



Data Engineering

DBT

ALL CONCEPTS TO GET STARTED



Shwetank Singh
GritSetGrow - GSGLearn.com

DBT (DATA BUILD TOOL)

A COMMAND-LINE TOOL THAT
ENABLES DATA ANALYSTS AND
ENGINEERS TO TRANSFORM DATA IN
THEIR WAREHOUSE MORE
EFFECTIVELY.

TRANSFORMING RAW DATA INTO
CLEANED, MODELED TABLES USING
SQL.



Shwetank Singh
GritSetGrow - GSGLearn.com



MODELS

**SQL FILES THAT CONTAIN
TRANSFORMATION LOGIC. MODELS
ARE SELECT STATEMENTS IN DBT.**

**MODELS/MY_MODEL.SQL WITH
SELECT * FROM RAW_DATA WHERE
CONDITION IS NOT NULL;**



Shwetank Singh
GritSetGrow - GSGLearn.com

dbt™

A large, stylized orange word "dbt" with a trademark symbol. To the right of the "t" is a thick, black, hand-drawn style arrow pointing to the right, with a green outline.

DBT PROJECT

A COLLECTION OF MODELS, TESTS,
DOCUMENTATION, AND MORE,
ORGANIZED IN A DIRECTORY
STRUCTURE.

PROJECT DIRECTORY WITH
SUBDIRECTORIES FOR MODELS,
MACROS, ETC.



Shwetank Singh
GritSetGrow - GSGLearn.com

dbt™

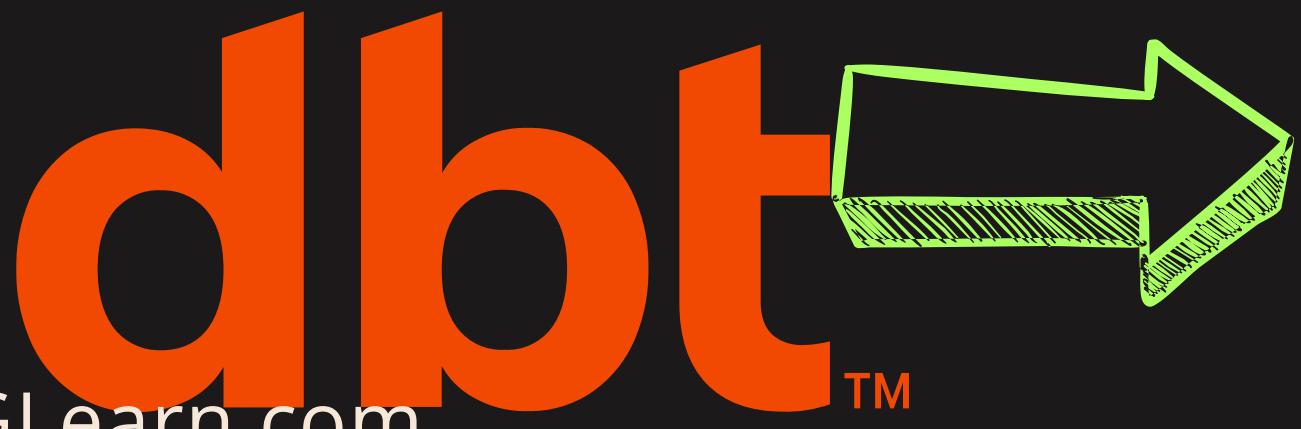
SOURCES

REPRESENT RAW DATA IN YOUR WAREHOUSE. DEFINED IN YAML FILES AND REFERENCED IN MODELS.

SOURCES.YML DEFINING A SOURCE TABLE AND USING IT IN A MODEL WITH {{ SOURCE('MY_SOURCE', 'TABLE') }}



Shwetank Singh
GritSetGrow - GSGLearn.com



SEEDS

**CSV FILES THAT ARE LOADED INTO
YOUR WAREHOUSE AS TABLES.**

**PLACING MY_DATA.CSV IN THE DATA
DIRECTORY AND RUNNING DBT SEED
TO LOAD IT INTO THE WAREHOUSE.**



Shwetank Singh
GritSetGrow - GSGLearn.com



TESTS

USED TO ENSURE DATA QUALITY.
THERE ARE BUILT-IN TESTS AND
CUSTOM TESTS.

ADDING A UNIQUENESS TEST IN
SCHEMA.YML: - NAME: MY_COLUMN
TESTS: - UNIQUE



Shwetank Singh
GritSetGrow - GSGLearn.com



DOCUMENTATION

DBT ALLOWS FOR IN-DEPTH DOCUMENTATION OF MODELS, SOURCES, AND MORE.

ADDING DESCRIPTIONS IN SCHEMA.YML AND GENERATING DOCUMENTATION WITH DBT DOCS GENERATE.



Shwetank Singh
GritSetGrow - GSGLearn.com

dbt™

A large, stylized orange word "dbt" with a trademark symbol. To the right of the "t" is a thick, black, hand-drawn style arrow pointing to the right. The arrow has a textured, hatched base and a thin, light green outline.

SNAPSHOTS

USED TO CAPTURE CHANGES IN
SLOWLY CHANGING DIMENSIONS.

DEFINING A SNAPSHOT IN
SNAPSHOTS/ DIRECTORY TO TRACK
CHANGES OVER TIME.



Shwetank Singh
GritSetGrow - GSGLearn.com



MACROS

SQL FUNCTIONS AND REUSABLE LOGIC THAT CAN BE USED IN MODELS AND TESTS.

CREATING A MACRO IN MACROS/DIRECTORY AND USING IT IN A MODEL WITH {{ MY_MACRO() }}



Shwetank Singh
GritSetGrow - GSGLearn.com



REF FUNCTION

A FUNCTION TO REFERENCE OTHER MODELS, ENSURING DEPENDENCIES ARE RESPECTED.

```
SELECT * FROM {{  
REF('MY_OTHER_MODEL') }}
```



Shwetank Singh
GritSetGrow - GSGLearn.com



CONFIGURING MODELS

MODELS CAN BE CONFIGURED WITH SPECIFIC ATTRIBUTES LIKE MATERIALIZATIONS, TAGS, ETC.

SETTING A MODEL TO BE MATERIALIZED AS A TABLE: {{
 CONFIG(MATERIALIZED='TABLE') }}



Shwetank Singh
GritSetGrow - GSGLearn.com



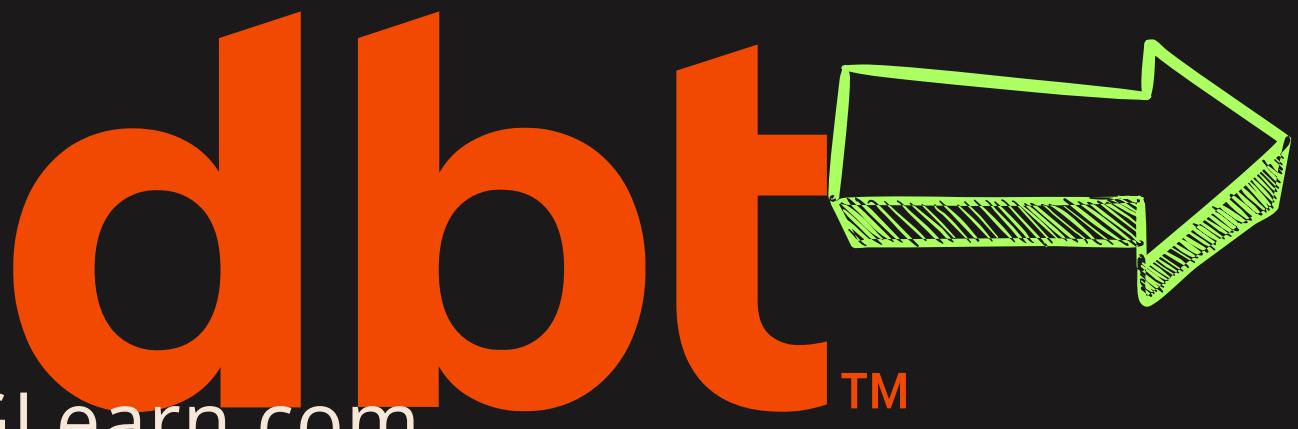
MATERIALIZATIONS

DEFINE HOW MODELS ARE BUILT IN THE DATABASE (TABLE, VIEW, INCREMENTAL, Ephemeral).

USING `{}
CONFIG(MATERIALIZED='INCREMENTAL')` TO CREATE AN INCREMENTAL MODEL.



Shwetank Singh
GritSetGrow - GSGLearn.com



INCREMENTAL MODELS

MODELS THAT ONLY PROCESS NEW OR CHANGED DATA SINCE THE LAST RUN.

DEFINING AN INCREMENTAL MODEL WITH A `UNIQUE_KEY` AND USING `IS_INCREMENTAL()` IN THE SQL.



Shwetank Singh
GritSetGrow - GSGLearn.com



EPHEMERAL MODELS

MODELS THAT ARE NEVER CREATED IN THE DATABASE BUT ARE USED IN SUBSEQUENT MODELS.

USING `{
 CONFIG(MATERIALIZED='EPHEMERAL'
)} }` TO CREATE A CTE-LIKE MODEL.



Shwetank Singh
GritSetGrow - GSGLearn.com



JINJA

A TEMPLATING LANGUAGE USED IN
DBT TO WRITE DYNAMIC SQL.

USING JINJA TO CREATE A DYNAMIC
FILTER: WHERE DATE = '{{
EXECUTION_DATE }}'



Shwetank Singh
GritSetGrow - GSGLearn.com



DBT RUN

A COMMAND TO EXECUTE ALL MODELS IN A DBT PROJECT.

RUNNING ALL MODELS WITH DBT RUN.



Shwetank Singh
GritSetGrow - GSGLearn.com

dbt™

DBT TEST

A COMMAND TO RUN ALL TESTS
DEFINED IN THE PROJECT.

EXECUTING DATA QUALITY CHECKS
WITH DBT TEST.



Shwetank Singh
GritSetGrow - GSGLearn.com

dbt™

DBT SEED

A COMMAND TO LOAD SEED FILES
INTO THE DATABASE.

LOADING CSV FILES WITH DBT SEED.



Shwetank Singh
GritSetGrow - GSGLearn.com

dbt™

DBT SNAPSHOT

A COMMAND TO EXECUTE ALL
SNAPSHOT FILES.

CAPTURING CHANGES WITH DBT
SNAPSHOT.



Shwetank Singh
GritSetGrow - GSGLearn.com

dbt™

DBT DOCS

A COMMAND TO GENERATE AND SERVE DOCUMENTATION FOR YOUR PROJECT.

GENERATING DOCS WITH DBT DOCS
GENERATE AND VIEWING THEM WITH DBT DOCS SERVE.



Shwetank Singh
GritSetGrow - GSGLearn.com

dbt™

PROFILES.YML

CONFIGURATION FILE FOR SETTING
UP CONNECTION DETAILS TO THE
DATA WAREHOUSE.

CONFIGURING DATABASE
CONNECTION IN PROFILES.YML FILE.



Shwetank Singh
GritSetGrow - GSGLearn.com

dbt™

SCHEMA.YML

CONFIGURATION FILE FOR DEFINING TESTS, DOCUMENTATION, AND SOURCES.

DEFINING A SOURCE TABLE AND ADDING TESTS IN SCHEMA.YML.



Shwetank Singh
GritSetGrow - GSGLearn.com

dbt™

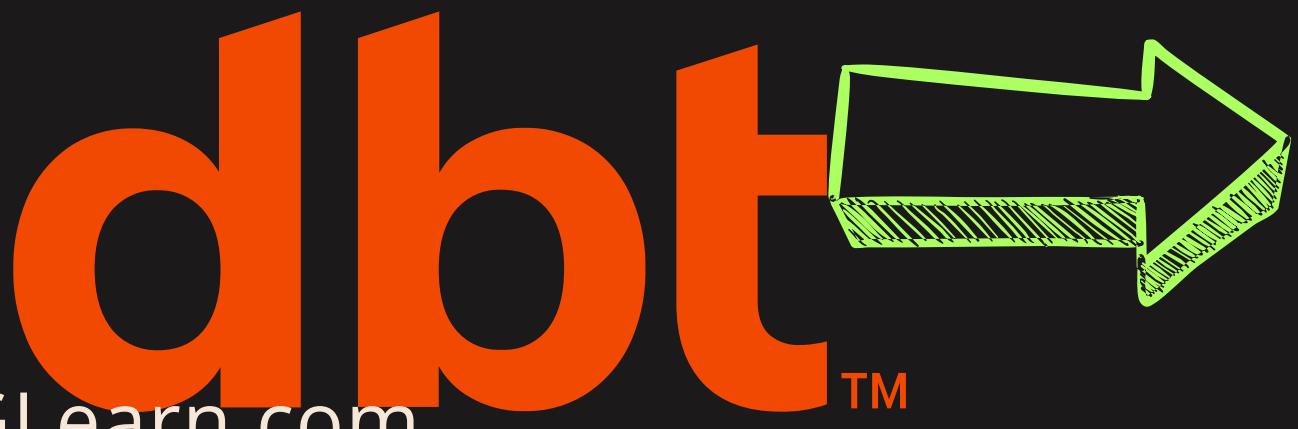
HOOKS

SQL STATEMENTS OR MACROS THAT RUN AT DIFFERENT POINTS DURING A DBT RUN.

CREATING PRE-MODEL AND POST-MODEL HOOKS FOR LOGGING.



