

Building a Data Warehouse for Publisher Ltd

By: Karina Jonina

Completed as part of a group assignment with Polly Lloyd and Stephen Quinn for module called ‘Database Warehousing & Business Intelligence’ while studying Higher Diploma in Science in Data Science in Dublin Business School. Submitted on the 1st November 2020.

Table of Contents

[Part 1 - Business Drivers 3](#_Toc157185043)

[Goals of Publisher Ltd 3](#_Toc157185044)

[Key Stakeholders of Publisher Ltd 3](#_Toc157185045)

[Questions to be answered/associated KPIs 4](#_Toc157185046)

[Part 2: Data Modelling 5](#_Toc157185047)

[Overview 5](#_Toc157185048)

[Step 1 – Business Process 5](#_Toc157185049)

[Step 2 – Grain 5](#_Toc157185050)

[Step 3 – Facts 5](#_Toc157185051)

[Step 4 – Dimensions 5](#_Toc157185052)

[Star Schema 6](#_Toc157185053)

[Part 3: ETL 8](#_Toc157185054)

[Run ETL within Visual Studio 2018 - SSIS 8](#_Toc157185055)

[Part 4: Reporting 9](#_Toc157185056)

# Part 1 - Business Drivers

## Goals of Publisher Ltd

The aim of the data warehouse is to enable Publisher Ltd to capture all relevant information pertaining to the **sales process** from their databases, for both current and historical data. The information should be captured at the grain that allows the information to be further interrogated and analysed by title, book type, time period, store, and publishing location. The data warehouse will capture only that information and nothing more, to ensure efficiency and optimisation.

## Key Stakeholders of Publisher Ltd

The following end users (who extract, and analyse data) and users of the analysis (to make business decisions) were identified as key stakeholders within the organisation:

**Chief Executive Officer (CEO)** – Responsible for every part of the daily operations, from selecting manuscripts to managing the finance. This stakeholder will need to see all performance measures.

**Business Operations Manager** – Responsible for overseeing the production and delivery of issues ensuring they are on schedule, budget and to a high quality and refining/improving the production and operations processes. The Business Operations Manager also provides operational input into change management processes. (S)he will need to monitor cost vs profit, tracking sales trends over time, and return on investment, with regard to process improvements.

**Chief financial officer (CFO)** – Responsible for managing the financial actions of a company, and include tracking cash flow and financial planning, as well as analysing the company's financial strengths and weaknesses and proposing corrective actions. (S)he will need to monitor, total sales, performance of each publishing location/staff performance.

**Publisher** – Responsible for the editorial and commercial direction of the company. (S)he makes decisions about the markets that their companies will serve and the type of content they will offer their audience. (S)he will need to monitor the sales trends over time, particularly by TitleName, TitleType and the locations of the publishing houses and that of the customers (Stores).

**Marketing Manager** – Responsible for product marketing, market research, product development, and customer feedback. The marketing manager will want to assess the success of their campaigns, by monitoring sales volumes and revenues over time, with particular regard to their product range (Title/Type) and customer (Stores).

**Public Relations Manager -** Responsible for PR campaigns via a variety of channels, and planning and executing events and launches with authors, trade contacts and press.

**Sales Representative** - Responsible for sales to the customers (Stores). The source database does not allow for Publisher Ltd to know the sales generated by a particular Sales Rep (where more than one Sales rep work in a Publishing House), so any metrics should be viewed as team performance for each Publishing House. They will be particularly interested in sales volumes, discounts, and revenue generated by Store (Name and location/area), Title and Type.

## Questions to be answered/associated KPIs

What is the best performing Title in terms of Revenue Total? What was the maximum/minimum revenue value in any order?

What is the Revenue per month per Title?

What is the best performing Title Type in terms of Revenue Total?

What Stores generate the most Revenue Total?

What Store Cities generate the most Revenue Total?

What Store Cities generate the most Revenue Total for Washington Publishing House?

What Store States generate the most Revenue Total?

What Publisher House City generates the most Revenue Total?

What is the YTD sales by Title for Washington Publishing House?

