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

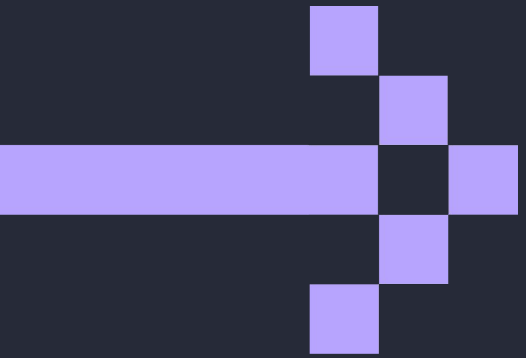
Data Engineer

dv01

a Fitch Solutions Company

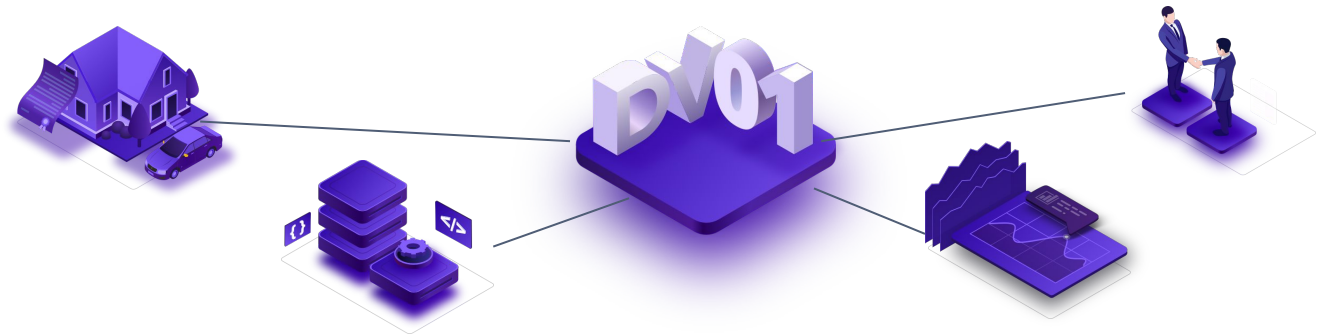
How to build cross functional data teams