Shwetank Singh
GritSetGrow - GSGLearn.com



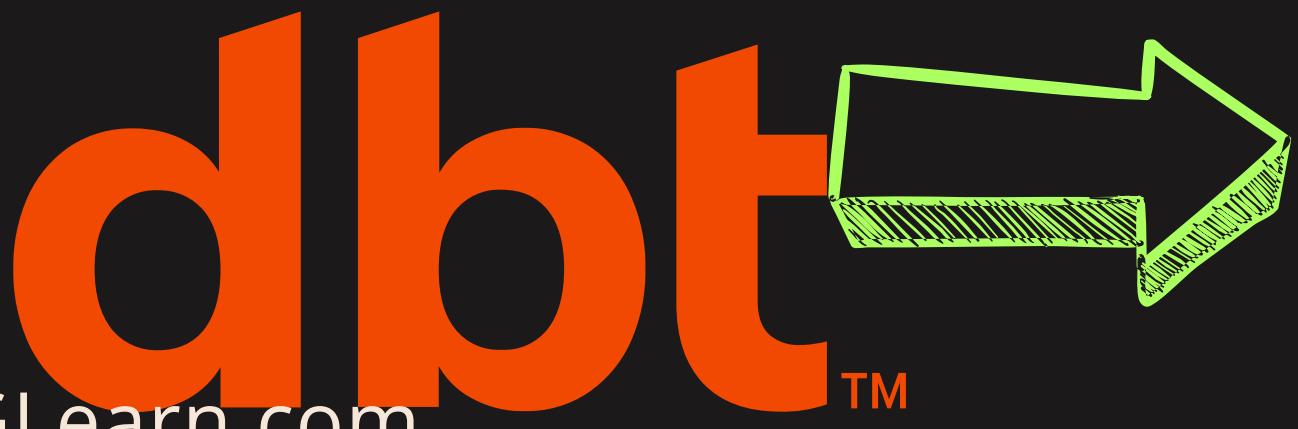
VARIABLES

PLACEHOLDER VALUES THAT CAN BE DEFINED IN THE DBT_PROJECT.YML FILE AND USED IN MODELS AND MACROS.

DEFINING A VARIABLE IN DBT_PROJECT.YML AND USING IT IN A MODEL WITH {{ VAR('MY_VARIABLE') }}



Shwetank Singh
GritSetGrow - GSGLearn.com



ENVIRONMENTS

DIFFERENT CONFIGURATIONS FOR DIFFERENT ENVIRONMENTS LIKE DEVELOPMENT, STAGING, AND PRODUCTION.

USING TARGET AND PROFILES.YML TO CONFIGURE DIFFERENT ENVIRONMENTS.



Shwetank Singh
GritSetGrow - GSGLearn.com



PACKAGES

THIRD-PARTY OR SHARED DBT
PROJECTS THAT CAN BE INCLUDED IN
YOUR PROJECT.

ADDING A PACKAGE TO
PACKAGES.YML AND RUNNING DBT
DEPS TO INSTALL IT.



Shwetank Singh
GritSetGrow - GSGLearn.com



DBT DEPS

A COMMAND TO INSTALL THE
DEPENDENCIES LISTED IN THE
PACKAGES.YML FILE.

RUNNING DBT DEPS TO INSTALL
PACKAGES.



Shwetank Singh
GritSetGrow - GSGLearn.com



CUSTOM MATERIALIZATIONS

CREATING CUSTOM MATERIALIZATIONS FOR SPECIFIC NEEDS.

DEFINING A CUSTOM MATERIALIZATION IN THE MACROS/DIRECTORY.



Shwetank Singh
GritSetGrow - GSGLearn.com



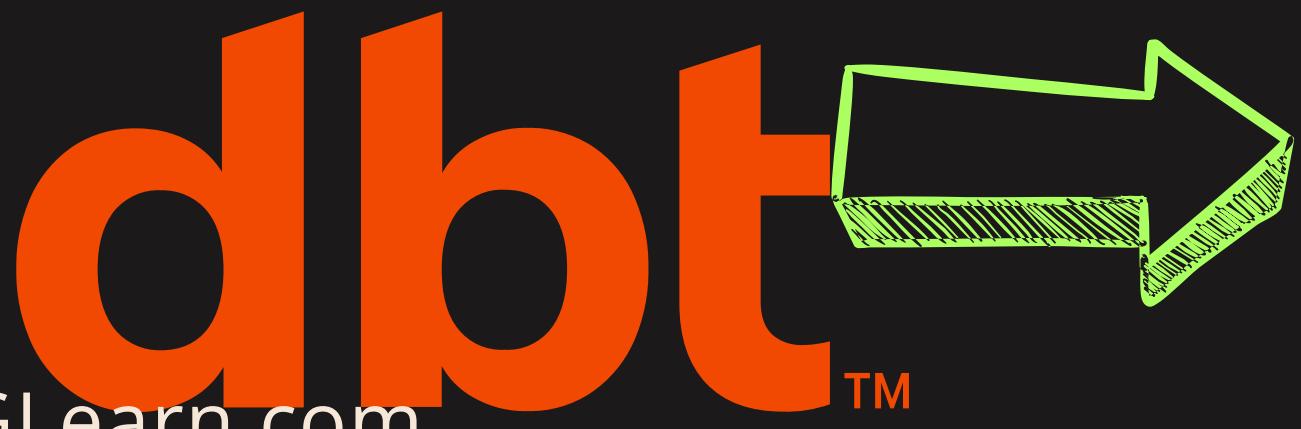
INCREMENTAL STRATEGIES

DIFFERENT STRATEGIES FOR
HANDLING INCREMENTAL UPDATES
LIKE INSERT_OVERWRITE OR MERGE.

USING INSERT_OVERWRITE STRATEGY
IN AN INCREMENTAL MODEL.



Shwetank Singh
GritSetGrow - GSGLearn.com



MODEL CONFIGURATION

SETTING CONFIGURATIONS AT THE
MODEL LEVEL, SUCH AS ALIAS,
SCHEMA, AND TAGS.

CONFIGURING A MODEL WITH {{
 CONFIG(ALIAS='NEW_NAME',
 SCHEMA='NEW_SCHEMA', TAGS=
 ['TAG1', 'TAG2']) }}



Shwetank Singh
GritSetGrow - GSGLearn.com



SNAPSHOT CONFIGURATION

CONFIGURING SNAPSHOTS WITH
STRATEGIES LIKE CHECK AND
TIMESTAMP.

USING CHECK STRATEGY IN A
SNAPSHOT DEFINITION.



Shwetank Singh
GritSetGrow - GSGLearn.com



PROJECT STRUCTURE

ORGANIZING THE DBT PROJECT WITH APPROPRIATE DIRECTORIES AND FILES FOR MODELS, SNAPSHOTS, SEEDS, AND MACROS.

CREATING SUBDIRECTORIES UNDER MODELS/ FOR BETTER ORGANIZATION.



Shwetank Singh
GritSetGrow - GSGLearn.com



DATA WAREHOUSE SUPPORT

DBT SUPPORTS MULTIPLE DATA WAREHOUSES LIKE SNOWFLAKE, BIGQUERY, REDSHIFT, AND POSTGRES.

CONFIGURING A SNOWFLAKE CONNECTION IN PROFILES.YML.



Shwetank Singh
GritSetGrow - GSGLearn.com



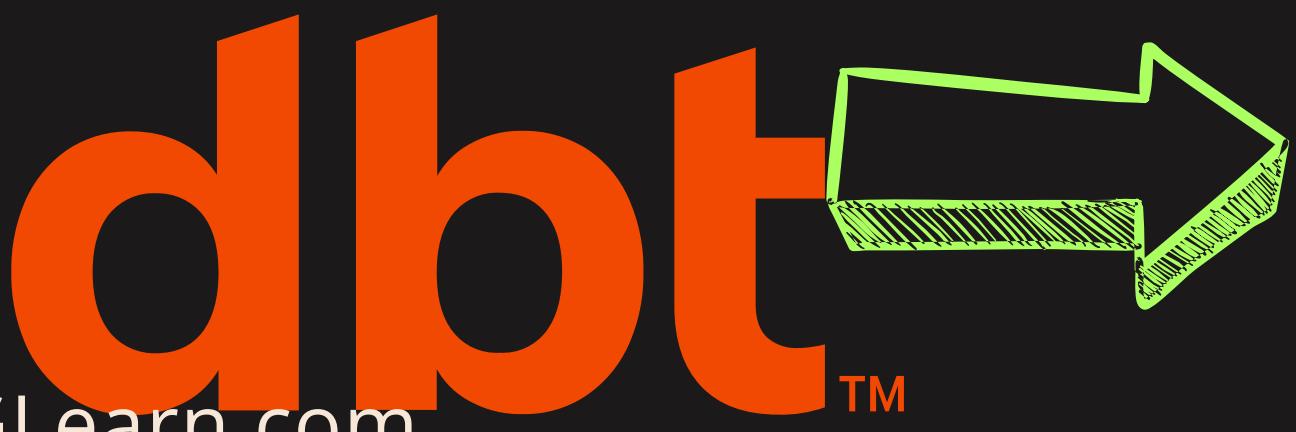
DAG (DIRECTED ACYCLIC GRAPH)

DBT AUTOMATICALLY CREATES A DAG OF YOUR MODELS, SHOWING DEPENDENCIES AND THE ORDER OF EXECUTION.

VISUALIZING THE DAG WITH DBT DOCS GENERATE AND DBT DOCS SERVE.



Shwetank Singh
GritSetGrow - GSGLearn.com



MODEL NAMING CONVENTIONS

USING CONSISTENT NAMING CONVENTIONS FOR MODELS, SOURCES, AND TESTS TO IMPROVE READABILITY AND MAINTAINABILITY.

NAMING MODELS WITH PREFIXES LIKE **STG_**, **INT_**, AND **DIM_**.



Shwetank Singh
GritSetGrow - GSGLearn.com



DBT CLOUD

A HOSTED SERVICE THAT PROVIDES
AN INTEGRATED DEVELOPMENT
ENVIRONMENT FOR DBT.

USING DBT CLOUD FOR SCHEDULING
RUNS AND MANAGING
ENVIRONMENTS.



Shwetank Singh
GritSetGrow - GSGLearn.com

dbt™

DBT CLI

THE COMMAND-LINE INTERFACE FOR DBT, USED TO EXECUTE VARIOUS DBT COMMANDS LOCALLY.

RUNNING DBT RUN FROM THE COMMAND LINE.



Shwetank Singh
GritSetGrow - GSGLearn.com

dbt™

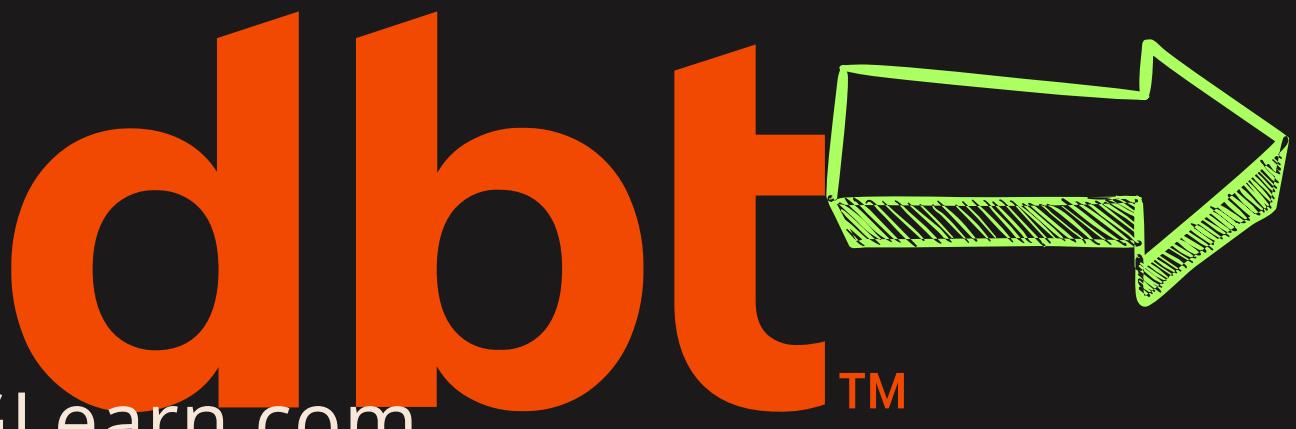
YAML CONFIGURATION

USING YAML FILES FOR
CONFIGURATION, WHICH PROVIDE A
HUMAN-READABLE FORMAT FOR
DEFINING SOURCES, TESTS, AND
MORE.

DEFINING SOURCES IN SOURCES.YML
AND TESTS IN SCHEMA.YML.



Shwetank Singh
GritSetGrow - GSGLearn.com



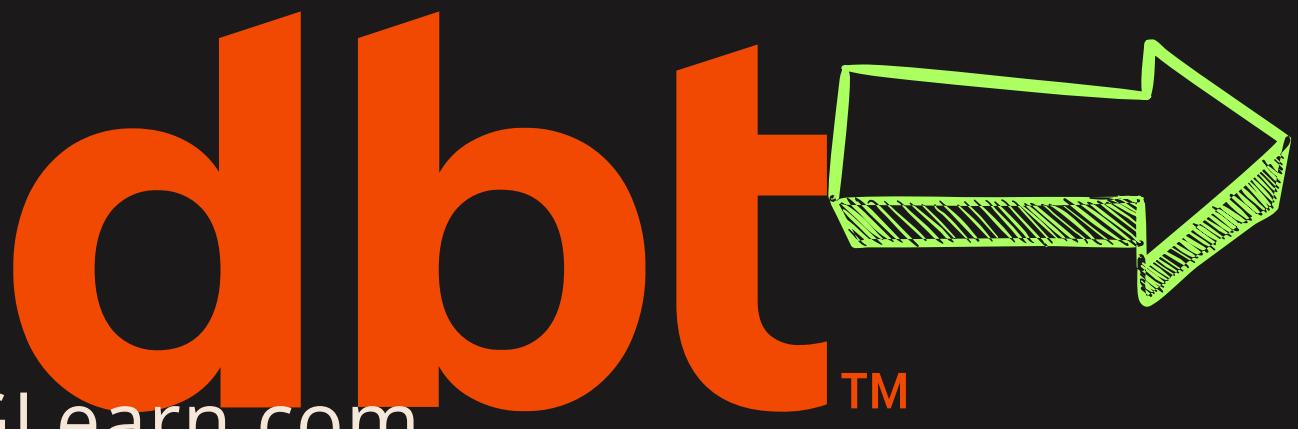
DEBUGGING

**USING DBT DEBUG TO
TROUBLESHOOT CONNECTION
ISSUES AND VALIDATE PROJECT
SETUP.**

**RUNNING DBT DEBUG TO CHECK
DATABASE CONNECTION.**



Shwetank Singh
GritSetGrow - GSGLearn.com



LOGGING

DBT PROVIDES DETAILED LOGS FOR EACH RUN, WHICH CAN BE FOUND IN THE LOGS/ DIRECTORY.

