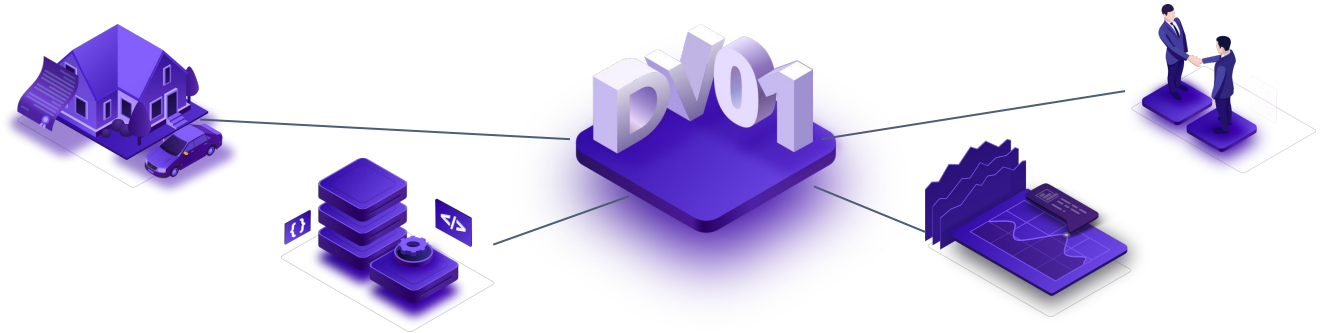
How to build cross functional data teams