Business Value = Engineering + Subject Matter Expertise



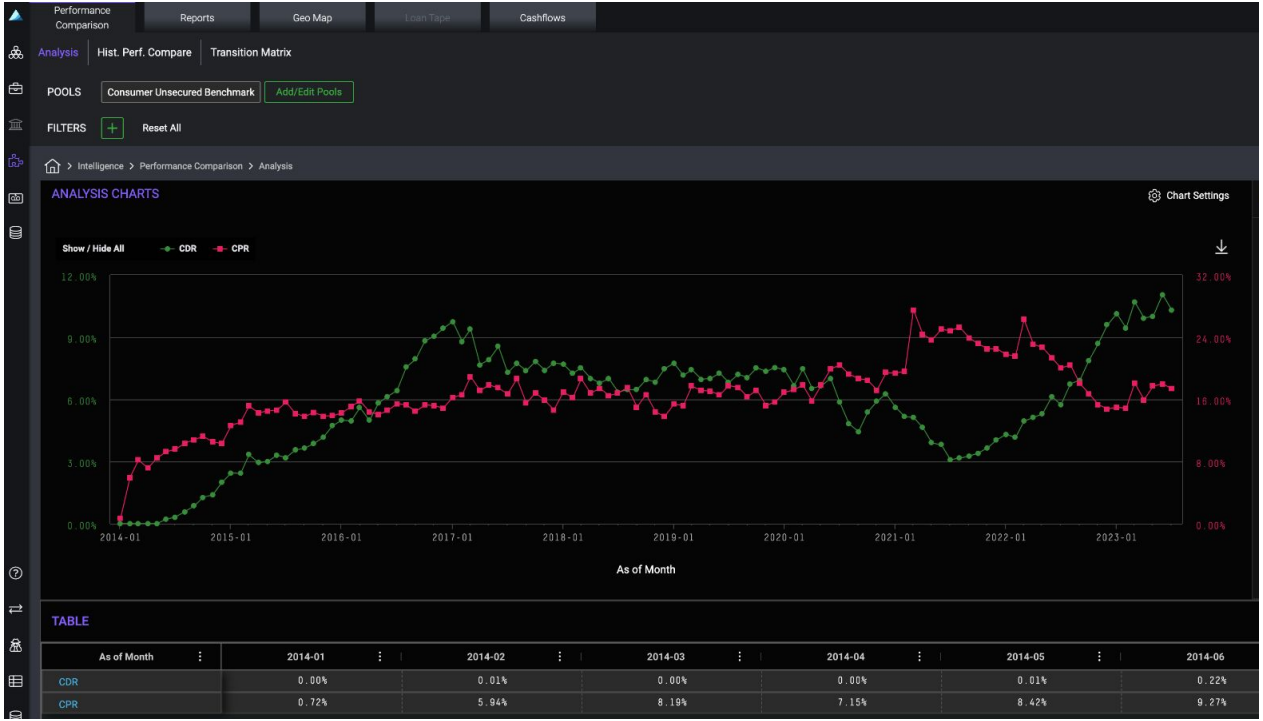
Background

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”).

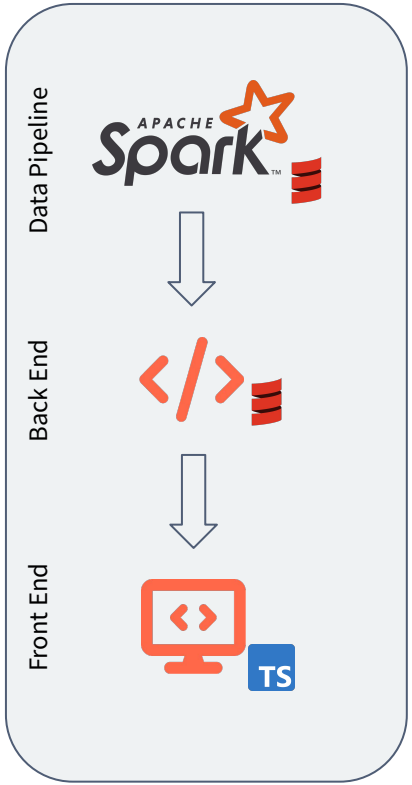
The dv01 Web App

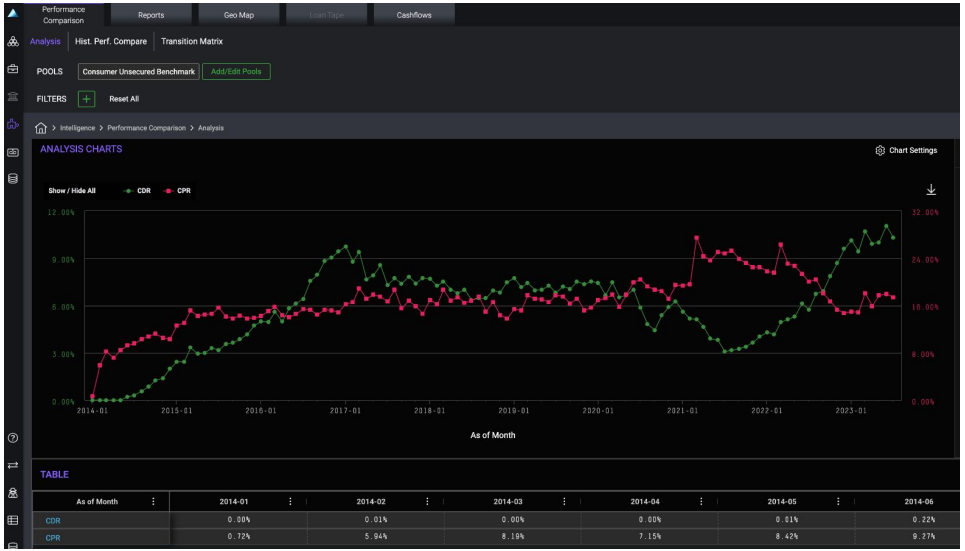


- dv01 standardizes loan data **and** offers users embedded analytics for analysis
- Data can also be accessed via SFTP files and BigQuery