REVIEWING LOGS IN LOGS/DBT.LOG TO DIAGNOSE ISSUES.



Shwetank Singh
GritSetGrow - GSGLearn.com



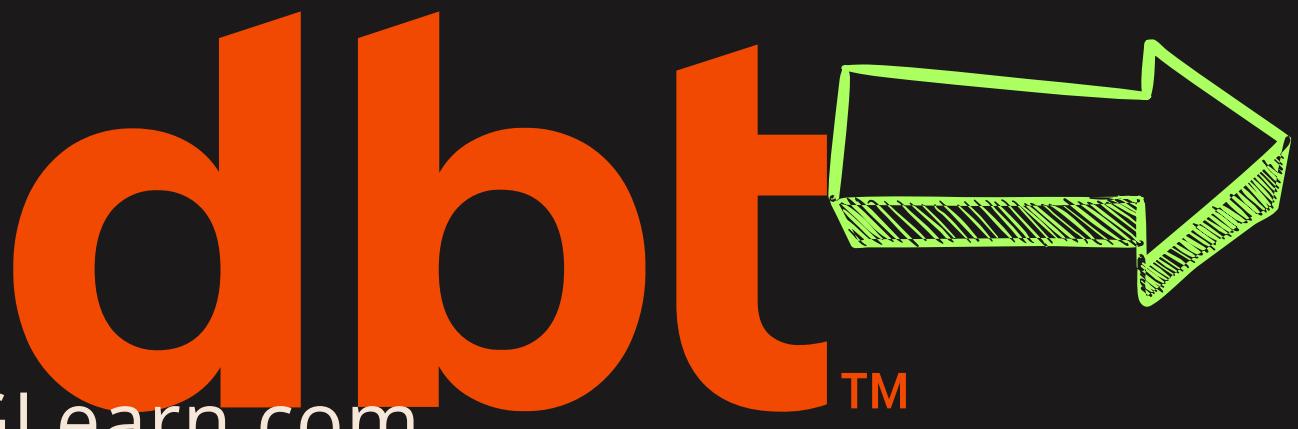
ERROR HANDLING

HANDLING ERRORS IN DBT MODELS
AND MACROS TO ENSURE SMOOTH
EXECUTION.

USING TRY...EXCEPT BLOCKS IN JINJA
FOR ERROR HANDLING.



Shwetank Singh
GritSetGrow - GSGLearn.com



STATE COMPARISON

COMPARING THE STATE OF THE WAREHOUSE BEFORE AND AFTER A DBT RUN USING DBT ARTIFACTS.

USING DBT RUN --STATE TO COMPARE CHANGES.



Shwetank Singh
GritSetGrow - GSGLearn.com



AUTOMATED TESTING

AUTOMATING DATA QUALITY CHECKS
BY SCHEDULING DBT TEST IN A CI/CD
PIPELINE.

SETTING UP A GITHUB ACTIONS
WORKFLOW TO RUN DBT TEST ON
EACH COMMIT.



Shwetank Singh
GritSetGrow - GSGLearn.com



CI/CD INTEGRATION

INTEGRATING DBT WITH CI/CD TOOLS
LIKE GITHUB ACTIONS, GITLAB CI, OR
JENKINS.

CONFIGURING A GITHUB ACTIONS
WORKFLOW TO RUN DBT
COMMANDS.



Shwetank Singh
GritSetGrow - GSGLearn.com



MODEL OPTIMIZATION

OPTIMIZING MODELS FOR
PERFORMANCE BY USING EFFICIENT
SQL PRACTICES AND DBT
CONFIGURATIONS.

USING CLUSTERED TABLES IN
BIGQUERY FOR BETTER
PERFORMANCE.



Shwetank Singh
GritSetGrow - GSGLearn.com



MODULAR SQL

**BREAKING DOWN COMPLEX SQL
TRANSFORMATIONS INTO SMALLER,
REUSABLE MODELS.**

**CREATING INTERMEDIATE MODELS
FOR EACH STEP OF THE
TRANSFORMATION PROCESS.**



Shwetank Singh
GritSetGrow - GSGLearn.com



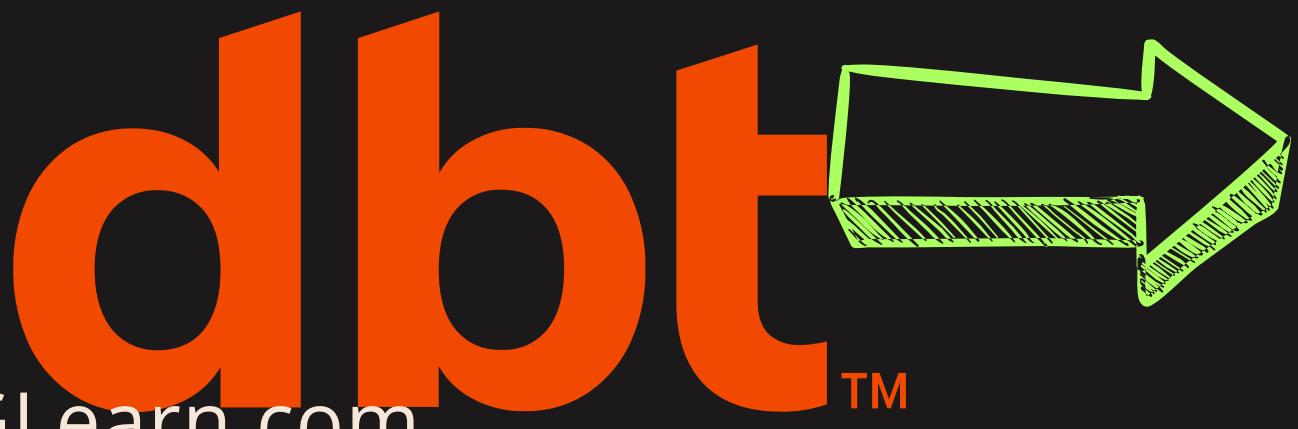
SQL STYLE GUIDE

**ADOPTING A SQL STYLE GUIDE FOR
CONSISTENT AND READABLE SQL
CODE IN DBT MODELS.**

**USING UPPERCASE FOR SQL
KEYWORDS AND LOWERCASE FOR
COLUMN NAMES.**



Shwetank Singh
GritSetGrow - GSGLearn.com



COLLABORATION

**USING GIT FOR VERSION CONTROL
AND COLLABORATION ON DBT
PROJECTS.**

**CREATING PULL REQUESTS FOR
CHANGES TO DBT MODELS AND
CONFIGURATIONS.**



Shwetank Singh
GritSetGrow - GSGLearn.com



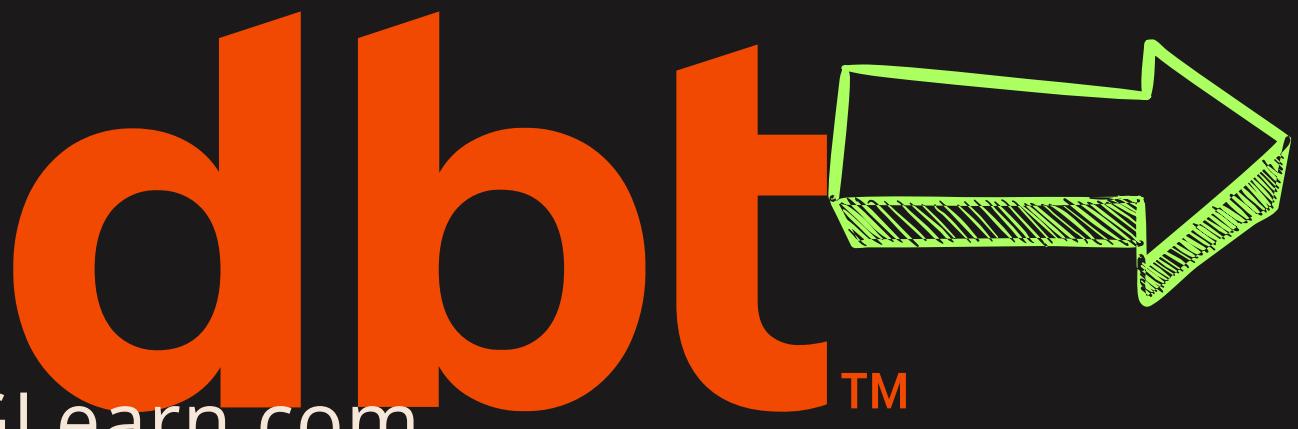
DATA LINEAGE

TRACKING DATA LINEAGE FROM SOURCE TO TARGET USING DBT'S DAG AND DOCUMENTATION FEATURES.

VIEWING THE DATA LINEAGE IN THE DBT DOCUMENTATION SITE.



Shwetank Singh
GritSetGrow - GSGLearn.com



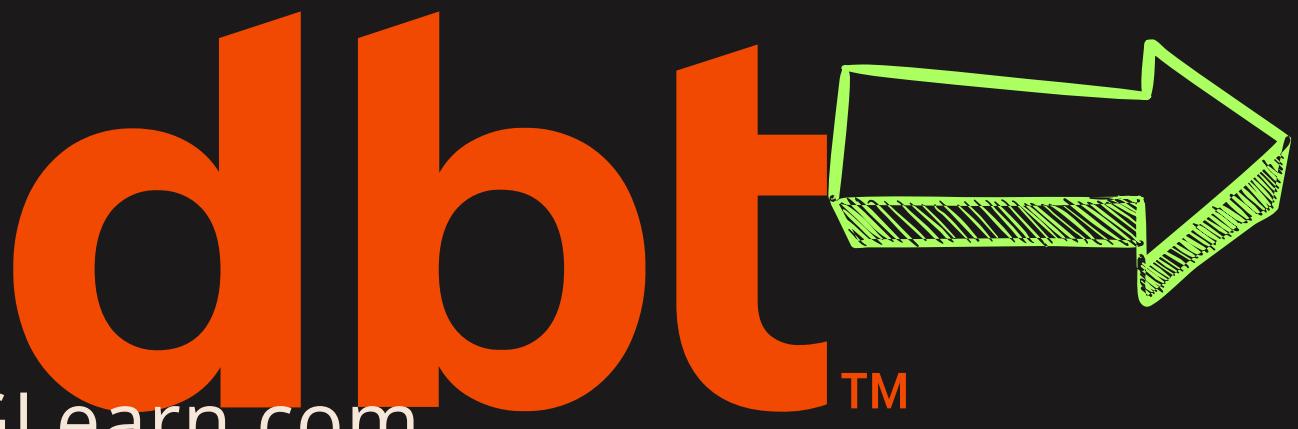
DATA GOVERNANCE

**ENSURING DATA GOVERNANCE BY
DEFINING OWNERSHIP,
DOCUMENTATION, AND TESTS FOR
ALL MODELS AND SOURCES.**

**ADDING OWNERSHIP AND
DESCRIPTIONS IN SCHEMA.YML.**



Shwetank Singh
GritSetGrow - GSGLearn.com



MODEL DEPENDENCIES

MANAGING DEPENDENCIES BETWEEN
MODELS USING THE REF FUNCTION.

REFERENCING UPSTREAM MODELS
WITH {{ REF('UPSTREAM_MODEL') }}.



Shwetank Singh
GritSetGrow - GSGLearn.com



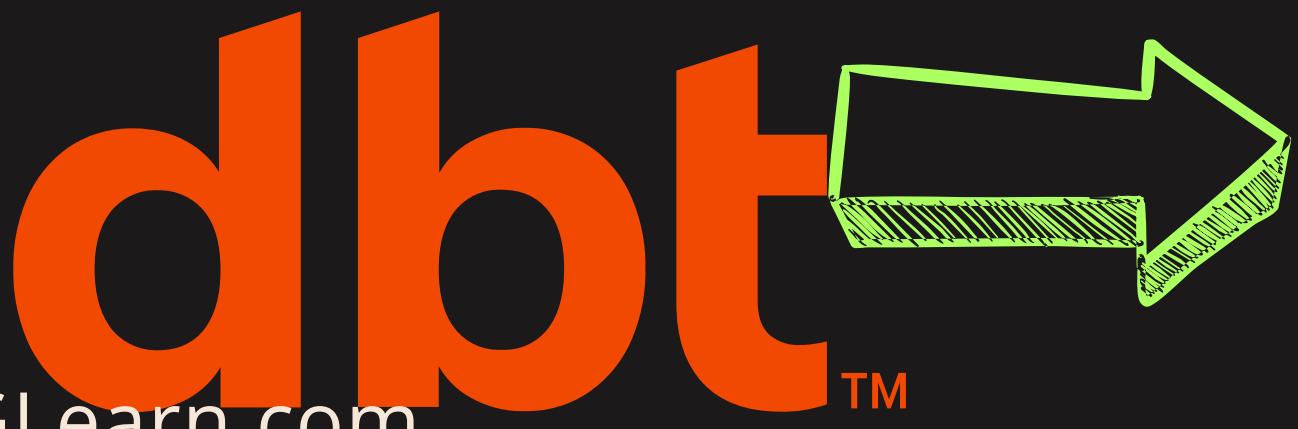
MODEL OVERRIDES

OVERRIDING MODELS FOR DIFFERENT ENVIRONMENTS OR USE CASES BY USING THE OVERRIDE FEATURE.

