**CSP595 - Assignment 12**

Name: Jagruti Vichare

Email ID: [jvichare@hawk.iit.edu](mailto:jvichare@hawk.iit.edu)

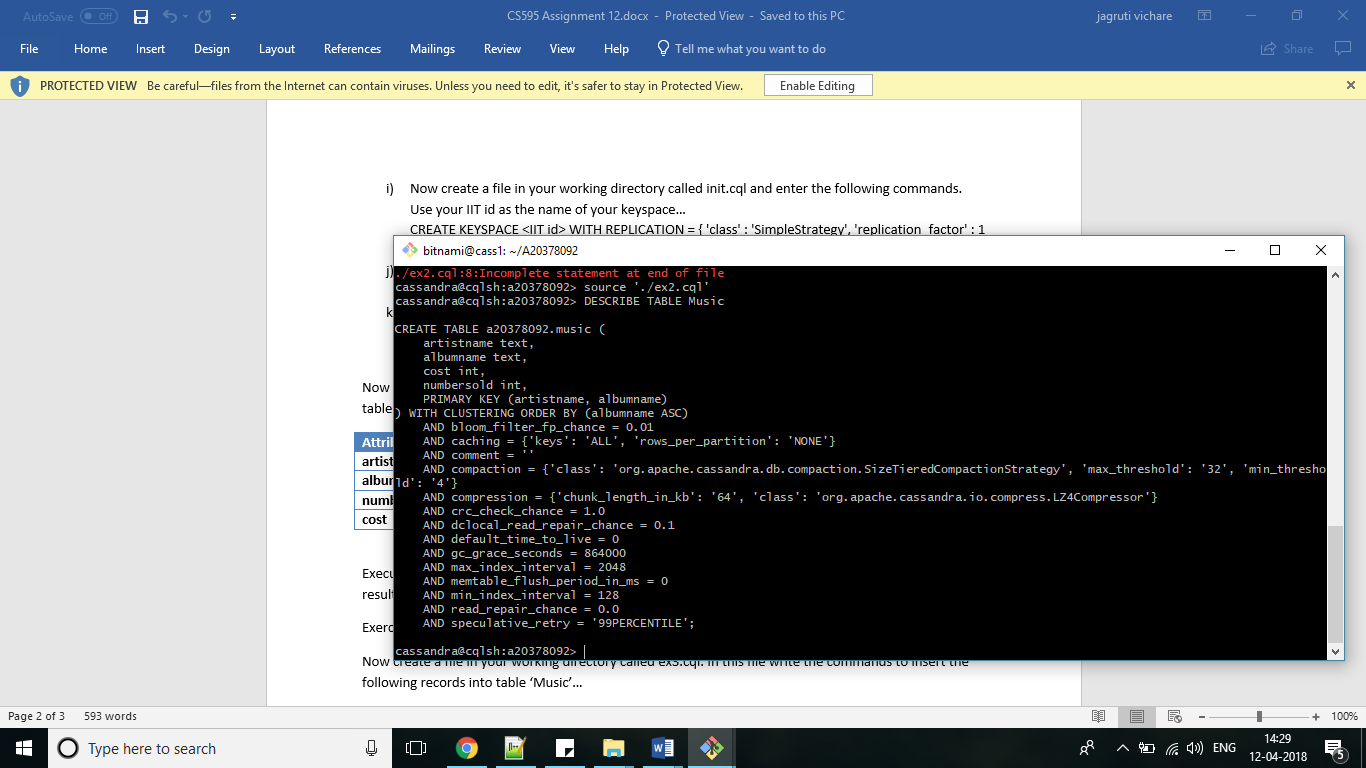
CWID: A20378092

**Exercise 1**

Apache Cassandra is a leading transactional, scalable, and highly-available distributed database. It is known to manage some of the world’s largest datasets on clusters with many thousands of nodes deployed across multiple data centres. Cassandra data management use cases include product catalogs and playlists, sensor data and Internet of Things, messaging and social networking, recommendation, personalization, fraud detection, and numerous other applications that deal with time series data. The wide adoption of Cassandra [3] in big data applications is attributed to, among other things, its scalable and fault-tolerant peer-to-peer architecture [4], versatile and flexible data model that evolved from the BigTable data model [5], declarative and user-friendly Cassandra Query Language (CQL), and very efficient write and read access paths that enable critical big data applications to stay always on, scale to millions of transactions per second, and handle node and even entire data center failures with ease.