Business Value = Engineering + Subject Matter Expertise



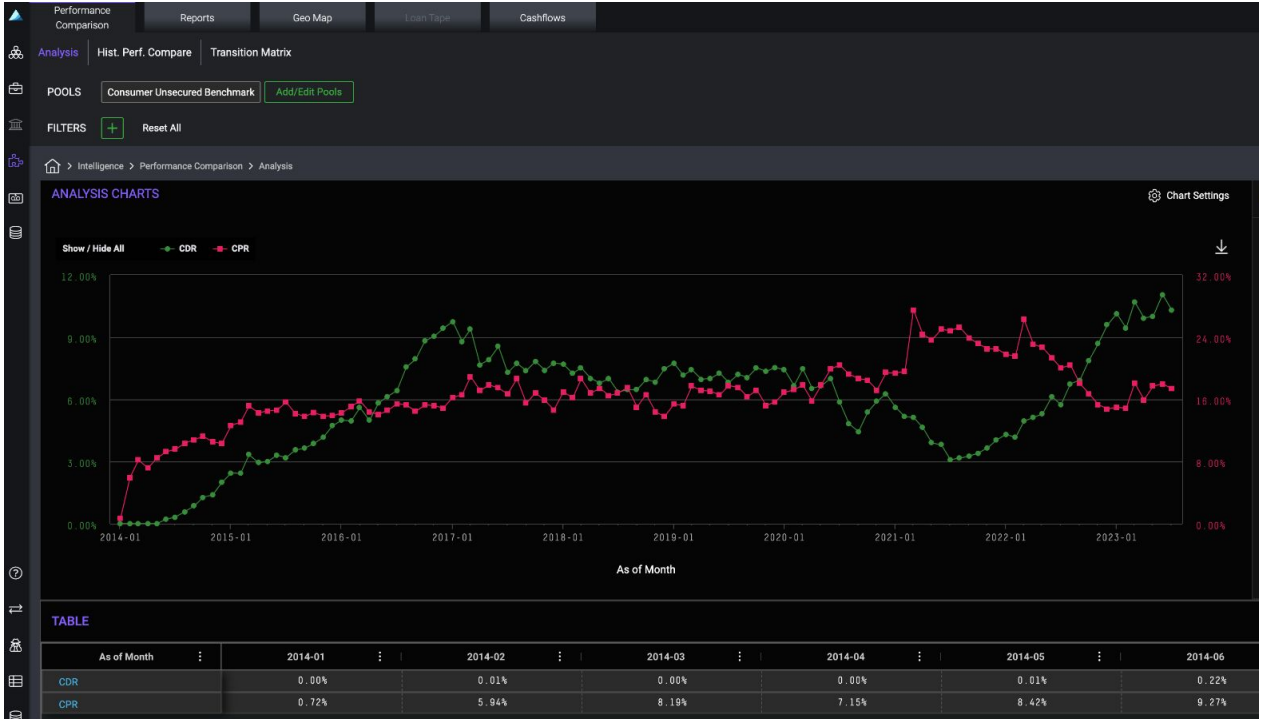
Background

dv01 is the Data Hub for Institutional Investors in Structured Finance



- A retail investor:
  - **Security Types**: Stocks, ETFs, Mutual Funds, bonds
  - **Performs:** Research and analysis
  - **Why:** To determine which to buy and sell
- A structured finance investor has the same motive as retail investor but the securities comprises pools of loans:
  - **Security Types:** Asset-Backed Securities (“ABS”) & Mortgage-Backed Securities (“MBS”).

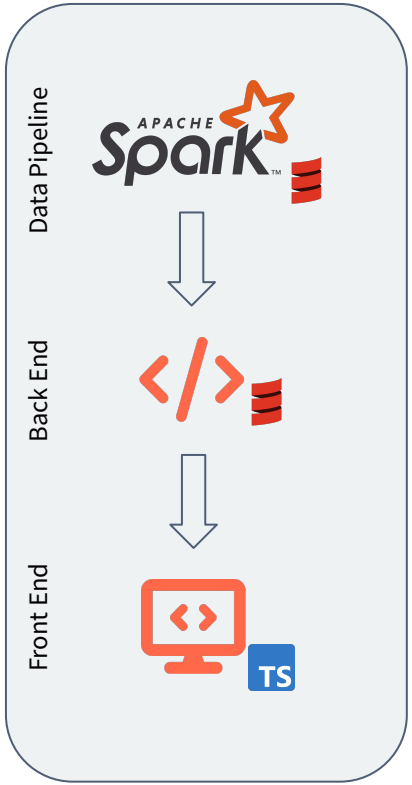
The dv01 Web App

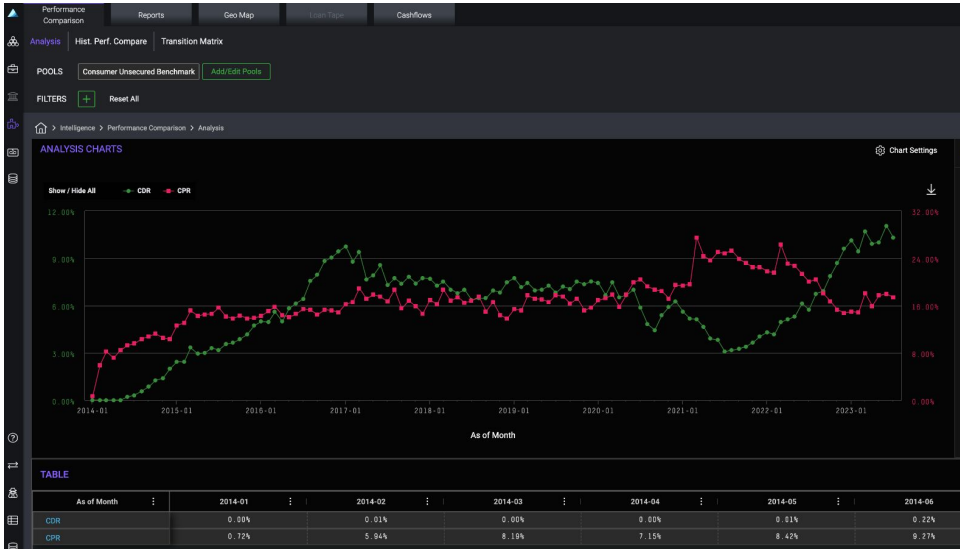


- dv01 standardizes loan data **and** offers users embedded analytics for analysis
- Data can also be accessed via SFTP files and BigQuery

