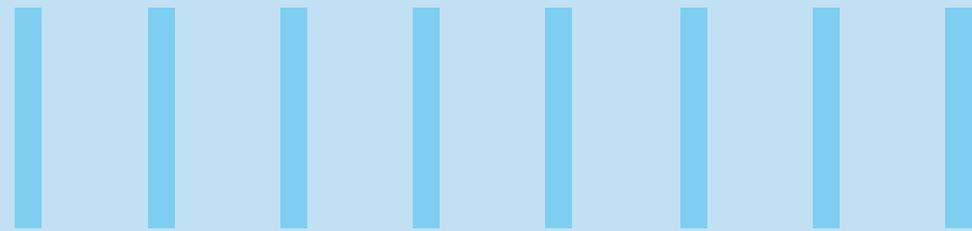




EVALUATION OF GRAPH DATABASE

Unlocking new possibilities with connected data



Chaitanya Deshpande – *Senior Technology Architect*

Pulak Senapati – *Technology Architect*

Sivakumar Kenchappa – *Technology Architect*

Sonia Bhardwaj – *Technology Lead*

Graceta Jesurajan – *Senior Consultant*

Shaurya Singh – *Specialist Programmer*

Contents

Introduction	3
Graph at a glance.....	3
Popular Graph Databases	4
Graph DB Evaluation.....	5
Graph Database Evaluation Findings	6
Graph database Use cases:	7

Introduction

Today, in the era of Big Data, characterized by 4 V's (volume, velocity, variety, and veracity), we have large volume datasets generated by many applications (Enterprise applications, social applications, Healthcare, E-commerce, sensor data, and devices like mobile, laptop, camera, etc.) The exponential growth of the volume of data generated by users, systems, and sensors are increasing the interdependency and complexity of data these days.

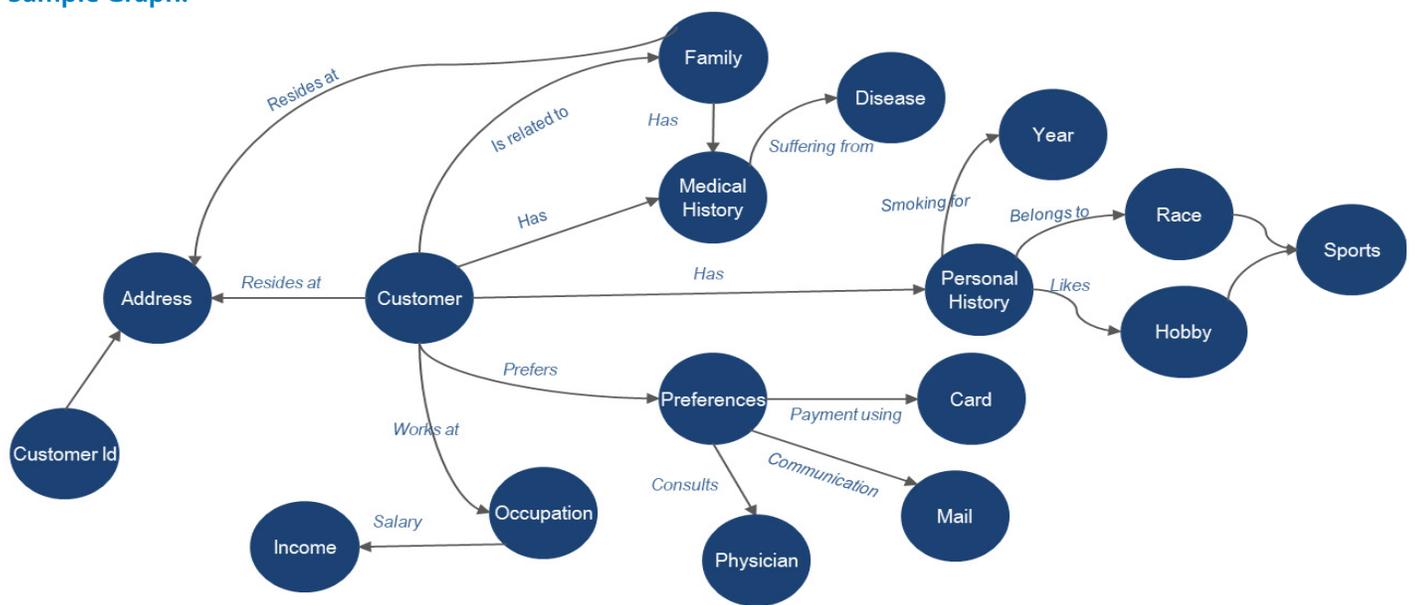
Highly connected entities of this type of unstructured data is not easy to be represented and modeled using traditional relational schemas, simply because the data is highly unstructured. Instead, using graph data structures can make it easy to represent connected data and to perform rapid analysis on these large datasets. The graph technique emphasizes on determining relationships between different data points and levels of relationship for analyzing the data. Graph Databases and graph-based

analysis is one of the most interesting and upcoming trends in Data Science for solving large and unstructured data problems. This whitepaper elaborates the applicability and advantages of using graph database and provides the comparison between the three major graph databases used widely by organizations. This paper will be a useful reference to architects, developers who may want to consider different options to analyze their connected data problems.

