**RESTRICTED**

Data Intelligence Service Centre

ZeeMovies – Technical Architecture and Logical Model

**Table of Contents**

0. Document Information 4

1. Overall Description 6

1.1 Purpose description 6

1.2 Use-Case Description of the Application 6

2. Data Architecture 7

2.1 Data Sources 7

**List of Tables**

**No table of figures entries found.**

**List of Figures**

**No table of figures entries found.**

# Document Information

**Versioning:** v0.1 Draft

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Version No. | Date | Author | Comment | Reviewer | Release |
| 0.1 | 24/10/2018 | Zeeshan Bajwa | Initial Draft | HR |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |

**Table 1: Amendment History**

|  |  |
| --- | --- |
| Abbreviation | Description |
| DISC | Data Intelligence Service Centre |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |

**Table 3: Abbreviations and Special Terms**

# Overall Description

## Purpose description

This document will allow the stake holders to get hold of the idea to create an graphical database which caters movie store csv data in graph database along with the relationship created between the nodes.

Purpose is to explain the complete end to end Data architecture along with the technology platforms used in the implementation.

## Use-Case Description of the Application

This model will explain creation of data pipeline created for movie store csv data into graph database i.e., Neo4j. Model will be created out of the understandings based on csv data provided.

**Architecture Assumptions:**

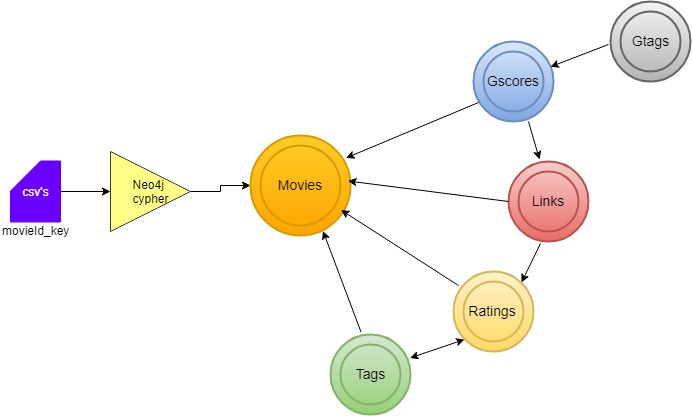
* Ingest data once only
* Graph Database selection based on popularity and share of market of Neo4j

**Major Key points / Requirements**

* Creating data pipeline between csv’s and the graph database
  + Cypher language used
  + Nodes and properties are created
  + Relationships between the nodes

# Data Architecture

This module of document highlights the data flow architecture of the propose solution for creating ingestion pipeline of data to graph-db.



Data Flow Architecture

## Data Sources

This module will discuss about the data layer, which is ingested by Cypher language used in Neo4j.

Following are the identified nodes, properties and relationships.

In the case of this paper, the database of a online movie store will be designed. Here are the nodes, the relationships that will be created with their properties.

**Nodes**

The nodes of the graph database will represent entities:

* Movies
* Gtags
* Score
* Ratings
* Tags
* Links

**Properties of Nodes**

The nodes of the graph database will represent entities:

* Movies - movieId,title,genres
* Gtags - tag id , tag
* Score - movie id ,taq id ,relevence
* Ratings - userId,movieId,rating,timestamp
* Tags - userId,movieId,tag,timestamp
* Links - movie id, imbd id , tmid

**Major Relationships of Nodes**

The nodes of the graph database will represent entities:

* Movies - [Links] - Links
* Gtags - [taged\_with] - Scores
* Scores - [Relevance] - Movies
* Ratings - [User\_Rated] - Movies
* Tags - [User\_Tagged] - Movies

**Technologies:**

* Neo4j Desktop version
* Cypher Language

**Example Queries:**

Query example for creating a Movie node:

MERGE (:Movies {movieId: 9, title:'Sudden Death (1995)', genres:'Action'})

Query example for creating a User node:

MERGE (:Ratings { movieId: '9’})

Query example for creating a Relationship between User and Movie nodes:

MATCH (u1:Ratings),(m1:Movies) WHERE m1.title=' Sudden Death (1995)'

CREATE (u1)-[:RATED {rate:2}]->(m1)