CREATING AN OVERRIDE MODEL WITH `{{ CONFIG(OVERRIDE='TRUE') }}`.



Shwetank Singh
GritSetGrow - GSGLearn.com



MODEL INHERITANCE

INHERITING PROPERTIES AND CONFIGURATIONS FROM BASE MODELS USING THE EXTEND FEATURE.

CREATING A MODEL THAT EXTENDS ANOTHER MODEL WITH `{{ EXTEND('BASE_MODEL') }}`.



Shwetank Singh
GritSetGrow - GSGLearn.com



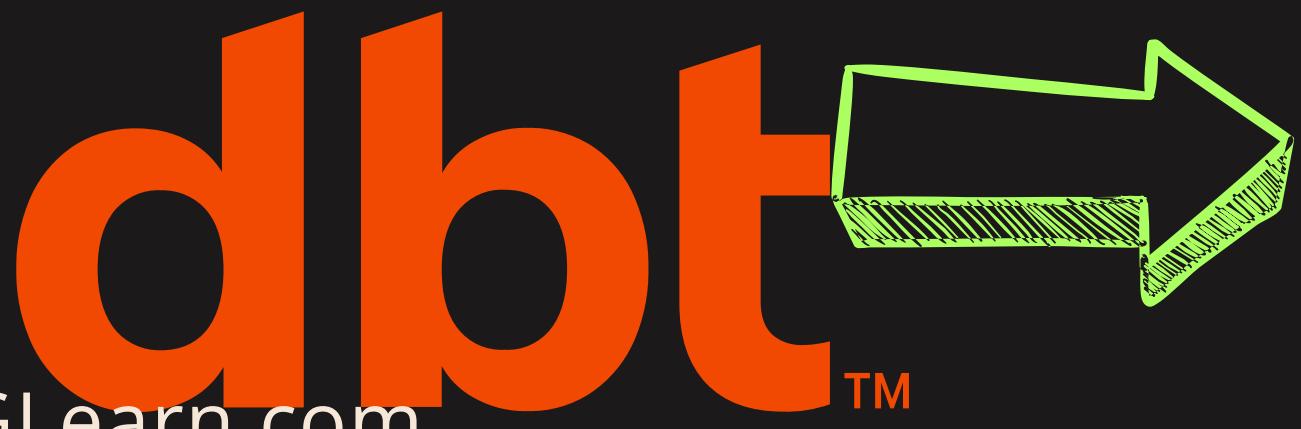
DATA QUALITY

**ENSURING DATA QUALITY BY
IMPLEMENTING TESTS, SNAPSHOTS,
AND DOCUMENTATION IN DBT.**

**USING BUILT-IN TESTS LIKE UNIQUE,
NOT_NULL, AND CUSTOM TESTS.**



Shwetank Singh
GritSetGrow - GSGLearn.com



PERFORMANCE MONITORING

MONITORING THE PERFORMANCE OF DBT MODELS AND RUNS USING DBT CLOUD OR CUSTOM LOGGING SOLUTIONS.

ANALYZING RUN TIMES AND QUERY PERFORMANCE IN DBT CLOUD.



Shwetank Singh
GritSetGrow - GSGLearn.com



SCALABILITY

**SCALING DBT PROJECTS BY
ORGANIZING MODELS, USING
PACKAGES, AND OPTIMIZING SQL
CODE.**

**BREAKING DOWN LARGE PROJECTS
INTO SMALLER, MANAGEABLE SUB-
PROJECTS.**



Shwetank Singh
GritSetGrow - GSGLearn.com



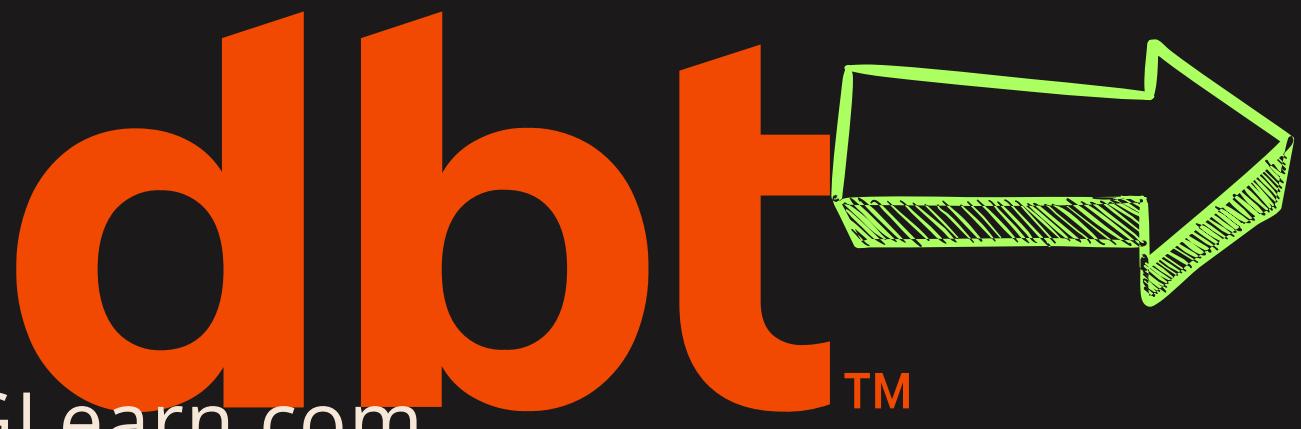
SECURITY

**ENSURING SECURITY BY MANAGING
ACCESS CONTROLS AND SENSITIVE
DATA IN DBT PROJECTS.**

**USING ENVIRONMENT VARIABLES TO
HANDLE SENSITIVE DATA LIKE
PASSWORDS.**



Shwetank Singh
GritSetGrow - GSGLearn.com



DATA MODELING

IMPLEMENTING DATA MODELING
BEST PRACTICES IN DBT, SUCH AS
STAR SCHEMA AND SNOWFLAKE
SCHEMA.

CREATING DIMENSION AND FACT
TABLES FOR A STAR SCHEMA.



Shwetank Singh
GritSetGrow - GSGLearn.com



CHANGE MANAGEMENT

**MANAGING CHANGES TO DBT
MODELS AND CONFIGURATIONS
USING VERSION CONTROL AND
DOCUMENTATION.**

**USING GIT BRANCHES AND PULL
REQUESTS FOR CHANGES.**



Shwetank Singh
GritSetGrow - GSGLearn.com



MODEL DEPLOYMENT

DEPLOYING DBT MODELS TO
DIFFERENT ENVIRONMENTS USING
CI/CD PIPELINES AND ENVIRONMENT
CONFIGURATIONS.

SETTING UP AUTOMATED
DEPLOYMENTS TO STAGING AND
PRODUCTION.



Shwetank Singh
GritSetGrow - GSGLearn.com



DOCUMENTATION WORKFLOW

CREATING AND MAINTAINING
DOCUMENTATION AS PART OF THE
DEVELOPMENT WORKFLOW.

REGULARLY UPDATING SCHEMA.YML
AND GENERATING DOCUMENTATION.



Shwetank Singh
GritSetGrow - GSGLearn.com

dbt™

COMMUNITY SUPPORT

LEVERAGING THE DBT COMMUNITY
FOR SUPPORT, BEST PRACTICES, AND
SHARED PACKAGES.

PARTICIPATING IN DBT SLACK
CHANNELS AND FORUMS.



Shwetank Singh
GritSetGrow - GSGLearn.com



ADVANCED MACROS

**CREATING ADVANCED MACROS FOR
REUSABLE SQL LOGIC AND COMPLEX
TRANSFORMATIONS.**

**WRITING A MACRO TO DYNAMICALLY
CREATE PARTITIONS IN A TABLE.**



Shwetank Singh
GritSetGrow - GSGLearn.com



METADATA MANAGEMENT

MANAGING METADATA BY DOCUMENTING SOURCES, MODELS, AND TESTS IN DBT.

USING SCHEMA.YML TO ADD DESCRIPTIONS AND METADATA FOR MODELS.



Shwetank Singh
GritSetGrow - GSGLearn.com

dbt™

SOURCE FRESHNESS

MONITORING THE FRESHNESS OF SOURCE DATA USING DBT'S SOURCE FRESHNESS FEATURE.

CONFIGURING SOURCE FRESHNESS CHECKS IN SOURCES.YML AND RUNNING DBT SOURCE FRESHNESS.



Shwetank Singh
GritSetGrow - GSGLearn.com



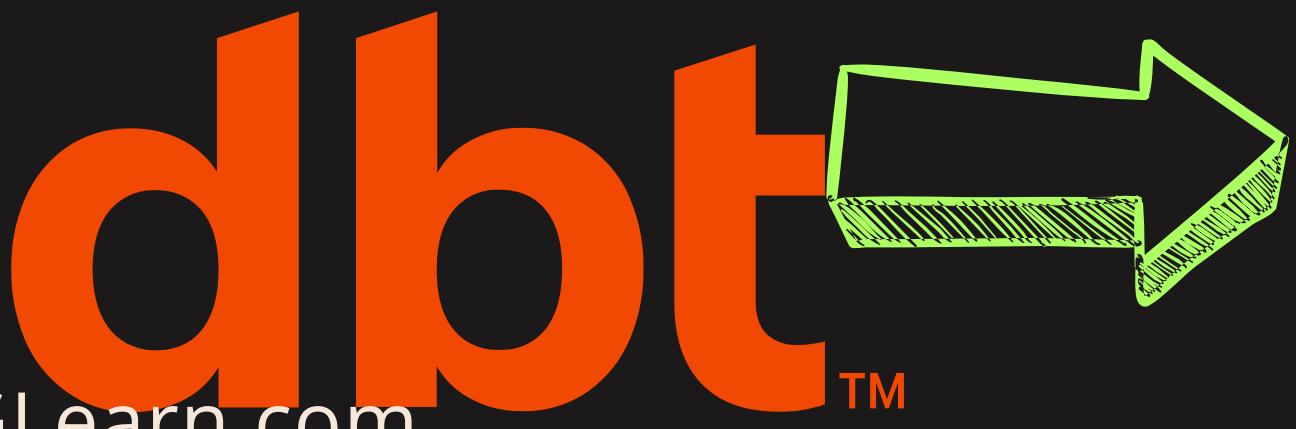
DATA WAREHOUSING BEST PRACTICES

IMPLEMENTING BEST PRACTICES FOR DATA WAREHOUSING, SUCH AS INDEXING, PARTITIONING, AND CLUSTERING.

APPLYING CLUSTERING TO BIGQUERY TABLES FOR PERFORMANCE IMPROVEMENT.



Shwetank Singh
GritSetGrow - GSGLearn.com



ANALYTICS ENGINEERING

INTEGRATING DBT WITH OTHER
ANALYTICS TOOLS FOR A COMPLETE
ANALYTICS ENGINEERING
WORKFLOW.

USING DBT WITH BI TOOLS LIKE
LOOKER OR TABLEAU FOR
DOWNSTREAM ANALYSIS.



Shwetank Singh
GritSetGrow - GSGLearn.com

dbt™

DATA INTEGRATION

INTEGRATING DBT WITH OTHER DATA
INTEGRATION TOOLS AND
PLATFORMS FOR A SEAMLESS
WORKFLOW.

USING DBT WITH FIVETRAN OR
STITCH FOR DATA INGESTION.



Shwetank Singh
GritSetGrow - GSGLearn.com



DBT SCHEDULER

SCHEDULING DBT RUNS USING DBT CLOUD OR OTHER SCHEDULING TOOLS.

SETTING UP A SCHEDULE IN DBT CLOUD TO RUN MODELS DAILY.



Shwetank Singh
GritSetGrow - GSGLearn.com



DATA OBSERVABILITY

IMPLEMENTING DATA
OBSERVABILITY PRACTICES TO
MONITOR AND ALERT ON DATA
ISSUES.

USING DBT AND TOOLS LIKE MONTE
CARLO OR DATABAND FOR DATA
OBSERVABILITY.



Shwetank Singh
GritSetGrow - GSGLearn.com



MODEL TESTING

WRITING AND RUNNING TESTS FOR MODELS TO ENSURE THEY MEET BUSINESS REQUIREMENTS AND QUALITY STANDARDS.

CREATING CUSTOM TESTS IN TESTS/DIRECTORY AND RUNNING DBT TEST.



Shwetank Singh
GritSetGrow - GSGLearn.com



SCHEMA MANAGEMENT

MANAGING DATABASE SCHEMAS AND
TABLES CREATED BY DBT MODELS.

USING SCHEMA CONFIGURATION IN
MODELS TO CONTROL WHERE TABLES
ARE CREATED.



Shwetank Singh
GritSetGrow - GSGLearn.com

dbt™

CODE REVIEW

CONDUCTING CODE REVIEWS FOR
DBT MODELS AND CONFIGURATIONS
TO ENSURE QUALITY AND
CONSISTENCY.

REVIEWING PULL REQUESTS IN
GITHUB BEFORE MERGING.



Shwetank Singh
GritSetGrow - GSGLearn.com

dbt™

DATA TRANSFORMATION

TRANSFORMING RAW DATA INTO
CLEAN, USABLE DATASETS USING DBT
MODELS.

WRITING SQL TRANSFORMATIONS IN
MODELS/ DIRECTORY.



Shwetank Singh
GritSetGrow - GSGLearn.com



MODEL REFACTORING

REFACTORING MODELS TO IMPROVE
PERFORMANCE, READABILITY, AND
MAINTAINABILITY.

BREAKING DOWN LARGE MODELS
INTO SMALLER, REUSABLE
COMPONENTS.