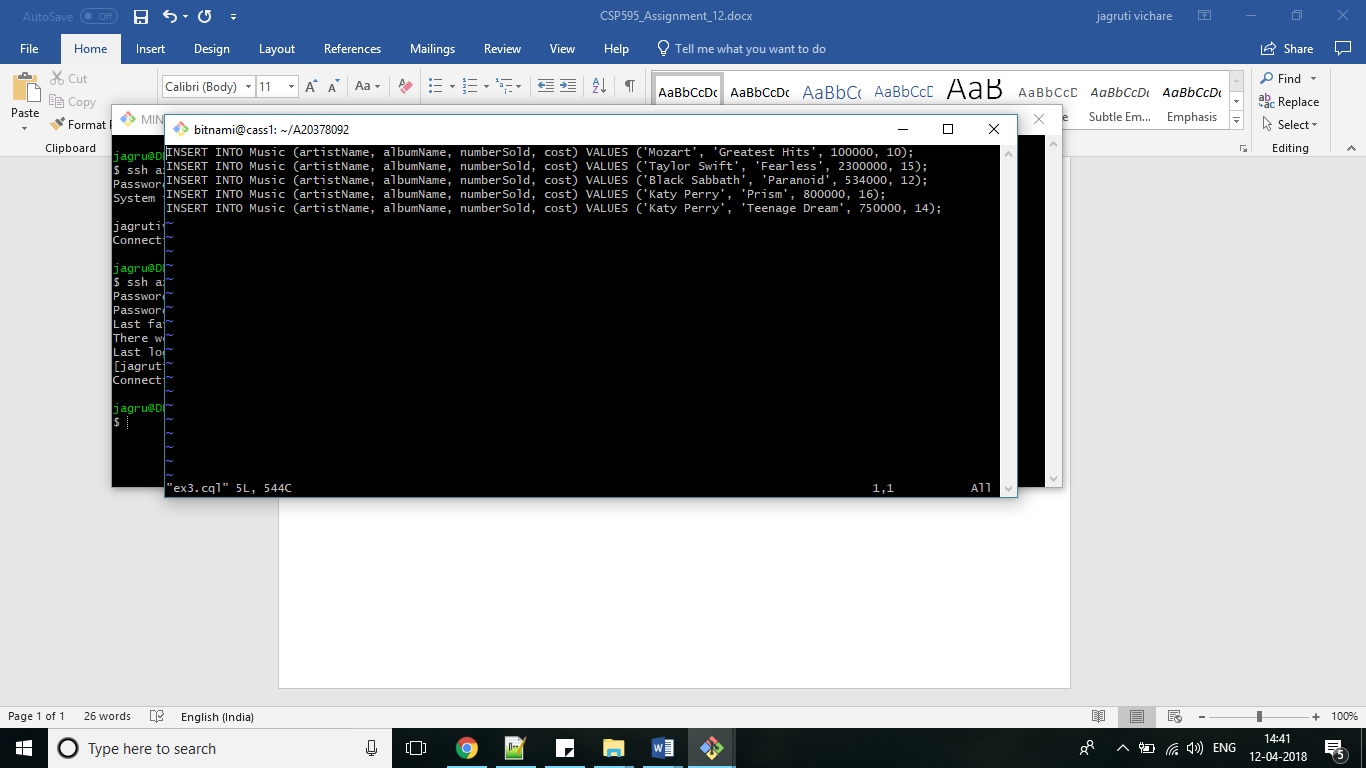
This paper introduces a rigorous query-driven data modelling methodology for Apache Cassandra. The methodology was shown to be drastically different from the traditional relational data modelling approach in a number of ways, such as query-driven schema design, data nesting and data duplication. This paper elaborates the fundamental data modelling principles for Cassandra and defines mapping rules and mapping patterns to transition from technology-independent conceptual data models to Cassandra-specific logical data models. It also explains the role of physical data modelling and proposed a novel visualization technique, called Chebotko Diagrams, which can be used to capture complex logical and physical data models. Finally, it presents a powerful data modelling tool, called KDM, which automates some of the most complex, error-prone, and time-consuming data modelling tasks, including conceptual-to-logical mapping, logical-to-physical mapping, and CQL generation.

