

University of Pisa

Master's degree in
Data Science & Business Informatics
Big Data Analytics



Project Assignment - Part 2

Group 17

Carlo Alberto Carrucciu (533967), Giacomo Lo Dico (600002),
Gian Maria Pandolfi (607268).

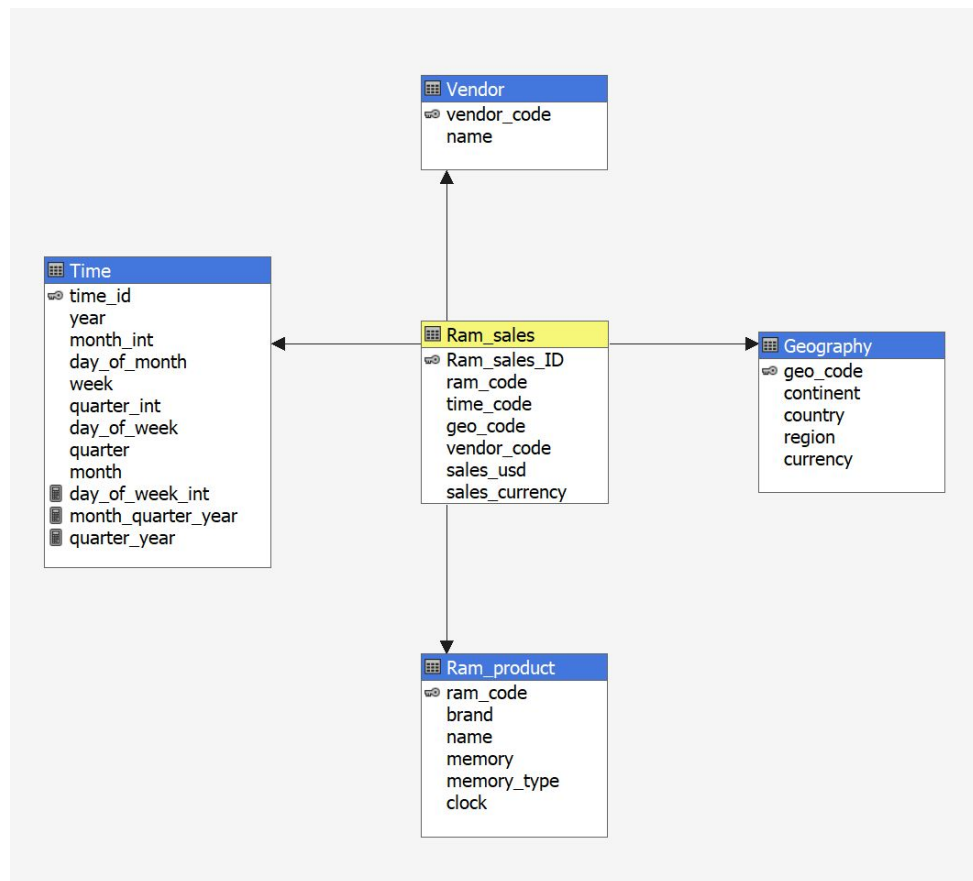
Costruzione cubo

Per la creazione del cubo, prima di tutto ci siamo connessi alla sorgente dati e selezionato le tabelle di nostro interesse per rispondere alle business question. Quindi, abbiamo creato una vista dell'origine dati con le tabelle: **Ram_sales, Time, Vendor, Ram_product e Geography**. Inoltre, sulla tabella **Time** abbiamo generato 3 colonne derivate che utilizzeremo per la creazione delle dimensioni.

Abbiamo generato una dimensione per ciascuna tabelle, tranne che per **Ram_sales** che è la nostra tabella dei fatti.

Per la dimensione **Time** abbiamo generato due gerarchie, la prima è stata generata in modo tale da ottimizzare i tempi di accesso ai dati e mette in relazione i seguenti attributi: **time_id, month_quarter_year, quarter_year e year**. La seconda gerarchia è composta dagli attributi: **day_of_week, quarter e week**. Inoltre, abbiamo aggiunto delle relazioni ai vari attributi che ci permettono di avere un corretto ordinamento, per esempio la relazione tra **month** (January, February, ...) e **month_int** (1, 2, ...). Per la dimensione **Geography** abbiamo generato una gerarchia tra gli attributi: **Region, Country e Continent**. Per la dimensione **Ram Product**, abbiamo creato una relazione gerarchica tra **brand** e **memory_name**. Mentre, per la dimensione **Vendor** non sono state generate delle gerarchie.

Una volta create le 4 dimensioni, abbiamo creato il cubo sulla tabella dei fatti **Ram_sales** utilizzando come metriche **sales_usd** e **sales_count**



Query MDX

Ora che il cubo è completo, siamo liberi di interrogarlo utilizzando il linguaggio per le query **MDX**, per poter ricavare informazioni specifiche sui dati salvati.

Per farlo, è prima però necessario connettersi al cubo *olap* attraverso *Microsoft SQL Server Management Studio*.

- ❑ Show the percentage increase in total sales with respect to the previous month for each ram brand and each country.
- ❑ Per risolvere questa query abbiamo utilizzato la funzione **LAG**, che permette di confrontare un record con quello precedente (o 'laggato' di un tot. di righe), e di ottenere quindi la differenza tra i membri desiderati.