Our data pipeline is designed for a highly engineered tech stack

SaaS Offerings

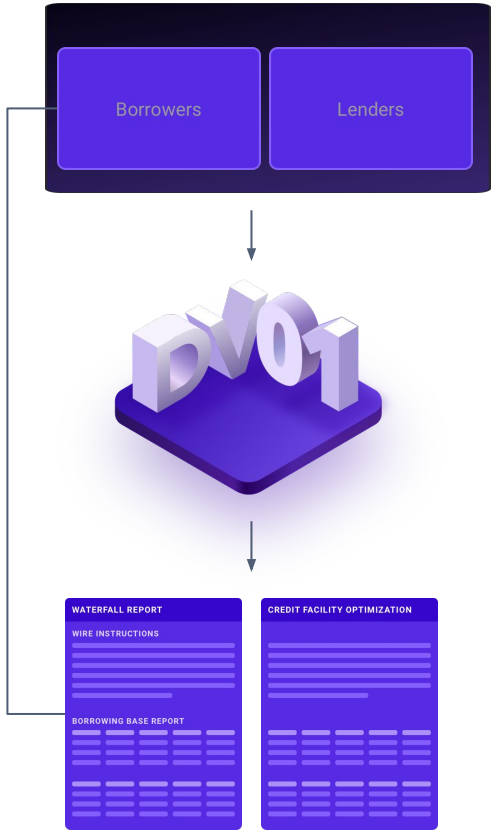




A new business line sprouts

Credit Facility Analytics

A credit facility is credit line for investors offered by a bank



- A borrower negotiates a revolving credit line from a lender. The borrower uses the credit line to purchase loans. The purchased loans become collateral used to make additional borrowing from the credit line.

- Agreements include stringent requirements on asset performance and financial covenants
  - Ex: The collateral pledged to a facility cannot have more than 20% of loans issued in a given state, delinquencies cannot exceed 10%

- Borrowers must produce reports to lenders that prove that facilities meet requirements

dv01's data workflow is designed for a front end, not bespoke reporting

SaaS Offerings

Data Pipeline

Back End

Front End

Siphon data from SaaS tech stack

Credit Facility Business Line

Consulting

Reporting

A new business line grows

Credit Facility Analytics

Financial Analysts are co-opted into data engineering

SaaS Offerings

Data Pipeline

Back End

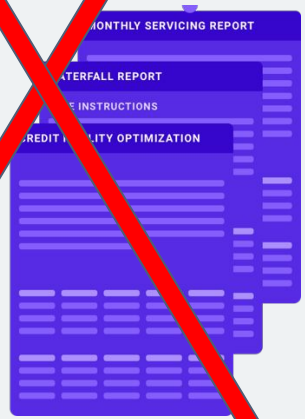
Front End

Siphon data from SaaS tech stack

Credit Facility Business Line

Consulting

Reporting



Financial Analysts are co-opted into data engineering

SaaS Offerings

Data Pipeline

Back End

Front End

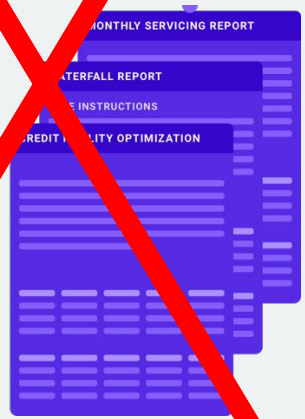
Siphon data from SaaS tech stack

- Detransform
- Pull dropped information from source data
- Rely on spark engineers to execute pipeline

Credit Facility Business Line

Consulting

Reporting



Problem: Spark data pipeline was built for SaaS offering which had conflicting requirements with our new business line

- SaaS Offering
  - **Standardized**: Allows apples-to-apples comparison
  - **Regular schedule**: Set cadence of data updates
  - **Owner:** Engineers

- Reporting Services
  - **Tailored:** Analysis based on idiosyncrasies of each facility
  - **Episodic schedule:** Each facility has unique requirements
  - **Owner:** Analysts