**Exercise 2**

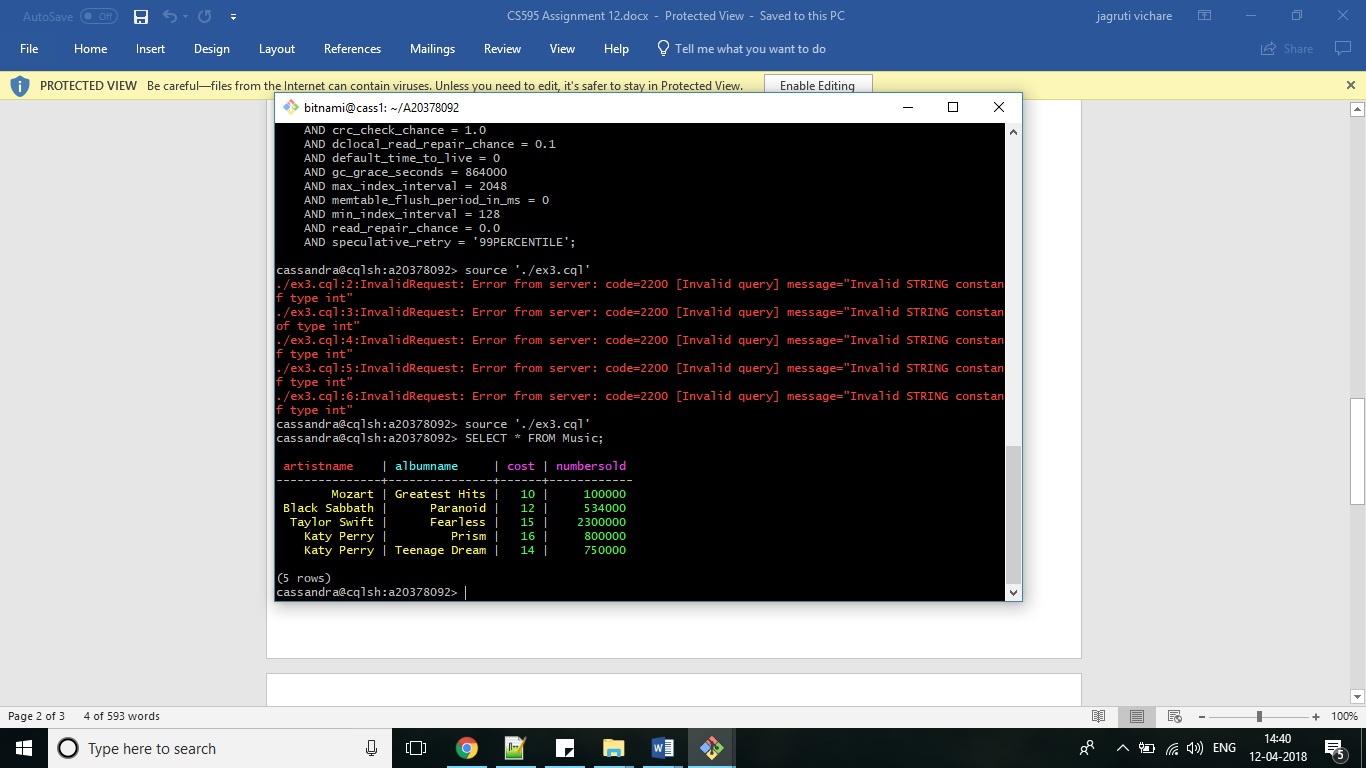


**Exercise 3**

Ex3.cql File Content

