**Assignment 11**

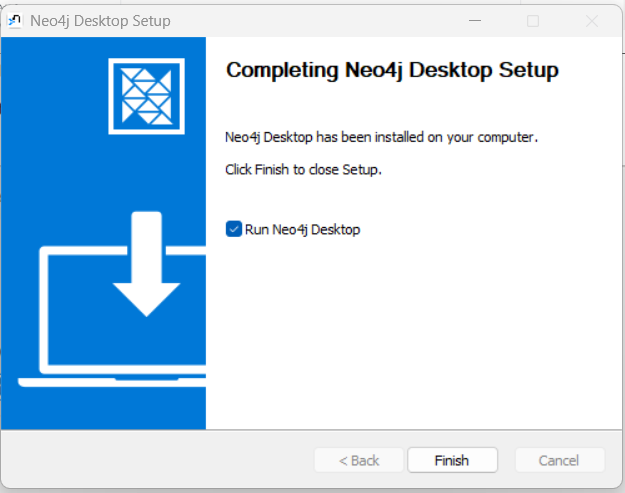
**Namw-Sujan Mujawar**

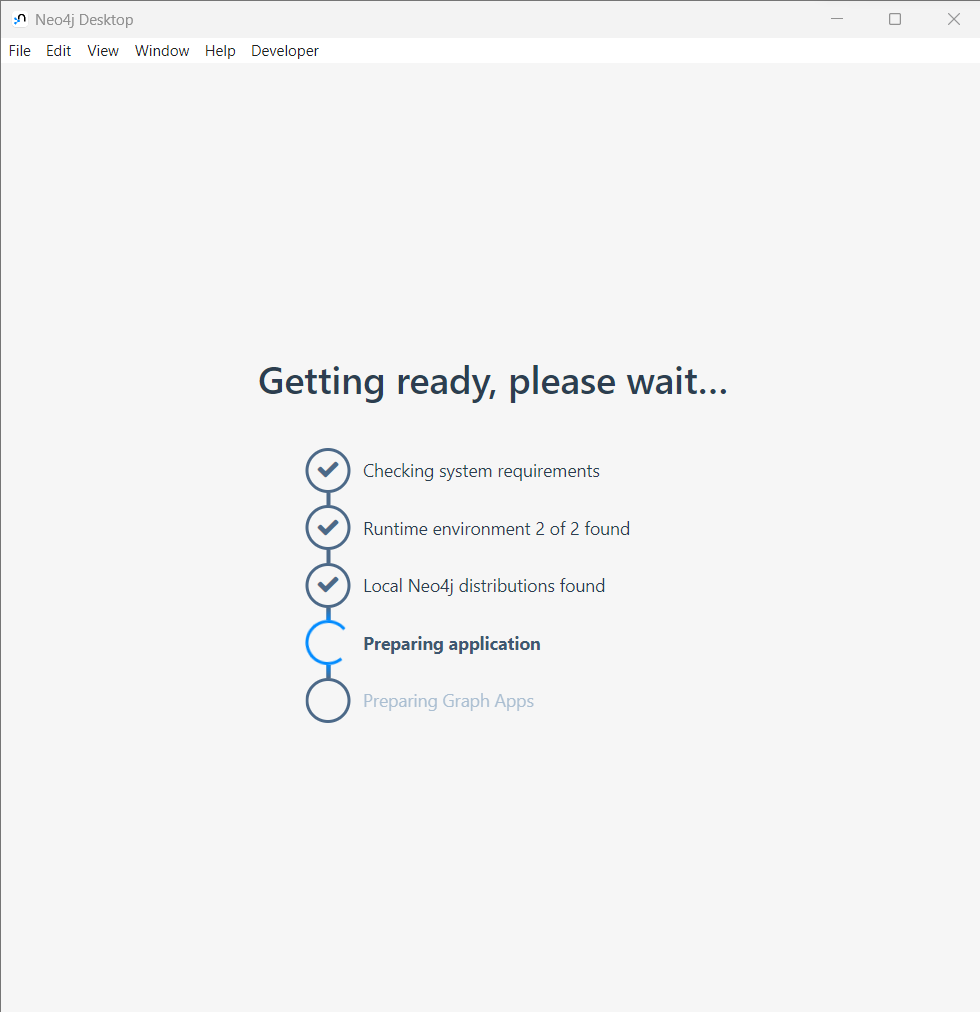
