Designing Advanced Data Architecture for Business Intelligence
Mini Project
IMDb Midterm Team Project

Team 6:
Abisha Vadukoot, NUID: 002371769
Bhagyashri Pagar, NUID: 002310690
Rahul Bothra, NUID: 002378790

STAGING SCHEME

```
-- Table for staging.name.basics.tsv
CREATE OR REPLACE TABLE STAGING_NAME_BASICS (
    NCONST VARCHAR(25),
    PRIMARYNAME VARCHAR(200),
    BIRTHYEAR NUMERIC(4),
    DEATHYEAR NUMERIC(4),
    PRIMARYPROFESSION VARCHAR(200),
    KNOWNFORTITLES VARCHAR(200),
    DI_JOB_ID VARCHAR(50),
    DI_LOAD_DT DATE
);

-- Table for staging.title.akas.tsv
CREATE OR REPLACE TABLE STAGING_TITLE_AKAS (
    TITLEID VARCHAR(50),
    ORDERING INTEGER,
    TITLE VARCHAR(300),
    REGION VARCHAR(50),
    LANGUAGE VARCHAR(30),
    TYPES VARCHAR(20),
    ATTRIBUTES VARCHAR(100),
    ISORIGINALTITLE NUMERIC(2),
    DI_JOB_ID VARCHAR(50),
    DI_LOAD_DT DATE
);

-- Table for staging.title.basics.tsv
CREATE OR REPLACE TABLE STAGING_TITLE_BASICS (
    TCONST VARCHAR(25),
    TITLETYPE VARCHAR(50),
```

```sql
    PRIMARYTITLE VARCHAR(350),
    ORIGINALTITLE VARCHAR(350),
    ISADULT NUMERIC(2),
    STARTYEAR NUMERIC(4),
    ENDYEAR NUMERIC(4),
    RUNTIMEMINUTES NUMERIC(20),
    GENRES VARCHAR(50),
    DI_JOB_ID VARCHAR(50),
    DI_LOAD_DT DATE
);

-- Table for staging.title.ratings.tsv
CREATE OR REPLACE TABLE STAGING_TITLE_RATINGS (
    TCONST VARCHAR(20),
    AVERAGERATING FLOAT(10),
    NUMVOTES NUMERIC(15),
    DI_JOB_ID VARCHAR(50),
    DI_LOAD_DT DATE
);

-- Table for staging.title.episodes.tsv
CREATE OR REPLACE TABLE STAGING_TITLE_EPISODES (
    TCONST VARCHAR(20),
    PARENTTCONST VARCHAR(20),
    SEASONNUMBER NUMERIC(5),
    EPISODENUMBER NUMERIC(10),
    DI_JOB_ID VARCHAR(50),
    DI_LOAD_DT DATE
);

-- Table for staging.title.crew.tsv
CREATE OR REPLACE TABLE STAGING_TITLE_CREW (
    TCONST VARCHAR(25),
    DIRECTORS VARCHAR(500),
    WRITERS VARCHAR(500),
    DI_JOB_ID VARCHAR(50),
    DI_LOAD_DT DATE
);

-- Table for staging.title.principals.tsv
CREATE OR REPLACE TABLE STAGING_TITLE_PRINCIPALS (
    TCONST VARCHAR(25),
    ORDERING NUMERIC(5),
    NCONST VARCHAR(15),
```

```sql
    CATEGORY VARCHAR(50),
    JOB VARCHAR(500),
    CHARACTERS VARCHAR(500),
    DI_JOB_ID VARCHAR(50),
    DI_LOAD_DT DATE
);

-- Table for staging.region_codes
CREATE OR REPLACE TABLE STAGING_REGION_CODES (
    REGION_ABBR VARCHAR(5),
    REGION_NAME VARCHAR(100),
    DI_JOB_ID VARCHAR(50),
    DI_LOAD_DT DATE
);

-- Table for staging.LANGUAGE_codes
CREATE OR REPLACE TABLE STAGING_LANGUAGE_CODES (
    LANGUAGE_NAME VARCHAR(100),
    LANGUAGE_ABBR VARCHAR(5),
    DI_JOB_ID VARCHAR(50),
    DI_LOAD_DT DATE
);
```

DIMENSIONAL MODELING SCHEMA

