Strategic Data and Dimensional Warehouse Modeling for University Analytics

**Satyanarayana Vinay Achanta (A02395874)**

**Hari Chandana Kotnani (A02396013)**

**Course: Advanced Database Systems**

**Table of Contents**

1. [**Introduction**](#_1._Introduction)
2. [**Project Objectives**](#_2._Project_Objectives)
3. [**Scope of the Project**](#_3._Scope_of)
4. [**Requirements Analysis**](#_4._Requirements_Analysis)
5. [**Conceptual Data Model**](#_5._Conceptual_Data)
6. [**Logical Data Model**](#_6._Logical_Data)
7. [**Physical Data Model**](#_7._Physical_Data)
8. [**Dimensional Model Design**](#_8.Dimensional_Model_Design)
9. [**Reporting and Analysis**](#_9.Reporting_and_Analysis)
10. [**Analytical Queries**](#_10._Analytical_Queries)
11. [**Challenges and Solutions**](#_11.Challenges_and_Solutions)
12. [**Conclusion and Future Work**](#_12._Conclusion_and)

## 1. Introduction

The goal of this project is to design and implement a comprehensive database and data warehouse for Utah State University to facilitate the analysis of admissions, enrollment, and educational data. This project focuses on several key areas:

* **Data Modeling**: Developing a robust data model to capture detailed information about students, admissions, programs, courses, enrollments, grades, departments, faculty, campuses, scholarships, and their relationships.
* **Dimensional Modeling**: Designing a data warehouse schema to support analytical queries and reporting, enabling efficient data retrieval for trend analysis and decision-making.

This database and data warehouse will support various departments, including Domestic Admissions, International Admissions, and Study Abroad programs. It will cover multiple education levels such as Undergraduate, Graduate, and PhD, providing a comprehensive platform for analyzing educational data, making informed decisions, and identifying trends.

The project includes the following critical components:

* **Conceptual Data Model**: Defines high-level entities and relationships to capture essential data elements and their interactions.
* **Logical Data Model**: Organizes data into normalized tables to ensure data integrity and reduce redundancy.
* **Physical Data Model**: Specifies the physical schema, indexing strategy, and partitioning requirements to optimize storage and query performance.
* **Dimensional Model Design**: Constructs fact and dimension tables, such as EnrollmentFact, GradeFact, and StudentDim, to support efficient data analysis and reporting.

**Analytical Queries**: To facilitate in-depth analysis, several key analytical queries are included, focusing on different aspects of educational data:

* **Student Performance Analysis**: Evaluates average grades per course.
* **Scholarship Distribution**: Analyzes total scholarship amounts awarded by department.
* **Enrollment Trends**: Tracks the number of enrollments in different programs over time.
* **Faculty Workload**: Measures the number of courses taught by faculty members.
* **Fee Collection Analysis**: Reviews the total fees collected for each course and semester.
* **Admission Status**: Reports on the status of student admissions by program.
* **Student Graduation Trends**: Examines graduation rates over the years.
* **Course Enrollment**: Counts the number of students enrolled in each course.

This system aims to enhance the university's ability to manage and analyze its data, providing valuable insights for strategic planning and operational improvements.

## 2. Project Objectives

* **Design a Normalized Database Schema:**
  + Develop a detailed, normalized schema to capture and manage educational data efficiently, ensuring data integrity and eliminating redundancy.
* **Develop a Denormalized Data Warehouse Schema:**
  + Create a denormalized data warehouse schema to facilitate efficient querying and reporting, optimized for analytical and business intelligence purposes.
* **Create Analytical Queries:**
  + Develop analytical queries and reports to support various stakeholders, including admissions, enrollment management, and program evaluation.

## 3. Scope of the Project

The project will cover the following areas:

* **Admissions Data:** Manage and analyze admissions data for domestic, international, and study abroad students.
* **Educational Levels:** Capture and analyze data across different educational levels, including undergraduate, graduate, and PhD.
* **Enrollment Details:** Track and manage enrollment details, including courses, grades, and academic records.
* **Database and Data Warehouse Development:** Develop a normalized operational database for detailed record-keeping and a denormalized data warehouse for efficient querying and reporting.