**PRN-21510048**

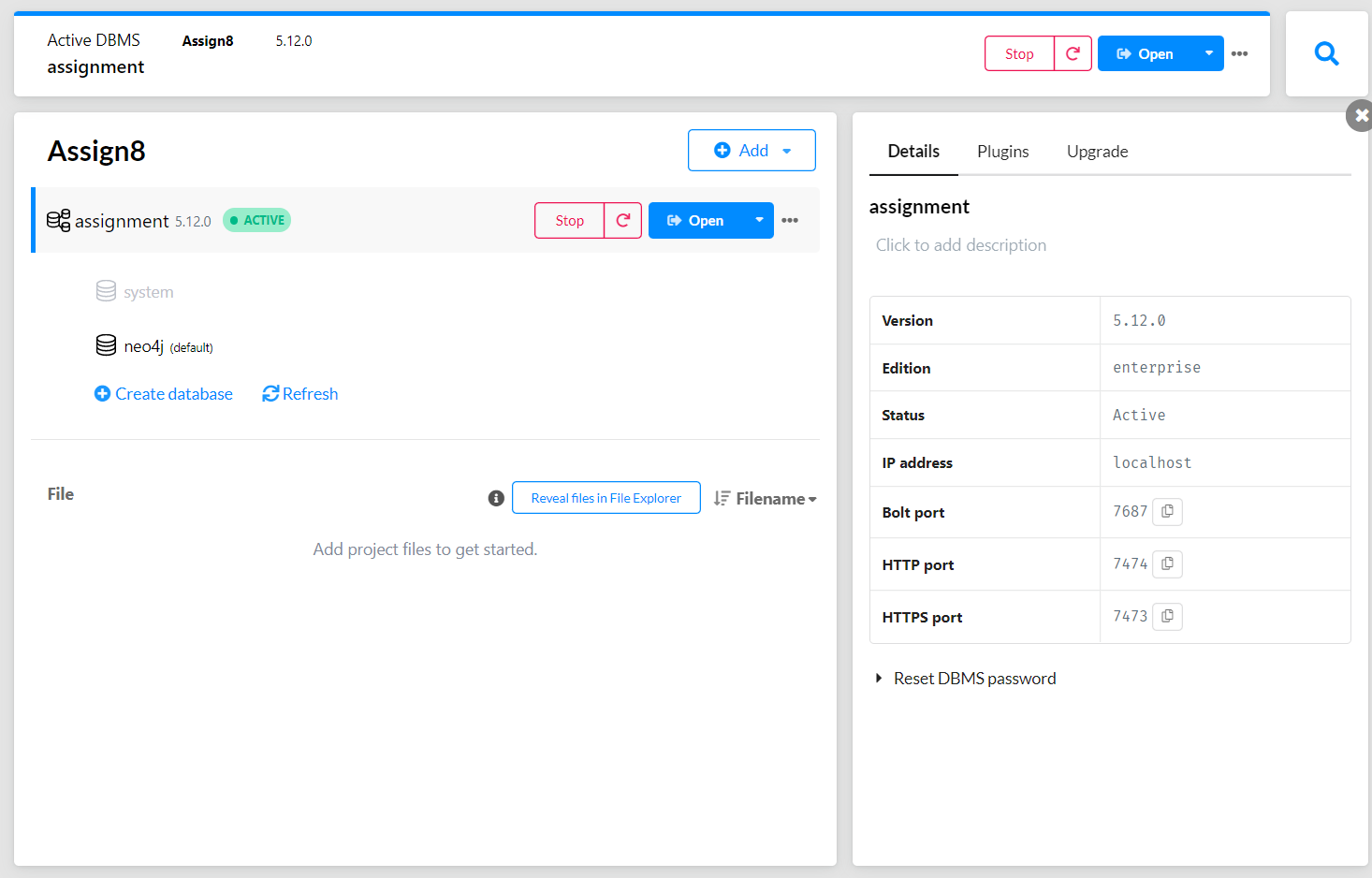
**Batch-T7**

Consider the “**Research Papers Database”** scenario as follows :

