**LEZIONE 1**

**OLAP:**

Un insieme di metodi e strumenti per l'analisi interattiva dei dati multidimensionali (OLAP - OnLine Analytical Processing) utilizzati per comprendere ed analizzare le performance aziendali al fine di ottenere informazioni utili (conoscenza) a supporto della presa decisioni non strutturata.

,

**DATA SOURCES:** EXETERNAL DATA SOURCES. DB OPERAZIONEALE

**DATA MOVEMENT STREAMING ENGINE:** **ETL**(EXTRACT, TRANSFORM, LOAD), COMPLEX EVENT PROCESSING ENGINE

**DATAWAREHOUSE SERVERS:**  RELATIONAL DBMS+ MAP REDUCE ENGINE

**MID-TIER SERVERS:** OLAP SERVER, ENTERPRISE SEARCH ENGINE, DATA MINING TEXT ANALYTIC ENGINES, REPORTING SERVER

**FRONT END APPLICATIONS:** SEARCH, SPREADSHEET, DASHBOASRD, AD HOC QUERY

**DATAWAREHOUSE**

Il data warehousing è il processo di portare i dati dalle fonti operative (OLTP) in un unico data warehouse per l'analisi (OLAP) con applicazioni di Business Intelligence

**FATTI**

Un fatto è un'osservazione delle prestazioni di un processo aziendale (oggetto di analisi) (ad esempio, una vendita).

**MISURE**

ATTIBUTI NUMERICI DI UN FATTO (ES. QUANTITY, PROFITTO), Ciò CHE HA SENSO AGGREGARE, PER ESEMPIO N. ORDINE NO.

**DIMENSIONI**

LA DIMENSIONE è IL CONTENSTO DEI FATTI (ES. PRODOTTI, MARKET, DATA)

**METRICHE**

KPI (KEY PERFORMANCE INDICATORS)

SONO I DATI AGGREGATI, (SOMMA, MEDIA, MINIMO, MASSIMO) DELLE MISURE CIRCA UN INSIEME DI FATTI DETERMINATO DA CONDIZIONI DSULLE DIMENSIONI. È Ciò CHE INTERESSA AI MANAGERS.

**GERARCHIE**

gli attributi di una dimensione possono essere correlati tramite una gerarchia di relazioni. Le gerarchie degli attributi sono un modo naturale per supportare l’esplorazione interattiva dei gatti. Gli utenti le comprendono intuitivamente perché sono abituati a guardare un rapporto riassunto e poi decidere di guardare un più dettagliato.

**DIMENSIONAL MODEL (MULTIDIMENSIONAL MODEL= CUBO)**

In questo modello, i dati sono strutturati in dimensioni e fatti. Le dimensioni rappresentano le categorie o gli aspetti dei dati che sono di interesse per l'analisi, mentre i fatti sono le misure numeriche o quantitative che vengono analizzate rispetto alle dimensioni.

l modello dimensionale facilita l'analisi dei dati attraverso l'utilizzo di cubi OLAP (OnLine Analytical Processing), che consentono di eseguire operazioni come il trapano (drill-down) per esaminare i dati a livelli di dettaglio più bassi o il roll-up per aggregare i dati a livelli più alti. Questo modello è ampiamente utilizzato nei data warehouse per supportare le esigenze di reporting e analisi dei dati aziendali.

**6 WS**

**SNOWFLAKE CHEMA**

**CONSTELLATION SCHEMA**

**CAMBI DI ATTRIBUTI DIMENSIONALI (ES. CUSTOMER ADDRESS):**

* overwrite the value(slow)
* add a dimension row (slow)
* preserving one or more verision of hisory (not recommended)
* add a new dimension (mini o profile) (fast)

**LEZIONE 2**

STAR JOIN

1)SRUTTURA DB DATA WAREHOUSE SU DBMS RELAZIONALE. SCHEMA RELAZIONE

TIPOLOGIA DI USO PRIVILEGIA SEMPLICITA DI QUERI CHE SONO MASSIVE E NON OPERAZIONALI.

NELL’ANALISI MULTIDIMENSIONALI AGGREGAIZONE SU GRANDI QU DI DATI . SEMPLICITà DI SCRITTURA RISPETTO A RIDONDANZA DEI DATI. I DATI SONO STORICI E NON VENGONO CAMBIATI.

