

Search Data Science Central

Search

- [Sign Up](#)
- [Sign In](#)

ice Central

THE ONLINE RESOURCE FOR BIG DATA PRACTITIONERS

HOME

DATAVIZ

HADOOP

BIG DATA

ANALYTICS

WEBINARS

DEEP LEARNING

AI

JOB\$

MEMBERSHIP

SEARCH

CLASSIFIED\$

CONTACT

Subscribe to DSC Newsletter

- [All Blog Posts](#)
- [My Blog](#)
- [Add](#)



## Lesson 8: Graph Databases

- [Posted by William Vorhies on September 12, 2014 at 8:26am](#)
- [View Blog](#)

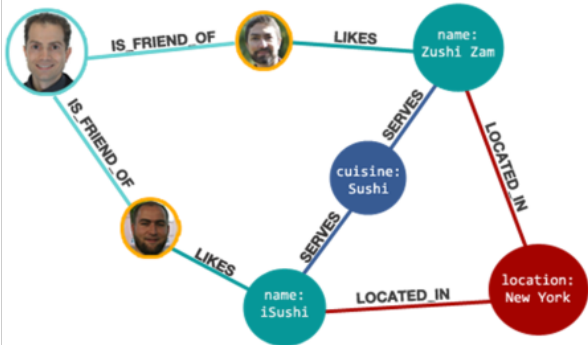
**Summary:** Graph databases are your go-to choice when a relationship among the data items is key.

Up to about 1999 web search engines evaluated each web page as a standalone entity, ranking them based on content without regard to any other pages. But in 1999 Google adopted PageRank, a graph-centered approach invented by co-founder Larry Page. PageRank evaluates web pages in relationship to other pages. Users quickly recognized that ranking pages based on their relationship to others resulted in much better recommendations and may be the single factor that moved Google rapidly ahead of its competitors.

We are so used to thinking of databases as tables or at least buckets of information that it can be a little challenging to wrap your head around the concepts of graph databases. That said, Graph DBs can do things that none of the other types of NOSQL or RDBMS DBs can do. Making the effort to understand and utilize this type can offer big returns.

### Characteristics

There are no classical indexes for Graph DBs. Rather, each object stored is mapped with “nodes” and



Graphic Source: Emil Eifrem and Philip Rathle, Neotechnology.com, 2013

“edges”. A node is a single record that has at least one and potentially many named properties. Edges define the relationship among nodes and both the nodes and their relationships have some predefined properties. Nodes can have multiple edges defining many different kinds of relationships they have with other nodes. Both nodes and relationships (edges) can be addressed with key values.

Search or query with Graph DBs is called “traversal”. These queries are designed to start at a specific node and explore its relationship with other nodes based on the relationships requested. A common example would be ‘what books are my friends reading that I haven’t yet read’. In this mode,

Graph DBs are often associated with 'recommender' engines widely used in social and ecommerce applications.

As Graph DBs become more dense, traversal search may require stopping at the same node several times which can slow the search. As a result Graph DBs learn and index these common relationships to speed up search.

#### Advantages

- Extremely fast for connected data. While RDBMS can be made to replicate graphical ones, the extensive use of joins would make the technique quite slow.
- Easy to query.
- Able to quickly handle complex queries involving multiple levels of related data.

#### Disadvantages

- Traditionally Graph DBs have been scaled vertically but not horizontally as searching nodes on different machines would dramatically slow the process. Vendors did not support distribution or sharding. This has made it difficult for Graph DBs to scale beyond a certain size. However, some vendors are challenging in this area.
- Requires a conceptual shift in thinking for developers so some learning curve will be required.

#### Particular Opportunities and Project Characteristics

- Traditionally recommender engines (any 'recommended for you' rating) have been based on Graph DBs. Note that some recommenders are now also being built using Column Oriented DBs.
- Use where objects have both dynamic properties and dynamic relationships among objects.
- Applications requiring very deep and complex joins in RDBMS can be replaced by Graph DBs typically with increases in speed greater than 100X.

Some sample use cases:

- Model and store 7 billion people objects and 3 billion non-people objects to provide an earth-view drill down from planet to sidewalk. (Neo4J)
- Tracking food sources from seed to table. (Objectivity, Inc.)
- Ad placement applications (Objectivity, Inc.)
- Network management.
- Genealogy
- Public Transport links, road maps

**Representative Vendors** (not a recommendation): [Neo4J](#), [Infinite Graph](#), [InfoGrid](#), [HyperGraphDB](#), [AllegroGraph](#), [BrightstarDB](#), and many others.

July 23, 2014

Bill Vorhies, President & Chief Data Scientist - Data-Magnum - © 2014, all rights reserved.