```sql
--
-- ER/Studio Data Architect SQL Code Generation
-- Project :    IMDB_Dimesional_Model.DM1
--
-- Date Created : Monday, March 10, 2025 23:28:06
-- Target DBMS : Snowflake
--


--
-- TABLE: BRIDGE_PERSONS_PROFESSION
--

USE SCHEMA GROUP_6_MIDTERM_PROJECT_EDW;

CREATE TABLE BRIDGE_PERSONS_PROFESSION(
    PERSON_SK           NUMBER(20, 0)   NOT NULL,
    PROFESSION_SK       NUMBER(20, 0)   NOT NULL,
```

```sql
    IS_CURRENT            NUMBER(2, 0),
    START_DATE           DATE,
    END_DATE             DATE,
    PEROSN_PROFESSION_DK   VARCHAR(50)      NOT NULL,
    JOB_ID               VARCHAR(50),
    LOAD_DT              DATE,
    CONSTRAINT PK15 PRIMARY KEY (PERSON_SK, PROFESSION_SK)
)
;



--
-- TABLE: BRIDGE_TITLE_AKAS
--

CREATE TABLE BRIDGE_TITLE_AKAS(
    TITLE_SK         NUMBER(20, 0)   NOT NULL,
    REGION_SK        NUMBER(20, 0)   NOT NULL,
    LANGUAGE_SK      NUMBER(20, 0)   NOT NULL,
    AKA_TITLE        VARCHAR(100),
    IS_ORIGINAL_TITLE   NUMBER(2, 0),
    CONSTRAINT PK13 PRIMARY KEY (TITLE_SK, REGION_SK, LANGUAGE_SK)
)
;



--
-- TABLE: BRIDGE_TITLE_GENRES
--

CREATE TABLE BRIDGE_TITLE_GENRES(
    TITLE_SK   NUMBER(20, 0)   NOT NULL,
    GENRE_SK   NUMBER(20, 0)   NOT NULL,
    CONSTRAINT PK9 PRIMARY KEY (TITLE_SK, GENRE_SK)
)
;



--
-- TABLE: BRIDGE_TITLE_LANGUAGE
--
```

```sql
CREATE TABLE BRIDGE_TITLE_LANGUAGE(
    TITLE_SK      NUMBER(20, 0)   NOT NULL,
    LANGUAGE_SK   NUMBER(20, 0)   NOT NULL,
    CONSTRAINT PK11 PRIMARY KEY (TITLE_SK, LANGUAGE_SK)
)
;




--
-- TABLE: BRIDGE_TITLE_REGIONS
--

CREATE TABLE BRIDGE_TITLE_REGIONS(
    TITLE_SK    NUMBER(20, 0)   NOT NULL,
    REGION_SK   NUMBER(20, 0)   NOT NULL,
    CONSTRAINT PK10 PRIMARY KEY (TITLE_SK, REGION_SK)
)
;




--
-- TABLE: DIM_GENRES
--

CREATE TABLE DIM_GENRES(
    GENRE_SK      NUMBER(20, 0)   NOT NULL,
    GENRE_NAME    VARCHAR(30),
    JOB_ID        VARCHAR(50),
    LOAD_DT       DATE,
    CONSTRAINT PK_GENRE_SK PRIMARY KEY (GENRE_SK)
)
;




--
-- TABLE: DIM_LANGUAGES
--

CREATE OR replace TABLE DIM_LANGUAGES(
    LANGUAGE_SK      NUMBER(20, 0)   NOT NULL,
```

```
    LANGUAGE_NAME    VARCHAR(100),
    LANGUAGE_ABBR    VARCHAR(5),
    JOB_ID           VARCHAR(50),
    LOAD_DT          DATE,
    CONSTRAINT PK_LANGUAGE_SK PRIMARY KEY (LANGUAGE_SK)
)
;



--
-- TABLE: DIM_PERSONS
--

CREATE OR REPLACE TABLE DIM_PERSONS(
    PERSON_SK        NUMBER(20, 0)   NOT NULL,
    NCONST           VARCHAR(30)     NOT NULL,
    PRIMARY_NAME     VARCHAR(200),
    BIRTH_YEAR       NUMBER(4, 0),
    DEATH_YEAR       NUMBER(4, 0),
    KNOWN_TITLE      VARCHAR(200),
    IS_CURRENT       NUMBER(2, 0),
    START_DATE       DATE,