## 4. Requirements Analysis

* **Stakeholders:**
  + **Admissions Department:** Requires data to manage and analyze applications, admissions processes, and trends for domestic, international, and study abroad students.
  + **Academic Departments:** Needs data on student enrollments, course performance, and academic records to support curriculum planning and academic advising.
  + **University Administration:** Requires comprehensive data to make informed decisions about university operations, resource allocation, and strategic planning.
* **Key Metrics:**
  + **Student Enrollment Trends:** Analyzing patterns and changes in student enrollments over time to support planning and resource management.
  + **Admission Rates and Acceptance Statistics:** Tracking admission rates, acceptance ratios, and application trends to assess the effectiveness of recruitment strategies.
  + **Academic Performance and Grade Distributions:** Evaluating academic performance across different programs and courses to identify trends, strengths, and areas for improvement.
  + **Course Enrollment Statistics:** Monitoring course enrollments to manage class sizes, staffing needs, and curriculum adjustments.

## 5. Conceptual Data Model

**Entities:**

1. Student
2. Admission
3. Program
4. Course
5. Enrollment
6. Grade
7. Department
8. Faculty
9. Campus
10. Scholarship
11. StudentScholarship
12. Fee

**Relationships:**

* Student to Admission: One-to-Many
* Admission to Program: Many-to-One
* Student to Enrollment: One-to-Many
* Enrollment to Course: Many-to-One
* Enrollment to Grade: One-to-One
* Program to Department: Many-to-One
* Course to Department: Many-to-One
* Course to Campus: Many-to-One
* Student to StudentScholarship: One-to-Many
* Scholarship to StudentScholarship: One-to-Many
* Faculty to Department: Many-to-One
* Program to Campus: Many-to-One
* Fee to Student (Many-to-One)
* Fee to Course (Many-to-One)

## 6. Logical Data Model

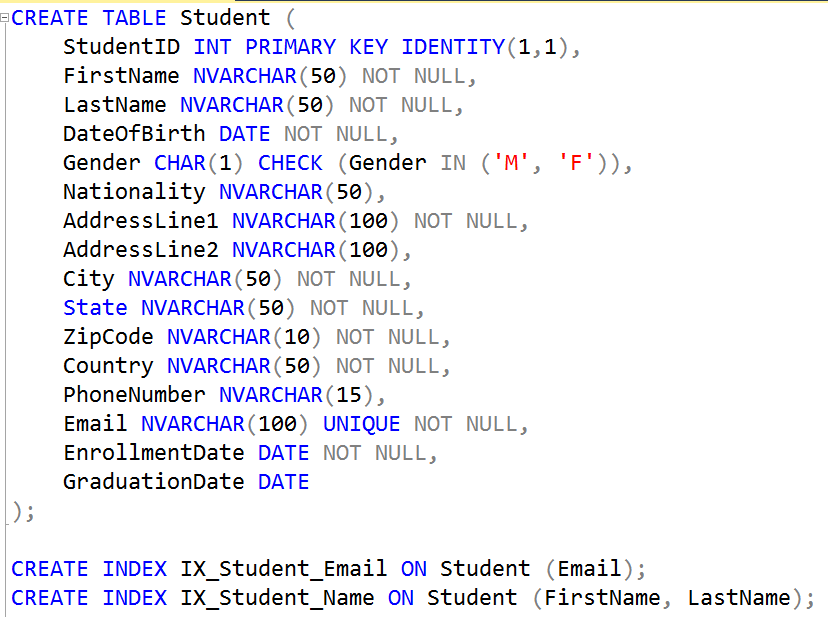
**Entity-Relationship Model**

A black screen with white text

Description automatically generated

## 7. Physical Data Model

**Normalized Schema (3NF):**

* **Student:**
  + 
* **Admission:**
  + **A screen shot of a computer

    Description automatically generated**
* **Program:** 
  + **A computer screen shot of a program

    Description automatically generated**
* **Course:**
  + **A screenshot of a computer program

    Description automatically generated**
* **Enrollment:**
  + **A screen shot of a computer

    Description automatically generated**
* **Grade:**
  + **A computer screen shot of a code

