A College Star Schema case study

BID3000 Fall 2024

- **This tutorial will create a data warehouse start schema from a database using Postgresql.**

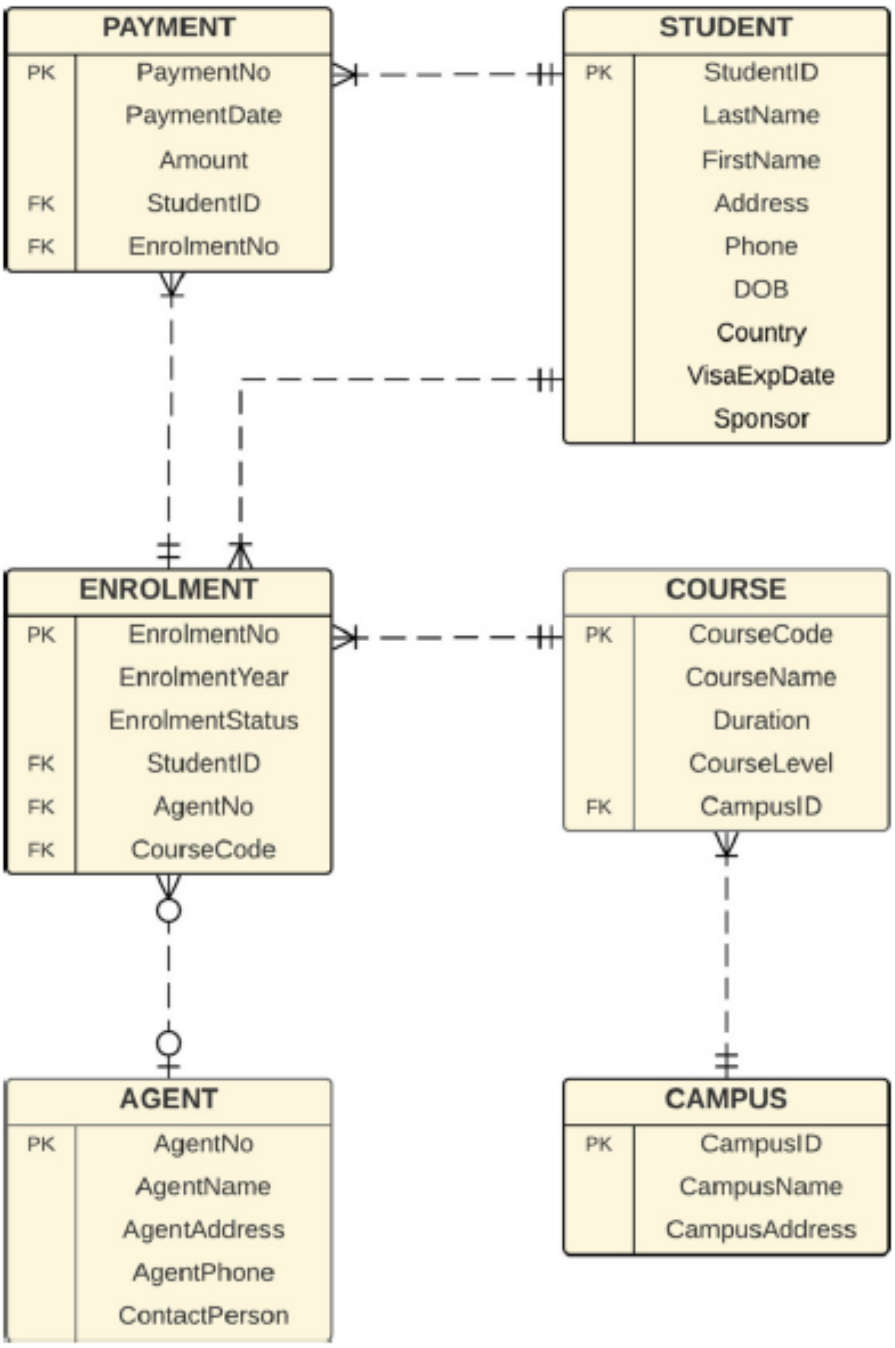
This case study involves an International College. The admissions office handles enrolment, payment and marketing campaigns to international students, often through educational agents located overseas. This admissions office has an operational system that maintains all the details of international students enrolled in the College. Payment details are also handled by this office. The operational system has the following features:

- All student details are kept in the database. This includes the courses in which the students have enrolled.
- As the College is a multi-campus university, some courses are offered at different campuses. The admissions office handles international students on all campuses.
- Some international students coming to the College are handled by an educational agent. This is particularly common for the first course in which a student enrols. Subsequent courses are not normally handled by an agent because the students themselves deal directly with the College.
- International students pay tuition fees several times per year, usually once each semester, for each course in which they are enrolled

The operational database that maintains the above system has the following tables. Note that the Primary Keys (PK) are underlined and the Foreign Keys (FK) are printed in italic. All the relationships in this E/R diagram are non-associative relationships as the entities do not share any common PK.