Shwetank Singh
GritSetGrow - GSGLearn.com

dbt™

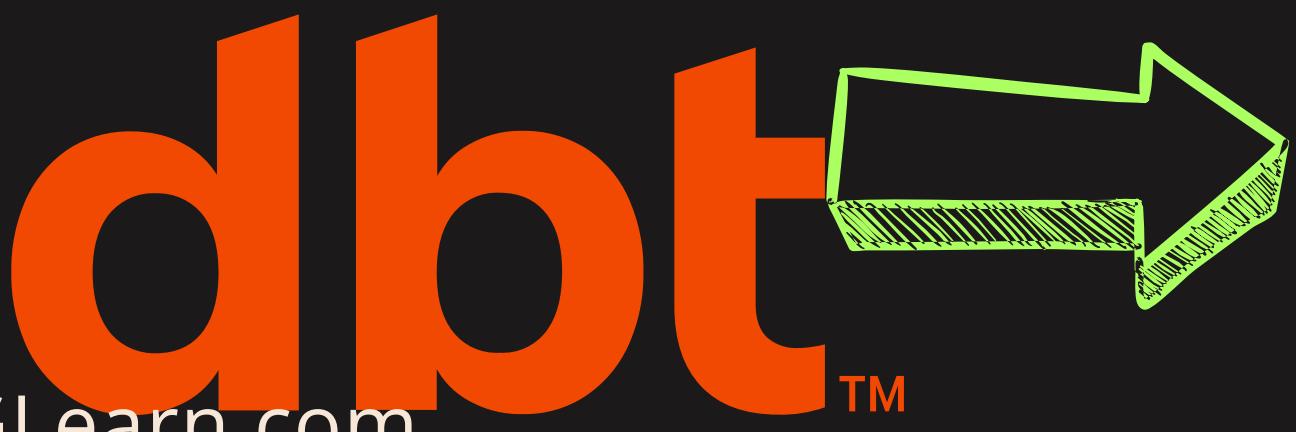
DATA ENRICHMENT

**ENRICHING DATA BY ADDING
CALCULATED FIELDS, DERIVED
METRICS, AND AGGREGATIONS IN
DBT MODELS.**

**ADDING A CALCULATED FIELD FOR
REVENUE PER USER IN A MODEL.**



Shwetank Singh
GritSetGrow - GSGLearn.com



DOCUMENTATION AUTOMATION

AUTOMATING THE GENERATION OF DOCUMENTATION AND ENSURING IT IS UP-TO-DATE WITH THE CURRENT STATE OF THE MODELS.

USING DBT DOCS GENERATE IN A CI/CD PIPELINE TO AUTOMATE DOCUMENTATION UPDATES.



Shwetank Singh
GritSetGrow - GSGLearn.com



DATA COMPLIANCE

**ENSURING DATA COMPLIANCE WITH
REGULATIONS AND STANDARDS
USING DBT'S TESTING AND
DOCUMENTATION FEATURES.**

**ADDING TESTS FOR DATA PRIVACY
AND SECURITY COMPLIANCE.**



Shwetank Singh
GritSetGrow - GSGLearn.com



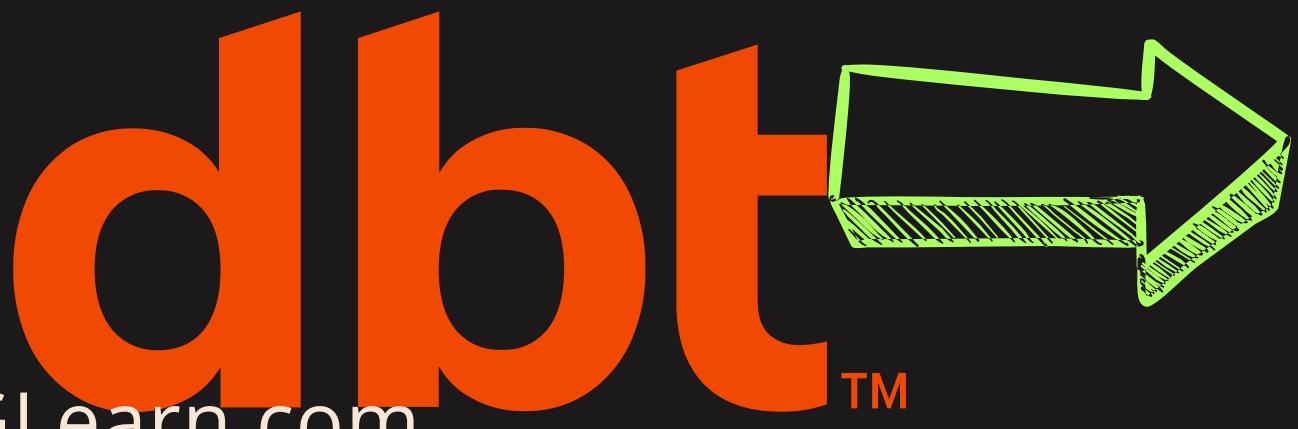
VERSIONING

MANAGING VERSIONS OF DBT MODELS AND CONFIGURATIONS TO TRACK CHANGES AND MAINTAIN HISTORY.

USING GIT TAGS AND BRANCHES TO MANAGE VERSIONS.



Shwetank Singh
GritSetGrow - GSGLearn.com



MODEL DEPLOYMENT STRATEGY

**DEVELOPING A STRATEGY FOR
DEPLOYING MODELS TO
PRODUCTION, INCLUDING ROLLBACK
PROCEDURES AND MONITORING.**

**PLANNING A PHASED DEPLOYMENT
WITH ROLLBACK OPTIONS.**



Shwetank Singh
GritSetGrow - GSGLearn.com

dbt™

A large, stylized orange lowercase word "dbt" with a trademark symbol. To the right of the "t" is a thick, hand-drawn style green arrow pointing to the right.

DATA TRANSFORMATION

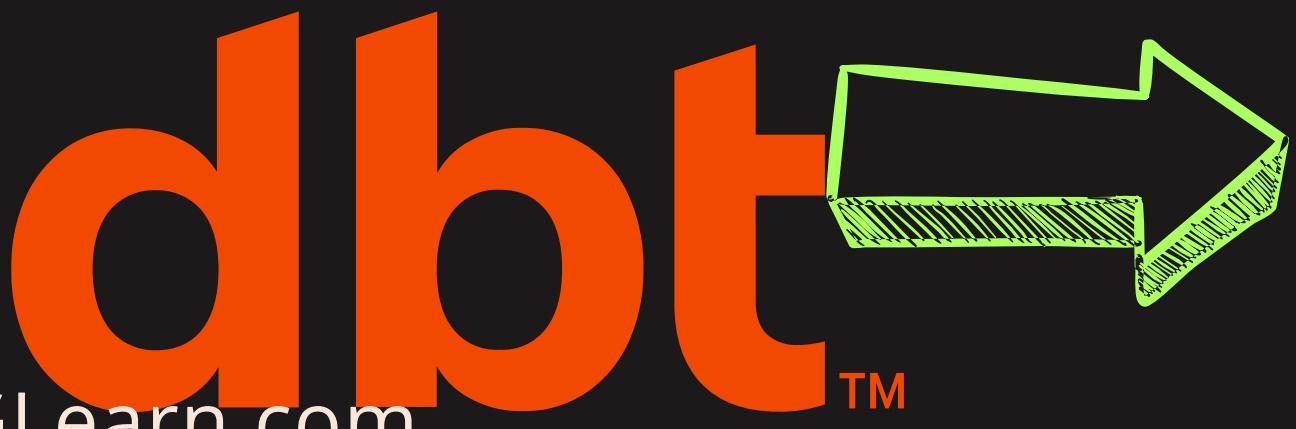
PIPELINE

BUILDING A COMPLETE DATA
TRANSFORMATION PIPELINE USING
DBT, FROM RAW DATA INGESTION TO
FINAL REPORTING TABLES.

INTEGRATING DBT WITH DATA
INGESTION AND BI TOOLS TO CREATE
A FULL PIPELINE.



Shwetank Singh
GritSetGrow - GSGLearn.com



USE CASE: SALES REPORTING

IMPLEMENTING DBT MODELS TO TRANSFORM SALES DATA FOR REPORTING AND ANALYSIS.

CREATING MODELS TO AGGREGATE SALES DATA BY REGION AND PRODUCT.



Shwetank Singh
GritSetGrow - GSGLearn.com

dbt™

USE CASE: CUSTOMER ANALYTICS

USING DBT TO MODEL CUSTOMER DATA FOR ANALYTICS AND INSIGHTS.

BUILDING MODELS TO SEGMENT CUSTOMERS AND ANALYZE BEHAVIOR.



Shwetank Singh
GritSetGrow - GSGLearn.com

dbt™

USE CASE: FINANCIAL REPORTING

DEVELOPING DBT MODELS TO
PREPARE FINANCIAL DATA FOR
REPORTING AND COMPLIANCE.

CREATING MODELS FOR FINANCIAL STATEMENTS AND COMPLIANCE REPORTS.



Shwetank Singh
GritSetGrow - GSGLearn.com

dbt™

A large, stylized orange lowercase word "dbt" with a trademark symbol. To the right of the "t" is a thick, hand-drawn style green arrow pointing to the right.

USE CASE: MARKETING ANALYTICS

IMPLEMENTING DBT MODELS TO
ANALYZE MARKETING CAMPAIGN
DATA AND MEASURE PERFORMANCE.

BUILDING MODELS TO TRACK
CAMPAIGN PERFORMANCE AND ROI.



Shwetank Singh
GritSetGrow - GSGLearn.com

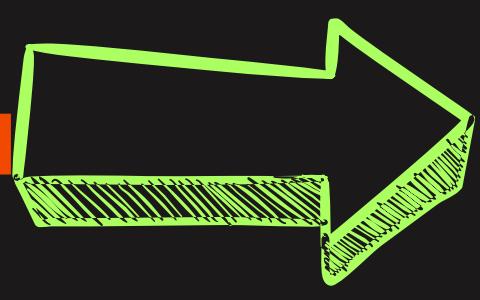
dbt™

USE CASE: OPERATIONAL **REPORTING** USING DBT TO TRANSFORM OPERATIONAL DATA FOR REAL-TIME REPORTING AND DECISION-MAKING.

CREATING MODELS TO MONITOR KEY
OPERATIONAL METRICS.



Shwetank Singh
GritSetGrow - GSGLearn.com

dbt™ 

DATA CLEANSING

MOVING DUPLICATES, HANDLING
MISSING VALUES, AND
STANDARDIZING DATA FORMATS.

CREATING A MODEL
CLEAN_DATA.SQL WITH SQL LOGIC
TO CLEAN THE RAW DATA.



Shwetank Singh
GritSetGrow - GSGLearn.com



DATA AGGREGATION

SUMMARIZING DATA AT DIFFERENT LEVELS OF GRANULARITY, SUCH AS DAILY, MONTHLY, OR YEARLY AGGREGATIONS.

AGGREGATED_SALES.SQL MODEL THAT GROUPS SALES DATA BY MONTH AND REGION.



Shwetank Singh
GritSetGrow - GSGLearn.com



CREATING DERIVED COLUMNS

ADDING NEW CALCULATED COLUMNS BASED ON EXISTING DATA, SUCH AS REVENUE PER USER OR PROFIT MARGINS.

DERIVED_COLUMNS.SQL WITH A NEW COLUMN REVENUE_PER_USER CALCULATED FROM SALES AND USER DATA.



Shwetank Singh
GritSetGrow - GSGLearn.com



BUILDING DIMENSION TABLES

CREATING DIMENSION TABLES FOR USE IN A STAR SCHEMA, LIKE CUSTOMER, PRODUCT, AND TIME DIMENSIONS.

DIM_CUSTOMER.SQL MODEL TO CREATE A CUSTOMER DIMENSION TABLE WITH CLEANED AND ENRICHED DATA.



Shwetank Singh
GritSetGrow - GSGLearn.com



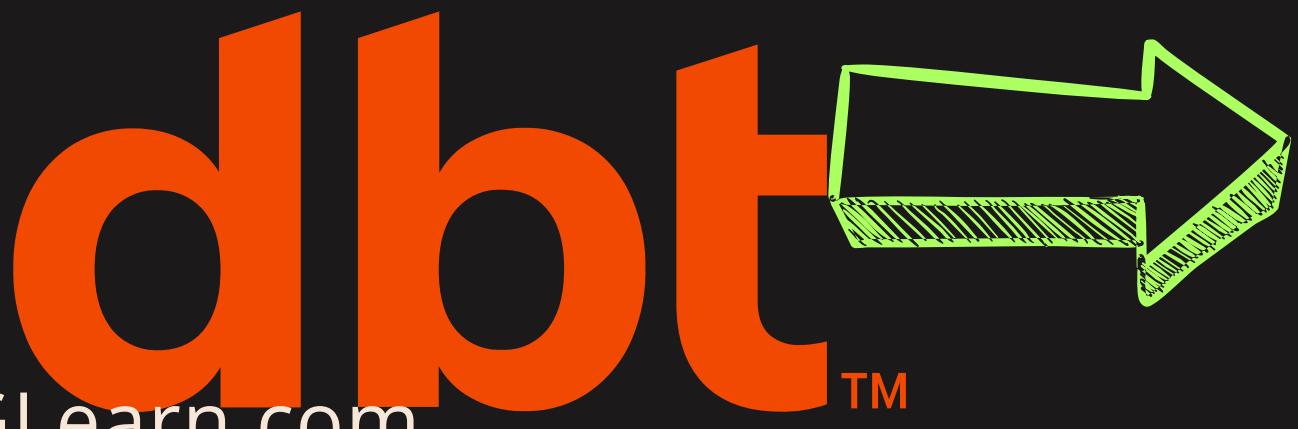
BUILDING FACT TABLES

CREATING FACT TABLES THAT STORE TRANSACTIONAL DATA, LIKE SALES, ORDERS, OR CLICKS.

FACT_SALES.SQL MODEL THAT AGGREGATES SALES TRANSACTIONS FOR ANALYSIS.



