TITOLO

Analisi degli incidenti registrati nella città di New York e correlazione con le condizioni meteorologiche nel periodo temporale 2010 - 2016

INTRODUZIONE

Obiettivo: implementazione di un data warehouse con incidenti e meteo nella città di new york

Le sorgenti dati di partenza sono:

Il flusso di lavoro è:

1. stabilire lo schema logico-relazionale delle sorgenti
2. reverse engineering per ottenere lo schema concettuale delle sorgenti dati
3. produrre una stima del carico di lavoro preliminare, un glossario dei requisiti ed effettuare un’analisi di valutazione della qualità dei dati disponibili
4. produrre uno schema concettuale riconciliato integrato dell’ODS (Operational Data Store), a partire dagli schemi concettuali delle sorgenti dati
5. definire lo schema logico-relazionale dell’ODS, a partire dallo schema concettuale riconciliato integrato dell’ODS
6. effettuare il processo di ETL (Extract, Transform, Load) dei dati dalle sorgenti operazionali all’ODS
7. effettuare la fase di progettazione concettuale del data warehouse, attraverso uno schema con il DFM (Dimensional Fact Model)
8. effettuare la fase di progettazione logica del data warehouse, producendo uno schema a stella oppure uno schema snowflake
9. effettuare il processo di ETL dei dati delle dimensioni e del fatto dall’ODS al data
10. analisi dei dati ottenuti
11. sviluppi futuri

CAPITOLO 1

SORGENTI DATI

1. Incidenti stradali

https://data.cityofnewyork.us/Public-Safety/NYPD-Motor-Vehicle-Collisions/h9gi-nx95

1. Condizioni meteo

https://www.wunderground.com/history/airport/KNYC/2018/4/12/DailyHistory.html?req\_city=New+York&req\_state=NY&req\_statename=New+York&reqdb.zip=10001&reqdb.magic=8&reqdb.wmo=99999