

PROGETTO MAADB

AZZURRA OBERTO

SOFIA TRUCCO



ARCHITETTURA DEL SISTEMA

- Main_Server
- MongoDB_Server
- Neo4J_Server

DATASET E PREPROCESSING

- I dati sono stati generati attraverso `ldbc_snbdatagen_spark`
- Prima di caricare i dati nei database, abbiamo pulito i dati, in particolare:
 - Standardizzazione del formato datetime
 - Conversione dei tipi di attributi per garantire la compatibilità

DISTRIBUZIONE DATI: MONGO DB

ENTITA':

- Comment
- Forum
- Organisation
- Person
- Place
- Post
- Tag
- TagClass

RELAZIONI:

- comment_hasTag_tag
- forum_hasTag_tag
- organisation_isLocatedIn_place
- person_hasInterest_tag
- person_isLocatedIn_city
- place_isPartOf_place
- post_hasTag_tag
- tagClass_isSubClassOf_tagClass
- tag_hasType_tagClass

DISTRIBUZIONE DATI: NEO4J

ENTITA':

- Comment
- Forum
- Organisation
- Person
- Post

RELAZIONI:

- comment_hasCreator_person
- comment_replyOf_comment
- comment_replyOf_post
- forum_containerOf_psot
- forum_hasMember_person
- forum_hasModerator_person
- person_knows_person
- person_likes_comment
- person_likes_post
- post_hasCreator_person
- person_studyAt_university
- person_workAt_Company

QUERY ANALITICHE

1. “Trovare la correlazione tra il numero di amici di una persona ed il numero di post che ha creato.” -> Neo4J
2. “Trovare la media di reply ai commenti.” -> Neo4J
3. “Trovare l’età media del genere femminile che frequenta l’università e l’età media del genere femminile che lavora in azienda.” -> Cross-database

QUERY PARAMETRICHE

1. “Trovare i nomi dei 3 forum più seguiti dagli universitari dell’università X.”
-> Neo4J
2. “Trovare tutti i giorni in cui una certa persona X ha messo like ad un commento in un certo range di date, utilizzando il browser X.”
-> Cross-database
3. “Trovare tutti i messaggi in cui è presente il tagClass X e ha location IP Y.”
-> MondoDB