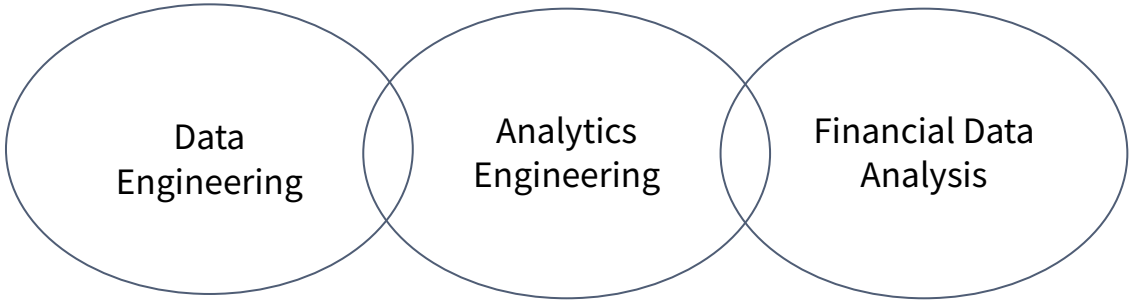
Solution: Build a team heavy on subject matter expertise with enough engineering to be self sufficient

- SQL
  - Language usable by engineers and financial analysts
- dbt
  - flexible orchestration
  - Transparent data pipelines

Credit Facilities Team Composition

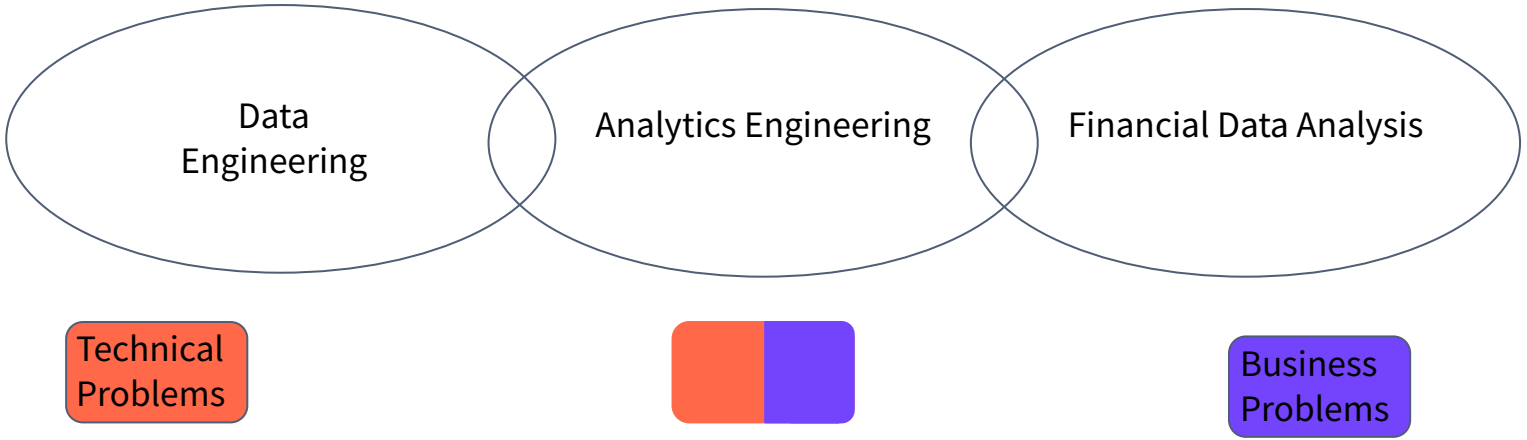
Building a Team Heavy on SME with Enough Engineering

