



5602201

5. Data Modeling & Management: Past, Present and Future

CHUTIPORN ANUTARIYA

Outline



Big Data and
Data Analytics



NoSQL Databases



Document Model



MONGODB



Data Privacy &
Data Governance

Data Era



Big Data



Data Analytics

Analyzing Big Data to uncover hidden patterns, identify opportunities and garner insights that drive business innovation and process improvements.

A top-down view of four people sitting around a wooden table in a meeting. They are looking at various devices like laptops, tablets, and smartphones. The image is overlaid with a semi-transparent teal filter.

Let's Discuss :

Sources of Data ...

Where do data come from?

Who/what generates data?

Three major sources of data...



Machines

Data generated from
real time sensors,
machines, vehicles,
web logs, etc.



People

Tweets, status
updates, social
media data, photos,
videos



Organizations

More traditional type
of data : transaction
information,
databases, data
warehouses

THE INTERNET IN 2023 EVERY MINUTE



2023 Internet Minute

Source: <https://ediscoverytoday.com/2023/04/20/2023-internet-minute-infographic-by-ediscovery-today-and-ltmg-ediscovery-trends/>



Big Data

Big Data : What is



Structured

Pre-defined data models like databases

Usually text only

Easy to search and filter

Examples:
Dates, phone numbers, transaction information



Semi-Structured

Both structured & unstructured qualities

Considerably easier to analyze than unstructured data

Examples:
Emails, CSV files, JSON files



Unstructured

No Pre-defined data models

Difficult to search through

Usually stored as different types of files

Examples:
Social media data, audio files, images