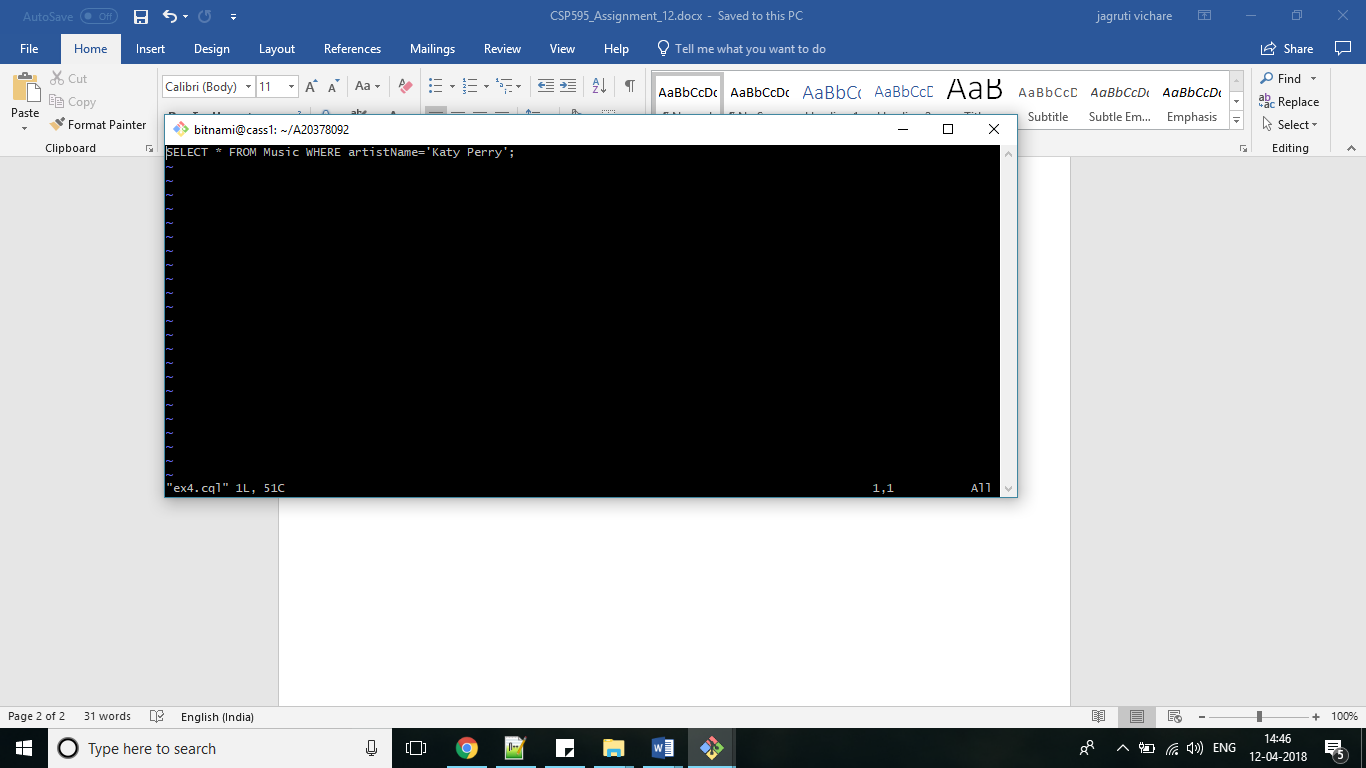


‘SELECT \* FROM Music;’ output

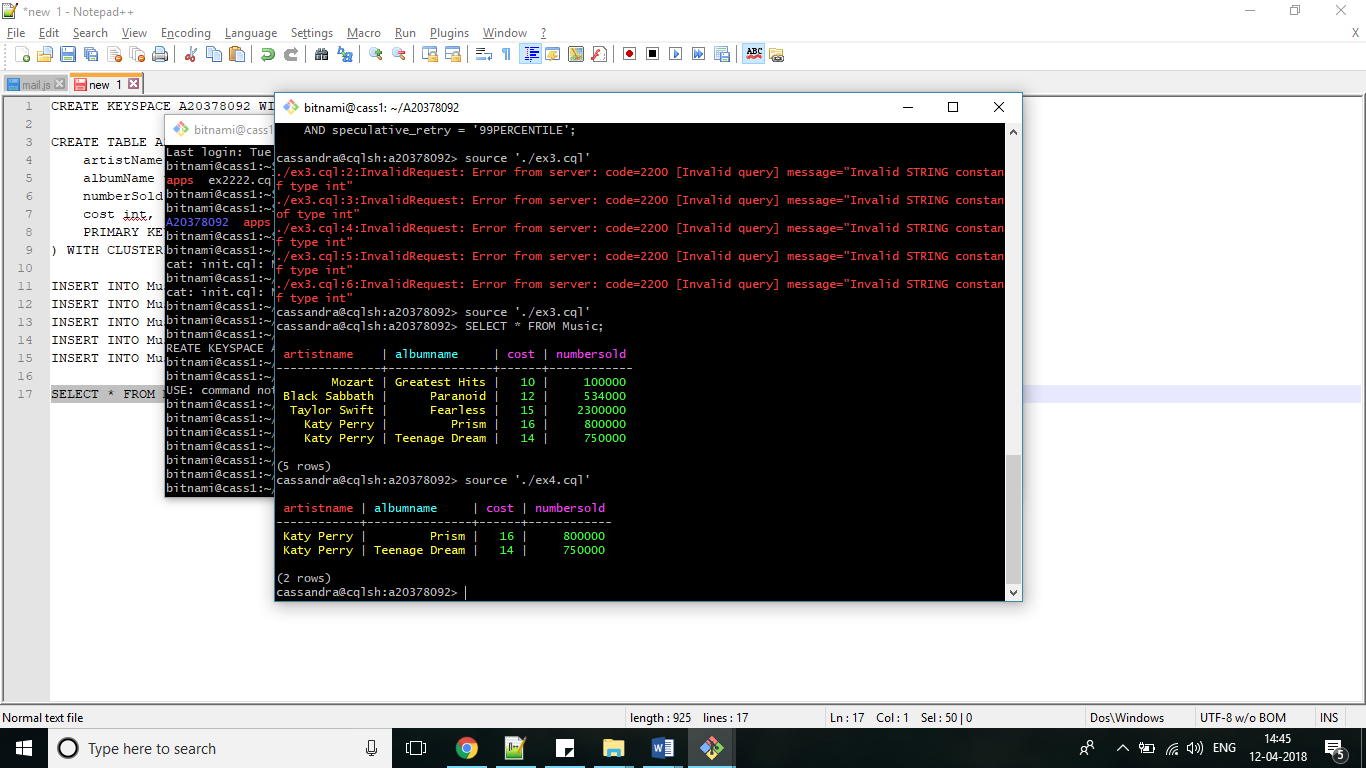


