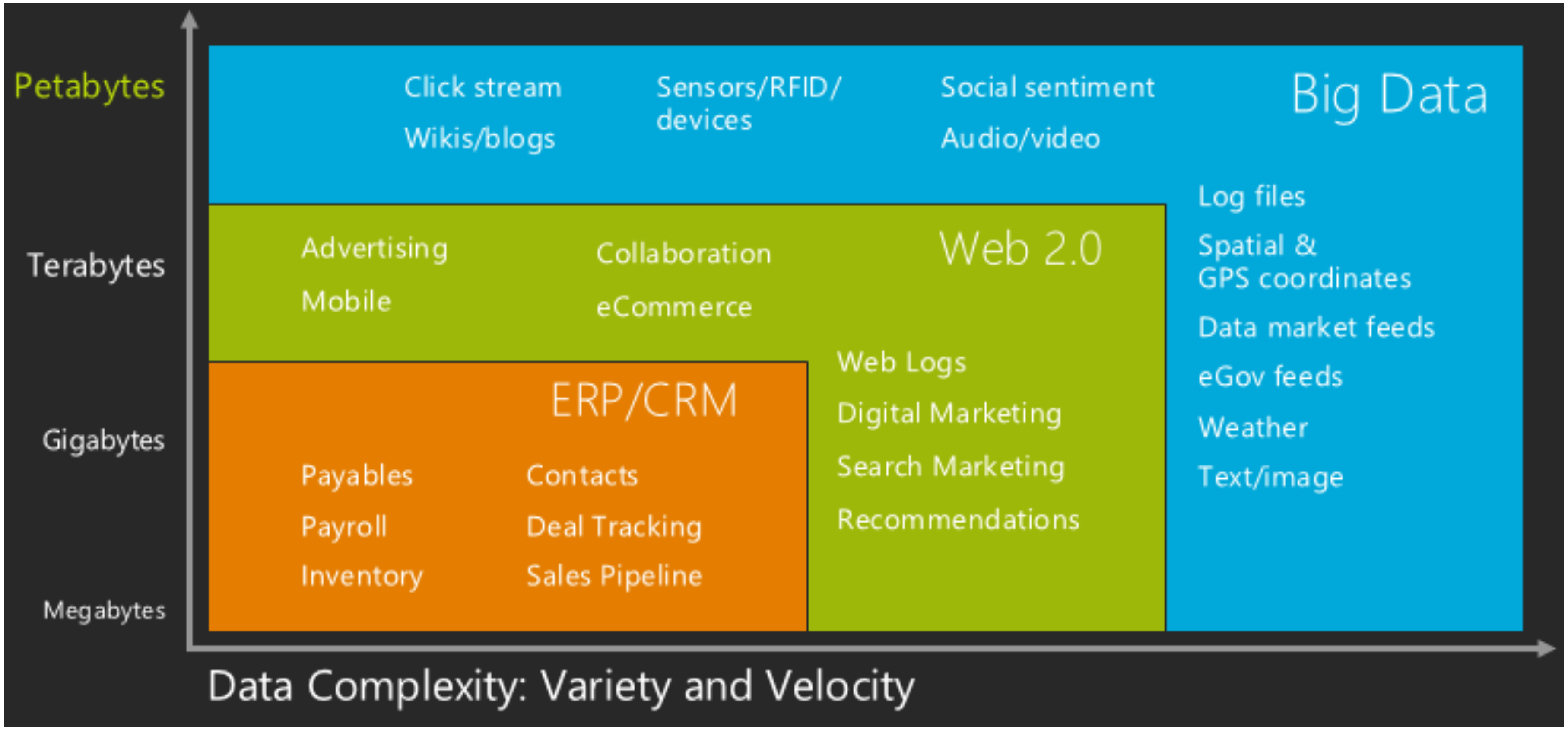
Beyond SQL Databases

Akitaka Matsuo

Big Data and Databases

University of Essex

- Your data can be really big
  - Gigabytes? Terabytes? Petabytes or more?
- And also very complicated



Database Solutions for Big Data

- Different types of databases (SQL vs NoSQL)
- Cloud solution using fully managed services

