**DATABASE SCHEMA DESIGNS**

### **1. Raw Staging Schema**

This schema will store the raw data as-is, as it is downloaded from the SEC filings, without any transformations. The main purpose is to retain the integrity of the raw data for future processing or reprocessing.

#### **Schema Design:**

* **Schema Name:** RAW\_STAGING
* **Tables:**
  + num\_table: Stores numeric data related to the financials.
  + sub\_table: Stores submission-related metadata.
  + tag\_table: Stores tags related to financial data (i.e., the financial concepts like "Assets", "Liabilities", etc.).
  + pre\_table: Stores presentation data (i.e., data related to the report structure).

#### **Table Designs:**

1. **RAW\_STAGING.num\_table**

Stores the numeric data, e.g., financial values such as revenue, expenses, etc.

|  |  |  |
| --- | --- | --- |
| **Column Name** | **Data Type** | **Description** |
| adsh | VARCHAR(20) | Unique identifier for the filing |
| tag | VARCHAR(50) | Financial tag (e.g., "Assets", "Revenue") |
| ddate | DATE | Date of the financial report |
| value | DECIMAL(18,2) | Financial value associated with the tag |
| qtrs | INTEGER | Number of quarters reported (if applicable) |
| uom | VARCHAR(50) | Unit of measure (e.g., USD, Percentage) |
| segments | VARCHAR(500) | Business segments information (optional) |
| coreg | BOOLEAN | Co-regulatory filing indicator (if applicable) |
| footnote | VARCHAR(512) | Footnote information (optional) |

1. **RAW\_STAGING.sub\_table**

Stores submission-related metadata, including company identifiers.

|  |  |  |
| --- | --- | --- |
| **Column Name** | **Data Type** | **Description** |
| adsh | VARCHAR(20) | Unique identifier for the filing |
| cik | INTEGER | Central Index Key for the company |
| name | VARCHAR(255) | Company name |
| sic | INTEGER | Standard Industry Classification code |
| countryba | VARCHAR(50) | Country of business address |
| stprba | VARCHAR(50) | State or province of business address |
| cityba | VARCHAR(50) | City of business address |
| zipba | VARCHAR(10) | Postal code of business address |
| countryma | VARCHAR(50) | Country of mailing address |
| stprma | VARCHAR(50) | State or province of mailing address |
| cityma | VARCHAR(50) | City of mailing address |
| zipma | VARCHAR(10) | Postal code of mailing address |

1. **RAW\_STAGING.tag\_table**

Stores tags that help define the financial data concepts.

|  |  |  |
| --- | --- | --- |
| **Column Name** | **Data Type** | **Description** |
| tag | VARCHAR(50) | Financial tag (e.g., "Revenue", "Assets") |
| datatype | VARCHAR(20) | Data type for the tag (e.g., "number") |
| tlabel | VARCHAR(255) | Label associated with the tag |

1. **RAW\_STAGING.pre\_table**

Stores presentation-related information about the report.

|  |  |  |
| --- | --- | --- |
| **Column Name** | **Data Type** | **Description** |
| adsh | VARCHAR(20) | Unique identifier for the filing |
| report | VARCHAR(50) | Report name (e.g., "10-K", "10-Q") |
| line | INTEGER | Line number in the presentation |
| stmt | VARCHAR(50) | Statement type (e.g., "BalanceSheet") |

### **2. JSON Transformation Schema**

In this schema, we will store the SEC data in a denormalized JSON format for easy querying, analysis, and faster data access. The JSON structure makes it easier to access the data programmatically and enables flexible querying through JSON functions in Snowflake.

#### **Schema Design:**

* **Schema Name:** JSON\_TRANSFORMED
* **Tables:**
  + json\_balance\_sheet: Stores the denormalized balance sheet data in JSON format.
  + json\_income\_statement: Stores the denormalized income statement data in JSON format.
  + json\_cash\_flow: Stores the denormalized cash flow data in JSON format.