Our data pipeline is designed for a highly engineered tech stack

SaaS Offerings

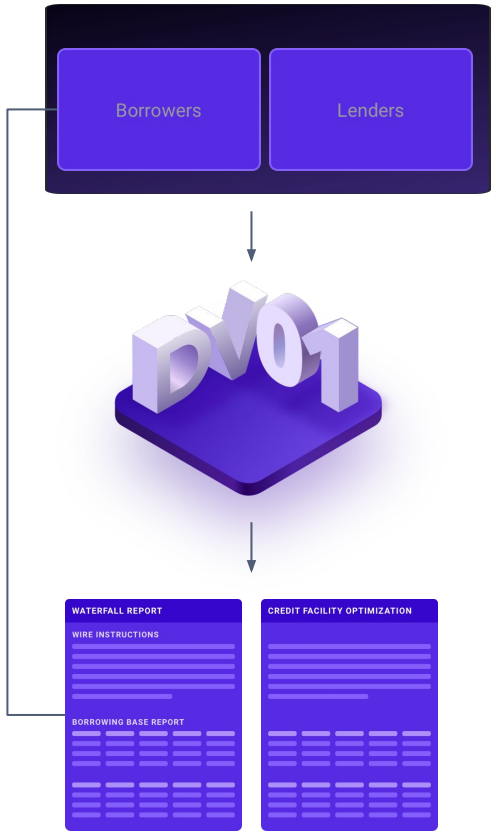




A new business line sprouts

Credit Facility Analytics

A credit facility is revolving credit for loan issuers offered by a bank



- A loan issuer (“borrower”) needs money to issue more loans to consumers. The borrower goes to a bank (“lender”) and takes out a credit facility, pledging the collateral. The lender will generate cash flows **and** accumulate loans that will later be bundled into ABS and MBS.
- 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

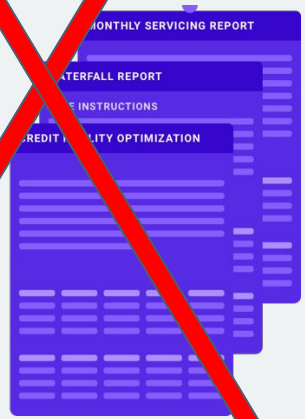
Front End

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

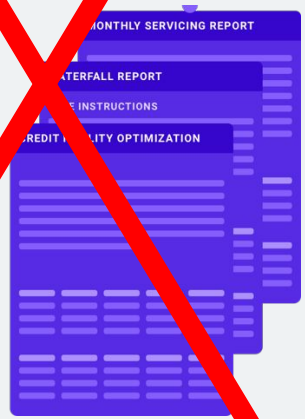
Siphon data from SaaS tech stack

- 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

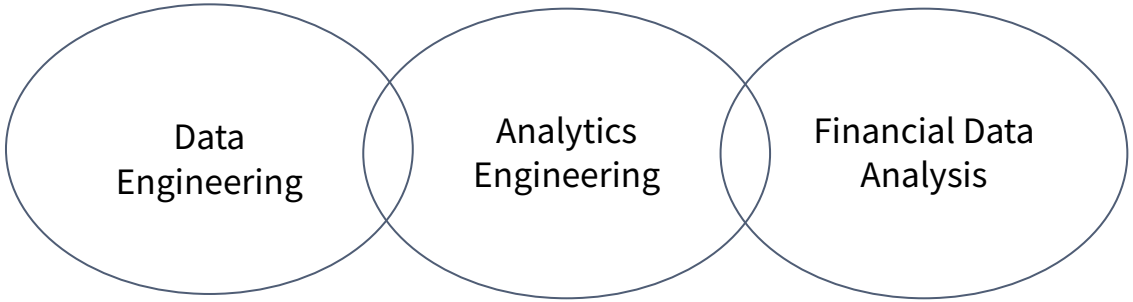
Solution: Build a team heavy on subject matter expertise with enough engineering to be self sufficient

