



Università della Calabria

Dipartimento di Matematica e Informatica
Corso di laurea in Informatica

Data Warehouse Project
CORDIS - EU research 2007-2013

Giovanni Brunetti
Matricola:193452

Anno accademico 2017/2018

Introduzione

Lo scopo del progetto è quello di costruire un sistema di Data Warehouse basato sui dati relativi all'assegnazione di progetti scientifici da parte dell'UE alle varie organizzazioni mondiali.

Per far ciò si provvederà alla creazione del/dei data-mart necessari per i fatti di interesse in modo da permettere lo svolgersi delle analisi OLAP su di essi. La progettazione ha seguito le varie fasi relativi all'approccio **source oriented**, includendo lo studio delle sorgenti, comprendendo operazioni di pulizia e di trasformazione dei dati per comprendere i dati ed il dominio di interesse. In seguito si è passato alla fase di alimentazione del database e dello studio dei dati ottenuti tramite analisi.

Per la fase di studio ,di ETL ed alimentazione del DB si è utilizzato il software Pentaho, mentre per lo studio dei dati e l'effettuazione delle analisi si è utilizzato il software Tableau.

Sorgente dei dati

La sorgente dei dati è reperibile all' url "https://data.europa.eu/euodp/data/dataset/cordisfp7projects" sotto forma di file csv.

La sorgente è composta da 6 file, essi contengono tutte le informazioni relativi ad ogni singolo progetto assegnato dal 2013 al 2017. La prima fase del progetto si è dunque concentrata nello studio di tutti i file per capire come essi erano connessi tra di loro. Essi infatti presentavano molte ripetizioni di attributi nei diversi file. Per ciascuna delle assegnazioni le principali informazioni espresse e rilevanti del il data mart sono:

- nome del progetto;
- le organizzazioni coinvolte nel progetto;
- la nazionalità di ogni organizzaione;
- programma di appartenenza del progetto;
- contributo economico dell'ue per ogni organizzazione;
- data di inizio e di fine.