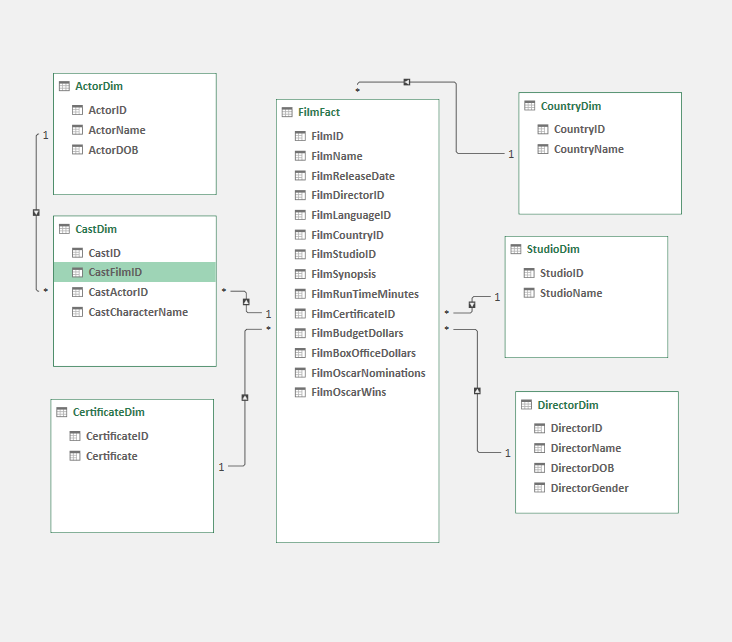
**Data Management Solution**

**Name:** SAHABI ALHAJI MUSA **Registration Number**: CST/21/COM/00833

**Department**: Computer Science

# Task 1.0: Data Modelling



(a) List of Tables, Columns, Primary Keys and Foreign Keys

|  |  |  |  |
| --- | --- | --- | --- |
| **Table Name** | **Nbr of rows** | **Fact or Dimension** | **Columns** |
| FilmFact | 160 | Fact | |  | | --- | | FilmID(PK) | | FilmName | | FilmReleaseDate | | FilmDirectorID(FK) | | FilmLanguageID(FK) | | FilmCountryID(FK) | | FilmFacttudioID(FK) | | FilmFactynopsis | | FilmRunTimeMinutes | | FilmCertificateID(FK) | | FilmBudgetDollars | | FilmBoxOfficeDollars | | FilmOscarNominations | | FilmOscarWins | |
| StudioDim | 20 | Dimension | |  | | --- | | StudioID(PK) | | StudioName | |
| DirectorDim | 81 | Dimension | |  | | --- | | DirectorID(PK) | | DirectorName | | DirectorDOB | | DirectorGender | |
| ActorDim | 333 | Dimension | |  | | --- | | ActorID(PK) | | ActorName | | ActorDOB | | ActorGender | |
| CastDim | 813 | Dimension | |  | | --- | | CastID(PK) | | CastFilmID(FK) | | CastActorID(FK) | | CastCharacterName | |
| CountryDim | 4 | Dimension | |  | | --- | | CountryID(PK) | | CountryName | |
| CertificateDim | 6 | Dimension | |  | | --- | | CertificateID(PK) | | Certificate | |
| LanguageDim | 2 | Dimension | |  | | --- | | LanguageID(PK) | | Language | |
| DateDim | 16899 | Dimension | |  | | --- | | Date(PK) | |

(b) List of Relationships

|  |  |  |  |
| --- | --- | --- | --- |
| **Table Name** | **Primary Key** | **Table** | **Foreign Key** |
| DirectorDim | DirectorID | FilmFact | FilmDirectorID |
| StudioDim | StudioID | FilmFact | FilmFacttudioID |
| CertificateDim | CertificateID | FilmFact | FilmCertificateID |
| CountryDim | CountryID | FilmFact | FilmCountryID |
| LanguageDim | LanguageID | FilmFact | FilmLanguageID |
| CastDim | CastID | FilmFact | CastFilmID |
| ActorDim | ActorID | CastDim | CastActorID |
| DateDim | Date | FilmFact | FilmReleasedDate |

#### (c) **Strengths of a Data Warehousing Solution**

1. **Scalability** – Cloud-based data warehouses allow for elastic scaling, accommodating large volumes of data efficiently.
2. **Performance Optimization** – Columnar storage and indexing improve query performance.
3. **Centralized Data Management** – Ensures data consistency and accuracy across departments.
4. **Flexible Integration** – Supports various ETL (Extract, Transform, Load) tools and APIs for data ingestion.
5. **Security & Compliance** – Many solutions offer built-in encryption, access control, and compliance features.

#### **Weaknesses Compared to SSIS**

1. **Complexity in ETL Development** – Unlike SSIS, which provides a drag-and-drop interface for ETL workflows, data warehouses often require scripting (SQL, Python, or Spark).
2. **Higher Cost for Cloud Services** – Some cloud-based data warehouses charge based on usage, which may be expensive compared to an on-premises SSIS solution.
3. **Learning Curve** – Implementing a full-fledged data warehouse requires expertise in database administration, whereas SSIS offers a more user-friendly approach.

#### **Comparison with SSIS**

SSIS is a robust ETL tool within Microsoft SQL Server that provides an intuitive interface for managing data pipelines. It is highly **cost-effective** for organizations already using Microsoft infrastructure and offers **seamless integration** with SQL Server databases. However, **SSIS lacks the scalability and flexibility** of modern cloud-based data warehouses, making it less suitable for handling **big data** workloads.

Task 2: Data Analysis

## (a) **List of DAX Formulae for Calculated Columns (CC) and Measures (M)**

|  |  |  |  |
| --- | --- | --- | --- |
| **SN** | **Name of CC or M** | **CC or M** | **Formula** |
| 1. | Years since release | CC | =YEAR(TODAY()) - YEAR(FilmFact[FilmReleaseDate]) |
| 2. | Number of PG Cert FilmFact | M | PGCertFilms := CALCULATE(COUNT(FilmFact[FilmID]), CertificateDim[Certificate] = "PG") |
| 3. | Film duration | CC | = IF(FilmFact[FilmRunTimeMinutes] > 150, "Long", IF(FilmFact[FilmRunTimeMinutes] > 100, "Medium", "Short")) |
| 4. | Profit | M | Profit:=sum([FilmBoxOfficeDollars])-sum(FilmFact[FilmBudgetDollars]) |
| 5. | Director Ranking | M | DirectorRanking:=RANKX(ALL(DirectorDim), FilmFact[Profit], , DESC) |
| 6. | YTD Profit | M | YTDProfit:=CALCULATE([Profit], DATESYTD(FilmFact[FilmReleaseDate])) |
| 7. | Female Percentage | M | FemalePercentage:=DIVIDE(CALCULATE(COUNT(CastDim[CastID]), ActorDim[ActorGender] = "Female"), COUNT(CastDim[CastID]), 0) |

1. **Answers to Questions**:

|  |  |  |
| --- | --- | --- |
| **Question** | **Answer** | **Evidence** |
| The actor with 10 casts | **Samuel L. Jackson** |  |
| What is the percentage of female casts | **20.42%** |  |
| What is the ranking, of director Steven Spielberg? | **2** |  |

1. **Screenshot of PivotTable 1, PivotChart 1 and PivotTable 2.**

|  |  |
| --- | --- |
| **Pivot Table 1** | **Pivot Chart 1** |
| **Pivot Table 2**  **YTD Profit for 2006** |  |