- 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

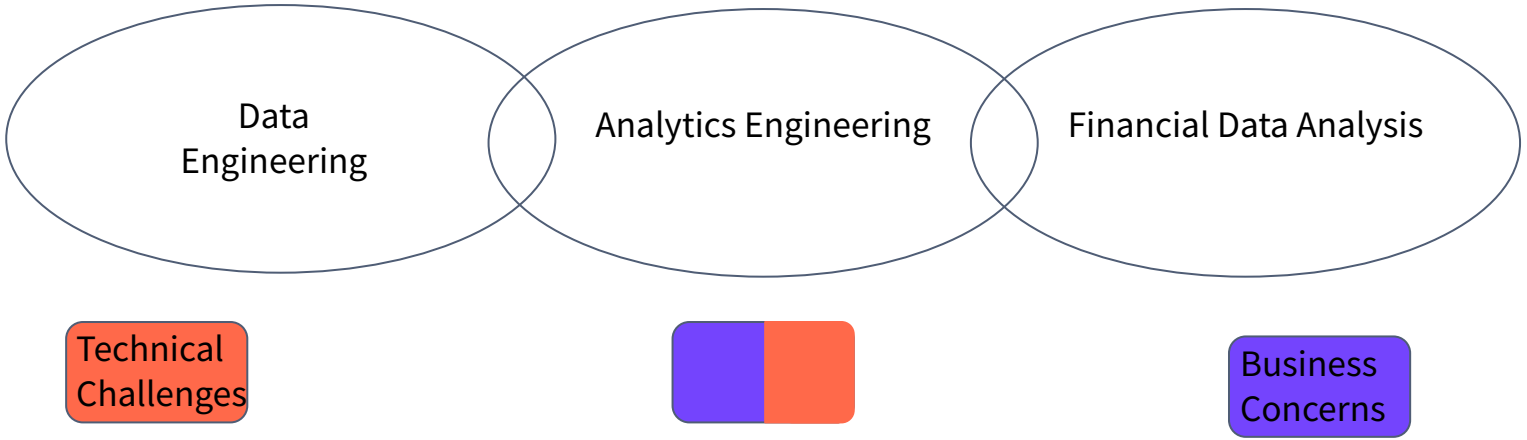
Job Functions



Technologies

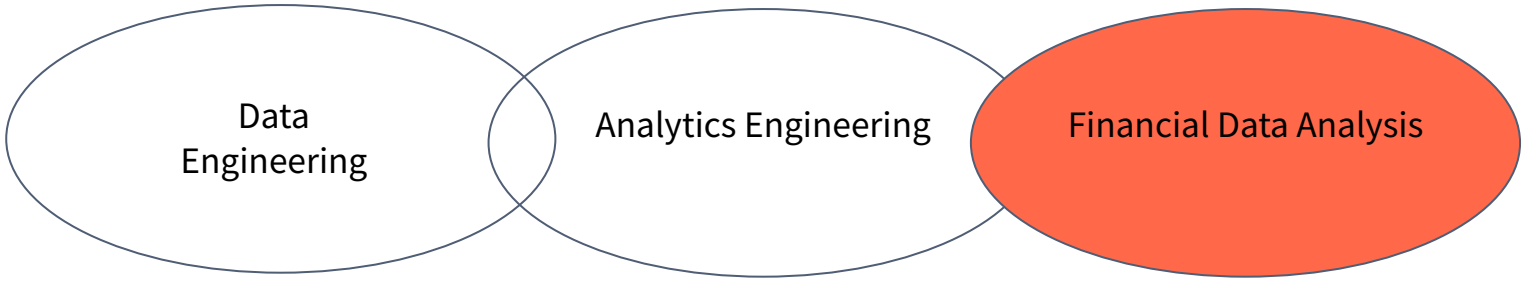
GCS Logo

Building a Team Heavy on SME with Enough Engineering



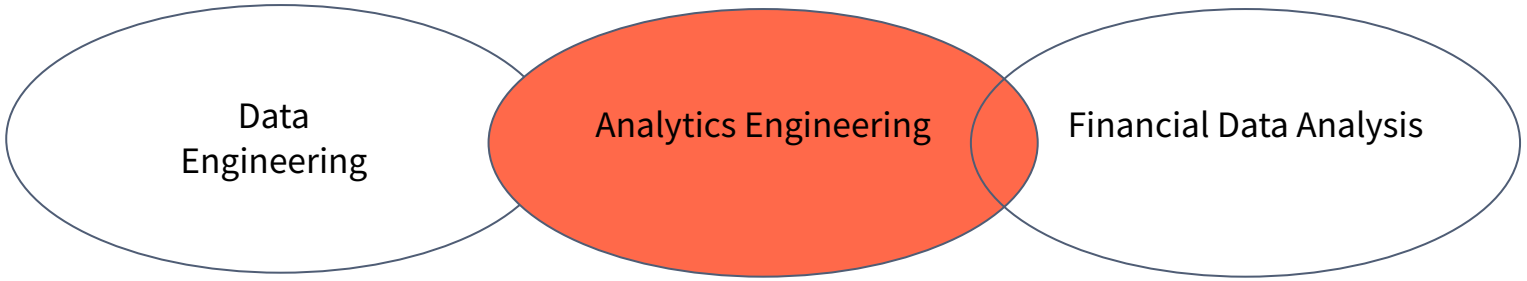
- Distinct roles with some overlap
- Overlap allows a lean team to handle volatile workloads

