INTRO to DATA SCIENCE
DATABASES

I. DATABASES
II. SQL LABS

I. DATABASES

- Tabular data
- Nested data
- Graph data

Tabular data

‣ Relational DataBase Management Systems (RDBMS) – broadly implementing SQL

‣ Serverless/embedded: SQLite, H2, etc.

‣ Client-Server: MySQL/MariaDB, PostgreSQL, Oracle, MSSQL, etc.

Nested data

‣ Document stores: MongoDB, CouchDB, etc.

‣ Serialization schemes, often on HDFS

‣ Key-value stores: Redis, Voldemort, etc.

‣ Blurred Lines: Object-Relational Mapping, Cassandra, column orientation, BigTable/Hbase, etc.

Graph data

‣ Client-server (general): Neo4j, perhaps more

‣ Client-server (triple store): AllegroGraph, etc.

‣ Query languages: SPARQL, RDF++, Prolog…

II. SQL LABS