

kokchun giang

developing in **dbt** to model and transform the data



there are several ways to **materialize** models in dbt

	WHAT IS IT?	WHEN TO USE?
TABLE	physical table	query performance, and when data doesn't change frequently
VIEW	virtual table that runs underlying query each time it is accessed	lightweight models and reflect current data
INCREMENTAL	loading and transforming new or changed data since last run	small data changes, large dataset
EPHEMERAL	temporary model that is not materialized as tables or views	used as subquery that other models can reference

reuse sql code through **jinja macros** in dbt

```
{% macro capitalize_first_letter(column) %}  
case  
  when {{ column }} is null then null  
  else upper(substr({{ column }}, 1, 1)) || lower(substr({{ column }}, 2))  
end  
{% endmacro %}
```

macro definition



```
WITH src_employer AS (SELECT * FROM {{ ref('src_employer') }})
```

```
SELECT  
  {{ capitalize_first_letter("coalesce(workplace_city, 'stad ej specificerat')") }} AS workplace_city  
FROM src_employer
```

use the macro



the compiled sql files are stored in **target** folder

```
target
├── compiled/dbt_code/models
│   ├── dim
│   ├── fct
│   ├── mart
│   └── src
├── run
├── graph_summary.json
├── graph.gpickle
├── manifest.json
├── partial_parse.msgpack
├── run_results.json
└── semantic_manifest.json
```

manifest.json

```
{
  "metadata": {
    "dbt_schema_version": "https://schemas.getdbt.com/dbt/manifest/v12.json",
    "dbt_version": "1.8.4",
    "generated_at": "2024-08-02T08:06:20.538299Z",
    "invocation_id": "8782e079-581d-4e1d-94f7-e5d20138f42f",
    "env": {},
    "project_name": "dbt_code",
    "project_id": "ee957c419bbda82850a7975238360289",
    "user_id": null,
    "send_anonymous_usage_stats": true,
    "adapter_type": "snowflake",
    "nodes": {
      "model.dbt_code.dim_job_details": {
        "database": "job_ads",
        "schema": "warehouse",
        "name": "dim_job_details",
        "resource_type": "model",
        "package_name": "dbt_code",
        "path": "dim/dim_job_details.sql",
        "original_file_path": "models/dim/dim_job_details.sql",
        "unique_id": "model.dbt_code.dim_job_details",
        "fqdn": ["dbt_code", "dim", "dim_job_details"],
        "alias": "dim_job_details",
        "checksum": {
          "name": "sha256",
          "checksum": "8088ee78ded29fe8fe41937f421512975dfef16679d2c26d948cddfe2a92b21e"
        },
        "config": {
          "enabled": true,
          "alias": null,
          "schema": "warehouse",
          "database": null,
          "tags": [],
          "meta": {},
          "group": null,
          "materialized": "view",
          "incremental_strategy": null,
          "persist_docs": {},
          "post_hook": [],
          "pre_hook": [],
          "quoting": {},
          "column_types": {},
          "full_refresh": null,
          "unique_key": null,
          "on_schema_change": "ignore",
          "on_configuration_change": "apply",
          "grants": {},
          "packages": [],
          "docs": {
            "show": true,
            "node_color": null
          },
          "contract": {
            "enforced": false,
            "alias_types": true
          },
          "access": "protected",
          "tags": []
        },
        "description": "",
        "columns": {},
        "meta": {},
        "group": null,
        "docs": {
          "show": true,
          "node_color": null
        },
        "patch_path": null,
        "build_path": null,
        "unrendered_config": {
          "materialized": "view",
          "schema": "warehouse"
        },
        "created_at": 1722245160.8502321,
        "relation_name": "job_ads.warehouse.dim_job_details",
        "raw_code": "WITH src_job_details AS\n(SELECT * FROM {{ ref('src_job_details') }})\n\n{{\n  dbt_utils.generate_surrogate_key(['id','headline'])\n}} AS job_details_id, \n\n  headline,\n  description,\n  description_html_formatted,\n  employment_type,\n  duration,\n  salary_type,\n  scope_of_work_min,\n  scope_of_work_max\nFROM src_job_details",
        "language": "sql",
        "refs": [
          {
            "name": "src_job_details",
            "package": null,
            "version": null
          }
        ],
        "sources": [],
        "metrics": [],
        "depends_on": {
          "macros": [
            "macro.dbt_utils.generate_surrogate_key"
          ],
          "nodes": [
            "model.dbt_code.src_job_details"
          ]
        },
        "compiled_path": null,
        "contract": {
          "enforced": false,
          "alias_types": true,
          "checksum": null
        },
        "access": "protected",
        "constraints": [],
        "version": null,
        "latest_version": null,
        "deprecation_date": null,
        "model.dbt_code.src_job_ads": {
          "database": "job_ads",
          "schema": "warehouse",
          "name": "src_job_ads",
          "resource_type": "model",
          "package_name": "dbt_code",
          "path": "src/src_job_ads.sql",
          "original_file_path": "models/src/src_job_ads.sql",
          "unique_id": "model.dbt_code.src_job_ads",
          "fqdn": ["dbt_code", "src", "src_job_ads"],
          "alias": "src_job_ads",
          "checksum": {
            "name": "sha256",
            "checksum": "8088ee78ded29fe8fe41937f421512975dfef16679d2c26d948cddfe2a92b21e"
          },
          "config": {
            "enabled": true,
            "alias": null,
            "schema": "warehouse",
            "database": null,
            "tags": [],
            "meta": {},
            "group": null,
            "materialized": "view",
            "incremental_strategy": null,
            "persist_docs": {},
            "post_hook": [],
            "pre_hook": [],
            "quoting": {},
            "column_types": {},
            "full_refresh": null,
            "unique_key": null,
            "on_schema_change": "ignore",
            "on_configuration_change": "apply",
            "grants": {},
            "packages": [],
            "docs": {
              "show": true,
              "node_color": null
            },
            "contract": {
              "enforced": false,
              "alias_types": true
            },
            "access": "protected",
            "tags": []
          },
          "description": "",
          "columns": {},
          "meta": {},
          "group": null,
          "docs": {
            "show": true,
            "node_color": null
          },
          "patch_path": null,
          "build_path": null,
          "unrendered_config": {
            "materialized": "view",
            "schema": "warehouse"
          },
          "created_at": 1722245160.8502321,
          "relation_name": "job_ads.warehouse.src_job_ads",
          "raw_code": "WITH src_job_details AS\n(SELECT * FROM {{ ref('src_job_details') }})\n\n{{\n  dbt_utils.generate_surrogate_key(['id','headline'])\n}} AS job_details_id, \n\n  headline,\n  description,\n  description_html_formatted,\n  employment_type,\n  duration,\n  salary_type,\n  scope_of_work_min,\n  scope_of_work_max\nFROM src_job_details",
          "language": "sql",
          "refs": [
            {
              "name": "src_job_details",
              "package": null,
              "version": null
            }
          ],
          "sources": [],
          "metrics": [],
          "depends_on": {
            "macros": [
              "macro.dbt_utils.generate_surrogate_key"
            ],
            "nodes": [
              "model.dbt_code.src_job_details"
            ]
          },
          "compiled_path": null,
          "contract": {
            "enforced": false,
            "alias_types": true,
            "checksum": null
          },
          "access": "protected",
          "constraints": [],
          "version": null,
          "latest_version": null,
          "deprecation_date": null
        }
      }
    }
  }
}
```