About the author: Bill Vorhies is President & Chief Data Scientist of Data-Magnum and has practiced as a data scientist and commercial predictive modeler since 2001. He can be reached at:

[Bill@Data-Magnum.com](mailto:Bill@Data-Magnum.com)

This original blog can be viewed at:

<http://data-magnum.com/lesson-8-graph-databases-including-object-dbs/>

All nine lessons can be downloaded as a White Paper at:

<http://data-magnum.com/resources/white-papers/>

Views: 5689

Like

1 member likes this

Share   Facebook

 Like 0

- [< Previous Post](#)
- [Next Post >](#)

Comment

You need to be a member of Data Science Central to add comments!

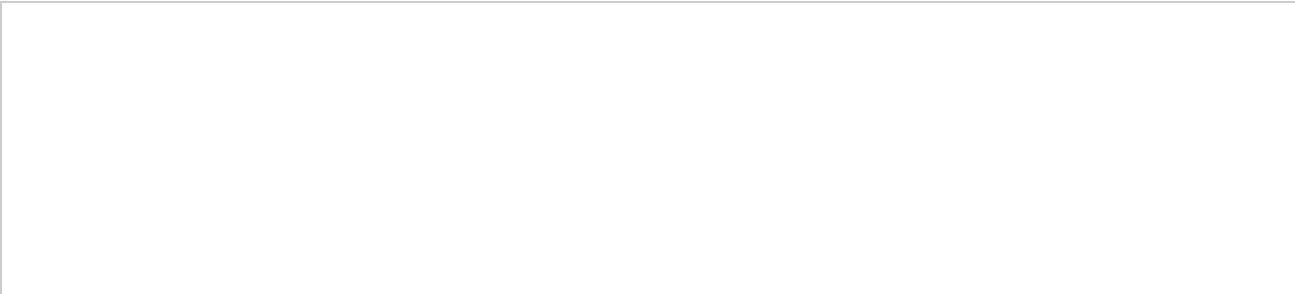
[Join Data Science Central](#)

Welcome to  
Data Science Central

[Sign Up](#)  
or [Sign In](#)

Or sign in with:


- 
- 
- 
- 





**FOLLOW US**


[@DataScienceCtrl](#) | [RSS Feeds](#)


TOP CONTENT


- 

1 Have You Heard About Unsupervised Decision Trees
- 

2 Difference between Machine Learning, Data Science, AI, Deep Learning, and Statistics
- 

3 Machine Learning Algorithm to Trade Bitcoin
- 

4 Non-traditional strategies for mid-career switch to #Datascience and #AI
- 

5 Salary history and career path of a data scientist
- 

6 To PhD or not to PhD?
- [RSS](#)
  - [View All](#)

ANNOUNCEMENTS

- [The Making of a Data Science Superhero](#)
- [12-month MS in Analytics Online from AU](#)
- [Decision Optimization with Prescriptive Analytics Applications](#)
- [10 Considerations When Building a Dashboard](#)
- [Earn your Master's in Applied Data Science in 18 Months](#)
- [The Keys to Enterprise-Ready Machine Learning](#)
- [Boost your business intelligence with smart data](#)
- [Data Science for Big Data with Anaconda](#)
- [The Fast Path to Success with AI](#)
- [The Data Scientist's Guide to Apache Spark™](#)

VIDEOS

- 
  - [Apache Spark and Agile Model Development](#)  
Added by Tim Matteson 0 Comments 0 Likes
- 
  - [Building Better Maps with Data and Data Viz](#)

Building Better Maps with Data and Data Viz

Added by Tim Matteson 0 Comments 0 Likes

- [Add Videos](#)
- [View All](#)

RESOURCES

- [Migrating an Excel Spreadsheet to MySQL and to Spark 2.0.1 \(Part 1\)](#)
- [Introduction to Programming in Stata](#)
- [Benchmarking 20 Machine Learning Models Accuracy and Speed](#)
- [Stata Cheat Sheet](#)
- [Selection of best articles from our past weekly digests](#)
- [Statistical Analysis Advisor Chart](#)
- [Selection of best articles from our past weekly digests](#)
- [Free Online Book: Forecasting, Principles and Practice](#)
- [38 Seminal Articles Every Data Scientist Should Read](#)
- [Black-box Confidence Intervals: Excel and Perl Implementation](#)

TOP CATEGORIES

- [Machine Learning](#)
- [R Programming](#)
- [Python for Data Science](#)
- [Visualization, Dashboards](#)
- [NoSQL and NewSQL](#)
- [Big Data](#)
- [Cheat Sheets](#)
- [Internet of Things](#)
- [Excel](#)

© 2017 Data Science Central Powered by NING

[Badges](#) | [Report an Issue](#) | [Privacy Policy](#) | [Terms of Service](#)

#ODSC

FUTURE PROOF YOUR CAREER

OPEN DATA SCIENCE CONFERENCE

AI, Deep Learning, Autonomous , NLP, Predictive Analytics, etc.

register

San Fran. NOV 2-4