- **Student** (StudentID, LastName, FirstName, Address, Phone, DOB, Country, VisaExpDate, Sponsor)
- **Campus** (CampusID, CampusName, CampusAddress)
- **Course** (CourseCode, CourseName, Duration, CourseLevel, *CampusID*)
- **Agent** (AgentNo, AgentName, AgentAddress, AgentPhone, ContactPerson)
- **Enrolment** (EnrolmentNo, EnrolmentYear, EnrolmentStatus, *StudentID, AgentNo, CourseCode*)
- **Payment** (PaymentNo, PaymentDate, Amount, *StudentID, EnrolmentNo*)

An E/R diagram showing the current operational system:


PAYMENT
PK PaymentNo
PaymentDate
Amount
FK StudentID
FK EnrolmentNo
STUDENT
PK StudentID
LastName
FirstName
Address
Phone
DOB
Country
VisaExpDate
Sponsor
ENROLMENT
PK EnrolmentNo
EnrolmentYear
EnrolmentStatus
FK StudentID
FK AgentNo
FK CourseCode
COURSE
PK CourseCode
CourseName
Duration
CourseLevel
FK CampusID
AGENT
PK AgentNo
AgentName
AgentAddress
AgentPhone
ContactPerson
CAMPUS
PK CampusID
CampusName
CampusAddress

The College now requires a data warehouse for analysis purposes. The analysis is needed to answer at least the following questions:

- How many students come from certain countries?
- What is the total income for certain postgraduate courses?
- How many students are handled by certain agents?
- How do enrolment numbers in particular courses fluctuate across the year?

Step 1: Analysis of the case

The first question could be used by management to identify countries that may be targeted for future international marketing campaigns. The second question could be used by the financial office for further planning. The third question could be used in conjunction with future international marketing campaigns.

Step 2: Choosing the fact table

From the above requirements, it is clear that there are two fact measures needed in the star schema. The first one is *Number of Students*, from the "How many students ...?" question. Number of Students is an aggregated numerical value; hence, it satisfies the two requirements of a fact measure. The second fact measure is *Total Income*, from the "What is the total income...?" question.

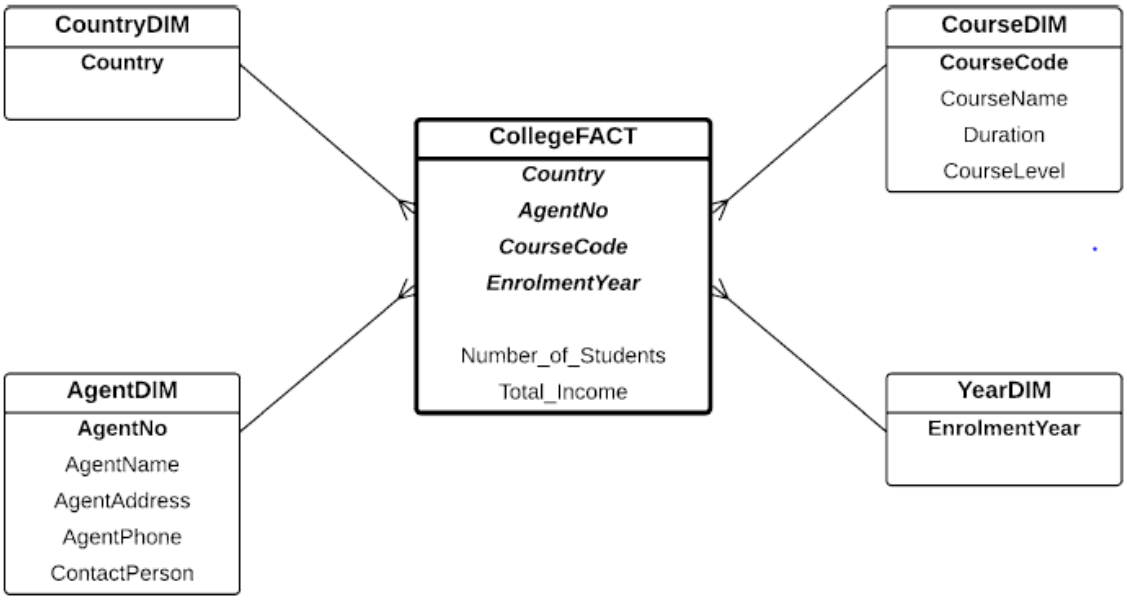
Step 3: Choosing the dimension tables

The dimensions are the perspectives from which the fact measures are viewed. So, "How many students come from certain countries", for example, is to look at the Number of Students from a country point of view; hence *Country* is a dimension.

We can deduce other dimensions from the other questions, such as Course, Agent and Year. As a result, the star schema has four dimensions (e.g. Country, Course, Agent and Year) and two fact measures (e.g. Number of Students and Total Income).