- ❑ For each region and ram brand show the total sales in percentage with respect to the total sales of the corresponding country.
- ❑ La principale difficoltà nella risoluzione di questa query è confrontare livelli diversi della stessa gerarchia. Questo è stato possibile utilizzando nella definizione del primo membro, la sintassi **currentmember.parent** che ha permesso di ricavare il primo parente dei membri richiamati nella principale.

- ❑ Show the ram memory types having a total sales greater than 10% of the totals sales in each continent by continent and year.
- ❑ Rispetto le altre due query, per questa è stato fondamentale cimentarsi con l'utilizzo della funzione **filter**. Abbiamo dato come argomento le righe a cui eravamo effettivamente interessati, mentre il filtro vero e proprio è stato definito su un membro precedentemente calcolato.

Graphic dossiers:

Come richiesto negli “*Assignment*” da 4 a 6, sono state realizzate delle specifiche dashboard interattive relative alle vendite del prodotto “RAM”.

Si è preferito utilizzare il software *Desktop* di *Microstrategy* in alternativa a *PowerBI*, il quale ci ha permesso di esplorare i dati seguendo un approccio intuitivo e moderno.

Come prima cosa, abbiamo collegato il software alla fonte del nostro database su server Microsoft attraverso un plug-in di default che implementa l'interfaccia client-server, nello specifico abbiamo usato il servizio di **Azure SQL Database** per la connessione su una nuova origine, specificando ovviamente le credenziali di accesso al server e il numero di porta per abilitare la connessione.

Da qui abbiamo creato una nuova origine dati dalla quale abbiamo importato i dati relativi alle sole tabelle di interesse che sono: **Ram_sales, Ram_product, Vendor, Time e Geography**. In seguito, solo per alcune di esse abbiamo modificato diverse colonne specifiche (attraverso l'uso di tools integrati) per cambiare ed adattare informazioni utili per le varie rappresentazioni grafiche.

Inoltre si è deciso di ampliare le librerie che permettono di specificare la modalità di visualizzazione con altre sviluppate per integrare ed espandere dell'ambiente *microstrategy*.

È stato creato un Dossier unico suddivisi in 4 capitoli:

- **Time & Geo**, all'interno del quale abbiamo rappresentato le pagine: **Calendar & KPI**, che prevede due quadranti completamente separati ma che riportano informazioni complementari che danno un quadro completo sulla situazione delle vendite durante l'anno; **Sales over Time**, suddivisa in 2 piani collegati tra di loro, nei quali vengono mostrati i “Sales by quarters” basati sulle vendite, mentre nel piano inferiore, un grafico a barre che mostra le vendite totali suddivise per workdays e weekend; in fine per l'ultima pagina viene inserita **Country Trend**, la quale mostra su tre livelli diversi la tendenza delle vendite in base al periodo temporale, con un filtering per ogni tasto nel quadrante principale.
- **Brand**, nel quale si vuole mostrare l'andamento dinamico delle vendita in base ai vari marchi che producono ram. Nella prima pagina vediamo: **Brand Race** dove viene mostrata l'evoluzione dinamica delle vendite per ogni brand in base allo scorrere del tempo. Invece nella seconda pagina vediamo una **Sales Map** che mostrata una mappa colorata in base ai guadagni che generano i vari brand, inoltre è possibile utilizzare un click bidirezionale sui quadranti a “torta” e a “barre” per avere un quadro più specifico sulle vendite di ogni stato.

- In **Vendor** dove includiamo una sola pagina ovvero **Top Vendors**, dove indichiamo una panoramica per tutti i venditori per i quali vediamo nel dettaglio l'ammontare delle vendite realizzate in base agli anni (con un grafico a "scatola"); servendosi invece di un grafico a "torta" sull'attributo *geografico*, si possono vedere i 10 top vendors (in base alle vendite totali realizzate su un bar chart verticale) basato sul periodo di vendita settimanale.
- Concludiamo con il capitolo **Memory**, per il quale abbiamo due pagine: **Type Trends**, per la quale vengono visualizzate una line chart in rappresentanza del tipo di "Memory RAM" venduta in un determinato periodo (mese e anno) in relazione alle entrate generate abbinata ad una column chart che permette di vedere le stesse informazioni ma con un aspetto diverso e per ultimo diamo la possibilità di vedere le informazioni di vendita con un column chart basato sulle caratteristiche più importanti della RAM. **Rams: Sales & Sold**, implementa la visualizzazione contemporanea di due grafici a bolle (uno in base alla velocità della ram e l'altro sulla versione) entrambi correlati a due grafici che indicano rispettivamente la quantità di denaro e di prodotto venduto in base alla versione della ram (caratteristica più rilevante), in fine il tutto viene regolato da due filtri sull'anno e sul continente geografico.

In tutti le pagine prima menzionate vengono implementati dei filtri utili per gestire la granularità delle informazioni o per focalizzare un aspetto più specifico, sostanzialmente sono stati importati due categorie di filtri diversi, una tipologia che rimane fissa sulla parte esterna del capitolo e permette di selezionare i dati degli attributi letti direttamente dal database. La seconda categoria di filtro viene direttamente implementata come quadrante integrale della visualizzazione, per cui permette di personalizzare specifiche sulla granularità delle informazioni visualizzati in quel momento.

