Istruzioni per avviare il db mongo

1. Preparazione ambiente per MongoDB

Bisogna creare una cartella dove MongoDb andrà a tenere i dati. Apri un terminale facendo “Tasto Windows + r” e scrivendo “cmd” nella finestrella che si apre, poi nel terminale digita in quest’ordine:

* 1. cd /
  2. mkdir data
  3. cd data
  4. mkdir shard1\_1 shard1\_2 shard1\_3 shard2\_1 shard2\_2 shard2\_3 shard3\_1 shard3\_2 shard3\_3 config1 config2 config3

Ora si può chiudere il terminale.

1. Ora copia tutti i files nella cartella configuration files dal drive ed incollali in questa nuova cartella “data” (percorso: C:\data)
2. Scarica la cartella “Batch files (inizializzazione sharded DB)” e salvala da qualche parte, segnandoti il percorso.
3. Apri il primo script, “1) Avvio\_DBkiva\_sharded-def”, nella prima cella inserisci il percorso della cartella “Batch files (inizializzazione sharded DB)”
4. Lancia il primo script, tutte le celle tranne l’ultima. Ci metterà un po’ (circa 5 minuti) perchè ho inserito uno sleep per permettere al db di avviarsi prima di continuare.
5. Apri un terminale (come prima), poi incolla e lancia in quest’ordine:
   1. cd /
   2. cd C:\Program Files\MongoDB\Server\3.6\bin
   3. mongo --host localhost --port 27018
   4. sh.addShard( "kiva\_s1/localhost:27118")
   5. sh.addShard( "kiva\_s2/localhost:27218")
   6. sh.addShard( "kiva\_s3/localhost:27318")

Puoi chiudere il terminale una volta fatto questo.

1. Ora lancia l’ultima cella del primo script. Il db ora è inizializzato e pronto a ricevere i dati attraverso la porta 27018. Dovresti avere 12 terminali mongod aperti + 1 terminale mongos. Lasciali aperti perchè loro sono il DB.

1. Ricezione dati dalla World Bank.

Apri lo script “2) World\_bank\_data” e lancialo (ci sono pezzi di codice non utilizzati, ma tanto solo le celle utili sono segnate come codice, quindi avvia tutto senza problemi). Attenzione, alla fine devi inserire il percorso della cartella dove vuoi salvare i CSV sui dati economici.

1. Avvio download e salvataggio dei dati di Kiva.

Avvia lo script “3) Download&Salvataggio\_kiva-def” e lascialo fare. Questa è la parte lunga del lavoro, a me ha impiegato circa 48 ore (continue) a scaricare tutti i dati, quindi se vuoi crearti il db in locale dovrai lasciarlo andare per un bel po’.

1. Data quality (Seba, questa è la parte da controllare un po’).

Apri lo script “4) Data\_quality”

Due passaggi in questa fase:

* 1. Ricerca ed eliminazione loans con ID duplicato. Se fai andare lo script in una botta sola non dovrebbe capitare, ma se lo blocchi e lo fai ripartire potresti avere dei duplicati.
  2. Ricerca degli stati con nome diverso nel DB Kiva rispetto al nome nel CSV della World Bank e riempimento dei rispettivi campi dei loans. In pratica, se un loan ha il nome dello stato diverso rispetto alla World Bank, lo script (3) lascia un nan, mentre ora lo andiamo a riempire correttamente.

1. Queries sul DB

Apri lo script “5) Queries\_MongoDB\_Kiva”.

Lo script risolve le 3 queries indicate, ma c’è da fare un po’ di statistica sulla terza query per vedere se c’è qualche correlazione tra il GNI e l’ammontare dei loans.

Metto i CSV nella cartella “output\_queries”, così potete farlo anche senza creare il db sul vostro pc.