**Exercise 4**

File Content

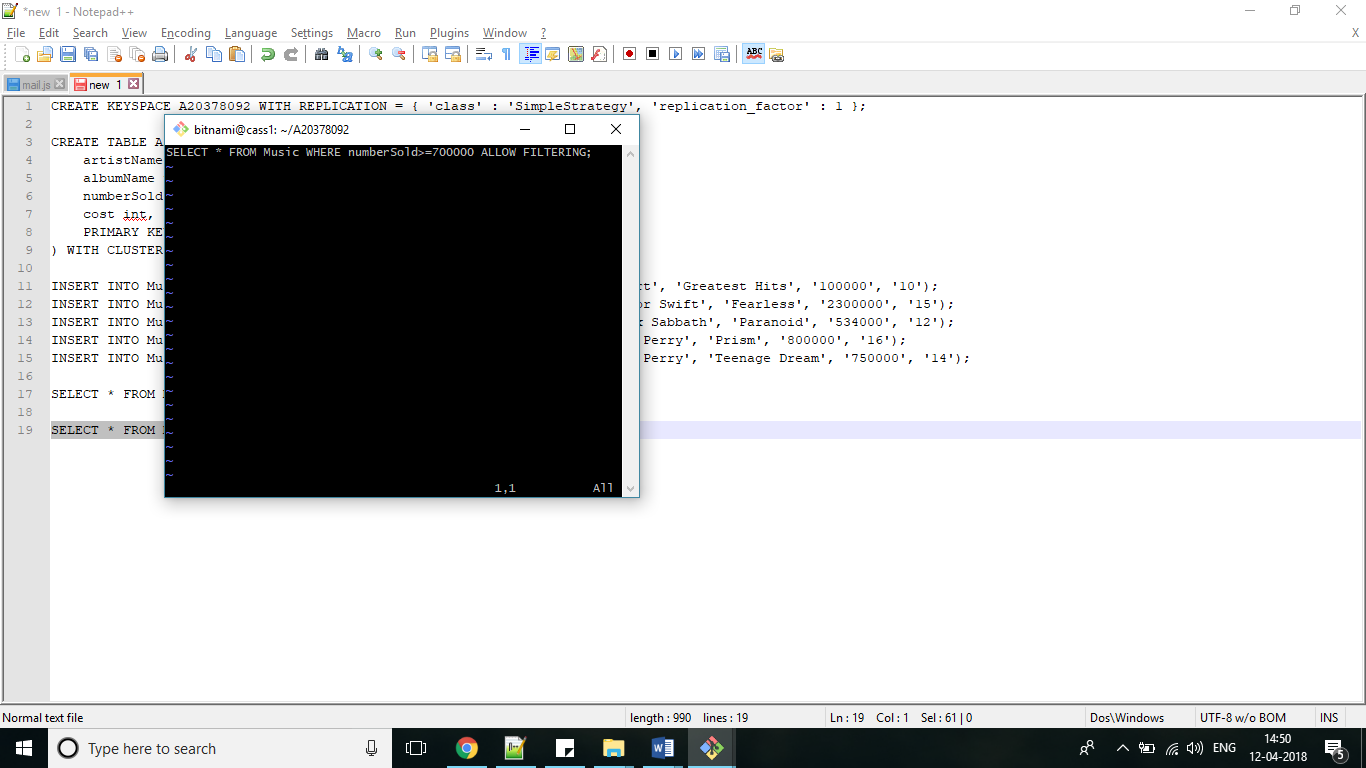


Output



**Exercise 5**

File Content



Output