The research papers have authors (often more than one). Most papers have a classification (what the paper is about). The classifications form a hierarchy in several levels (for example, the classification “Databases” has the subclassifications “Relational” and “Object-Oriented”). A paper usually has a list of references, which are other papers. These are called citations.

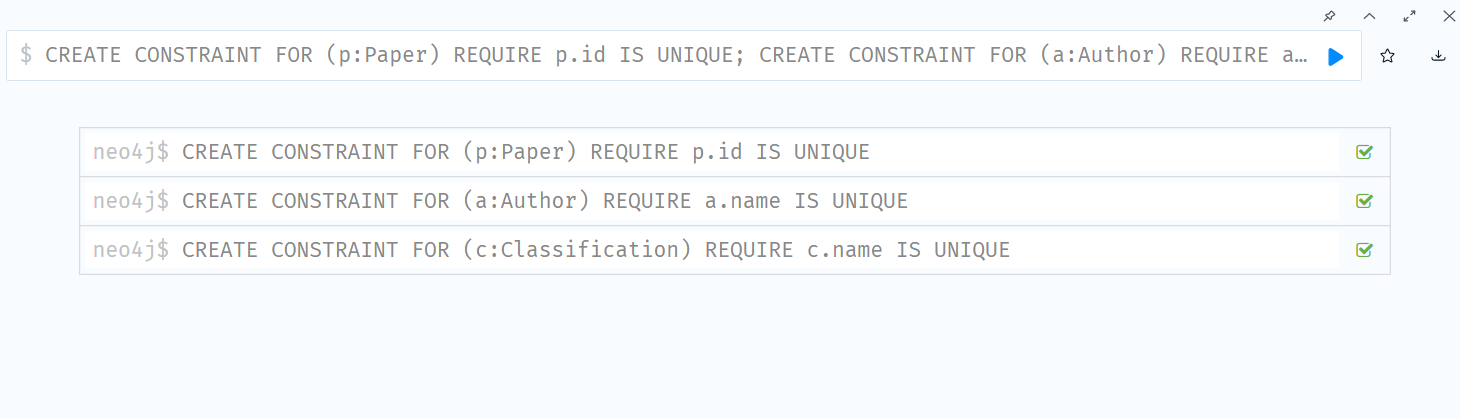


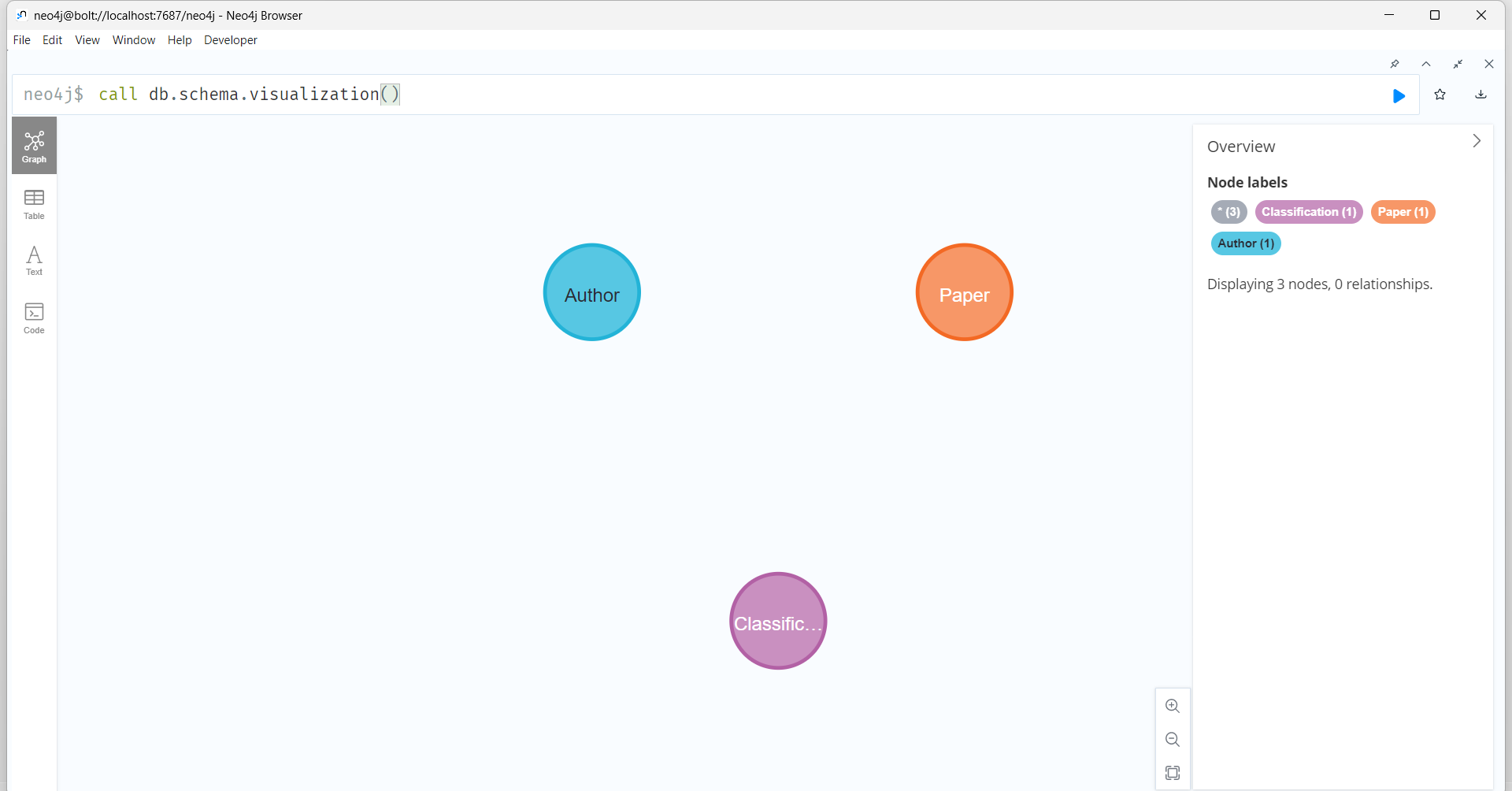


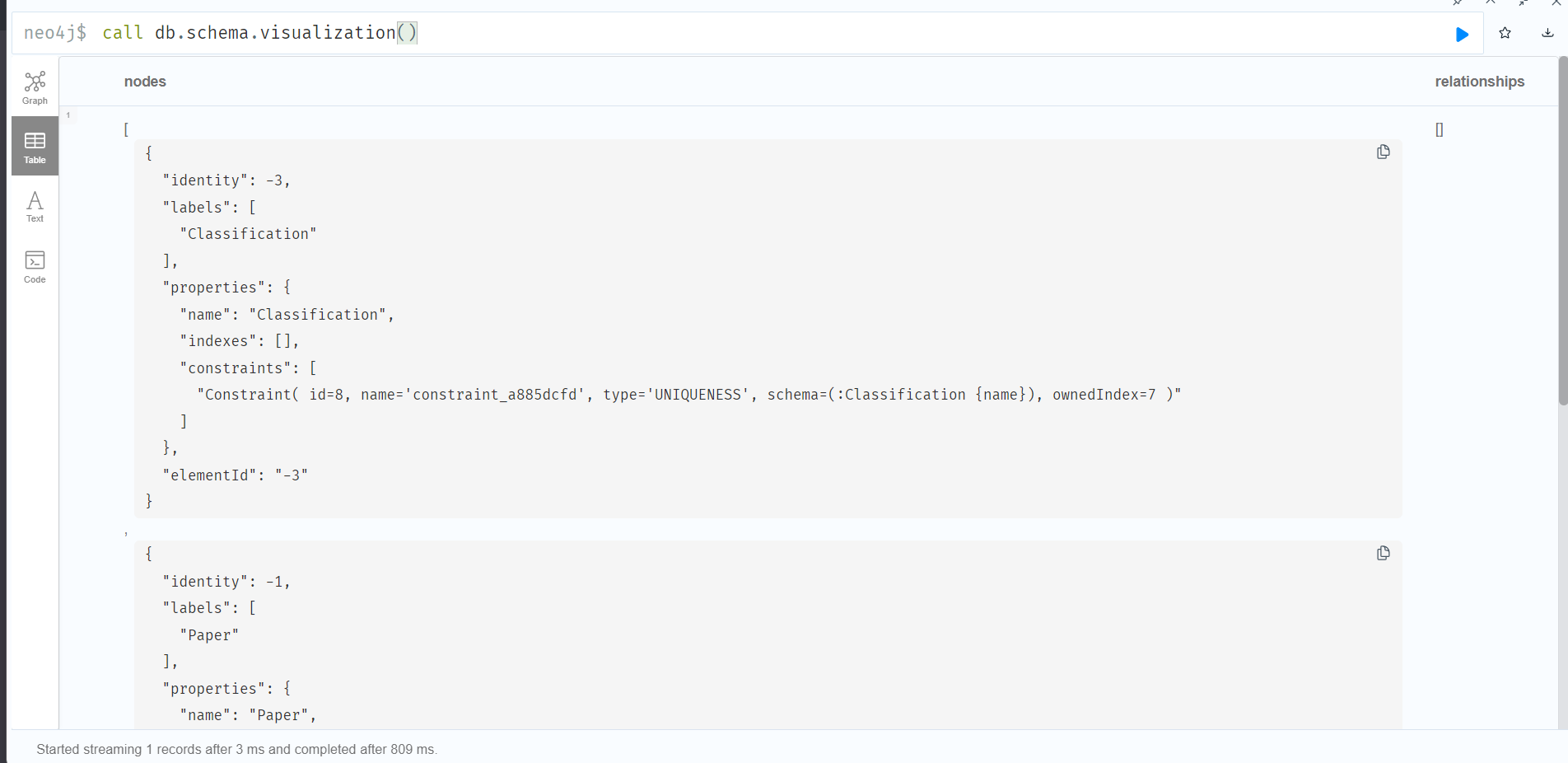


1. Design/model the graph database using Neo4j for above scenario.

1. Download the raw data from **Cora Research Paper Classification Project** : <http://people.cs.umass.edu/~mccallum/data.html>The database contains approximately 25,000 authors, 37,000 papers and 220,000 relationships.









1. Load this data using Neo4j Data Browser

LOAD CSV WITH HEADERS FROM

'https://raw.githubusercontent.com/ngshya/datasets/master/cora/cora\_content.csv'

AS line FIELDTERMINATOR ','

CREATE (:Paper {id: line.paper\_id, class: line.label})



LOAD CSV WITH HEADERS FROM

'https://raw.githubusercontent.com/ngshya/datasets/master/cora/cora\_cites.csv'

