Phase 0: Initialization

Start ETL

Establish MySQL Connection

Connection Successful? — No → Log Error & Exit → End ETL

Yes

Disable Database Constraints

Phase 1: Read Shared Data

title.basics.tsv.gz

Read Title Basics

df_basics (kept in memory)

Phase 2: Dimension Tables

Truncate Dim_Time

Generate Time Dimension

Bulk Insert into Dim_Time

Dim_Time ✓

Truncate Dim_Genre

Extract & Deduplicate Genres from df_basics

Bulk Insert into Dim_Genre

Dim_Genre ✓

Truncate Dim_Person

name.basics.tsv.gz

Read Name Basics

title.crew.tsv.gz

Read Crew Data & Extract Missing IDs

title.principals.tsv.gz

Read Principals & Extract Missing IDs

Build Person Records

Bulk Insert into Dim_Person

Dim_Person ✓

df_names (kept in memory)

Truncate Dim_Title

Parse Title Data from df_basics

Bulk Insert into Dim_Title

Dim_Title ✓

Phase 3: Dependent Dimensions & Bridge Tables

Truncate Bridge_Title_Genre

Query Dim_Genre for Keys

Parse Title Genres from df_basics

Build Title-Genre Relationships

Bulk Insert into Bridge_Title_Genre

Bridge_Title_Genre ✓

Truncate Dim_Episode

title.episode.tsv.gz

Read Episode Data

Query Valid Title IDs

Episode tconst Valid? — No → Skip record (loop to next)

Yes

Parent tconst Valid? — Yes → Use parentTconst value / No → Set parentTconst = NULL

Loop Done

Bulk Insert into Dim_Episode

Dim_Episode ✓

yellow-outlined shape sections are the loops for dealing with referential integrity constraints (loop condition is just simplified to "Loop Done" to save space in the diagram)

Truncate Bridge_Person_KnownFor

Use Stored df_names

Query Valid Title IDs

Parse knownForTitles

Title Valid? — No → Skip record (loop to next)

Yes

Add (nconst, tconst) to data

Loop Done

Bulk Insert into Bridge_Person_KnownFor

Bridge_Person_KnownFor ✓

Truncate Bridge_Title_Person

Query Valid IDs

Initialize Unique Records Dictionary

title.crew.tsv.gz

Read Crew Data

Process Directors & Writers

title.principals.tsv.gz

Read Principals Data

Both IDs Valid? — No → Skip record (loop to next)

Yes

Add to Unique Records

Loop Done

Bulk Insert into Bridge_Title_Person

Bridge_Title_Person ✓

Phase 4: Fact Table

Truncate Fact_Title_Performance

title.ratings.tsv.gz

Read Ratings Data

Query Time Keys

Query Title Years

Title Exists in Dim_Title? — No → Skip record (loop to next)

Yes

Map Year to Time Key

Build Fact Record

Loop Done

Bulk Insert into Fact_Title_Performance

Fact_Title_Performance ✓

Phase 5: Finalization

Commit All Transactions

Re-enable Constraints

Print Load Summary

Close Database Connection

ETL Complete ✓