

1. Progettazione Data Warehouse

Il Ministero dell'Università e della Ricerca vuole svolgere un'analisi del tempo dedicato dai ricercatori ai vari progetti di ricerca. Ogni ora dedicata alle attività di un progetto di ricerca da parte di un ricercatore ha un corrispettivo costo orario nominale, ottenuto dividendo lo stipendio annuale dello specifico ricercatore per il suo monte ore nominale annuo.

Ogni progetto di ricerca è costituito da insiemi disgiunti di attività denominati Work Package. Le attività di un progetto di ricerca appartengono ad uno dei seguenti tre tipi: Management, Ricerca & Sviluppo, Altro. Ogni Work Package raggruppa attività omogenee tra loro per tipo.

I progetti di ricerca sono suddivisi per argomento (es. Big Data, Smart City); ogni argomento appartiene ad una disciplina (es. informatica, urbanistica). Ogni progetto di ricerca è identificato da un nome, ha una durata prefissata (uno o più anni), e nasce in risposta ad un bando di finanziamento. Ogni bando ha una specifica modalità di finanziamento (es. fondo perduto, cofinanziamento, venture capital), è pubblicato da uno specifico ente finanziatore (es. Comunità Europea, Ministero, fondazioni), e appartiene ad uno specifico settore (es. ICT, Nanotechnologies, Energy).

I singoli ricercatori compilano ogni mese un resoconto, denominato timesheet, con l'indicazione di quante ore di attività hanno svolto per ogni Work Package di ogni progetto nel quale sono coinvolti. I singoli ricercatori appartengono a delle Unità di Ricerca (es. i dipartimenti universitari), le quali a loro volta appartengono ad uno specifico Ente di Ricerca (es. Politecnico di Torino, Istituto Superiore Mario Boella).

Il Ministero vuole analizzare il numero totale di ore, il costo totale, e il costo medio orario, al variare di:

- progetto e relativi Work Packages;
- tipo di attività, argomento del progetto e disciplina coinvolta;
- durata del progetto (i valori di interesse sono 1, 2, 3, 4, 5, o pi anni);
- bando e settore del bando;
- modalità di finanziamento ed ente finanziatore;
- Unità di Ricerca ed Ente di Ricerca;
- mese, mese dell'anno, bimestre, trimestre, semestre, anno, anno accademico (inizia a settembre, finisce ad agosto).

Progettazione

- (a) (6 Punti) Progettare il data warehouse (modello concettuale e tabelle dei fatti e delle dimensioni) in modo da soddisfare le richieste descritte nelle specifiche del problema. Il data warehouse progettato deve inoltre permettere di rispondere in modo efficiente alle seguenti interrogazioni.
- (b) (8 Punti) Esprimere le due interrogazioni seguenti utilizzando il linguaggio SQL esteso. Indicare la risposta nello spazio sottostante ciascuna domanda.

- (a) Per ogni bimestre e per ogni progetto, calcolare il costo medio orario e il numero medio di ore per Work Package. Calcolare inoltre il cumulativo per ogni anno accademico delle ore totali bimestrali e dei costi totali bimestrali, separatamente per ogni progetto.

```
SELECT PROJECT, 2MONTH, ACADEMIC_YEAR,
SUM(EURO) / SUM(HOURS),
SUM(HOURS) / COUNT(DISTINCT WORK_PACKAGE),
( SUM(SUM(HOURS)), SUM(SUM(EURO)) )
  OVER (PARTITION BY ACADEMIC_YEAR, PROJECT
        ORDER BY 2MONTH
        ROWS UNBOUNDED PRECEDING )
WHERE <JOINS>
GROUP BY PROJECT, 2MONTH, ACADEMIC_YEAR
```

- (b) Calcolare la percentuale di ore che ogni progetto ha dedicato ai vari tipi di attività rispetto al totale di ogni progetto, la percentuale di ore di ogni Work Package rispetto al totale di ogni progetto, e il costo orario medio per ogni Work Package.

```
SELECT PROJECT, ACTIVITY_TYPE, WORK_PACKAGE
100 * SUM(SUM(HOURS)) OVER
  (PARTITION BY PROJECT, ACTIVITY_TYPE) /
  SUM(SUM(HOURS)) OVER
  (PARTITION BY PROJECT),
100 * SUM(HOURS) /
  SUM(SUM(HOURS)) OVER
  (PARTITION BY PROJECT),
SUM(EURO) / SUM(HOURS)
WHERE <JOINS>
GROUP BY PROJECT, ACTIVITY_TYPE, WORK_PACKAGE
```