FIT3003 S2 2022

Lecture 7 Activity

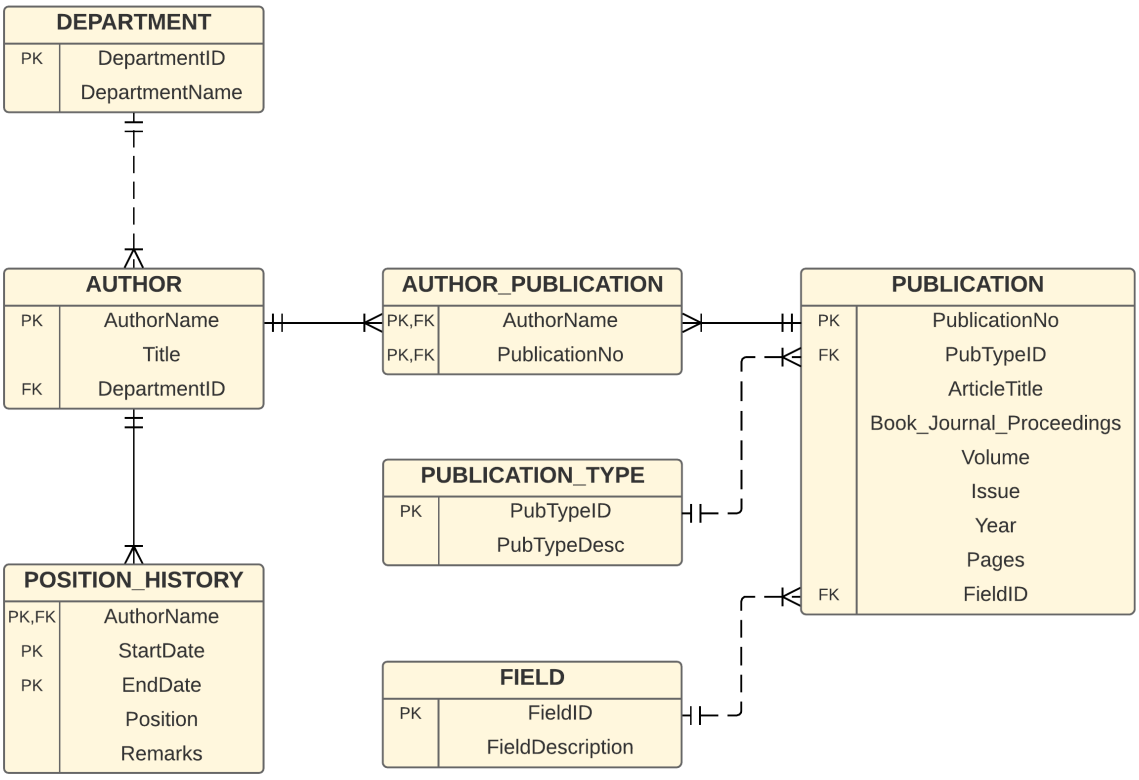
Publication Case Study

It is common that professors, lecturers, and all academics working in a university to do research, and publish their research in journals, conference proceedings, and edited books. The university keeps track all publications that were published by their academics (e.g. professors, researchers, lecturers). When a publication is entered into the system, the system will generate a publication number as an identifier. Other details, such as author names, article title, journal volume, issue, year and page numbers are recorded. It is common that an article (a paper) has several authors. There are several publication types. Each paper is categorized into one publication type, whether it is a *journal paper*, or a *conference paper*, or a *book chapter*.

For classification purposes, the university allocates a research field for each paper, such as computer science, economics, law, etc. If the paper is a multi disciplinary paper, the university will choose the main discipline for identifying the research field.

Academic jobs in the university have many different ranks (e.g. lecturer, senior lecturer, associate professor, and professor), in which an academic may rise from one level to another; often is based on their publications.

The operational database for publications is maintained by the university, and is shown by the following E/R diagram.



You are required to build a data warehouse to analyse number of publications for each year, research field, publication type, and academic/author. Note the history of each academic job rank in the university is maintained in the database.