Building a Team Heavy on SME with Enough Engineering



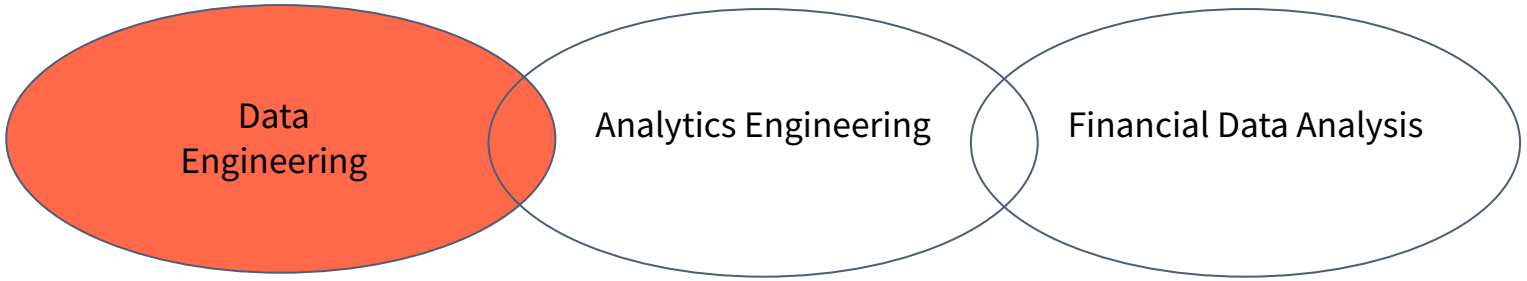
- Focused fully on business problems
- Significant client contact
- Trained to understand dbt codebase and ad hoc SQL queries
  - Do not make major contributions to dbt codebase
- Freed from juggling different roles

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

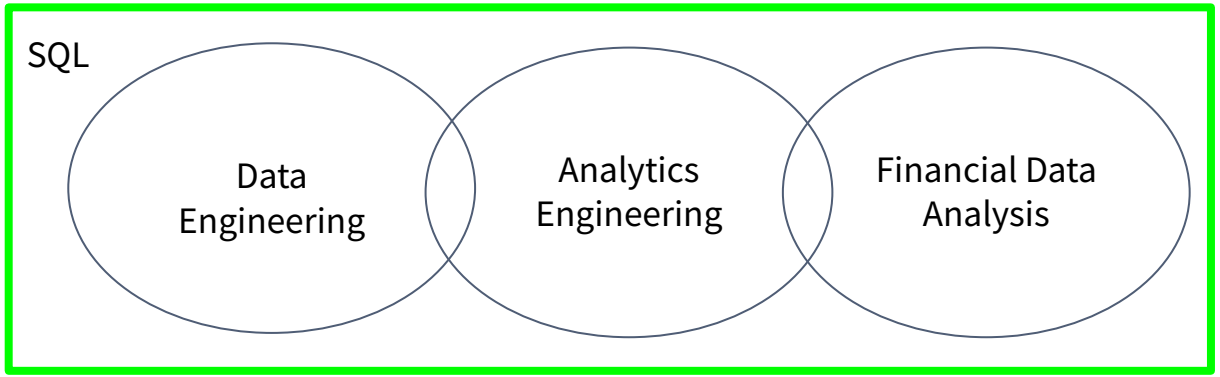
Building a Team Heavy on SME with Enough Engineering



- 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?