Shwetank Singh
GritSetGrow - GSGLearn.com



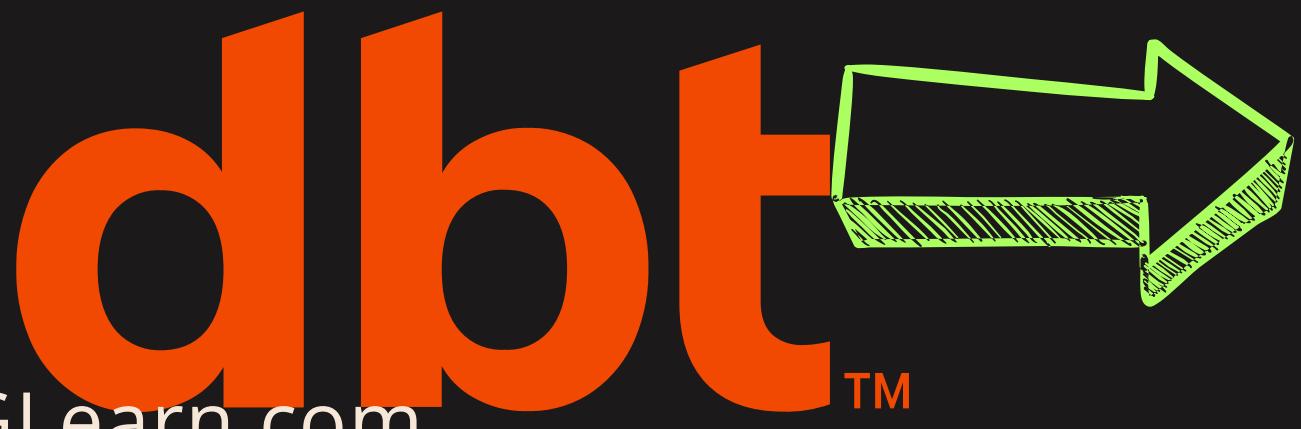
DATA ENRICHMENT

ENRICHING DATA BY JOINING IT
WITH EXTERNAL DATASETS OR
ADDING LOOKUP VALUES.

JOINING SALES DATA WITH
GEOGRAPHIC DATA TO ADD REGION
INFORMATION IN
ENRICHED_SALES.SQL.



Shwetank Singh
GritSetGrow - GSGLearn.com



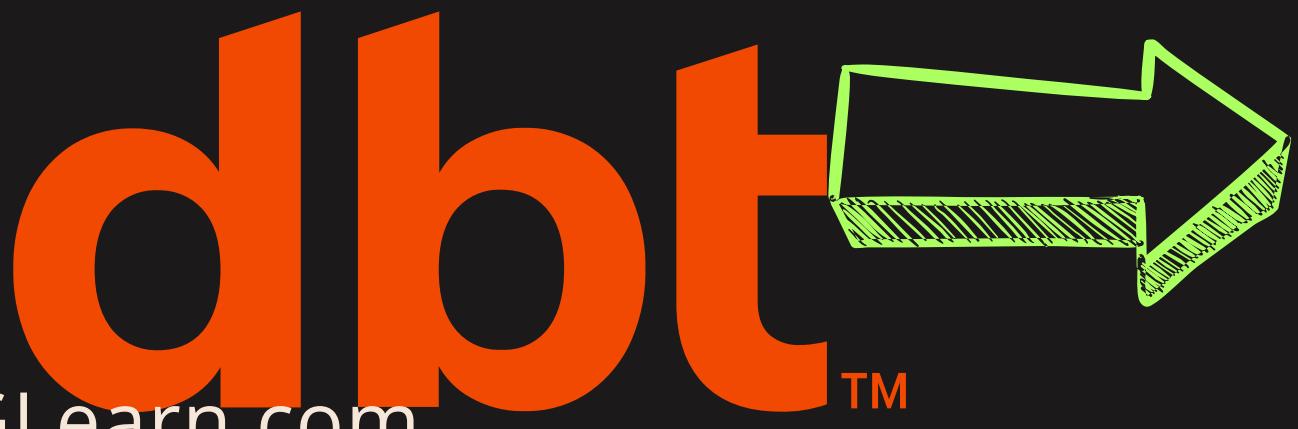
IMPLEMENTING SLOWLY CHANGING DIMENSIONS

CAPTURING CHANGES IN DIMENSION
DATA OVER TIME USING SNAPSHOTS.

DEFINING A SNAPSHOT
`CUSTOMER_SNAPSHOT.SQL` TO TRACK
CHANGES IN CUSTOMER
INFORMATION.



Shwetank Singh
GritSetGrow - GSGLearn.com



DATA QUALITY TESTING

IMPLEMENTING TESTS TO ENSURE DATA QUALITY, SUCH AS UNIQUENESS, NON-NULL VALUES, AND RELATIONSHIPS BETWEEN TABLES.

ADDING A UNIQUENESS TEST FOR CUSTOMER IDS IN SCHEMA.YML.



Shwetank Singh
GritSetGrow - GSGLearn.com



SCHEMA DOCUMENTATION

DOCUMENTING THE SCHEMA OF THE MODELS, INCLUDING DESCRIPTIONS OF COLUMNS AND TABLES.

USING SCHEMA.YML TO ADD DESCRIPTIONS AND GENERATE DOCS WITH DBT DOCS GENERATE.



Shwetank Singh
GritSetGrow - GSGLearn.com

dbt™

SOURCE FRESHNESS MONITORING

MONITORING THE FRESHNESS OF SOURCE DATA TO ENSURE IT IS UP-TO-DATE AND TIMELY.

CONFIGURING SOURCE FRESHNESS CHECKS IN SOURCES.YML AND RUNNING DBT SOURCE FRESHNESS.



Shwetank Singh
GritSetGrow - GSGLearn.com



AUTOMATING DATA PIPELINES

AUTOMATING THE EXECUTION OF DATA TRANSFORMATION PIPELINES USING DBT CLOUD OR CI/CD TOOLS.

SETTING UP A DAILY RUN SCHEDULE IN DBT CLOUD FOR THE DATA TRANSFORMATION PIPELINE.



Shwetank Singh
GritSetGrow - GSGLearn.com



DATA LINEAGE AND DEPENDENCY

TRACKING

VISUALIZING THE DEPENDENCY

GRAPH OF MODELS TO UNDERSTAND
THE DATA LINEAGE.

USING DBT DOCS GENERATE AND DBT
DOCS SERVE TO VIEW THE DAG.



Shwetank Singh
GritSetGrow - GSGLearn.com

dbt™

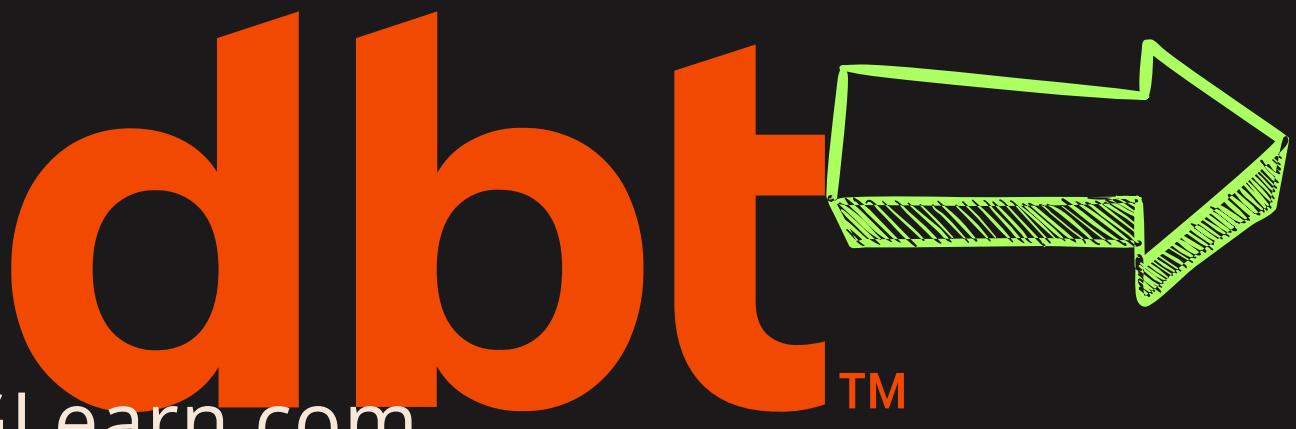
CREATING REUSABLE MACROS

WRITING SQL MACROS TO ENCAPSULATE REUSABLE LOGIC FOR TRANSFORMATIONS.

DEFINING A MACRO
`CALCULATE_GROWTH.SQL` TO
CALCULATE GROWTH RATES ACROSS
DIFFERENT MODELS.



Shwetank Singh
GritSetGrow - GSGLearn.com



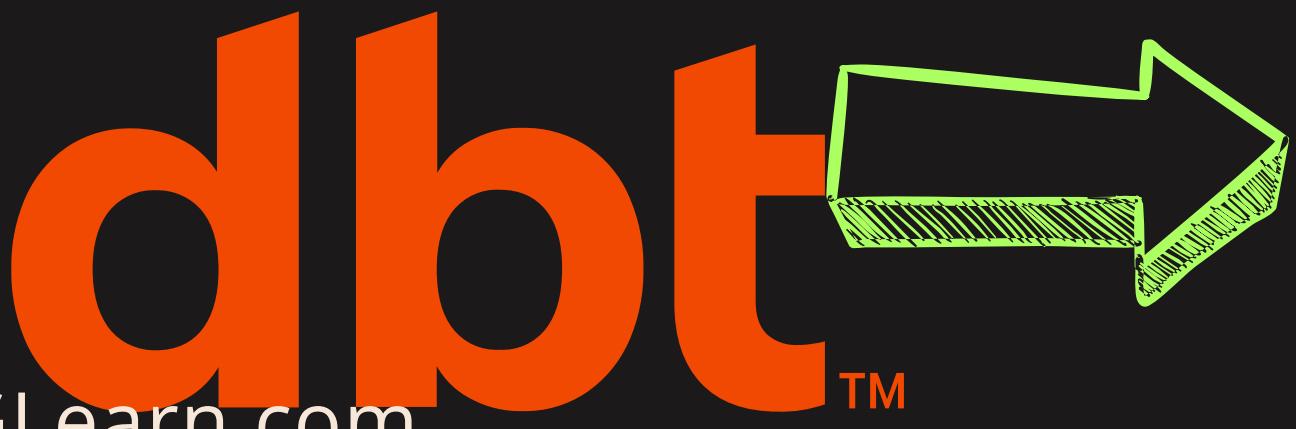
USING INCREMENTAL MODELS

IMPLEMENTING INCREMENTAL MODELS TO PROCESS ONLY NEW OR UPDATED DATA, IMPROVING PERFORMANCE.

CREATING AN INCREMENTAL MODEL `INCREMENTAL_SALES.SQL` WITH `UNIQUE_KEY` AND `IS_INCREMENTAL()`.



Shwetank Singh
GritSetGrow - GSGLearn.com



IMPLEMENTING CUSTOM **MATERIALIZATIONS**

DEFINING CUSTOM MATERIALIZATIONS FOR SPECIFIC NEEDS, SUCH AS PARTITIONED TABLES OR SPECIALIZED VIEWS.

WRITING A CUSTOM MATERIALIZATION IN MACROS/CUSTOM_MATERIALIZATION.SQL.



Shwetank Singh
GritSetGrow - GSGLearn.com



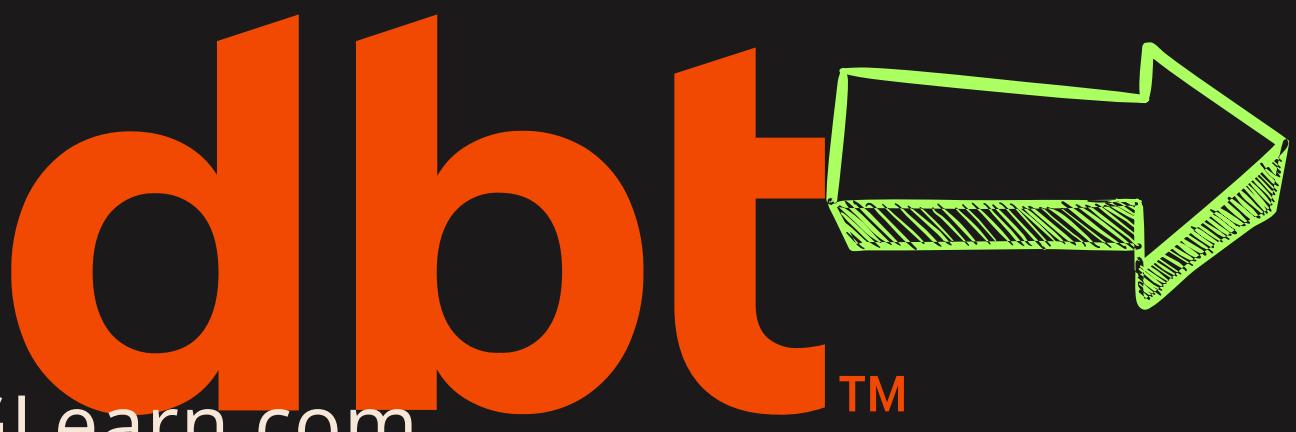
MANAGING ENVIRONMENT CONFIGURATIONS

CONFIGURING DIFFERENT ENVIRONMENTS (DEVELOPMENT, STAGING, PRODUCTION) FOR MODEL RUNS.

USING TARGET AND PROFILES.YML TO MANAGE DIFFERENT ENVIRONMENT CONFIGURATIONS.