What Publisher House Country generates the most Revenue Total?

What Title Type generates the most Revenue in a specific Store (e.g. Bookbeat)?

What are the Titles in the Top Quartile for Revenue Total? & Bottom Quartile?

What are the Top Titles sold by State in Revenue Total?

What is the best-selling Title by Title Type by Revenue Total?

What is the best-selling Title purchased by each store, by Revenue Total?

What is the best performing Title in terms of Gross Revenue? What was the maximum/minimum revenue value in any order?

What Stores generate the most Gross Revenue?

What is the Discount Value for each Store?

Which Title is most ordered? What were the minimum and maximum order quantities per Title?

Which TitleType is most ordered? What were the minimum and maximum order quantities per Title?

Which TitleType is ordered by day of the week – ordered by the day of the week with the most orders?

What quantity of books did each Store buy?

What quantity of books were purchased in each State?

What quantity of books were purchased in each State by Quarter and Year?

What quantity of books did each Publisher Country sells?

What quantity of books did each Publisher City sell during a particular week this year?

What TitleType generates the most Revenue in a specific Store (e.g. Bookbeat)?

How many Titles are there in each Type?

Which Stores changed location, ordered by descending date? Report showing all the changes to Store Locations in most recent date order. DATA WAREHOUSE

# Part 2: Data Modelling

## Overview

Using Kimble’s 4-step process, a data warehouse was designed.

## Step 1 – Business Process

**Sales** is identified as a business process.

## Step 2 – Grain

The level of granularity was identified as *TitleName*, which is the lowest atomic level.

## Step 3 – Facts

The Facts that we wish to measure to assess the performance of the business are:

Quantity/volumes:

* Quantity

Financial:

* *GrossTotal* - is the product of ‘Qty’ and ‘Price’ (enhanced fact).
* *DiscountTotal* - is the product of ‘Qty’ and ‘Discount’ (enhanced fact).
* *NetTotal (AKA Revenue Total)* - is the product of ‘Qty’ and ‘Price’ less the product of ‘Qty’ and ‘Discount’ (enhanced fact).

## Step 4 – Dimensions

The Dimensions are Title, Store and Calendar. These are what we measure the facts by, how we slice the data.

## Star Schema

The star schema below shows the design of the process and all the associated Facts and Dimensions.



DATA WAREHOUSE

Tracking Changes within the Dimension Tables

|  |  |  |  |
| --- | --- | --- | --- |
| Dimension | Attribute | Type | Comments |
| Title | TitleName | 1 | The title of a book will not be changed. If it’s an updated version, such as a second edition, it’s a different book so will have its own TitleID in any case. |
| Title | TitleType | 1 | It may be categorized but there’s no need to retain what type it was previously. |
| Title | TitlePrice | 3 | The price will rarely change. |
| Title | PublisherCity | 3 | The location will change rarely, if at all. |
| Title | PublisherCountry | 3 | The location will change rarely, if at all. |
| Store | All attributes | 3 | All attributes will rarely change, less than annually, but the most recent change should be retained. |
| Calendar | All attributes | 1 | All attributes will not change as these are generated from the original order date. |

Type 1 - *No history to be kept, simply overwritten.*

Type 3 - *Previous value retained along with the current value.*

# Part 3: ETL

Create the Data Warehouse using SQL Server Management Studio 2018

See the folder SQL.zip for the SQL files as listed and execute.

* Create\_DW.sql - *to build the data warehouse*
* Update\_Publisher.sql - *for changes to be made to the original database to see Type 1 and 3 changes.*

## Run ETL within Visual Studio 2018 - SSIS

See the folder SSIS.zip for the SSIS files as listed to extract, transform and load the data from the Publisher operational database into the Dimension and Fact tables of the schema.

* Calendar\_Type1.dtsx
* Store\_Type13.dtsx
* Title\_Type13.dtsx
* Fact.dtsx

# Part 4: Reporting

Run the SQL Views to generate reports for the Sales Process

See the folder for the following SQL file:

* SQL\_reports.sql

This contains all Views and Select statements to call each View, providing reports to answer the above questions.