SQL: Review

SQL: **S**tructured **Q**uery **L**anguage

- We have learned how to run query in SQL databases
- Example:

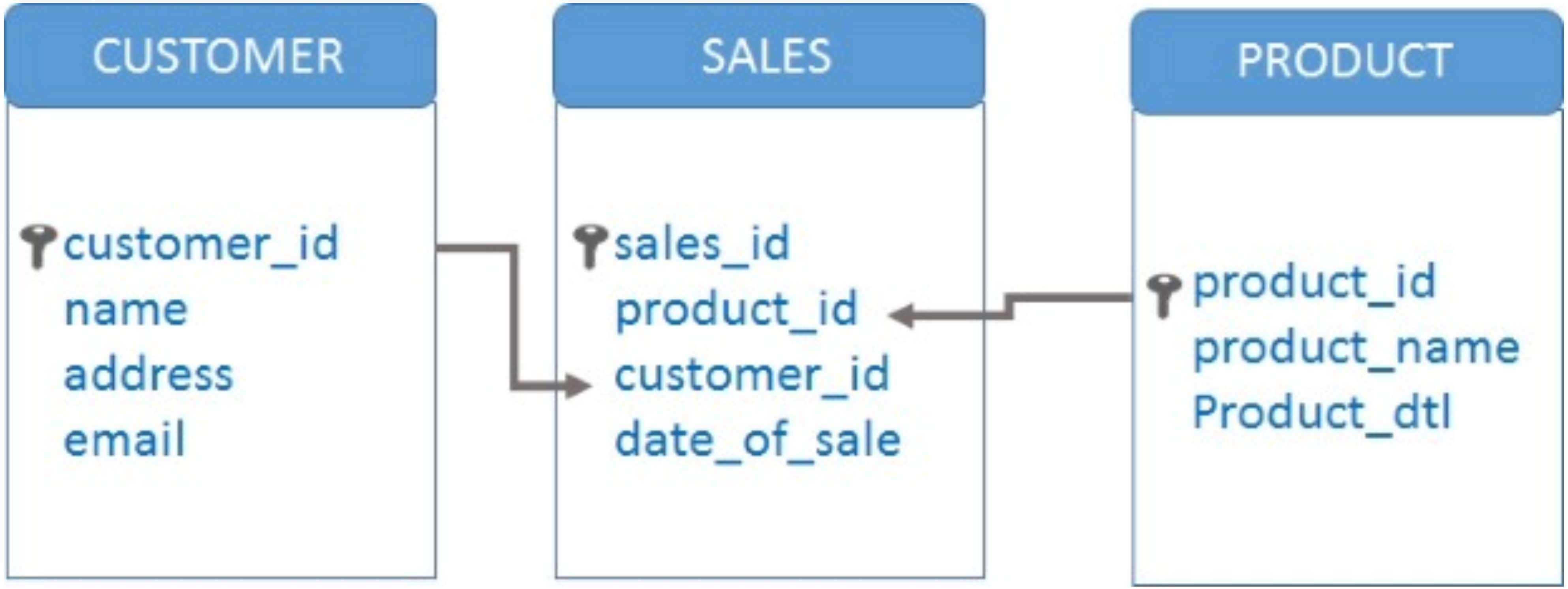
```
SELECT name, party FROM congress;
```

- SQL databases has a strict structure

SQL Structure

It’s all about relations

A simple e-commerce example



CUSTOMER
customer_id
name
address
email

SALES
sales_id
product_id
customer_id
date_of_sale

PRODUCT
product_id
product_name
Product_dtl

SQL: Review

- The result of an SQL query is always a table
- There are a number of commands, and many of them are about the relations between tables SELECT columns
  - **FROM** a table in a database
  - **WHERE** rows meet a condition
  - **GROUP** BY values of a column
  - **ORDER** BY values of a column when displaying results
  - **LIMIT** to only X number of rows in resulting table
- Always required: **SELECT** and **FROM**.
- Rest are optional.
  - **SELECT** can be combined with operators such as **SUM, COUNT, AVG**…
  - To merge multiple tables, you can use **JOIN**

