# CA270 Assignment 4 report – Joseph Oluwasanya

**Domain:** Gaming (Game development company)

In this project, I want to design a data warehouse schema for a game company and use it to answer some common questions that a business user or analyst may have. *The company wants to understand the trends in order periods and order amounts to make better estimates on the numbers of physical copies of games needed to be produced. They also want to know what promotional conditions are yielding the highest profits.*

**Section 1: Warehouse Design**

The process I want to model is the order transactions for the company’s products. I will develop a **Star Schema** for orders process with a **separate periodic snapshot fact table** for monthly summary statistics as a pre-computing optimisation.

The monthly snapshot is for total orders of each product per buyer and its facts are computed as aggregates from the transaction fact table (**Rollup on date**). The most useful observations in this table are orders made by stores.

**Order fact table**

**Note:** the below data lattice cube diagram shows the structure of the warehouse concerning the order transactions fact table and not the periodic monthly snapshot fact table.

* **Grain**: one row for an individual order of a particular product on a date to a customer in a town in a state/province/county in a named Country.

**(**conceptual hierarchies: Country -> … -> Town, Year -> … ->Day**)**

* **Dimensions:** product, date, buyer, promotion
* **Measured facts:** units\_sold, sale\_price, production\_cost
* **Computed facts:** total\_profit

**Monthly orders fact table**

* **Grain:** one row for a months’ worth of orders of a particular product by a particular buyer.
* **Dimensions:** product, date, buyer. **(augmented date** dimension table. Instead of drawing another date table, I used parentheses to surround the attributes that are not present in the monthly orders data
* **Computed facts:** all

Diagram

Description automatically generated

Chart, radar chart

Description automatically generated

**Section 2: OLAP query specifications**

**QUERY 1**

The company is trying to extract information on the trends of orders for their games. Some games regain popularity seasonally and as well as this, unseen trends in sales data often arise. Having good estimates for production quotas per period will save money for the company and further optimise the production process of physical copies. The first OLAP query I will write will extract the monthly orders data for physical copies of games by stores. (That is, total units sold, total production cost, total sales and total profit per product) I will pull this data from the monthly snapshot fact table. Each observation will be the total monthly orders of physical copies of a game made by the store. The output table of this query will have many attributes to accommodate for further analysis; (how many games in total were bought by a store each month? how many copies of a particular game in Spanish were bought in 2020?) The precomputing monthly orders table makes this query run fast. Also, the use of conforming dimensions allows for cleaner joining syntax:

“a JOIN b USING (key)”

**Facts required:**

total\_units\_sold, total\_production\_cost, total\_month\_profit, total\_sales

**Dimension attributes required:**

Product – product\_name, game\_category, game\_rating, platform

Date – year, month

Buyer – buyer\_name, country, state, town, primary\_language

**QUERY 2**

The second query I will write will extract a table showing the total units sold during and total profit made by each promotional condition. (Note that this is irrespective of the games that were included in the promotion. Information on this is provided in the promo\_details field of promotion dimension) With this, the company can gain insight on what promotions yield the highest returns and adapt the best working strategies into future promotion events.

**Facts required:**

total\_profit, total\_units\_sold

**Dimension attributes required:**

Promotion – promo\_name, promo\_start, promo\_end, promo\_details

**Data cube interpretation of the queries**

A 3-Dimensional (product, buyer, date) cuboid is being extracted in QUERY 1 with aggregated facts. (This is essentially the makeup of the monthly orders snapshot)

A 1-Dimensional (promotion) cuboid is being extracted in QUERY 2 with aggregated facts. (Shown in QUERY 2 code below)

**Section 3: Coding the queries**

**Text

Description automatically generatedQUERY1:**

**Text

Description automatically generatedQUERY2:**

I talked about dimensionality and the structure of the cube in section 2 above.