BIG DATA PROJECT PROPOSAL

Project topic #02

Radhika Malhotra (A20491601)
Aastha Dhir (A20468022)
Aman Singh (A20491333)
Rohit Kumar (A20501314)

Neo4j is the leading database for linked data that was designed from the ground up to leverage data and data connections. Neo4j is fundamentally distinct from other data stores since it is a native graph database. Neo4j is built around a straightforward yet effective optimization. Direct pointers to all the nodes that it is connected to are contained in each data record or node. Relationships are what we term these direct pointers. The node itself contains all the data required to identify the subsequent node in the sequence. A connected graph serves as the native storage layer. Native implies exactly that. Neo4j doesn't have to calculate the relationships between your data at query time as a result of this approach. The database already has the connections saved. This leads to orders of magnitude faster inquiries on highly connected data. Direct pointers between records are not saved in other databases. They must compute links by looking through an index, a different data structure. This lookup method is quite expensive and becomes exponentially slower as the number of data and the complexity of the queries increase. It must be repeated to discover each link. As a result, they are fundamentally slower for relationship-intensive queries than Neo4j. So, we decided to move forth with this topic.

We have chosen to visualize and increase our understanding of the 'Bollywood' industry via our graphs. Hollywood and Bombay were combined to create the hybrid name 'Bollywood'. The Hindi/Urdu film business is located in Bombay, India, as the name implies. The movies likewise cover a wide range of subjects; some of them even deal with delicate political and religious issues, while others are heartwarming. The happy dance scenes in many movies help to brighten the atmosphere. Bollywood actors and actresses are well-known worldwide because of their work in the film business. A Bollywood movie dataset compiled by P. Premkumar includes roughly 1700 Bollywood films from a two-year period and currently looks like a good choice of dataset for us.

REFERENCES:

- https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=7960078
- https://www.computer.org/csdl/proceedings-article/icis/2017/07960078/12OmNzBOi3I
- https://neo4j.com/developer/get-started/