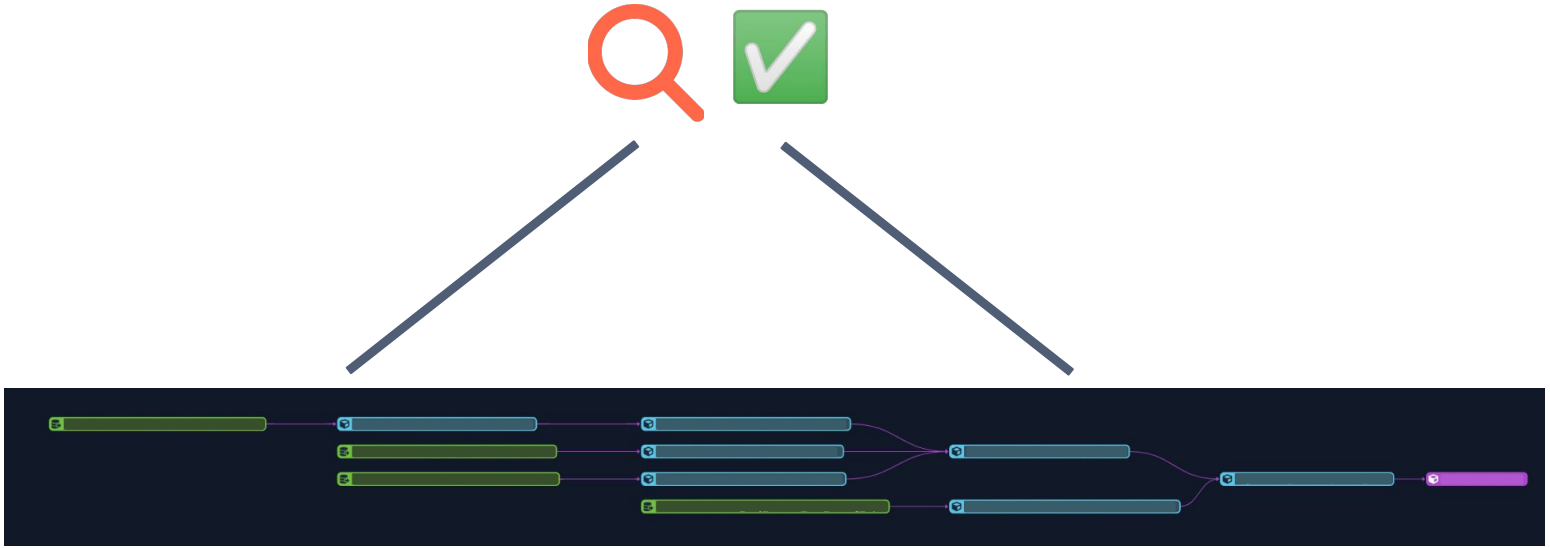
SQL breaks down silos

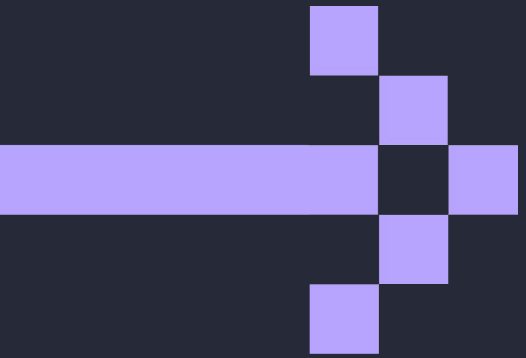


BigQuery consolidates source and final data in a single location

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

- Data Analysts were already using R & spark
- Following of BigQuery data lake, we ran a SQL training class for all analysts in the company, including the credit facility analysts
- **Goals**
  - Give every analyst the ability to interrogate data with SQL on an ad hoc basis
  - Provide a foundation for analysts who want to go deeper
- Tailored curriculum to our existing tool box and industry
  - Comparative approach - driven by use cases currently accomplished in R/Excel
  - Prior to designing curriculum, we interviewed analysts from different teams about their work flows
  - All examples were concrete problems our analysts solve every day. We did not use any of the typical sales analysis examples, which are not relevant to our company

SQL Training

Migrated to BigQuery

Lorem ipsum dolor sit amet, consectetur adipiscing.

Survey Analyst Workflows

Lorem ipsum dolor sit amet, consectetur adipiscing.

Design Tailored Curriculum

Lorem ipsum dolor sit amet, consectetur adipiscing.

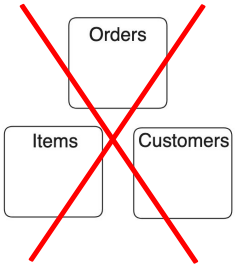
Start Course

Lorem ipsum dolor sit amet, consectetur adipiscing.