Job Functions



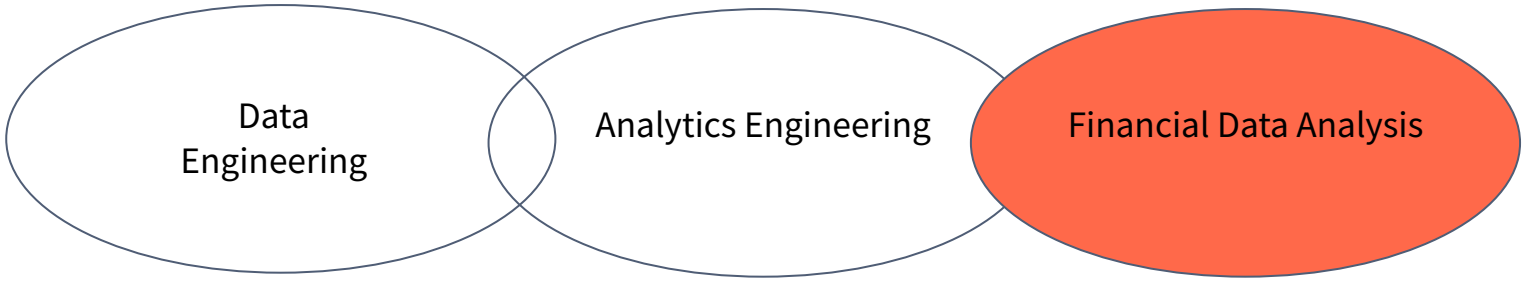
Technologies

Building a Team Heavy on SME with Enough Engineering



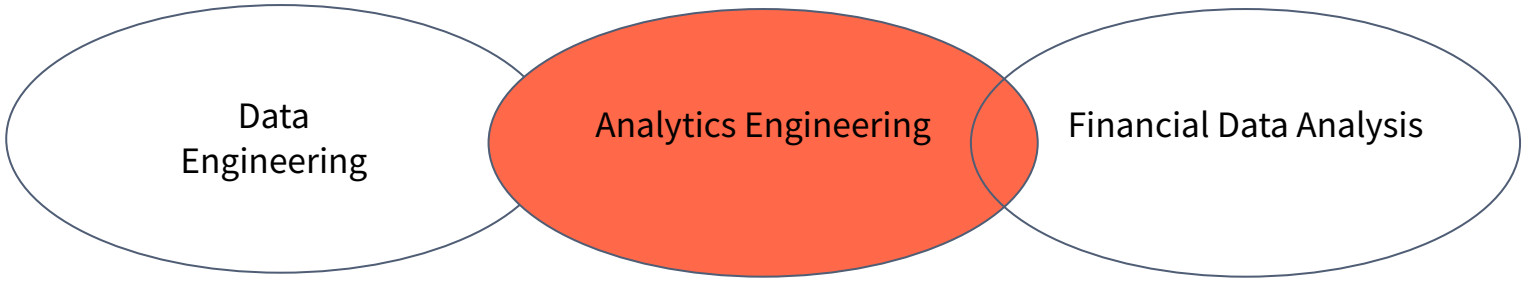
- Distinct roles with some overlap
- Overlap allows a lean team to handle volatile workloads

Building a Team Heavy on SME with Enough Engineering



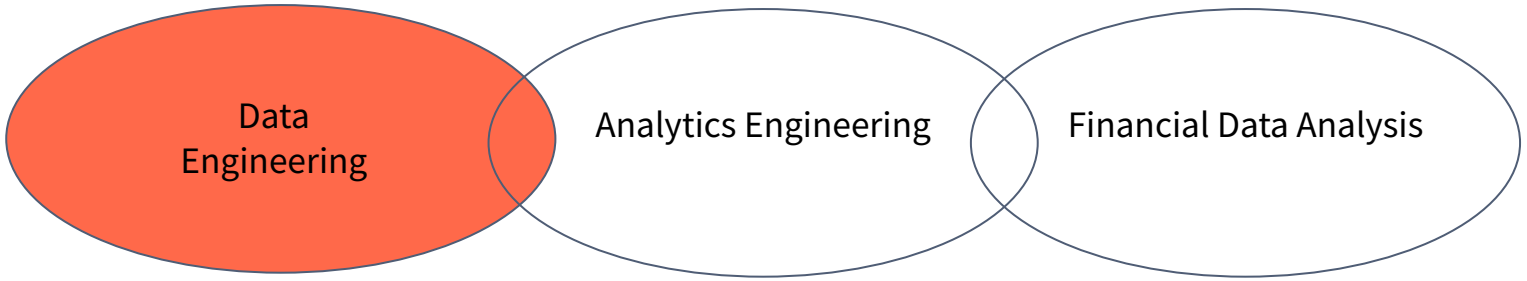
- Focused fully on business problems
- Significant client contact
- Trained to understand dbt codebase and ad hoc SQL queries
  - Do not make major contributions to dbt codebase
- Freed from juggling different roles

