**UdemyDwH Data Warehouse Documentation**

**Purpose**: This document describes the design and structure of the **UdemyDwH** data warehouse, which supports analytical reporting for a Udemy-like e-learning platform. The schema follows a **hybrid star schema** with conformed dimensions and optimized fact tables for enrollment, orders, and course analysis.

**1. Design Overview**

**1.1 Core Components**

* **Dimensions**: DimDate, DimUsers, DimCourses.
* **Sub-Dimensions**: SubDimCrsReq, SubDimSection, SDimQuiz (normalized details linked to courses).
* **Fact Tables**: FactEnrollment, FactOrder.
* **Schema Type**: Hybrid star schema with conformed dimensions.

**1.2 SCD Handling**

* **Type 2 SCD**: DimUsers (tracks historical user profile changes via StartDate/EndDate).
* **Type 1 SCD**: DimCourses (overwrites changes; no history).

**1.3 Key Metrics**

* Student enrollment trends and completion rates.
* Course revenue, discounts, and payment methods.
* Course content structure (sections, quizzes, requirements).
* Instructor performance and student demographics.

**2. Data Model**

**2.1 Core Dimensions**

**DimDate**

* **Purpose**: Standard time dimension for temporal analysis.
* **Key Fields**:
  + Date\_id\_sk: Surrogate key (e.g., YYYYMMDD).
  + HolidayText: Flags holidays for promotional analysis.

**DimUsers**

* **Purpose**: Track students, instructors, and admins with historical changes.
* **Key Fields**:
  + UserId\_SK: Surrogate key.
  + UserId\_BK: Business key from AspNetUsers.Id.
  + Social media flags (HasFacebook, HasX).
  + Role flags (IsStudent, IsInstructor, IsAdmin).
  + Wallet: Student/instructor earnings or spending balance.

**DimCourses**

* **Purpose**: Current state of courses.
* **Key Fields**:
  + CourseId\_SK: Surrogate key.
  + Category/SubCategory: Denormalized hierarchy.
  + OriginalPrice/CurrentPrice: Track pricing changes.
  + BestSeller: Manual curation flag.

**2.2 Sub-Dimensions**

**SubDimCrsReq**

* **Purpose**: Course prerequisites (e.g., "Basic Python knowledge").
* **Link**: 1:M with DimCourses via CourseId\_SK.

**SubDimSection**

* **Purpose**: Course content structure.
* **Key Fields**:
  + NoVideo/NoArticle: Media type distribution.
  + Duration: Total section runtime.

**SDimQuiz**

* **Purpose**: Quiz complexity analysis.
* **Key Fields**:
  + MultipleChoiceCount/TrueOrFalseCount: Question-type distribution.

**2.3 Fact Tables**

**FactEnrollment**

* **Grain**: One row per student-course enrollment.
* **Measures**:
  + ProgressPercentage: % of course completed.
  + Rating: Student feedback (1–5).
  + Grade: Quiz/exam score (if applicable).
* **Foreign Keys**:
  + UserId\_SK → DimUsers.
  + CourseId\_SK → DimCourses.
  + StartDateKey/CompletionDateKey → DimDate.

**FactOrder**

* **Grain**: One row per course purchase within an order.
* **Measures**:
  + TotalAmount: Number of courses in the order.
  + OrderPrice: Price at the time of purchase.
  + Discount: Applied discount percentage.
* **Foreign Keys**:
  + UserId\_SK → DimUsers.
  + CourseId\_SK → DimCourses.
  + OrderDateKey → DimDate.

**3. Key Features**

**3.1 Student-Centric Analysis**

* Track enrollment progress by demographics (DimUsers.CountryName, Age).
* Correlate social media presence (HasLinkedIn) with course preferences.

**3.2 Course Performance**

* Identify bestsellers by Category and Rating.
* Analyze pricing strategies (OriginalPrice vs. CurrentPrice).

**3.3 Instructor Insights**

* Link instructors (DimUsers.IsInstructor) to course revenue (FactOrder).

**4. ETL Workflow**

**4.1 Source Tables**

* **Operational Database**:
  + AspNetUsers, Students, Instructors.
  + Courses, Enrollments, Orders, Sections, QuizQuestions.

**4.2 Transformations**

1. DimUsers:
   * Merge AspNetUsers with Students/Instructors.
   * Derive social media flags from SocialMedias.
   * Track Type 2 changes using CreatedDate/ModifiedDate.
2. DimCourses:
   * Denormalize Categories and Subcategories.
3. FactEnrollment:
   * Calculate ProgressPercentage from Progresses table.

**4.3 Incremental Loading**

* Use CreatedDate from source tables to load new records.
* For Type 2 SCD (DimUsers), detect changes via checksum comparison.