

Advanced topics: big data (self-study)

Many important innovations to **store and process increasingly complex and heterogeneous data**, especially unstructured data

- **Data lake**: stores data in its raw form, schema on read which means load before transform and structure (in contrast to traditional data warehouse)
- **NoSQL (“not only SQL”)**: non-relational databases designed to handle unstructured and semi-structured data
- **Vector database**: popularized in recent years due to LLMs, optimized for storing embeddings (large vectors) generated by machine learning models, such as image, text, or audio embeddings, and used in similarity searches
- **Graph database**: stores data in the form of nodes and edges, ideal for representing network-based relationships