NOSQL (Not Only SQL)

Primarily meant for unstructured data.
- e.g. Facebook post, tweet, email, blog, documents

All info is stored as key-value pairs.

map in Java.

NOSQL Systems
- MongoDB
- Cassandra
- CouchDB

Graph schema — "Titan" - IBM

Big Data

A very large collection of data (mostly unstructured)

Requires a very long processing time

To solve this problem, a distributed processing is needed. $2^{31}$ YottaByte

This must be scalable and reliable.

Need a partitioning scheme (Sharding)

partition the big data into shards

To organize and reference shards, a Distributed File System is essential.

Google: Google Distributed File System (GDFS)

Apache: Hadoop Distributed File System (HDFS)
Skillset:
- Python scripting
- R Statistics
- Bridge Management requirements and big data relationships.

Map-Reduce Framework

[Diagram]

References:
Big Data
by Nathan Marz

Lambda Architecture
Live data → Speed → Final result
Archived data → Service batch