    END_DATE         DATE,
    PERSON_DK        VARCHAR(50) NOT NULL,
    JOB_ID           VARCHAR(50),
    LOAD_DT          DATE,
    CONSTRAINT PK_PERSON_SK PRIMARY KEY (PERSON_SK)
)
;



--
-- TABLE: DIM_PROFESSION
--

CREATE OR REPLACE TABLE DIM_PROFESSION(
    PROFESSION_SK      NUMBER(20, 0)   NOT NULL,
    PROFESSION_NAME    VARCHAR(50),
    IS_CURRENT         NUMBER(2, 0),
    START_DATE         DATE,
    END_DATE           DATE,
    PROFESSION_DK      VARCHAR(50) NOT NULL,
```

```
   JOB_ID          VARCHAR(50),
   LOAD_DT         DATE,
   CONSTRAINT PK_PROFESSION_SK PRIMARY KEY (PROFESSION_SK)
)
;




--
-- TABLE: DIM_REGIONS
--

CREATE OR REPLACE TABLE DIM_REGIONS(
   REGION_SK     NUMBER(20, 0)   NOT NULL,
   REGION_NAME    VARCHAR(50),
   REGION_ABBR    VARCHAR(5),
   JOB_ID        VARCHAR(50),
   LOAD_DT       DATE,
   CONSTRAINT PK_REGION_SK PRIMARY KEY (REGION_SK)
)
;




--
-- TABLE: DIM_TITLES
--

CREATE OR REPLACE TABLE DIM_TITLES (
   TITLE_SK NUMBER(22,0) PRIMARY KEY,
   TCONST VARCHAR(20) NOT NULL,
   PRIMARY_TITLE VARCHAR(255),
   ORIGINAL_TITLE VARCHAR(255),
   TITLE_TYPE VARCHAR(50),
   IS_ADULT NUMERIC(2,0),
   JOB_ID VARCHAR(50),
   LOAD_DT DATE
);




--
-- TABLE: FACT_EPISODES
```

```sql
--

CREATE TABLE FACT_EPISODES(
    EPISODE_SK        NUMBER(20, 0)   NOT NULL,
    SEASON_NUMBER     INTEGER,
    EPISODE_NUMBER    INTEGER,
    NUM_EPISODES      INTEGER,
    TITLE_SK          NUMBER(20, 0)   NOT NULL,
    PARENT_TITLE_SK   NUMBER(20, 0),
    JOB_ID            VARCHAR(50),
    LOAD_DT           DATE,
    CONSTRAINT PK8 PRIMARY KEY (EPISODE_SK)
)
;



--
-- TABLE: FACT_MOVIES
--

CREATE OR REPLACE TABLE FACT_MOVIES(
    MOVIE_SK          VARCHAR(50)   NOT NULL,
    START_YEAR        NUMBER(4, 0),
    END_YEAR          NUMBER(4, 0),
    RUNTIME_MINUTES   INTEGER,
    IS_ADULT          NUMBER(2, 0),
    TITLE_SK          NUMBER(20, 0)   NOT NULL,
    GENRE_SK          NUMBER(20, 0),
    REGION_SK         NUMBER(20, 0),
    LANGUAGE_SK       NUMBER(20, 0),
    JOB_ID            VARCHAR(50),
    LOAD_DT           DATE,
    CONSTRAINT PK_MOVIE_SK PRIMARY KEY (MOVIE_SK)
)
;



--
-- TABLE: FACT_PERSONS
--

CREATE OR REPLACE TABLE FACT_PERSONS(
```

```
    FACT_PERSON_SK    VARCHAR(100)    NOT NULL,
    PERSON_SK        NUMBER(20, 0)    NOT NULL,
    CATEGORY         VARCHAR(200),
    JOB            VARCHAR(200),
    CHARACTER        VARCHAR(200),
    ORDERING         NUMBER(10, 0),
    TITLE_SK        NUMBER(20, 0)    NULL,
    PROFESSION_SK    NUMBER(20, 0),
    JOB_ID          VARCHAR(50),
    LOAD_DT          DATE,
    CONSTRAINT PK_FCT_PERSONS_SK PRIMARY KEY (FACT_PERSON_SK,
PERSON_SK)
)
```