Questions:

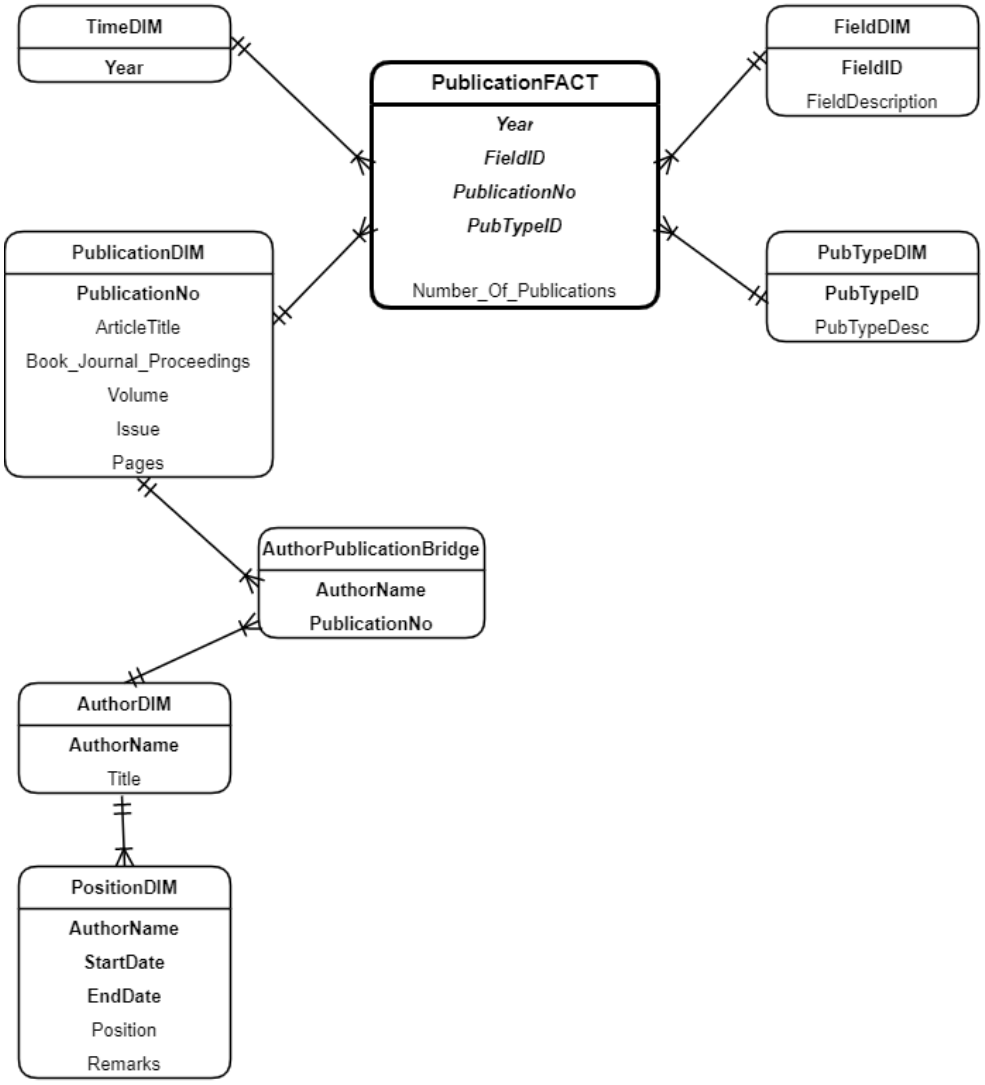
(a) Draw a star schema for the "Publication" data warehouse, following the above requirements.

(b) Write the SQL statements to create (and populate with records) the dimension and fact tables.

(c) Assume that the star schema you created in Question *(a)* is in the highest level of aggregation. Create another version of star schema that is in level 0, which means no aggregation.

Write your answer here:

Question – a:



Continue your answer here:

Question – b:

```
create table TimeDIM
as select distinct Year
from Publication;

create table FieldDIM
as select distinct F.FieldID, F.FieldDescription
from Field F, Publication P
where F.FieldID = P.FieldID;

create table PubTypeDIM
as select distinct T.PubTypeID, T.PubTypeDesc
from Publication_Type T, Publication P
where T.PubTypeID = P.PubTypeID;

create table PublicationDIM
as  select  PublicationNo,  ArticleTitle,  Book_Journal_Proceedings,
Volume, Issue, Pages
from Publication;

create table AuthorPublicationBridge
as select * from Author_Publication;

create table AuthorDIM
as select AuthorName, Title
from Author;

create table PositionDIM
as select * from Position_History;

create table PublicationFACT
as select P.Year, P.FieldID, P.PubTypeID, AP.PublicationNo,
   Count(*) as Number_of_Publications
from Publication P, Author_Publication AP
where P.PublicationNo = AP.PublicationNo
group by P.Year, P.FieldID, P.PubTypeID, AP.PublicationNo;
```

Question – c:

The star schema created in Question *(a)* is already in the lowest level of aggregation (Level-0), hence there is no change in the star schema design.

THE END