Building a Team Heavy on SME with Enough Engineering



- Work end to end **within** dbt pipeline
- Implement business logic and tests in SQL & dbt
- Recruited from data analysts with ambition to grow technically
- Possess significant SME and have some client interaction
- Example technical responsibilities
  - Code review and approval processes (PRs)
  - Manage CI/CD
  - Manage orchestration

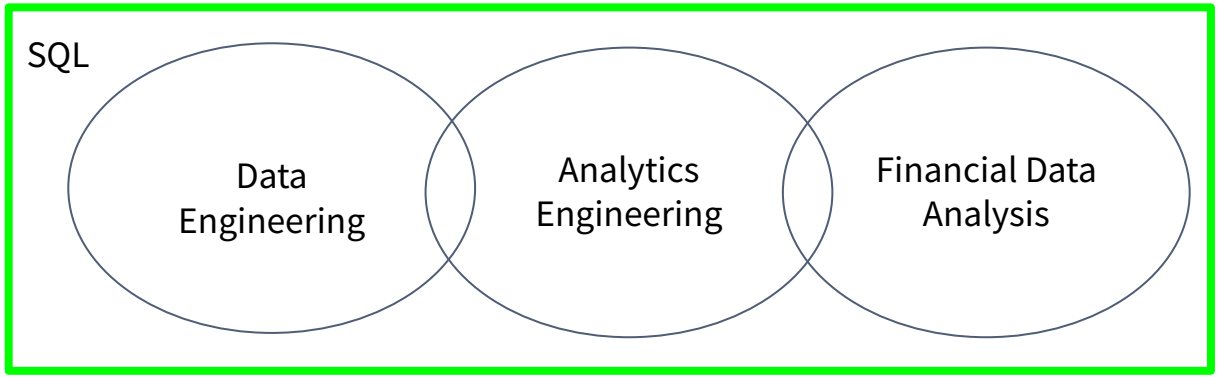
Building a Team Heavy on SME with Enough Engineering



- Some overlap with analytics engineers on day to day
  - Focus engineering heavy problems, allowing analytics engineers to focus on the confluence of business and engineering
- More collaboration with engineers from other parts of the company than analytics engineers
- Mentored on SME by financial analysts
  - Less SME compared to data analysts and engineers but broader technical skills

Why SQL + BigQuery?

