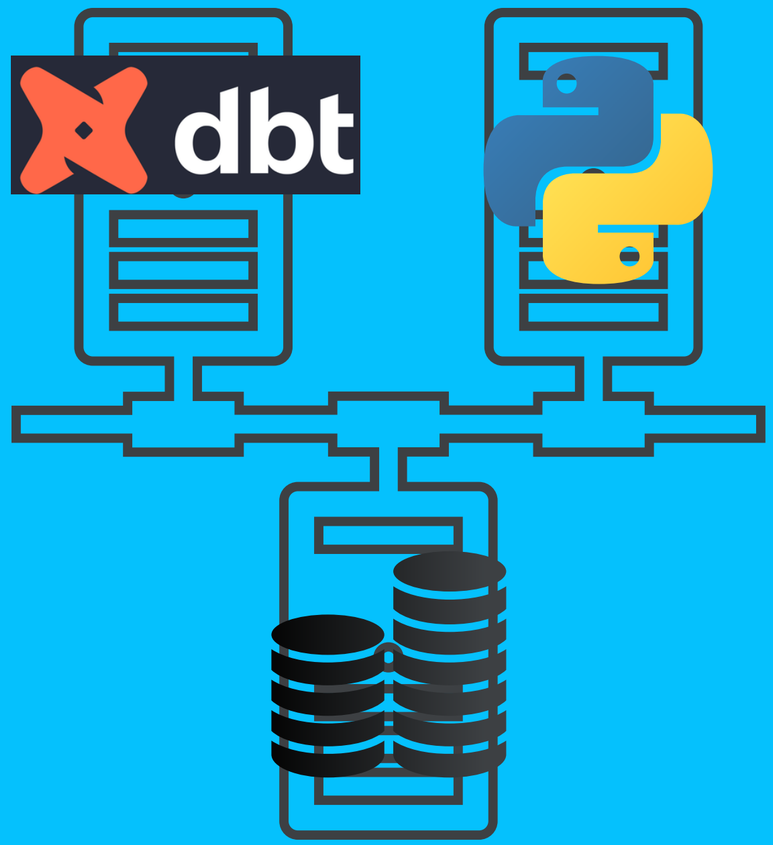
TRANSFORMING YOUR DATA WORKFLOW WITH DBT



From Raw to BI Ready:using DBT

https://github.com/insightbuilder

WHAT PROBLEM WE'RE FACING

TRACKING DATA MODEL ACROSS DEV

As the raw data enters the data pipeline, data lakes or data warehouse there are many members accessing the data. Tracking the data schema across multiple schema is not available

SERIES OF PROBLEM

- ETL PIPELINE DOCUMENTATION
- UNABLE TO IMPLEMENT CI/CD PIPELINE
- SCHEMA VERSION CONTROL NOT POSSIBLE
- CREATING MULTIPLE TABLES IN DATABASE

PYTHON ECOSYSTEM : DBT

HOW TO GET IT INTO MY OS?

- Install Python
- The do "pip install dbt-core"

WHERE CAN I LEARN ABOUT STREAMLIT

- DBT Site

https://docs.getdbt.com/docs/introduction

- SQL Basics

https://sqlzoo.net/

GUEST STARRER

***

HTTPS://GITHUB.COM/INSIGHTBUILDER

WHERE TO START

01

START WITH RAW DATA

- Connect DBT core by setting up the yaml files
- Init the project and check the connection
- Write your first schema model file and run

02

USING SQL TO CHECK

- DBT model executes the query and writes it to Database only.
- Need to manually check if the queries executed

HTTPS://GITHUB.COM/INSIGHTBUILDER

DBT MODELS : OUR HERO

HOW DOES DBT HELP

01 SIMPLE WAY TO CONNECT

- Profile.yml file in your OS keeps track of the database your are connected with
- dbt provides tools like init and debug to check connection