2) UTENTE TECNICO CHE VUOLE COSTRUIRE REPORT USANDO Q SWL SU SCHEMA SNOWFLAKE DI FOODMAR.

**SQL.** LING. DI INTERROGAZIONE MANIPOLAZIONE E CREAIZIONE DI DATI PER DB RELAZIONALI.

**MARGIN BY BRAND AND BY PRODUCT YEAR 2009**

SELECT BRAND, PRODUCT, SUM(REVENUE) AS REVENUE,

SUM(MARGIN) AS MARGIN

ROUND(100\*SUM(MARGIN)/SUM(REVENUE))) AS MARGIN%

FROM SALES

WHERE YEAR(DATE)=2009

GROUP BY BRAND, PRODUCT

ORDER BY BRAND, PRODUCT

**SU FOOD MART:**

SELECT COUNTRY, CITYM SUM (STORE SALES ) AS TOT SALES

FROM SALES FACT, CUSTOMER

WHERE SCUSTOMERID= CUSTOMER ID

GROUP BY COUNRY, CITY

**GROUP BY ROLLUP:**

**TOT SALES BY COUNTRY AND CITY WITH SUBTOTAL**

(SELECT COUNTRY, CITYM SUM (STORE SALES ) AS TOT SALES

FROM SALES FACT, CUSTOMER

WHERE SCUSTOMERID= CUSTOMER ID

GROUP BY COUNRY, CITY)

**UNION ALL**

(SELECT COUNTRY, NULL AS CITY, SUM (STORE SALES ) AS TOT SALES

FROM SALES FACT, CUSTOMER

WHERE SCUSTOMERID= CUSTOMER ID

GROUP BY COUNTRY)

**UNION ALL**

(SELECT NULL AS COUNTRY, NULL AS CITY, SUM (STORE SALES ) AS TOT SALES

FROM SALES FACT, CUSTOMER

WHERE SCUSTOMERID= CUSTOMER ID)

ORDER BY COUTNRY, CITY

**GROUP BY ROLLUP:**

(A,B)

1. SUBTOTALS

() TOTALS

SELECT COUNTRY, CITY, SUM (STORE SALES ) AS TOT SALES

FROM SALES FACT, CUSTOMER

WHERE SCUSTOMERID= CUSTOMER ID

**GROUP BY ROLLUP** (COUNRY, CITY)

ORDER BY COUNTRY, CITY

CROSS- TABULATION CON **GROUP BY CUBE**: (A,B) (A) (B) ()

SELECT COUNTRY, CITY, SUM (STORE SALES ) AS TOT SALES, **GROUPING** (COUNTRY), **GROUPING** (CITY) \* AGGIUNGE COLONNA CON 1/0 1= NON STA RAGGRUPPANDO PER QUELLO SCELTO, 0= STA RAGGRUPPANDO PER QUELLO.

FROM SALES FACT, CUSTOMER

WHERE SCUSTOMERID= CUSTOMER ID

**GROUP BY CUBE** (COUNRY, CITY)

ORDER BY **GROUPING** (COUNTRY), COUNTRY, **GROUPING** (CITY), CITY

**GROUPING (A)=1 QUANDO A NON è PRESENTE NELLA GROUP BY GENERATA DAL ROLLUP**

**SQL ANALITICO:**

**COUNT (\*) OVER (PARTITION BY B) AS N. OF….**

WITH TEMP AS (

SELECT COUNTRY, CITY, SUM (STORE SALES ) AS TOT SALES

FROM SALES FACT, CUSTOMER

WHERE SCUSTOMERID= CUSTOMER ID

GROUP BY COUNTRY, CITY

)

SELECT COUNTRY, CITY, TOT SALES,

SUM(TOTSALES) OVER (PARTIION BY COUNTRY) AS TOT SALES COUNTRY

FROM TEMP

ORDER BY COUNTRY

1:26

**RANK**

**5 MIGLIORI PRODOTTI:**

**SELECT** Customer, Product, SUM(Revenue) **AS**TotalRev,

**FROM** Sales

**WHERE** Customer **IN**(‘C1’, ‘C2’)

**GROUP BY** Customer, Product **ORDER BY** TotalRev **DESC**;

**SELECT** Customer, Product, SUM(Revenue) **AS** TotalRevenue,

**RANK** ( ) **OVER** (**PARTITION BY** Customer **ORDER BY** SUM(Revenue) **DESC**) **AS** Rank

**FROM** Sales

**WHERE** Customer **IN** (‘C1’, ‘C2’)

**GROUP BY** Customer, Product;

RANK: (1;2;2;4;4;6)