    Description automatically generated**
* **Department:**
  + **A computer screen shot of a computer code

    Description automatically generated**
* **Faculty:**
  + **A screenshot of a computer program

    Description automatically generated**
* **Campus:**
  + **A close up of a computer code

    Description automatically generated**
* **Scholarship:**
  + **A screenshot of a computer code

    Description automatically generated**
* **StudentScholarship:**
  + **A screen shot of a computer

    Description automatically generated**
* **Fee**
  + **A screenshot of a computer code

    Description automatically generated**

## 8.Dimensional Model Design

**Dimension Tables:**

* Student: Describes students.
* Faculty: Describes faculty members.
* Department: Describes departments.
* Campus: Describes campuses.
* Course: Describes courses.
* Scholarship: Describes scholarships.
* Program: Describes programs.
* Date: Will be used to manage date-related attributes for various events (e.g., enrollment, fee payment).

**Fact Tables:**

* Enrollment: Stores information about student enrollments.
* Grade: Stores grades received by students.
* Fee: Stores fee payments made by students.
* Admission: Stores admission details.

**Code:**

* **DateDim**
  + A screenshot of a computer code

    Description automatically generated
* **StudentDim**
  + **A screenshot of a computer program

    Description automatically generated**
* **FacultyDim**
  + **A screenshot of a computer code

    Description automatically generated**
* **DepartmentDim**
  + **A screenshot of a computer code

    Description automatically generated**
* **CampusDim**
  + **A computer screen shot of a table

    Description automatically generated**
* **CourseDim**
  + **A screenshot of a computer code

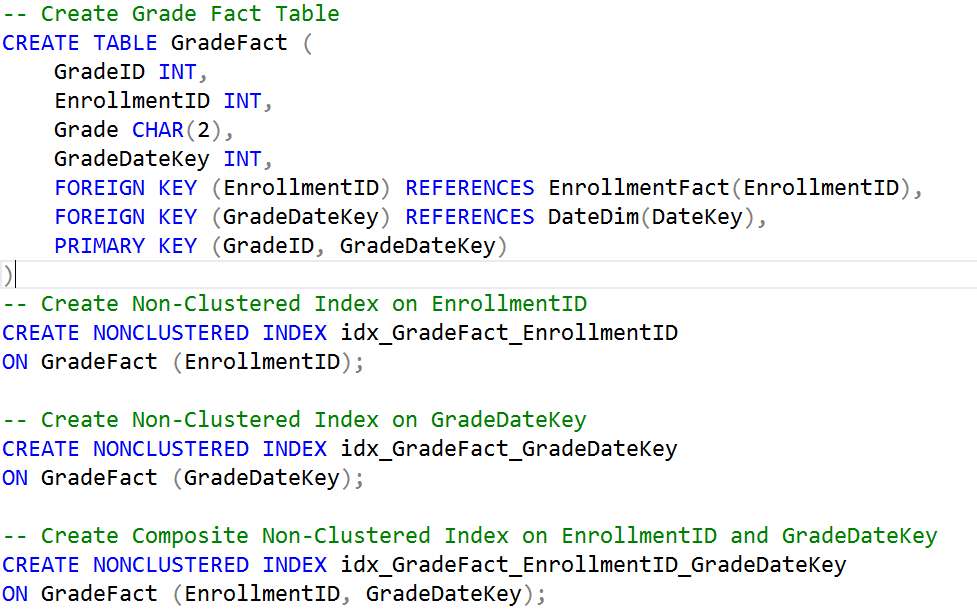
    Description automatically generated**
* **ScholarshipDim**
  + **A computer screen shot of a computer program

    Description automatically generated**
* **ProgramDim**
  + **A computer screen shot of a program

    Description automatically generated**
* **AdmissionDim**
  + **A computer screen shot of a program

    Description automatically generated**
* **Partition Function and Scheme**
  + **A computer code with numbers and letters

    Description automatically generated**
* **EnrollmentFact**
  + **A screenshot of a computer program

    Description automatically generated**
* **GradeFact**
  + 
* **FeeFact**
  + A screenshot of a computer code

    Description automatically generated
* **AdmissionFact**
  + A screenshot of a computer program

    Description automatically generated
* **ScholarshipAwardFact**
  + A screenshot of a computer program

    Description automatically generated

## 9.Reporting and Analysis

**Types of Reports**

* **Performance Reports**: Analyzing student performance and faculty workload.
* **Financial Reports**: Reviewing fee collections and scholarship distributions.
* **Enrollment Reports**: Tracking enrollment trends and admission statuses.