#### **Table Designs:**

1. **JSON\_TRANSFORMED.json\_balance\_sheet**

Stores balance sheet data as JSON, including company identifiers, financial period, and key financial figures.

|  |  |  |
| --- | --- | --- |
| **Column Name** | **Data Type** | **Description** |
| adsh | VARCHAR(20) | Unique filing identifier |
| cik | INTEGER | Central Index Key (Company ID) |
| company\_name | VARCHAR(255) | Company Name |
| report\_date | DATE | Date of the report |
| assets | VARIANT | JSON data for assets (denormalized) |
| liabilities | VARIANT | JSON data for liabilities (denormalized) |
| equity | VARIANT | JSON data for equity (denormalized) |

1. **JSON\_TRANSFORMED.json\_income\_statement**

Stores income statement data as JSON, including revenue, expenses, and net income.

|  |  |  |
| --- | --- | --- |
| **Column Name** | **Data Type** | **Description** |
| adsh | VARCHAR(20) | Unique filing identifier |
| cik | INTEGER | Central Index Key (Company ID) |
| company\_name | VARCHAR(255) | Company Name |
| report\_date | DATE | Date of the report |
| revenue | DECIMAL(18,2) | Revenue for the period |
| net\_income | DECIMAL(18,2) | Net income for the period |
| expenses | VARIANT | JSON data for expenses (denormalized) |

1. **JSON\_TRANSFORMED.json\_cash\_flow**

Stores cash flow data as JSON, including operating, investing, and financing cash flows.

|  |  |  |
| --- | --- | --- |
| **Column Name** | **Data Type** | **Description** |
| adsh | VARCHAR(20) | Unique filing identifier |
| cik | INTEGER | Central Index Key (Company ID) |
| company\_name | VARCHAR(255) | Company Name |
| report\_date | DATE | Date of the report |
| operating\_cash\_flow | DECIMAL(18,2) | Operating cash flow |
| investing\_cash\_flow | DECIMAL(18,2) | Investing cash flow |
| financing\_cash\_flow | DECIMAL(18,2) | Financing cash flow |

### **3. Denormalized Fact Tables Schema**

This schema will store the data in **three denormalized fact tables**: Balance Sheet, Income Statement, and Cash Flow. These tables consolidate the data for efficient querying by key identifiers (company, fiscal period, account).

#### **Schema Design:**

* **Schema Name:** DENORMALIZED\_FACT\_TABLES
* **Tables:**
  + balance\_sheet\_fact: Contains consolidated balance sheet data.
  + income\_statement\_fact: Contains consolidated income statement data.
  + cash\_flow\_fact: Contains consolidated cash flow data.

#### **Table Designs:**

1. **DENORMALIZED\_FACT\_TABLES.balance\_sheet\_fact**

Stores consolidated balance sheet data with important company and period identifiers.

|  |  |  |
| --- | --- | --- |
| **Column Name** | **Data Type** | **Description** |
| adsh | VARCHAR(20) | Filing identifier |
| cik | INTEGER | Central Index Key (Company ID) |
| company\_name | VARCHAR(255) | Company Name |
| report\_date | DATE | Report date |
| fiscal\_year | INTEGER | Fiscal Year |
| fiscal\_period | VARCHAR(2) | Fiscal period (Q1, Q2, etc.) |
| assets | DECIMAL(18,2) | Total assets |
| liabilities | DECIMAL(18,2) | Total liabilities |
| equity | DECIMAL(18,2) | Total equity |

1. **DENORMALIZED\_FACT\_TABLES.income\_statement\_fact**

Consolidated income statement data.

|  |  |  |
| --- | --- | --- |
| **Column Name** | **Data Type** | **Description** |
| adsh | VARCHAR(20) | Filing identifier |
| cik | INTEGER | Central Index Key (Company ID) |
| company\_name | VARCHAR(255) | Company Name |
| report\_date | DATE | Report date |
| revenue | DECIMAL(18,2) | Total revenue |
| net\_income | DECIMAL(18,2) | Net income |
| operating\_income | DECIMAL(18,2) | Operating income |
| expenses | DECIMAL(18,2) | Total expenses |

1. **DENORMALIZED\_FACT\_TABLES.cash\_flow\_fact**

Consolidated cash flow data.

|  |  |  |
| --- | --- | --- |
| **Column Name** | **Data Type** | **Description** |
| adsh | VARCHAR(20) | Filing identifier |
| cik | INTEGER | Central Index Key (Company ID) |
| company\_name | VARCHAR(255) | Company Name |
| report\_date | DATE | Report date |
| operating\_cash\_flow | DECIMAL(18,2) | Operating cash flow |
| investing\_cash\_flow | DECIMAL(18,2) | Investing cash flow |
| financing\_cash\_flow | DECIMAL(18,2) | Financing cash flow |