DENSE RANK: (1;2;2;3;3;4)

ROW NUMBER : (1;2;3;4;5;6)

**LAG (ROW BEFORE)**

**LEAD (FOW AFTER)**

**WINDOWING** A moving average of total revenue, with a moving window of 3 months, by month

**LEZIONE 3**

**OVER (partition by A )**

**COUNT(\*) OVER (partition by A )**

**RANK () OVER (partition by A order by E)**

**LEAD (B,1) OVER (partition by A order by E)**

**Calcolare DELTA come diff. Di revenue rispetto al II anno**

**Re. Di anno, rev di anno prima e join.**

**FULL OUTER JOIN ( riporta anche i null) > varianti LEFT OUTER JOIN / RIGHT OUTER JOIN**

**A product may have been sold in one year, but not in the other !**

**SOLUTION USING LAG-LEAD (and NO JOIN)**

WITH temp AS (

SELECT Brand, Product, Year(Date) AS Year, SUM(Revenue) AS Revenue

FROM Sales

WHERE Year(Date) IN (2008, 2009)

GROUP BY Brand, Product, Year(Date)

),

laglead AS (

SELECT Brand, Product, Year, Revenue,

LAG(Revenue, 1, 0) OVER(PARTITION BY Brand, Product ORDER BY Year) AS PrevR,

LEAD(Revenue) OVER(PARTITION BY Brand, Product ORDER BY Year) AS NextR

FROM temp

)

SELECT Brand, Product, Revenue AS Revenue2009, PrevR AS Revenue2008,

CASE WHEN Year = 2008 THEN -100

ELSE Round(100\*(Revenue-PrevR)/Revenue) END AS Delta

FROM laglead

WHERE Year = 2009 OR (Year = 2008 AND NextR IS NULL)

ORDER BY Brand, Product

**LEZIONE 4**

GAIN= PER UN SINGOLO FATTO è (DIFFERENZA TRA STORE\_SALES E SORE\_COST) MOLTIPLICATO PER UNIT\_SALES

GAIN PER UN SINGOLO PRODOTTO= SOMMA DEI GAIN PER TUTTE LE VENDITE DI QUEL PRODOTTO

PRODURRE UN FILE CSV CON I PRODOTTI ORDINATI EPR GAIN DISCENDENTE ( DAL + VENDUTO AL \*)

***SALES\_FACT***

***TRASFORMAZIONI PER CALCOLARE GAIN DI SING VENDITA, AGGREGARE GAIN PER PRODOTTO, ORDINARE PER GAIN DISCENDENTE.***

IN OUT UT. PRODUCT NAME (ANCEH DI TAB. PRODUCT

* PIPELINE CON LOOK UP
* STRATIFIED SAMPLING Conditional split
* GESTIONE SLOWLY CHANGING DIMENSIONS (

TABELLA DIMENSIONALE SU CUI COPIARE DATI DI OPERATIONAL CUSTOMER DIM

1. CREARE TABELLA DIMENSIONALE SU DBEAVER(UNO STESSO CUSTOMER Può AVERE + ID A SECONDA DI QUANTI INDIRIZZI)
2. CREARE FLUSSO SISS CHE LEGGA DA TAB OPERAZ E AGGIUNGA A QUELLA DIM CLIENTI CHE NON CI SONO.

ALL’INIZIO è VUOTA E DEVE AGGIUNGERE TUTTI.

**LEZIONE 5**

POPOLAMENTO PERIODICO DI DATI

1. SE E’ PRES NELLA OPERAZ E NON NELLA DIM LO INSERISCE
2. SE UNO CAMBIA NOME NELLA OPERAZ VA SOSTITUITO CON NUOVO NOME
3. CREARE CHIAVE SURROGATA CON NUOVO INDIRIZZO

TRAFORMAZIONI NUOVE COLONNE:

FULL NAME

FULL ADDRESS

SURROGATE\_KEY, COLONNE AUTOINCREMENTANTI

* INSERIRE NELLA TABL DIM. CLIENTI CHE NON CI SONO (TUTTI LA PRIMA VOLTA)

DEST. DI FLUSSO: OLEDB DESTINATION

36.19