**Key Performance Indicators (KPIs)**

* **Average Grade**: Measures overall student performance.
* **Total Scholarship Amount**: Evaluates scholarship distribution by department.
* **Enrollment Count**: Assesses program popularity and trends.

**Reporting Tools and Technologies**

* **Reporting Tools**: Tools like Power BI or Tableau for visualizing data.
* **Technologies**: SQL for querying, ETL tools for data integration.

## 10. Analytical Queries

* **Student Performance Analysis**
  + A screenshot of a computer

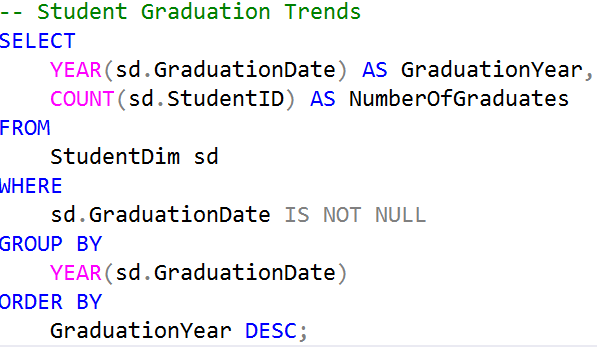
    Description automatically generated
* **Scholarship Distribution**
  + **A screenshot of a computer program

    Description automatically generated**
* **Enrollment Trends**
  + **A screenshot of a computer code

    Description automatically generated**
* **Faculty Workload**
  + **A screenshot of a computer program

    Description automatically generated**
* **Fee Collection Analysis**
  + **A screenshot of a computer code

    Description automatically generated**
* **Admission Status**
  + **A screenshot of a computer program

    Description automatically generated**
* **Student Graduation Trends**
  + ****
* **Course Enrollment**
  + **A screen shot of a computer program

    Description automatically generated**

## 11.Challenges and Solutions

**Data Quality Issues**

* **Challenge:** Ensuring data consistency and accuracy across different sources and entities within the data model.
* **Solution:** Implemented thorough data validation rules and constraints within the data model to ensure data integrity. Developed comprehensive data dictionaries to standardize data definitions and formats.

**Performance Optimization**

* **Challenge:** Optimizing query performance for large datasets within the data warehouse.
* **Solution:** Designed indexing strategies and optimized schema design to improve query efficiency. Applied best practices in dimensional modeling, such as star and snowflake schemas, to facilitate faster data retrieval.

**Scalability Concerns**

* **Challenge:** Designing a data model that can efficiently handle increasing volumes of data over time.
* **Solution**: Adopted a scalable schema design with consideration for future data growth. Implemented partitioning strategies and designed the data model to support potential expansion in data volume and complexity.

## 12. Conclusion and Future Work

**Summary of Achievements**

* **Data Warehouse Design**: Successfully designed a comprehensive data warehouse for Utah State University, including well-structured dimensional models and a robust schema for data analysis.
* **Dimensional Modeling**: Developed detailed star and snowflake schemas to support efficient analytical queries, facilitating enhanced reporting and decision-making capabilities.

**Lessons Learned**

* **Data Modeling Best Practices**: Emphasized the importance of a well-designed data model in ensuring data consistency and integrity. Identified how effective dimensional modeling can significantly improve query performance and ease of use.
* **Performance Optimization**: Learned the value of applying indexing and schema design best practices to optimize performance for complex queries and large datasets.

**Recommendations for Future Enhancements**

* **Real-Time Data Processing**: Consider exploring the integration of real-time data processing capabilities to enable more dynamic and up-to-date reporting and analysis.
* **Advanced Analytics**: Investigate the potential of incorporating advanced analytics and machine learning techniques to derive predictive insights and further enhance decision-making processes.