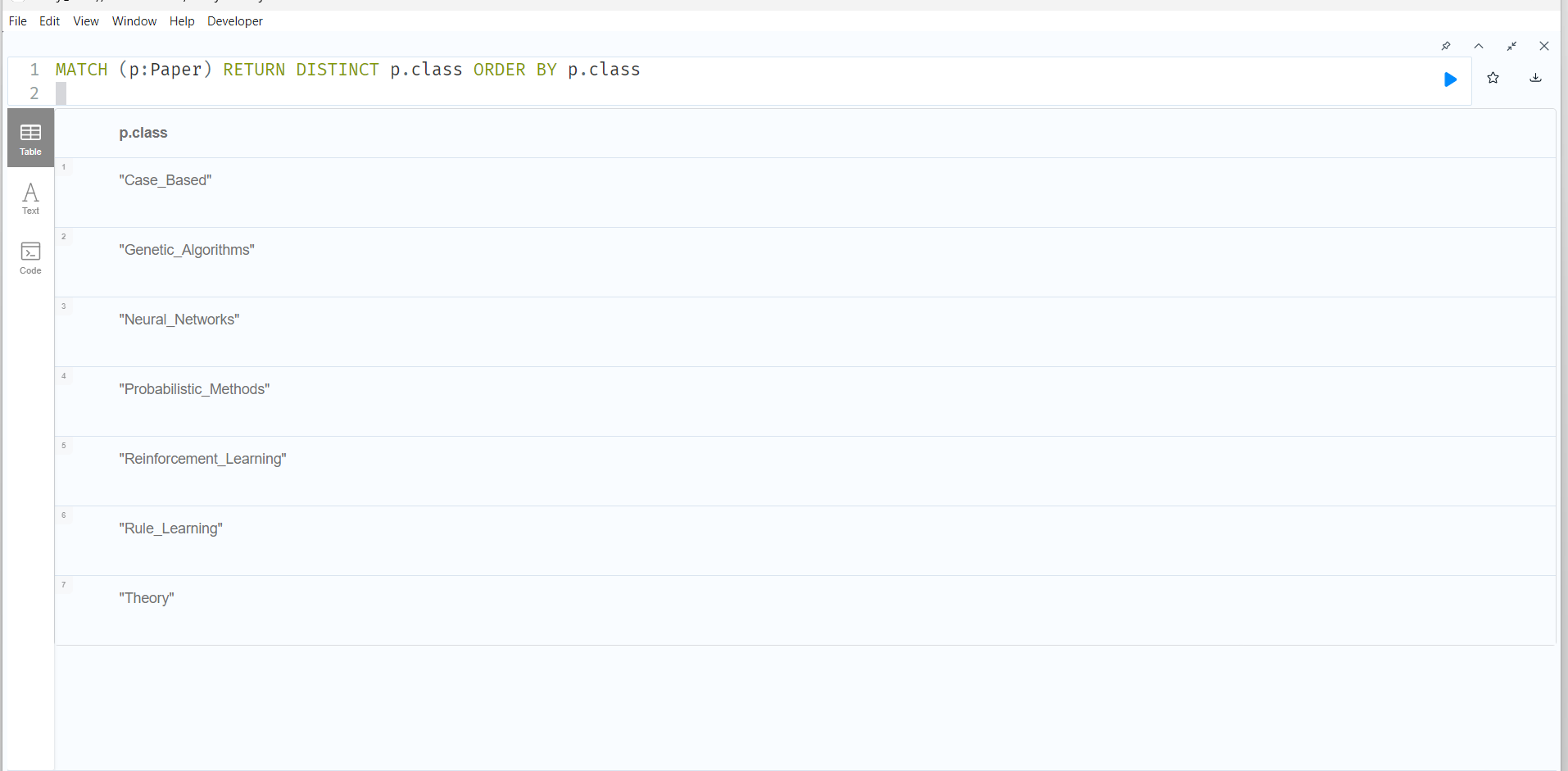
AS line FIELDTERMINATOR ','

MATCH (citing\_paper:Paper {id: line.citing\_paper\_id}),(cited\_paper:Paper {id: line.cited\_paper\_id})