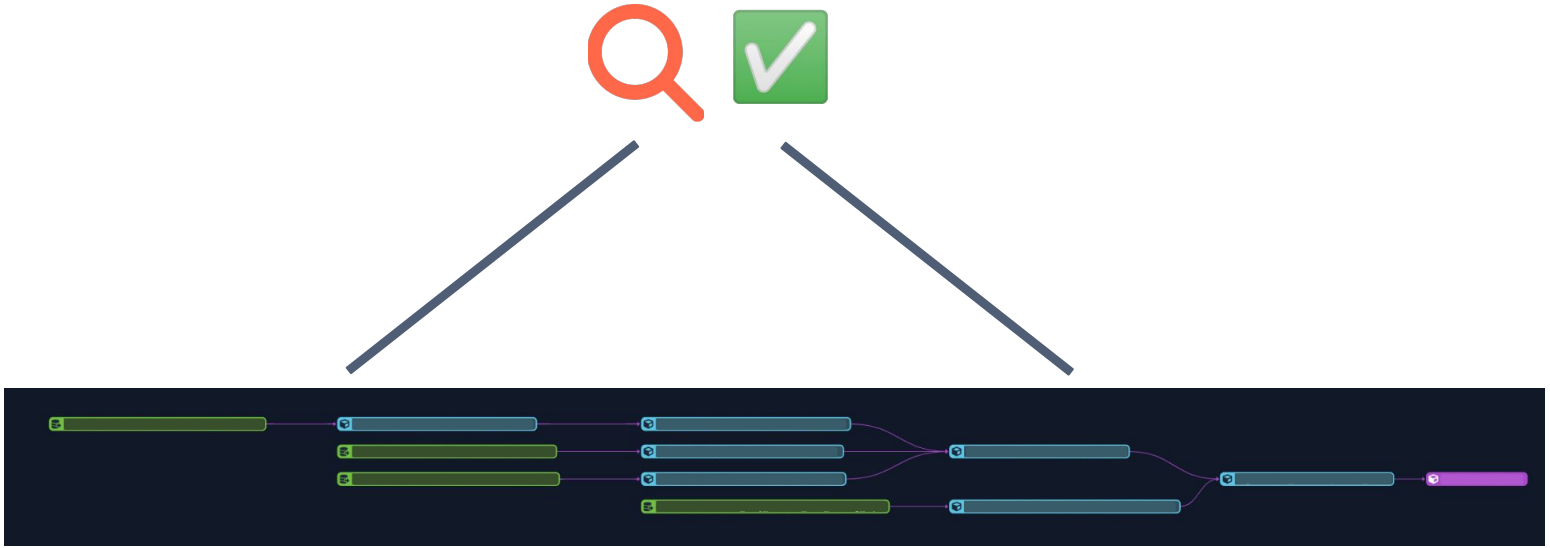
SQL breaks down silos

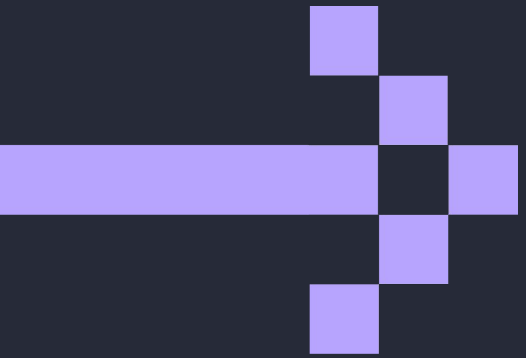


BigQuery consolidates source and final data in a single location

Why dbt?

- dbt makes SQL **organized**, **modular** and **testable**
- dbt models make intermediate states **transparent**
  - It is challenging to inspect intermediate states of lazy evaluated spark pipelines, often requiring a debugger and expensive collect statements





Laying the foundation for a cross functional data team

SQL Training

Goals

- Teach all analysts to interrogate data with SQL
- Provide foundation for analysts who want to go deeper

Migrated to BigQuery

Survey Analyst Workflows

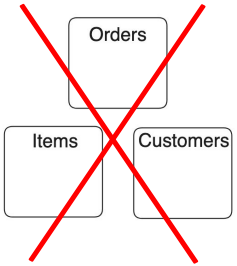
Design Tailored Curriculum

Start Course

SQL Curriculum Guiding Principles

Comparative approach

Practical examples



Cookie-cutter exercises

Growing Analysts to Analytics Engineers

Why	Who	How
Niche domain knowledge Technically sophisticated analysts	Analysts from SQL training with: 1. High technical aptitude 2. Interest	Mentoring Program Analysts <> Experienced Engineers

Keys to successful mentoring program

Incentive Alignment

Recognize mentors in performance reviews

Core Competencies

Performant SQL

Code Review & PR Practices

CICD!!

Personal Learning Plans

Mentors help design

Mentees engage in self study

Pair Programming

I do

I watch you do

You do

Instituting a Culture of Test Driven Development

- Testing starts with analysts
  - Best equipped to define business logic assertions and nuanced client requirements
  - Analysts suggest tests for every client requirement
- Engineers implement tests
  - Implement tests on sources and models
  - Set of standard column level tests & relational tests on dbt DAG

Impact of Testing Culture

Reactive → **Proactive**

Data Quality Report Card

Data Layer

Columns Affected

97% 1% 2%

Column	Test Level (%)	Rule	Values
UUID (PK)	Error (0.1%)	Not NULL	NA, NA, NA
Interest rate	Error (1%)	Positive	-0.05, -0.06
State	Warn (6%)	NULL < 5%	NA, NA, NA

Reporting Layer

Calculations Affected

49% 1% 50%

Calculation	Test Level	Column	Rule
# Delinquent Loans	Error	UUID (PK)	Not NULL
Balance by rate bucket	Error	Interest rate	Positive
# Loans by state	Error	UUID (PK)	Not NULL
# Loans by state	Warn	State	Not NULL

Conclusion

- Business Value = Engineering + Subject Matter Expertise
- SQL **unifies** domain knowledge & engineering
- dbt makes SQL **modular** and **testable**

Slides available at

Thank you