**Preliminary Research into NoSQL and Graph Databases**

* NoSQL
  + Types of different services - some top mainstream services
    - MongoDB
    - Cassandra
    - ElasticSearch
    - Amazon DynamoDB
    - Hbase
  + Not only SQL
  + Main types of NoSQL databases
    - Document
    - Key-value
    - Wide column
    - Graph
      * Graph DB = type of NoSQL DB
      * Data is stored in nodes and edges
        + Nodes contain info about people, places and things
        + Edges contain info about the relationships between the nodes
  + Refers to any non-relational database
  + Data varies in shape and size
    - Structured
    - Semi-structured
    - Polymorphic
    - Unstructured
  + NoSQL can handle lots of unstructured data
  + Features
    - Flexible schemas
    - Horizontal scaling
      * Add more machines to distribute querying of data
      * High scaling
    - Fast queries due to the data model
    - Ease of use for developers
  + Relationship data is stored but differently compared to RDBMS
    - Related data to be nested within a single data structure
  + Queries are less flexible
  + Primary key value stored/known
    - You need to know the key your looking for before you query
      * Like a Hash map?
  + Access
    - Rest APIs
    - Create Read Update Delete (CRUD) in languages specific to service being used
  + NoSQL API's

**Sources**

* <https://www.mongodb.com/nosql-explained>
* <https://www.mongodb.com/nosql-explained/nosql-vs-sql#what-are-the-benefits-of-nosql-databases>
* <https://www.youtube.com/watch?v=ruz-vK8IesE>
* <https://www.youtube.com/watch?v=dGHSKpx4Xjs>
* <https://www.analyticsvidhya.com/blog/2020/09/different-nosql-databases-every-data-scientist-must-know/>
* <https://learning.oreilly.com/library/view/next-generation-databases/9781484213292/9781484213308_Ch03.xhtml#Sec4>

* Graph DB
  + A collection of nodes and edges, where nodes contain info and edges contain relationships
  + Types of different graph database services:
    - ArangoDB
    - Dgraph
    - Neo4j
      * Open sourced
      * Supported programming languages
        + .Net
        + Clojure
        + Elixir
        + Go
        + Groovy
        + Haskell
        + Java
        + JavaScript
        + Perl
        + PHP
        + Python
        + Ruby
        + Scala
      * Server operating systems
        + Linux
        + OS X
        + Solaris
        + Windows
      * Queries
        + New querying language

Cypher

To interact with data on spark

* + - * Popular graph DB but lacks performance
    - DataStax
    - Amazon Neptune
      * Server operating system
        + Hosted
      * Supported programming languages
        + C#
        + Go
        + Java
        + JavaScript
        + PHP
        + Python
        + Ruby
        + Scala
      * Aims for high availability and durability
      * Designed to offer >99.99% availability and automatically detects and recovers from most database failures in < 30 seconds
      * Secure
      * Dual-nature
        + Supports different graph db models

Common graph database models

Resource Description Framework - RDF

Property Graphs - PG

* + - * + However different models can't be used interchangeably due to differences in models
    - OrientDB
    - Cayley
    - FlockDB
    - Cassandra
    - Titan
    - TigerGraph
      * Implementation language
        + C++
        + Also supports java
      * Server operating system
        + Linux
      * Queries
        + Done through a SQL similar language called GSQL
      * Compared to Neo4j
        + Has a longer loading time than Neo4j, however when pre-processing time is considered its faster
        + Is efficient because

Needs less storage space than

Faster on one-hop and three-hop path query

Can compete six-hop path query

This kind of processing kills Neo4j as after 2 hours its run out of memory

* + Graph DB focuses
    - Processing highly connected data
    - Be flexible in usage data models behind graphs used
    - Exceptional performance for local reads, by traversing the graph

**Sources**

* <https://learning.oreilly.com/library/view/graph-databases/9781449356255/ch01.html>
* <https://www.g2.com/categories/graph-databases?tab=highest_rated>
* <https://towardsdatascience.com/comparing-graph-databases-5475bdb2e65f>
* <https://db-engines.com/en/system/Amazon+Neptune%3BNeo4j%3BTigerGraph>
* <https://doi.org/10.1007/978–3-030–63924-2_5> (Hoferet al., 2020)
* <https://www.emerald.com/insight/content/doi/10.1108/IJWIS-03-2021-0023/full/html>
* <https://link.springer.com/chapter/10.1007/978-3-319-24369-6_5>