Step 4: Drawing the star schema

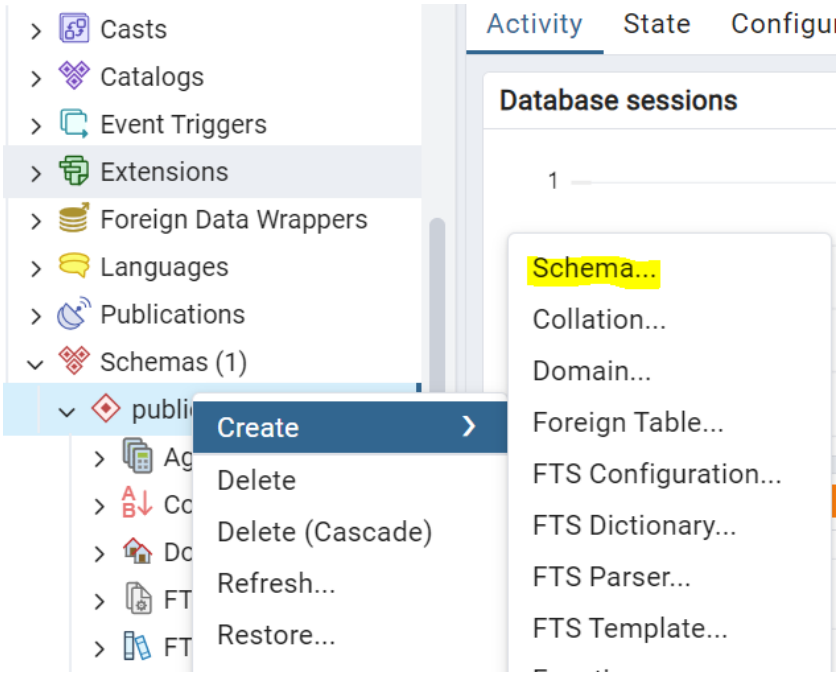
For simplicity, some of the dimensions have only one attribute, which is the DimensionID attribute. For example, the Country dimension has only the Country Name attribute and the Year dimension has only the Enrolment Year attribute. In a real case, there will be other attributes in these dimensions. But for this first case study, the scope of the attributes is limited to the information in the tables in the operational database



Once the star schema is drawn, it is time to implement it.

Step 5: Creating the College Database

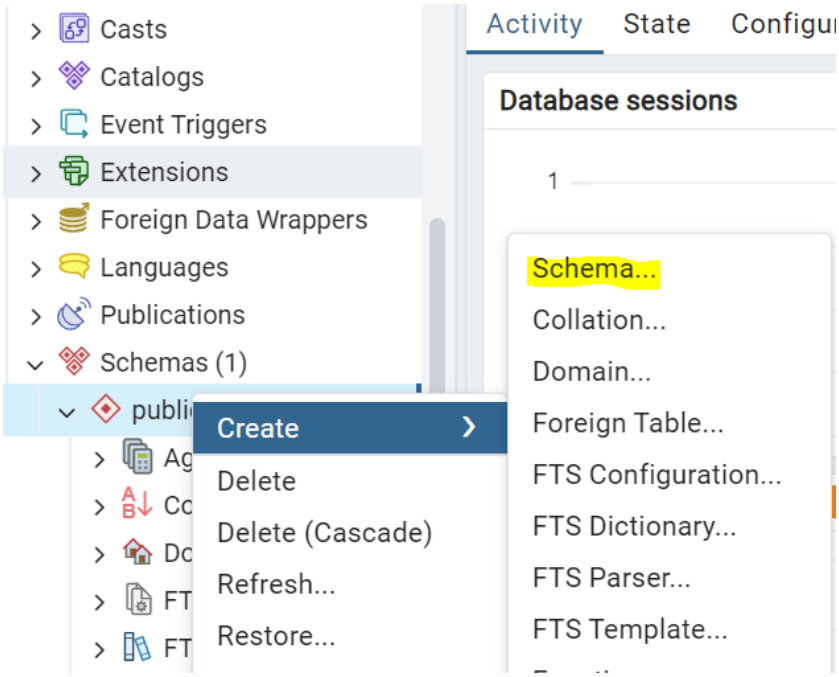
1. Create a database called College
2. Open the script CollegeDB.sql
3. Create a schema in Postgresql called “CollegeDB”



4. Run the script CollegeDB.sql to set up the database. It contains all the tables and all the data from the tables.

Step 6: Building the College Data Warehouse

1. Open the script CollegeDW.sql
2. Create a schema in Postgresql called “CollegeDW”



3. Run the script CollegeDW.sql to set up the database. It contains all the tables and fills the data from the operational database CollegeDB.

Congratulations! You have created the College data warehouse.

Reference:

Taniar, D., & Rahayu, W. (2022). *Data warehousing and analytics: fueling the data engine*. Springer Nature.