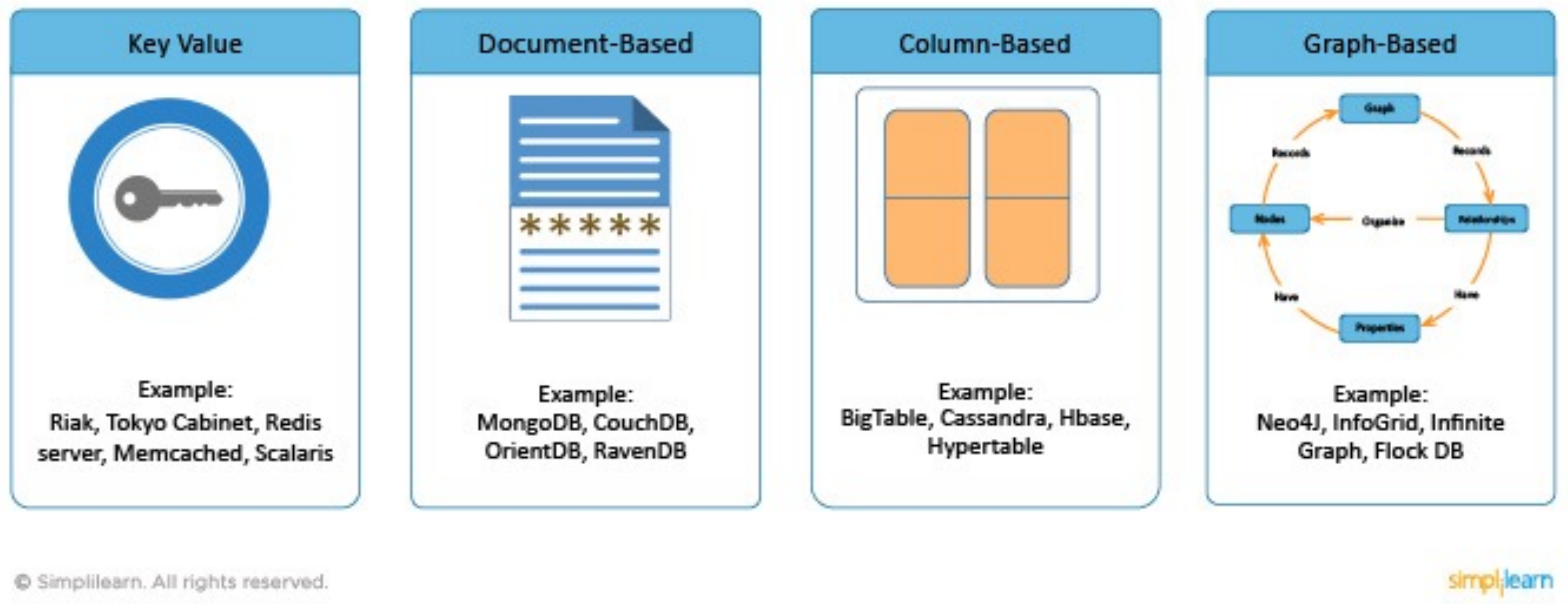
noSQL

A **noSQL** (originally referring to “non SQL”, “non relational” or “not only SQL”) database provides a mechanism for storage and retrieval of data which is modeled in means other than the tabular relations used in relational databases. (

noSQL databases are good for the data with:

- – High data **velocity** – lots of data coming in very quickly
- – Data **variety** – data can be structured, semi-structured and unstructured
- – Data **volume** - total size of data
- – Data **complexity** - stored in many locations

noSQL: types



Key Value

Example:
Riak, Tokyo Cabinet, Redis server, Memcached, Scalaris

Document-Based

Example:
MongoDB, CouchDB, OrientDB, RavenDB

Column-Based

Example:
BigTable, Cassandra, Hbase, Hypertable

Graph-Based

Example:
Neo4J, InfoGrid, Infinite Graph, Flock DB

From: Simplelern

noSQL: Pros and Cons

PROS	CONS
Massive scalability	Limited query capabilities
High availability	Not standardized
Schema flexibility	Not matured
Sparse and semistructured data	Developer heavy