comprehensive representation of the dbt project

dbt power user is an extension in vscode to simplify dbt development

Preview 67 rows in 4.6s SQL History Bookmarks + New query Clear results Open in Tab Help

query result 67 x 4

COLUMN_NAME	DATA_TYPE	TABLE_SCHEMA	TABLE_NAME
OCCUPATION_FIELD__CONCEPT_ID	TEXT	STAGING	DATA_FIELD_JOB_ADS
APPLICATION_DETAILS__VIA_AF	BOOLEAN	STAGING	DATA_FIELD_JOB_ADS
WORKING_HOURS_TYPE__LEGACY_AMS_TAXONOMY_ID	TEXT	STAGING	DATA_FIELD_JOB_ADS
SALARY_TYPE__LEGACY_AMS_TAXONOMY_ID	TEXT	STAGING	DATA_FIELD_JOB_ADS
EXPERIENCE_REQUIRED	BOOLEAN	STAGING	DATA_FIELD_JOB_ADS
WEBPAGE_URL	TEXT	STAGING	DATA_FIELD_JOB_ADS
WORKPLACE_ADDRESS__COUNTRY_CODE	TEXT	STAGING	DATA_FIELD_JOB_ADS
WORKPLACE_ADDRESS__CITY	TEXT	STAGING	DATA_FIELD_JOB_ADS
OCCUPATION__LABEL	TEXT	STAGING	DATA_FIELD_JOB_ADS
_DLT_ID	TEXT	STAGING	DATA_FIELD_JOB_ADS
DRIVING_LICENSE_REQUIRED	BOOLEAN	STAGING	DATA_FIELD_JOB_ADS
WORKPLACE_ADDRESS__MUNICIPALITY_CODE	TEXT	STAGING	DATA_FIELD_JOB_ADS
WORKING_HOURS_TYPE__CONCEPT_ID	TEXT	STAGING	DATA_FIELD_JOB_ADS
HEADLINE	TEXT	STAGING	DATA_FIELD_JOB_ADS
SALARY TYPE LABEL	TEXT	STAGING	DATA FIELD JOB ADS

can see the query directly in vscode