Shwetank Singh
GritSetGrow - GSGLearn.com



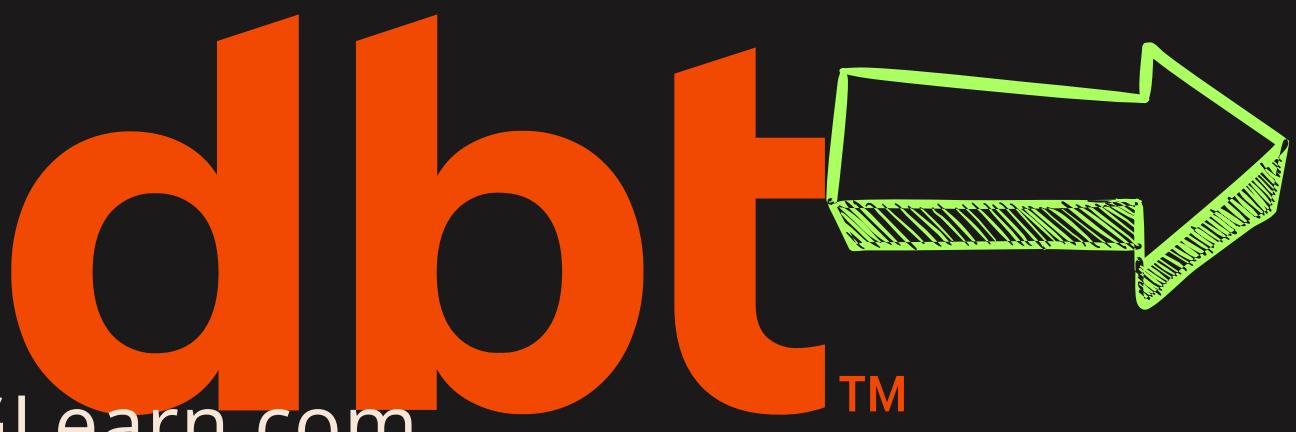
VERSION CONTROL AND COLLABORATION

USING GIT FOR VERSION CONTROL,
COLLABORATION, AND CODE REVIEW
PROCESSES.

SETTING UP A GIT REPOSITORY FOR
THE DBT PROJECT AND USING PULL
REQUESTS FOR CODE REVIEWS.



Shwetank Singh
GritSetGrow - GSGLearn.com



DATA TRANSFORMATION

LOGIC

WRITING COMPLEX

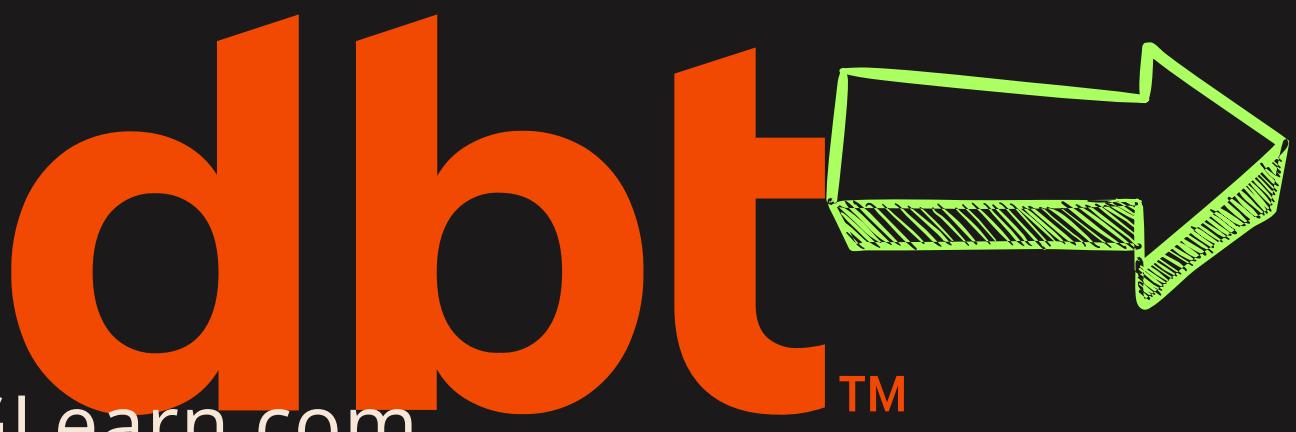
TRANSFORMATION LOGIC IN SQL TO
RESHAPE, FILTER, AND JOIN
DATASETS.

CREATING A MODEL

TRANSFORM_ORDERS.SQL TO JOIN
ORDER, CUSTOMER, AND PRODUCT
DATA.



Shwetank Singh
GritSetGrow - GSGLearn.com



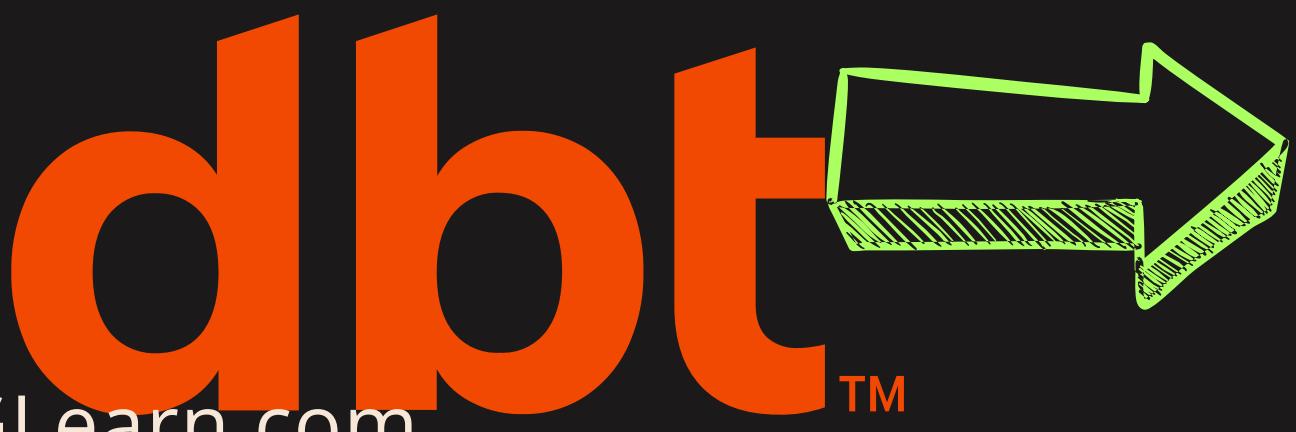
HANDLING SCHEMA CHANGES

MANAGING SCHEMA CHANGES, SUCH AS ADDING NEW COLUMNS OR CHANGING DATA TYPES, WITH MINIMAL DISRUPTION.

USING DBT'S ALTER COMMANDS OR CREATING NEW MODELS TO REFLECT SCHEMA CHANGES.



Shwetank Singh
GritSetGrow - GSGLearn.com



CREATING CUSTOM TESTS

WRITING CUSTOM TESTS TO VALIDATE SPECIFIC BUSINESS LOGIC OR DATA QUALITY REQUIREMENTS.

DEFINING A CUSTOM TEST IN TESTS/CUSTOM_TEST.SQL TO CHECK FOR VALID EMAIL FORMATS.



Shwetank Singh
GritSetGrow - GSGLearn.com



BUILDING ANALYTICAL MODELS

CREATING MODELS TO SUPPORT ADVANCED ANALYTICS, SUCH AS COHORT ANALYSIS, CHURN PREDICTION, OR TREND ANALYSIS.

BUILDING A COHORT_ANALYSIS.SQL MODEL TO ANALYZE USER RETENTION OVER TIME.



Shwetank Singh
GritSetGrow - GSGLearn.com

dbt™

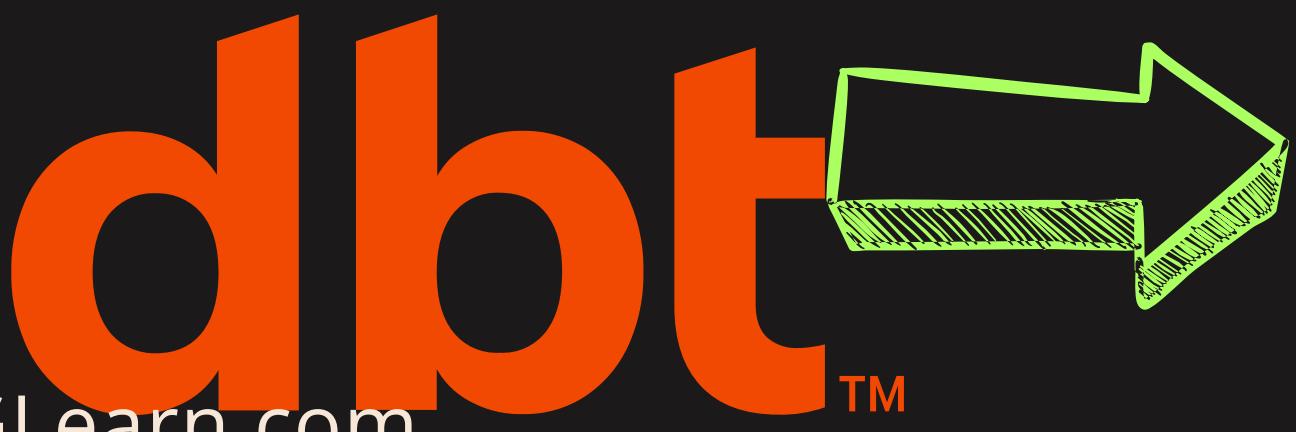
IMPLEMENTING BUSINESS LOGIC

EMBEDDING BUSINESS RULES AND
LOGIC INTO DATA TRANSFORMATION
PROCESSES TO ENSURE
CONSISTENCY.

WRITING SQL TRANSFORMATIONS
THAT APPLY DISCOUNT RULES AND
TAX CALCULATIONS IN
`BUSINESS_LOGIC.SQL`.



Shwetank Singh
GritSetGrow - GSGLearn.com



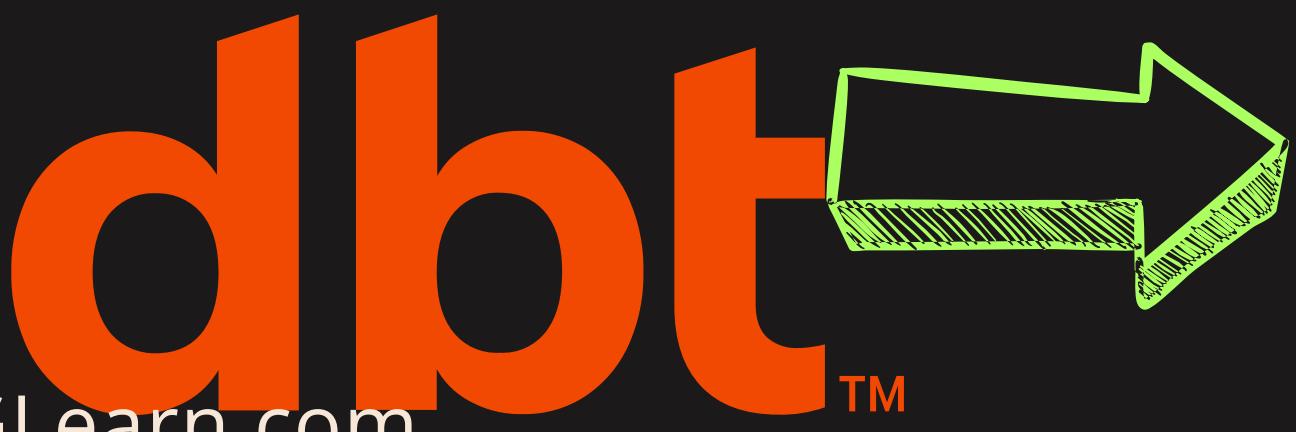
MONITORING AND ALERTING

SETTING UP MONITORING AND ALERTING FOR DATA PIPELINE FAILURES OR DATA QUALITY ISSUES.

USING DBT CLOUD OR THIRD-PARTY TOOLS TO MONITOR PIPELINE RUNS AND SEND ALERTS ON FAILURES.



Shwetank Singh
GritSetGrow - GSGLearn.com



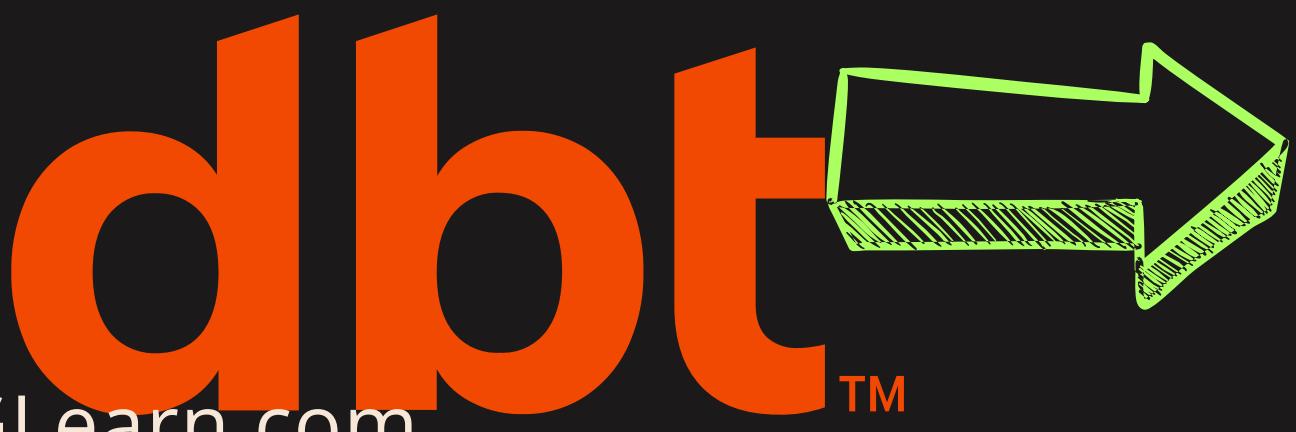
DATA GOVERNANCE

**ENSURING DATA GOVERNANCE BY
DOCUMENTING, TESTING, AND
MONITORING DATA
TRANSFORMATIONS AND LINEAGE.**

**ADDING OWNERSHIP AND
DESCRIPTION TAGS IN SCHEMA.YML
AND GENERATING DOCUMENTATION.**



Shwetank Singh
GritSetGrow - GSGLearn.com



DATA MIGRATION

**USING DBT TO MIGRATE DATA FROM
ONE SYSTEM TO ANOTHER,
ENSURING DATA CONSISTENCY AND
INTEGRITY.**

**WRITING MODELS TO TRANSFORM
AND LOAD DATA INTO A NEW DATA
WAREHOUSE.**



Shwetank Singh
GritSetGrow - GSGLearn.com



INTEGRATING WITH BI TOOLS

CONNECTING DBT-TRANSFORMED DATA TO BI TOOLS LIKE LOOKER, TABLEAU, OR POWER BI FOR VISUALIZATION AND REPORTING.