02 VERSION CONTROL

- The version control support is provided by GIT
- All changes are pushed to GIT

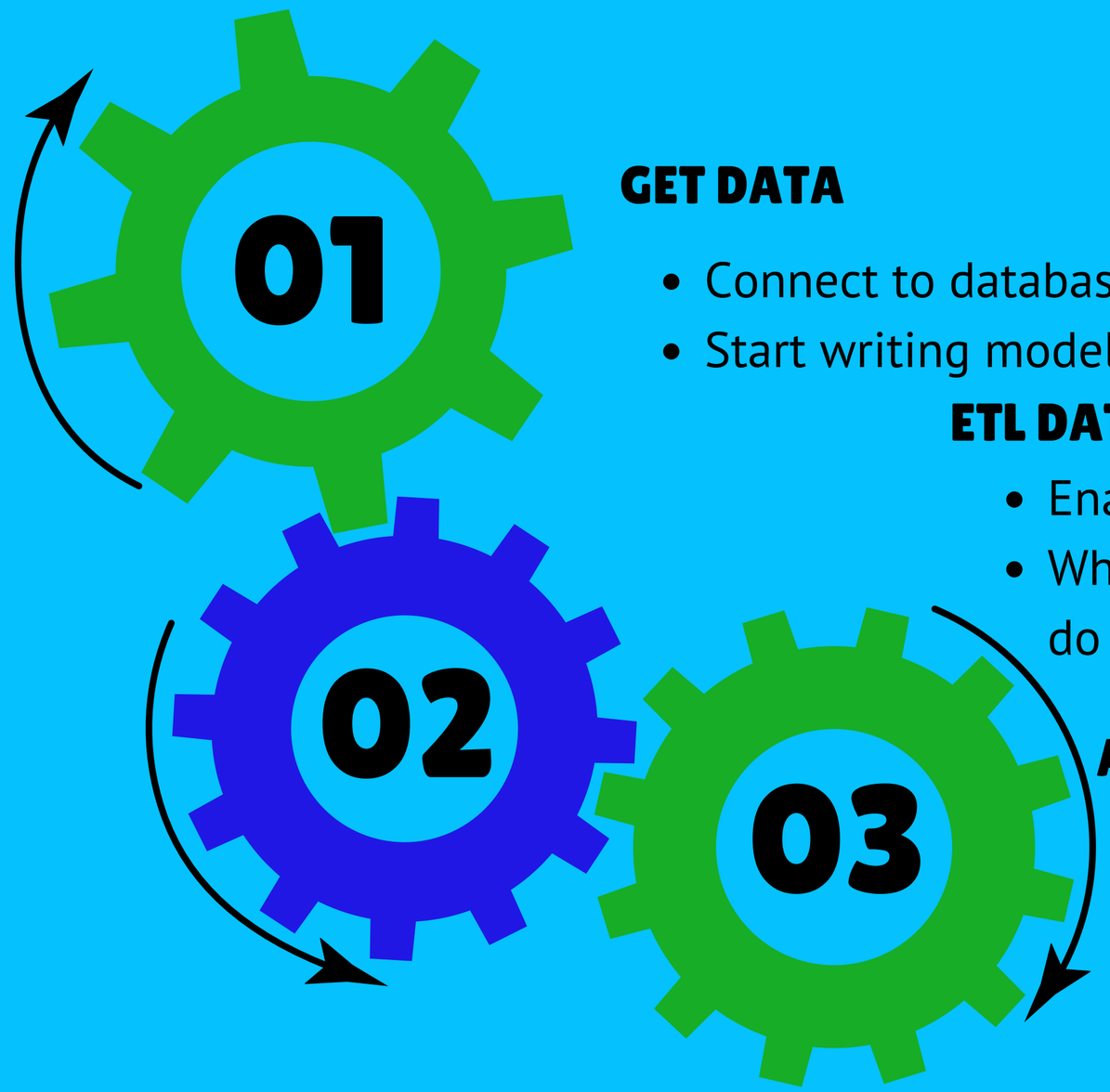
03 EXTENSIVE TRANSFORMATIONS

- All the SQL commands can be executed depending on the database you are connecting to.

04 ADVANCED SCHEMA MODELING

- The models created can be seen like python code, so easy to understand
- This leads to efficient modeling methods and use of staging and marts

DBT CORE PROCESS



GET DATA

- Connect to database
- Start writing models

ETL DATA

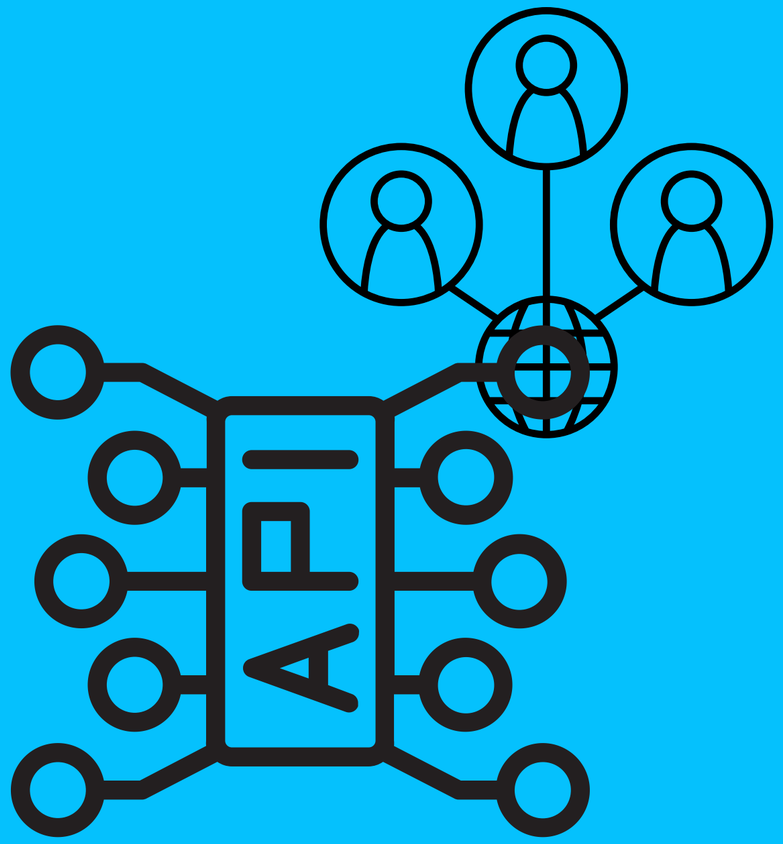
- Enable the models as gate keepers
- When new data arrives, the models do the transformaiton

AUTOMATE THE PROCESS

- dbt documents the models automatically
- The document can be accessed by the team
- Document is shareable

QUESTIONS AND COMMENTS

WHAT NEXT ???



PYTHON DATA APIS

STARRING MULTIPLE LIBRARIES