



SSIS

[Laboratory for Data Science Project Pt2]

Authors: *Davide Ricci, Alessandro Mastrorilli*

Professors: *Roberto Pellungrini, Anna Monreale, Cristiano Landi*
Master's degree: *Data Science & Business Informatics*

Introduction

Lo scopo di questa seconda parte progettuale è quello di risolvere alcuni problemi interrogando il database creato nella prima parte. In particolare si richiede di risolvere 3 assignments utilizzando SQL Server Integration Services (SSIS) definendo 3 diversi pacchetti costituiti da opportuni control e data flow mediante i nodi messi a disposizione dal servizio. Si precisa che il risultato dei 3 assignments viene scritto in appositi file csv memorizzati nella cartella "Assignment_results" la quale contiene anche una sottocartella Assignment_errors in cui verranno memorizzati invece opportuni file di testo relativi agli errori scaturiti a tempo di esecuzione nei vari flussi di lavoro. Terminata l'esecuzione di ogni pacchetto si è deciso di aggiungere un nodo "Attività esegui processo" che va automaticamente ad aprire in locale l'eseguibile di Excel per mostrare il risultato dell'esecuzione del relativo pacchetto.

Assignment 0 - For every year, the participant ordered by the total number of custodies.

Per visualizzare meglio il risultato di questa interrogazione si è deciso di effettuare un ordinamento su year e successivamente su total custodies.

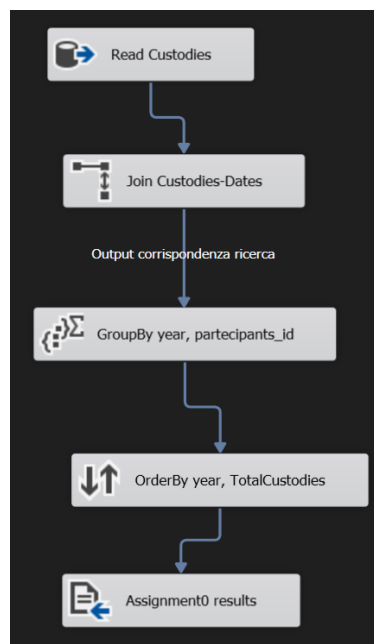


Figura 1: Dataflow assignment 0

Assignment 1 - For each state, compute the stolen gravity index defined as the ratio between the total gravity of custodies involving stolen guns divided by the overall gravity of custodies.

Per questo esercizio si è deciso di inserire un nodo che cattura potenziali errori derivanti dal rapporto per calcolare la metrica in esame. Si precisa che durante il calcolo del total gravity index, il denominatore è stato castato a float in modo da ottenere risultati più attendibili. Come anticipato precedentemente gli errori vengono all'occorrenza memorizzati in un opportuno file di testo nella cartella Assignment_errors.



Figura 2: Dataflow assignment 1

Assignment 2 - For each month, compute the total gravity in percentage with respect to the annual total.

La struttura di questo ultimo flusso di dati è molto simile a quella del precedente assignment. In particolare è stato inserito un opportuno nodo per catturare eventuali errori derivanti dal calcolo delle percentuali per ogni mese. Per una questione di visualizzazione si è deciso di ordinare il risultato prima per anno e poi per mese, ottenendo così una più chiara visualizzazione delle percentuali di total gravity, per ogni mese di ogni anno.



Figura 3: Dataflow assignment 2