Graph at a glance

- Graph is a data storage technology that stores facts about data and its relationships
- Graph simplifies understanding how people, places and things are related. e.g. links on the web, social networks, customer journeys, etc.
- Graph is used primarily to select the right content and context for the Insight. It helps in disambiguating and recognizing entities in context
- Graph database is not a replacement of traditional data warehouse but it complements it. Graph can be used for Exploratory Data Analysis to validate that we are even asking the right questions prior to bringing those into the warehouse

Sample Graph:



Customer Identity Graph

Popular graph databases

Neo4j

Network Exploration and Optimization 4 Java is a graph data base management system developed by Neo4j, Inc. With an ACID-compliant transactional database which is native graph storage and processing, Neo4j is also a GPL3-licensed open-source "community edition", along with online backup and high availability extensions licensed under a closed-source commercial licensee and closed-source commercial terms.

TigerGraph

TigerGraph is kind of graph database, which is built for loading massive amounts of data in minimum duration of hours. It can also analyze as many as 10 or more hops deep into relationships in real-time environment. TigerGraph is ACID compliant and supports transaction as well as analytical workloads. With an proven technology TigerGraph also supports applications such as fraud detection, customer360, IoT, AI and machine learning with ever changing big data.

Amazon Neptune

Amazon Neptune is a managed graph database used as web service as part of Amazon Web Services (AWS). It supports popular graph models property graph and W3C's RDF, and their respective query languages Apache TinkerPop Gremlin and SPARQL, including other Amazon Web Services products.



Graph DB evaluation

Neo4J	Neptune	Tiger Graph
Criteria	Comments	Score
Business Requirement Fitment	<ul style="list-style-type: none"> Neo4j Connector for BI is available at no extra charge for Enterprise Edition customers. Neo4j desktop comes with a package of ETL, Visualization, monitoring and analyzer tools. There is no in-built IDE available with AWS Neptune. User defined functions and stored procedures are not supported in Neptune. TigerGraph comes with Graph studio which handles everything from schema design, query development, visualization and execution. 	4.6
		2.5
		4.3
Technical Requirement Fitment	<ul style="list-style-type: none"> Neo4j supports triggers and Kafka connections It supports explain plan and profile plan to analyze the query cost . Neptune does not support triggers but supports Kafka connection with API for real time streaming. Supports Explain plan/profiling. Tiger graph does not support triggers but supports Kafka connection with API for Real Time Streaming . It doesn't support explain plan or profiling. 	4.7
		4.2
		4.1
Security and System Integration Fitment	<ul style="list-style-type: none"> Neo4j does not support any out of the box integration for encryption of data. It can be implemented in the application layer. Neptune encrypted instances provide an additional layer of data protection The Tiger Graph data store uses a proprietary encoding scheme which both compresses the data and obscures the data . 	4.5
		5.0
		5.0
Core Functionality	<ul style="list-style-type: none"> Neo4j supports triggers out-of-box, using the APOC libraries With Neptune, triggers are not supported out-of-box, but changes can be identified, and the action can be taken with external orchestration With TigerGraph, triggers are not supported out-of-box, but it can be implemented with external orchestration of finding the CDC with API. 	5.0
		2.5
		3.0
Performance, Reliability, Scalability	<ul style="list-style-type: none"> Neo4j supports index creation to improve performance. Neo4j Aura is designed to build, deploy applications rapidly. With Neptune, you do not need to create custom indexes over your graph data. Repairs disk failures and fix database crashes automatically. Tiger graph supports secondary index , however no out-of-box support for zero copy cloning and self-healing. 	4.0
		4.1
		3.8
Change Management Impact and Business Continuity Risk	<ul style="list-style-type: none"> Neo4j database can be deployed on the major cloud platform of your choosing, including AWS, Azure and GCP. Kubernetes and Docker are supported. Neptune is available only on AWS. Container orchestration is not required as it is a managed service by AWS. Tiger graph is available in AWS, Azure and GCP marketplace TigerGraph available on docker image but not on Kubernetes Neo4j is widely used whereas TigerGraph and Neptune are still evolving. Hence, more skilled professionals are available for Neo4j compared to the other two. 	5.0
		2.5
		4.5