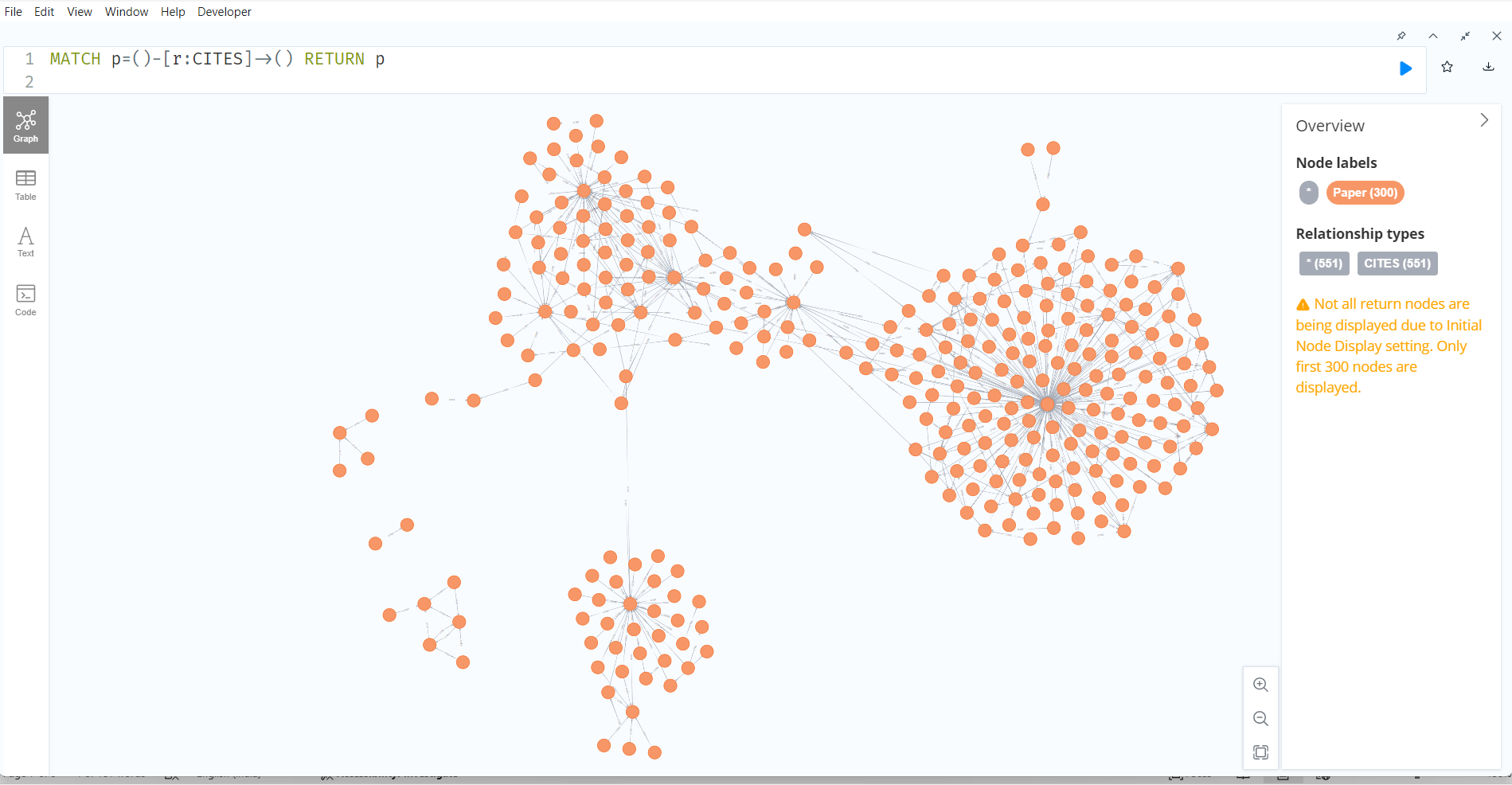
CREATE (citing\_paper)-[:CITES]->(cited\_paper)



MATCH (p:Paper) RETURN DISTINCT p.class ORDER BY p.class



MATCH p=()-[r:CITES]->() RETURN p



1. Design the python-based desktop application for any kind of search on above database. The application should able to answer queries like



PYTHON APPLICATION: 