SQL Curriculum Guiding Principles

Comparative approach

Practical examples



Cookie-cutter exercises

Growing Analysts to Analytics Engineers

- Why grow analysts into analytics engineers
  - Credit facilities is an idiosyncratic niche of finance that requires deep SME
  - Need engineers with deep understanding of the business
- Curious students from the SQL training & Credit Facility analysts already tackling the harder data problems
- Paired budding analytics engineers with data engineers
  - Shadowing
  - Pair programming
  - I do, I watch you do, you do
- Must align incentives
  - Mentoring takes time!
  - Must be recognized in performance evaluations or mentors will be disadvantaged relative to non-mentors

Focuses of mentoring relationships

- Mentors help mentees create a learning plan for self study
- Key areas of learning
  - Writing performant SQL
  - Code review & PR process
  - Devops - CICD!!

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
- Alerting allows us to catch issues early and proactively
- We created a data quality report card, which analysts inspect and include in client deliverables

Instituting a Culture of Test Driven Development

Prior to dbt

Strategy
Analysts produce reports with client data and investigate any issues in aggregate reported values

Pain points
- Diagnosing issues in final aggregated values is tedious and time consuming
- Not all data issues can be easily caught from aggregate values
- Dependent on central data platform team to fix problems
- Opaque intermediate steps in spark scala pipeline

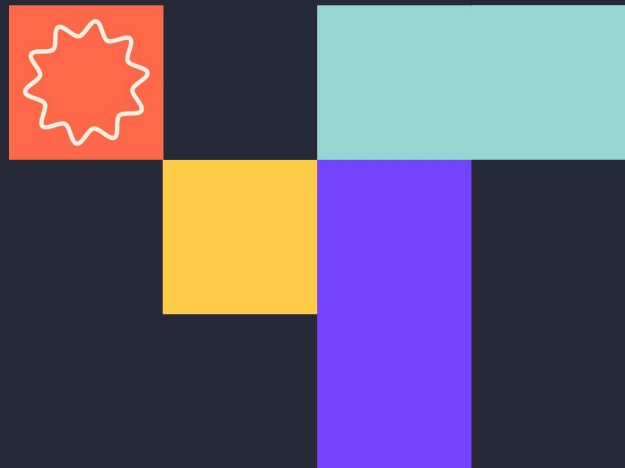
With dbt

Strategy
Testing suite is executed before any work or reporting is done on client data

Benefits
- Issues are identified and isolated early at row level
- Sources and intermediate models in data pipeline are easily interrogated with SQL
- Team has the autonomy to fix problems

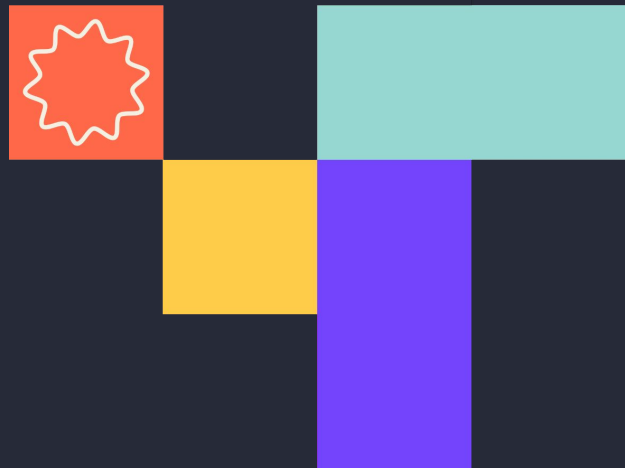
Conclusion

Thank you



Icon Library

A collection of frequently used icons.



Reference Slide: Coalesce Pixel Icons

Be mindful not to mix FontAwesome icons with Coalesce pixel icons. Use on light backgrounds.

Asterisk

At

Brackets

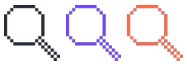
Heart

Hi

Nerd

Prompt

Search



User

Online

San Diego

London

Sydney

Berlin

Reference Slide: FontAwesome Navy Icon Library

New icons not included can be requested from #design-team. Use on light backgrounds.

Reference Slide: FontAwesome Orange Icon Library

New icons not included can be requested from #design-team. Use on dark or light backgrounds.

Reference Slide: FontAwesome White Icon Library

New icons not included can be requested from #design-team. Use on dark backgrounds.