Graph database evaluation findings

Neo4J

Strength:

- Matured and widely used
- Rich ML Library
- Contains UI based Neo4j browser
- to execute Cypher commands
- Supports API connectors for many

Limitation:

- No native support for data encryption

Tiger Graph

Strength:

- In built Rich Visualization
- Rich ML Library
- Distributed Parallel Processing
- Supports compressed data storage

Limitations:

- No support for triggers
- No support for profiling and performance techniques
- Self-recoverability not supported

Amazon Neptune

Strength:

- DB as service
- Provides Dynamic scalability
- Self healing

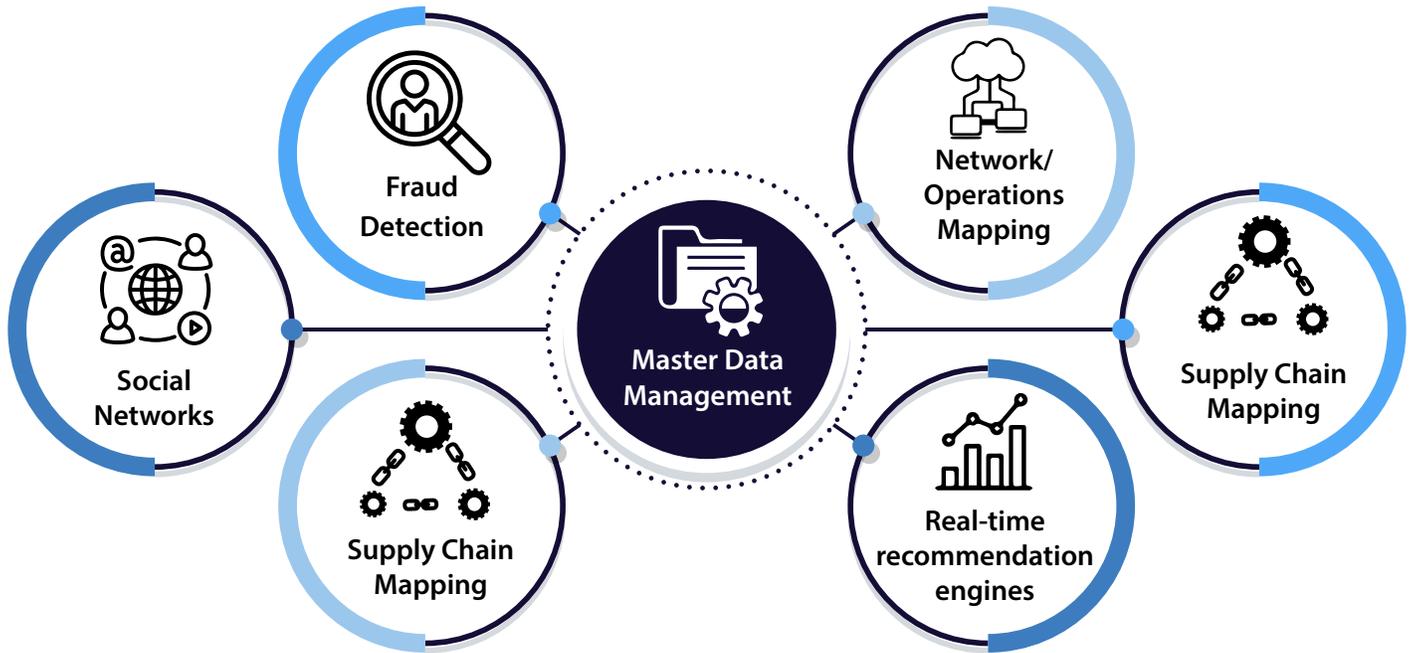
Limitations:

- Can be installed only on AWS
- No support for triggers
- Visualization is still in initial phase
- Limitation of 15 read replicas
- No support for indexes



Graph database use cases:

Organizations everywhere are turning to graph technology. Below is a list of some of the popular uses of graph, organized across the following industries and categories:



References

- <https://neo4j.com/docs/>
- <https://www.tigergraph.com/resources/>
- <https://docs.aws.amazon.com/>
- Internal case studies

For more information, contact askus@infosys.com



© 2021 Infosys Limited, Bengaluru, India. All Rights Reserved. Infosys believes the information in this document is accurate as of its publication date; such information is subject to change without notice. Infosys acknowledges the proprietary rights of other companies to the trademarks, product names and such other intellectual property rights mentioned in this document. Except as expressly permitted, neither this documentation nor any part of it may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, printing, photocopying, recording or otherwise, without the prior permission of Infosys Limited and/ or any named intellectual property rights holders under this document.