Data Warehouse Data Dictionary

Overview

This data dictionary documents all tables, columns, and relationships in the DuckDB data warehouse. The schema follows a dimensional modeling approach with fact and dimension tables to support analysis of job skills, technical competencies, and labor market data.

Fact Tables

factJobSkills

Purpose: Central fact table storing the relationship between jobs, skills, and technical competencies. Each row represents a mapping of a specific skill or technical skill to a job classification.

Primary Key: ID

Column Name	Data Type	Nullable	Description
ID	INTEGER	NO	Unique identifier for each job-skill relationship record. Primary key.
ONET_SOC_CODE	TEXT	NO	O*NET Standard Occupational Classification code that uniquely identifies a job category. Links to dimension tables.
Skill_Element	INTEGER	YES	Identifier for a general skill element associated with the job. Foreign key reference to dimSkills.
Skill_Name	TEXT	YES	Human-readable name of the skill element.
Tech_Skill_Key	INTEGER	YES	Unique identifier for a technical skill. Foreign key reference to dimTechSkills.
Tech_Skill_Name	TEXT	YES	Human-readable name of the technical skill (e.g., "Python", "SQL").
Job_Title	INTEGER	YES	Identifier for the job title. Foreign key reference to dimJobInfo.
created_at	TIMESTAMP	YES	System-generated timestamp recording when the record was inserted. Defaults to current timestamp.

Relationships:

- ONET_SOC_CODE → dimJobInfo.ONET_SOC_Code

- ONET_SOC_CODE → dimTechSkills.ONET_SOC_Code
- ONET_SOC_CODE → dimSkills.ONET_SOC_Code

Dimension Tables

dimJobInfo

Purpose: Dimension table containing detailed information about job classifications. Provides descriptive attributes for jobs identified by O*NET SOC codes.

Primary Key: ONET_SOC_Code (implicit)

Column Name	Data Type	Nullable	Description
ONET_SOC_Code	TEXT	YES	O*NET Standard Occupational Classification code. Unique identifier for a job classification.
Title	TEXT	YES	Job title or occupation name (e.g., "Software Developer", "Data Analyst").
Description	TEXT	YES	Detailed description of the job, responsibilities, and typical work activities.
created_at	TIMESTAMP	YES	System-generated timestamp recording when the record was inserted.

Usage: Use this table to look up job titles and descriptions by O*NET code. Typically joined with fact tables to add context to skill analysis.

---

dimTechSkills

Purpose: Dimension table listing technical skills, their classifications, and demand indicators. Provides a curated reference for technology-related competencies mapped to jobs.

Primary Key: ONET_SOC_Code + Tech_Skill_Key (implicit composite key)

Column Name	Data Type	Nullable	Description
ONET_SOC_Code	TEXT	YES	O*NET Standard Occupational Classification code associated with the technical skill.
Example	TEXT	YES	Example or use case demonstrating the technical skill in context.
Commodity_Code	INTEGER	YES	Industry or commodity classification code categorizing the technical skill.

Commodity_Title	TEXT	YES	Human-readable title or name of the commodity classification.
Hot_Tech	TEXT	YES	Indicator (yes/no or boolean) marking the skill as emerging or trending technology.
In_Demand	TEXT	YES	Indicator (yes/no or boolean) marking the skill as currently high-demand in the job market.
created_at	TIMESTAMP	YES	System-generated timestamp recording when the record was inserted.

Usage: Use this table to identify which technical skills are in demand, emerging, or associated with specific job categories.

