DUBLIN BUSINESS SCHOOL

UNIT TITLE: Data Warehousing & Business Intelligence

UNIT CODE: B8IT104

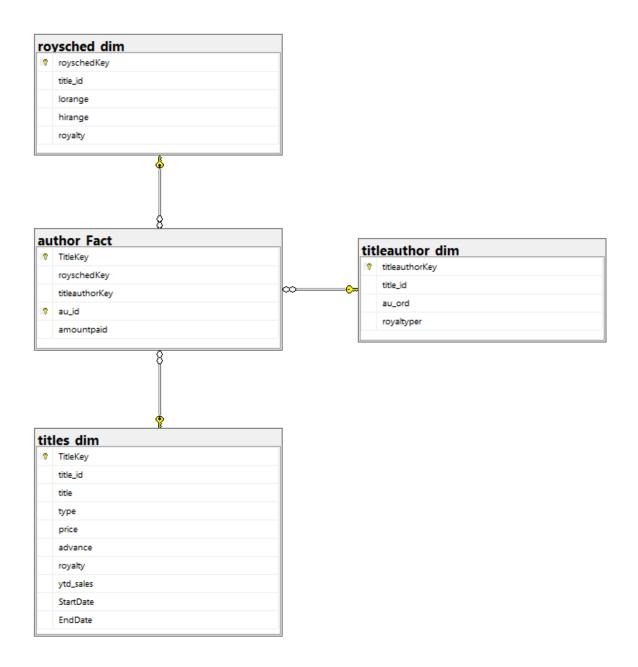**UNIT LEADER: Kunwar Madan**

LEVEL: 8

ASSESSMENT TITLE: Continuous Assessment One
STUDENT NAME: NICHOLAS UDOMBOSO

STUDENT ID: 10547333

PART 1 (DATA MODELLING)

I used Kimball's dimensional modelling approach to design a star schema for publisher's payment to authors business process, the below star schema was designed:

roysched_dim

🔑	royschedKey
	title_id
	lorange
	hirange
	royalty

author_Fact

🔑	TitleKey
	royschedKey
	titleauthorKey
🔑	au_id
	amountpaid

titleauthor_dim

🔑	titleauthorKey
	title_id
	au_ord
	royaltyper

titles_dim

🔑	TitleKey
	title_id
	title
	type
	price
	advance
	royalty
	ytd_sales
	StartDate
	EndDate

I acknowledged that prices of the books could change in future, hence introduced StartDate and EndDate columns in my titles_dim dimensional table to keep track of the prices as it changes. In other words, I made the price column in titles_dim to be SCD (slowly changing dimension) Type 2 and other columns of my dimension tables in the data warehouse to be SCD Type 1.

The schema was implemented and also data was extracted, transformed and loaded to the dimensions and fact tables of the schema.

The SSIS packages have been submitted.

PART 3 (Reporting)

Explaining the Schema in Part 1:

Using Kimball's model, I designed the star schema of PublisherDW. PublisherDW is the name I gave my datawarehouse. My data warehouse star schema has three dimension tables, namely titles_dim, titleauthor_dim and roysched_dim, and a fact table, named author_Fact.

My star schema models publisher's payment to different authors. In other words, how payments are being made to authors. In creating the schema, I considered the type of columns that each dimension table should have, i.e. SCD Type 1, SCD Type 2 or SCD Type 3.

So I made the price column of titles _dim table an SCD Type 2 as I want to track every changing price and the dates that the new price started. Then I made every other column to be SCD Type 1.

In achieving SCD Type 2 for price column I introduced two new columns, StartDate and EndDate columns to titles_dim dimension tables. The StartDate holds information on the date that a particular price for every book started whereas EndDate column has the date that the price was ended. If the price of a book is still active, the EndDate of the book title remains null.

To describe the Titles_dim dimension table, the table provides information on the titles of the 17 books being published. The price column shows the price of each book. Authors receive advances per book they write. The value of the advances is in the advance column. Royalty is also paid to the authors based on year-to-date sales per book title. The value of the royalty is given in percentages. A royalty with a value of 10, means 10% is paid in royalties on yearly basis based on the sales amount. Ytd_sales column gives information on the sales amount made per year per book title.

Roysched_dim dimension table specifies royalty scheme for all the book titles. It guides on how much royalty would be paid for sales made for a particular title. That means the royalty does not always remain the same. From the table for example, if sales are between 0 to 5000 royalty is 10%. For sales between 501 to 50000, royalty is 12%. So, it depends on the sales range. In other words, the Roysched_dim dimension table provides the agreed royalty value based on the sales range.