CREATING VIEWS IN DBT MODELS TO BE USED DIRECTLY IN BI TOOL DASHBOARDS.



Shwetank Singh
GritSetGrow - GSGLearn.com



BUILDING DATA MART

**CREATING A DATA MART FOR A
SPECIFIC BUSINESS UNIT OR
FUNCTION, WITH CURATED DATASETS
AND MODELS.**

**DEVELOPING MODELS TO AGGREGATE
AND PREPARE SALES DATA FOR THE
MARKETING TEAM.**



Shwetank Singh
GritSetGrow - GSGLearn.com

dbt™

A large, stylized orange lowercase word "dbt" with a trademark symbol. To the right of the "t" is a thick, hand-drawn style green arrow pointing to the right. The arrow has a textured, hatched base and a solid black top.

PARAMETERIZING MODELS

USING VARIABLES TO PARAMETERIZE MODELS FOR DYNAMIC BEHAVIOR AND CUSTOMIZATION.

DEFINING A VARIABLE IN DBT_PROJECT.YML AND USING IT IN PARAMETERIZED_MODEL.SQL.



Shwetank Singh
GritSetGrow - GSGLearn.com



PERFORMANCE TUNING

OPTIMIZING DBT MODELS AND SQL QUERIES FOR BETTER PERFORMANCE AND FASTER EXECUTION.

ADDING INDEXES AND OPTIMIZING SQL LOGIC IN PERFORMANCE-SENSITIVE MODELS.



Shwetank Singh
GritSetGrow - GSGLearn.com



DATA LAKE INTEGRATION

USING DBT TO TRANSFORM DATA
STORED IN A DATA LAKE AND LOAD
IT INTO A DATA WAREHOUSE.

CREATING MODELS TO PROCESS AND
TRANSFORM DATA FROM S3 OR
AZURE DATA LAKE.



Shwetank Singh
GritSetGrow - GSGLearn.com

dbt™

A large, stylized orange word "dbt" with a trademark symbol. To the right of the "t" is a thick, hand-drawn style green arrow pointing to the right.

REAL-TIME DATA PROCESSING

IMPLEMENTING REAL-TIME DATA
PROCESSING WITH DBT BY
INTEGRATING IT WITH STREAMING
DATA SOURCES.

SETTING UP A WORKFLOW TO
PROCESS REAL-TIME EVENT DATA
AND UPDATE MODELS FREQUENTLY.



Shwetank Singh
GritSetGrow - GSGLearn.com



COMBINING DATA FROM MULTIPLE SOURCES

MERGING DATA FROM DIFFERENT SOURCES, SUCH AS DATABASES, APIs, AND FILES, INTO A UNIFIED DATASET.

WRITING SQL MODELS TO JOIN DATA FROM DIFFERENT SOURCE TABLES.



Shwetank Singh
GritSetGrow - GSGLearn.com



ADVANCED ANALYTICAL TECHNIQUES

APPLYING ADVANCED ANALYTICAL
TECHNIQUES, SUCH AS MACHINE
LEARNING FEATURE ENGINEERING,
IN DBT MODELS.

CREATING MODELS THAT GENERATE
FEATURES FOR MACHINE LEARNING
ALGORITHMS.



Shwetank Singh
GritSetGrow - GSGLearn.com

dbt™

A large, stylized orange lowercase word "dbt" with a trademark symbol. To the right of the "t" is a thick, black, hand-drawn style arrow pointing to the right. The arrow has a textured, hatched base and a thin, green outline.

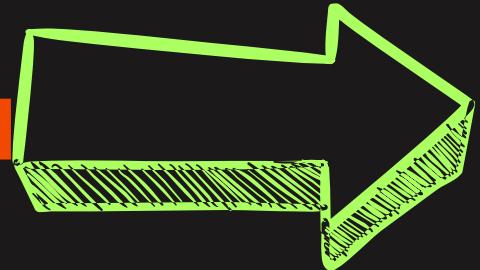
MANAGING LARGE DATASETS

HANDLING LARGE DATASETS
EFFICIENTLY USING PARTITIONING,
CLUSTERING, AND INCREMENTAL
PROCESSING.

IMPLEMENTING PARTITIONED MODELS TO MANAGE LARGE VOLUMES OF DATA.



Shwetank Singh
GritSetGrow - GSGLearn.com

dbt™ 

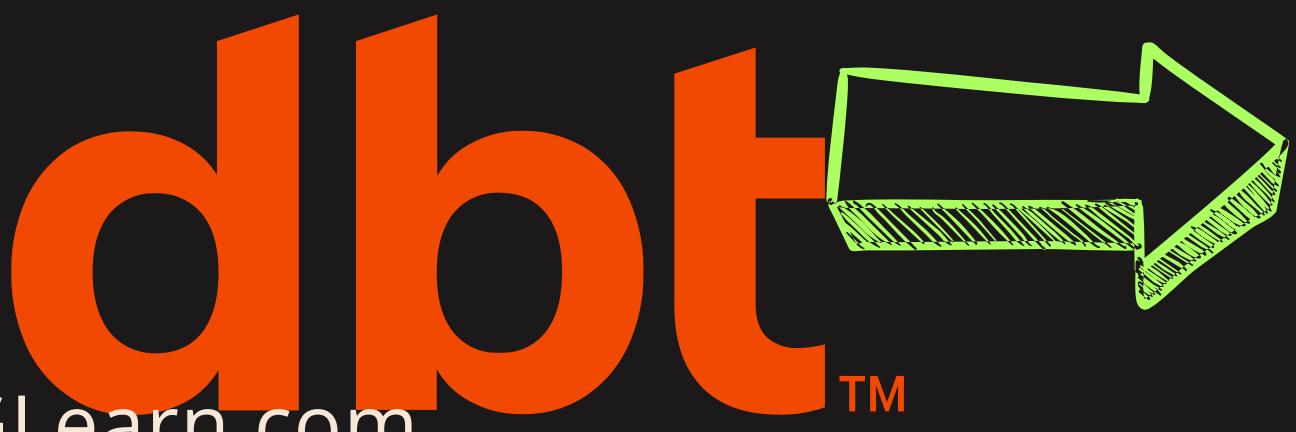
DATA INTEGRATION WORKFLOWS

DESIGNING END-TO-END DATA
INTEGRATION WORKFLOWS USING
DBT AND OTHER ETL TOOLS.

INTEGRATING DBT WITH FIVETRAN
TO AUTOMATE DATA EXTRACTION,
TRANSFORMATION, AND LOADING.



Shwetank Singh
GritSetGrow - GSGLearn.com



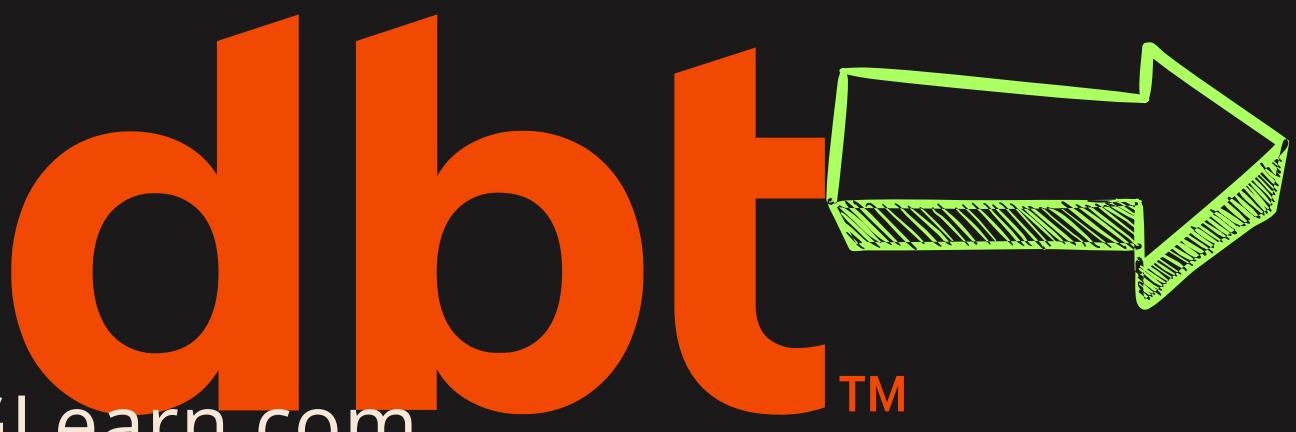
ETL PIPELINE ORCHESTRATION

ORCHESTRATING ETL PIPELINES WITH
DBT AND SCHEDULING TOOLS TO
AUTOMATE DATA PROCESSING
WORKFLOWS.

USING AIRFLOW OR PREFECT TO
MANAGE DBT RUN SCHEDULES AND
DEPENDENCIES.



Shwetank Singh
GritSetGrow - GSGLearn.com



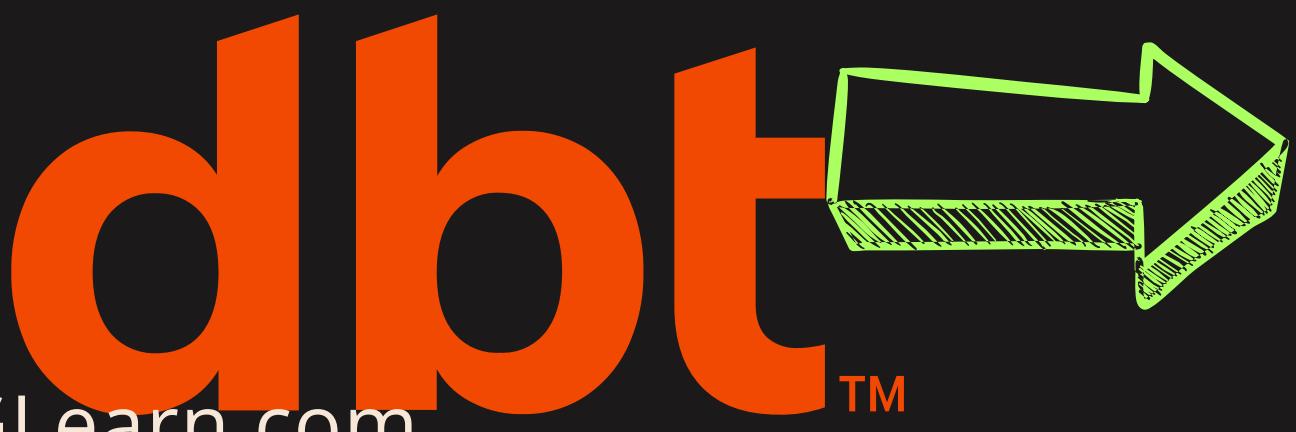
ENHANCING DATA SECURITY

IMPLEMENTING SECURITY MEASURES TO PROTECT SENSITIVE DATA DURING TRANSFORMATIONS.

MASKING SENSITIVE COLUMNS AND USING SECURE CONNECTIONS IN DBT PROFILES.



Shwetank Singh
GritSetGrow - GSGLearn.com



SUPPORTING DATA SCIENCE PROJECTS

PREPARING DATA FOR DATA SCIENCE PROJECTS BY CREATING FEATURE-RICH DATASETS AND ENSURING DATA QUALITY.

BUILDING MODELS TO GENERATE TRAINING DATASETS FOR MACHINE LEARNING MODELS.



Shwetank Singh
GritSetGrow - GSGLearn.com

dbt™

A large, stylized orange lowercase word "dbt" with a trademark symbol. To the right of the "t" is a thick, hand-drawn style green arrow pointing to the right.

CUSTOM LOGGING

**ADDING CUSTOM LOGGING TO TRACK
SPECIFIC EVENTS OR METRICS
DURING DBT RUNS.**

**CREATING A MACRO FOR CUSTOM
LOGGING AND USING IT IN MODELS.**



Shwetank Singh
GritSetGrow - GSGLearn.com



DATA MASKING

APPLYING DATA MASKING
TECHNIQUES TO PROTECT SENSITIVE
INFORMATION IN DBT MODELS.

USING SQL FUNCTIONS TO
ANONYMIZE OR OBFUSCATE
SENSITIVE DATA FIELDS.



Shwetank Singh
GritSetGrow - GSGLearn.com



MULTI-REGION DEPLOYMENTS

SETTING UP DBT PROJECTS TO
DEPLOY MODELS ACROSS MULTIPLE
REGIONS FOR REDUNDANCY AND
PERFORMANCE.

CONFIGURING DIFFERENT PROFILES
AND ENVIRONMENTS FOR EACH
REGION.



Shwetank Singh
GritSetGrow - GSGLearn.com



AUTOMATING DOCUMENTATION UPDATES

USING CI/CD PIPELINES TO AUTOMATE THE GENERATION AND DEPLOYMENT OF DBT DOCUMENTATION.

SETTING UP A CI PIPELINE TO RUN DBT DOCS GENERATE ON EACH COMMIT.



Shwetank Singh
GritSetGrow - GSGLearn.com



HANDLING NULL VALUES

IMPLEMENTING LOGIC TO HANDLE AND REPLACE NULL VALUES IN DATA TRANSFORMATIONS.

USING COALESCE OR IFNULL FUNCTIONS IN SQL MODELS TO HANDLE NULL VALUES.



Shwetank Singh
GritSetGrow - GSGLearn.com