---

dimSkills

Purpose: Dimension table containing general skills (soft skills, core competencies) with proficiency or importance scale ratings. Provides skill level assessments tied to specific jobs.
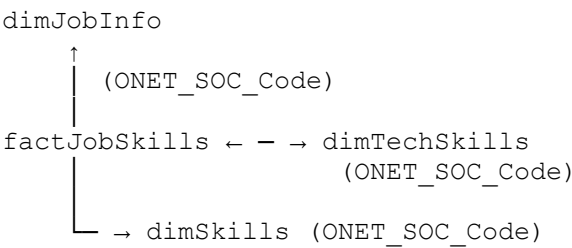
Primary Key: ONET_SOC_Code + Element_ID (implicit composite key)

Column Name	Data Type	Nullable	Description
ONET_SOC_Code	TEXT	YES	O*NET Standard Occupational Classification code for the job associated with the skill.
Element_ID	TEXT	YES	Unique identifier for the skill element or competency.
Skill	TEXT	YES	Name or description of the skill (e.g., "Problem Solving", "Communication").
Scale_ID	TEXT	YES	Identifier indicating the scale or rating system used to measure skill importance (e.g., "1-5", "low-medium-high").
Data Value	REAL	YES	Numeric value representing the skill's importance, proficiency level, or rating on the specified scale.
created_at	TIMESTAMP	YES	System-generated timestamp recording when the record was inserted.

Usage: Use this table to analyze required skills and their importance levels across different job classifications.

Data Relationships

Entity Relationship Overview

```
dimJobInfo
    ↑
    │  (ONET_SOC_Code)
    │
factJobSkills ← ─ → dimTechSkills
    │                   (ONET_SOC_Code)
    │
    └─ → dimSkills (ONET_SOC_Code)
```

All dimension tables are connected to the central fact table through the **ONET_SOC_Code**, which serves as the primary joining key across the schema.

Key Identifiers and Codes

O*NET SOC Code (ONET_SOC_Code)

- **Source:** U.S. Department of Labor O*NET database
- **Format:** Typically an alphanumeric code (e.g., "15-1132.00")
- **Purpose:** Standard occupational classification used across all tables for job identification
- **Usage:** Primary join key across all tables

Element_ID

- **Purpose:** Identifies specific skill elements or competencies
- **Format:** String identifier
- **Scope:** Local to dimSkills table

Commodity_Code & Commodity_Title

- **Purpose:** Classification system for grouping technical skills by industry or technology category
- **Format:** Integer for code, text for title
- **Scope:** Local to dimTechSkills table

Timestamp Conventions

All tables include a `created_at` column that captures when records are inserted into the warehouse. This supports:

- **Data lineage tracking:** Identifying when data was loaded

- **Historical analysis:** Understanding data warehouse growth
- **Data quality monitoring:** Detecting unexpected data loads or gaps

Default Behavior: created_at automatically defaults to the current timestamp if not explicitly provided during insertion.

Data Quality Notes

- **Nullable Columns:** Most columns allow NULL values, which may indicate missing source data or incomplete mappings
- **Primary Keys:** factJobSkills has an explicit PRIMARY KEY; dimension tables have implicit or composite keys
- **Relationships:** No explicit foreign key constraints are defined in the schema; enforce referential integrity at the application layer
- **Scale Values:** The "Data Value" column in dimSkills contains numeric ratings; verify scale definitions during analysis

---

Common Query Patterns

Find all technical skills for a job

```
SELECT DISTINCT Tech_Skill_Name, In_Demand, Hot_Tech
FROM factJobSkills
WHERE ONET_SOC_Code = '15-1132.00'
ORDER BY In_Demand DESC;
```

Find job details and required skills

```
SELECT j.Title, j.Description, s.Skill, s.Data_Value
FROM dimJobInfo j
JOIN dimSkills s ON j.ONET_SOC_Code = s.ONET_SOC_Code
WHERE j.ONET_SOC_Code = '15-1132.00'
ORDER BY s.Data_Value DESC;
```

Identify in-demand technical skills across all jobs

```
SELECT DISTINCT Tech_Skill_Name, Commodity_Title, COUNT(*) as Job_Count
FROM factJobSkills
WHERE In_Demand = 'yes'
GROUP BY Tech_Skill_Name, Commodity_Title
ORDER BY Job_Count DESC;
```

Version History

Version	Date	Changes
1.0	Current	Initial schema documentation

Last Updated: 10/15/2025

Contact: Kevin