Titleauthor_dim dimension table gives information on which author is working on a title and on which. Looking at the table, one can say that if a title_id occurs more than once, that means the title has more than an author. It also means if a title_id occurs once, that means the book title has only one author. Royaltyper column specifies royalty percentage per author for titles with more than one author. Au_ord column gives the list of authors in the order of their contribution to writing a book. Royalty on each of the books is split between or among the authors depending on the number of them per book. An author

who wrote a book alone takes the entire royalty on the book. Another observation on the table is that it shows that different authors can be working on different book titles at d same time. That means a title can have multiple authors and a single author can also be working on multiple titles at the same time.

Author_Fact table is the fact table in the schema. The dimension tables (Titles_dim, Roysched_dim, Titleauthor_dim) are all connected to the Author_Fact table in the data warehouse.

I defined au_id column to identify the individual authors and the amountpaid column to specify the value of payment made to individual authors. With this I can calculate the payment made to the authors. This is the business process I have modelled.

I followed the below steps to develop the dimensional model:

a) I Identified the business process. In this case, it is about modelling payment to different authors.
b) I identified the grain i.e the lowest granular level to measure the data of the business process. In this case I can measure amount paid to authors by yearly sales per title, amount paid to authors by royalty per title.
c) I identified the dimensions which are Titles_dim, Roysched_dim, Titleauthor_dim. Titles_dim provides the measurement for amount paid by yearly sales amount of the book titles. Roysched_dim and Titleauthor_dim provides measurement for amount paid by royalty per title.
d) I identified author as the fact. That formed the author_Fact table.

Analysis/Results of Analysis:

To calculate payment to the authors, I represent the calculation as follows:

Royalty from sales amount = royalty column in titles_dim table * ytd_sales column in titles_dim table

= value of the royalty * 0.01 * ytd_sales

Royalty per author = royaltyper column in titleauthor_dim table * Royalty from sales amount

= royaltyper value * 0.01 * value of the royalty * 0.01 * ytd_sales

That means, Amountpaid to authors = Royalty per author + advance

= (royaltyper value * 0.01 * value of the royalty * 0.01 * ytd_sales) + advance

SQL Reports to support the analysis:

1.The below query calculates the amount paid to each author

```
select ti.title_id, ta.au_id,
((ti.ytd_sales * ti.royalty * 0.01 * ta.royaltyper * 0.01) + ti.advance) as amountpaid
from titles ti inner join titleauthor ta
on ti.title_id = ta.title_id
```

⊞ Results ▦ Messages

	title_id	au_id	amountpaid
1	PS3333	172-32-1176	2407.2000
2	BU1032	213-46-8915	5163.8000
3	BU2075	213-46-8915	14618.2800
4	PC1035	238-95-7766	8404.8000
5	BU1111	267-41-2394	5155.0400
6	TC7777	267-41-2394	8122.8500
7	BU7832	274-80-9391	5409.5000
8	BU1032	409-56-7008	5245.7000
9	PC8888	427-17-2319	8204.7500
10	TC7777	472-27-2349	8122.8500
11	PC9999	486-29-1786	NULL
12	PS7777	486-29-1786	4333.6000
13	TC4203	648-92-1872	6113.4400
14	TC7777	672-71-3249	8163.8000
15	MC2222	712-45-1867	243.8400
16	MC3021	722-51-5454	19004.2800
17	BU1111	724-80-9391	5232.5600
18	PS1372	724-80-9391	7009.3750
19	PS1372	756-30-7391	7028.1250
20	TC3218	807-91-6654	7037.5000
21	PC8888	846-92-7186	8204.7500
22	MC3021	899-46-2035	16334.7600
23	PS2091	899-46-2035	2397.7000
24	PS2091	998-72-3567	2397.7000
25	PS2106	998-72-3567	6011.1000

2.The below query shows how much each author is being paid. This is queried from the Fact table in the schema.

```
select au_id as Author, amountpaid as "Amount Paid" from author_Fact
```

	Author	Amount Paid
1	267-41-2394	5155
2	724-80-9391	5233
3	213-46-8915	14618
4	274-80-9391	5410
5	712-45-1867	244
6	722-51-5454	19004
7	899-46-2035	16335
8	238-95-7766	8405
9	486-29-1786	NULL
10	724-80-9391	7009
11	756-30-7391	7028
12	899-46-2035	2398
13	998-72-3567	2398
14	998-72-3567	6011
15	172-32-1176	2407
16	486-29-1786	4334
17	807-91-6654	7038
18	267-41-2394	8123
19	472-27-2349	8123
20	672-71-3249	8164
21	213-46-8915	5164
22	409-56-7008	5246
23	427-17-2319	8205
24	846-92-7186	8205
25	648-92-1872	6113

References:

1.Madan, K. (2021) Data Warehousing & Business Intelligence. Dublin Business School. Available at: https://elearning.dbs.ie/course/view.php?id=12709 (Accessed: 17 February 2021).