Structure of Data



Importance & Role of Data Analytics in Businesses & Organizations

Importance & Role of Data Analytics in Organizations



Make Sense Of Data



**Transform Raw Data Into
Insights And Trends**



**Use Various Tools
And Techniques:
Coding, Math &
Statistics, Machine
Learning**



**Drive Data-driven
Decision Making**

Goals & Benefits



Better insight into customer needs and expectations



Faster, more effective decision making



Improvement of process and cost efficiency



Streamlined business operations

Business/ Career Opportunities



Data / Business Consultancy & Services



Data Infrastructure / Cloud Service Providers



Tech Startups providing data analytics tools/products/services



Skilled & Talent workforce with Data Analytics / Data Science Skills



Professional Trainings and Education Programs



Data Professionals



Data
Scientist



Big Data
Engineer



Big Data
Analyst



Data
Visualization
Developer



Machine
Learning
Engineer

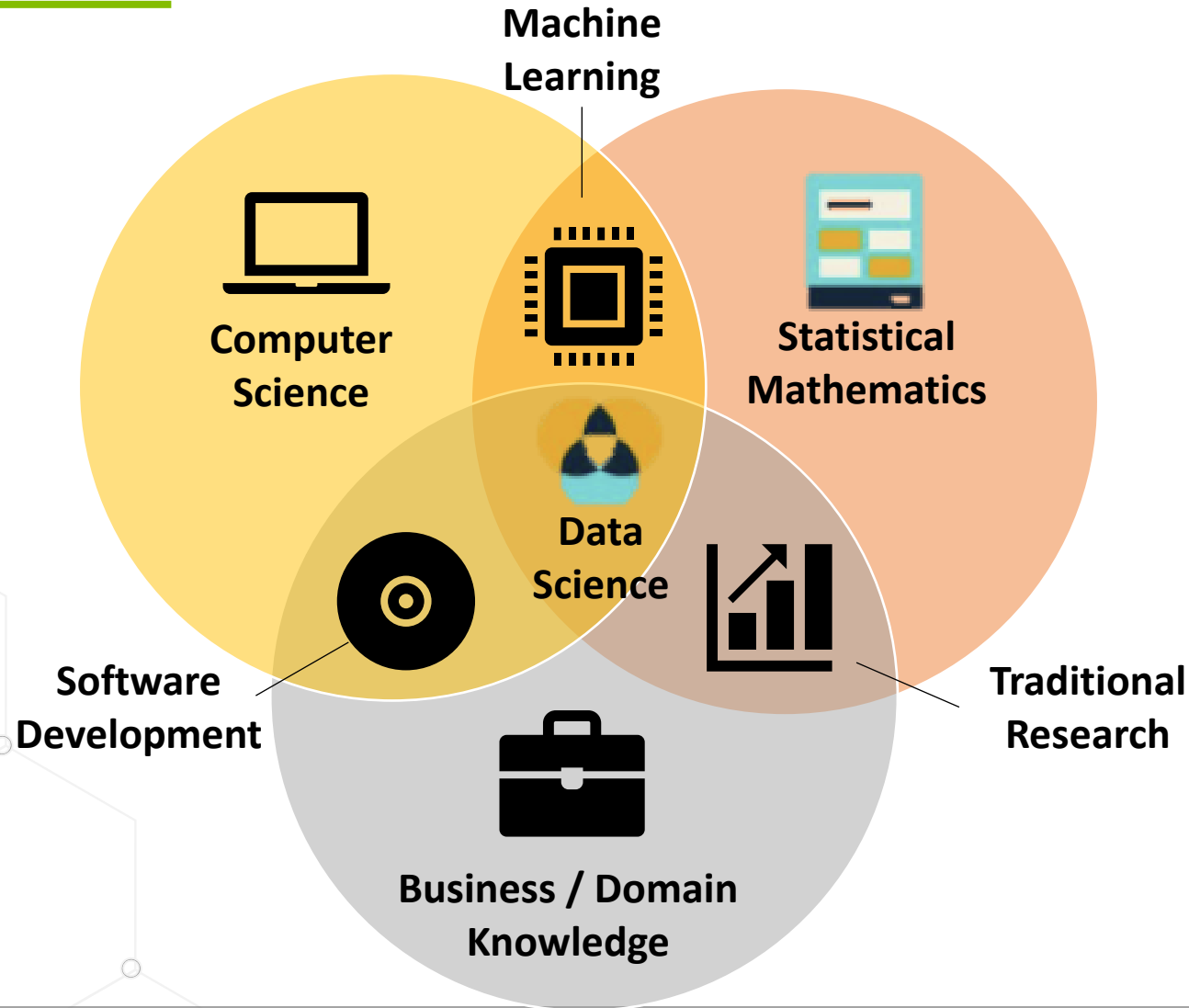


Business
Intelligence
Engineer



Business
Analytics
Specialist

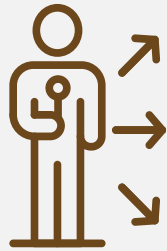
Data Analytics/Data Science Knowledge & Skills Landscape





Data is a starting point...

Why is DATA IMPORTANT?



Data-driven
Decision
Making and
Planning



Data-driven
Business
Operation



Data-driven
Business
Strategic
Development

Characteristics of Data Quality



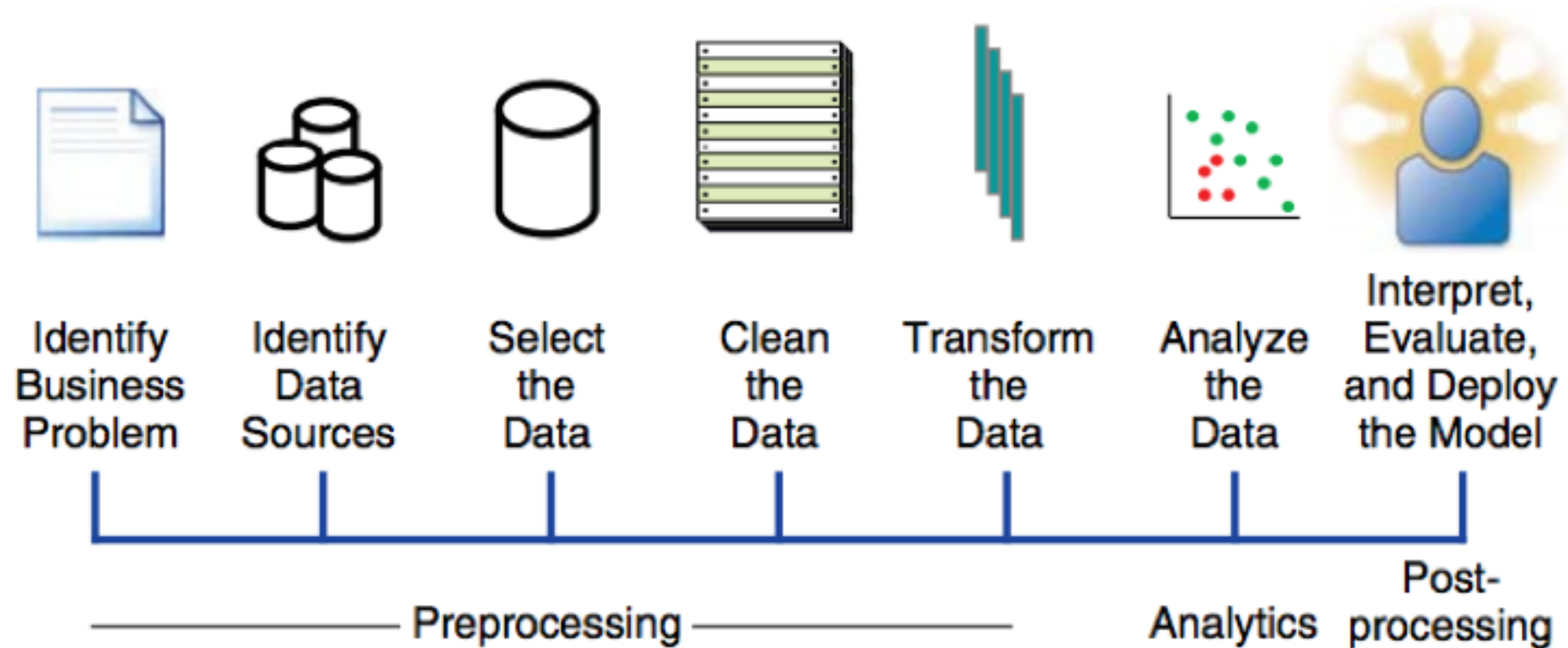
Characteristics of Data Quality

Characteristic	How to measure
Accuracy	Is the information correct in every detail?
Completeness	How comprehensive is the information?
Reliability	Does the information contradict other trusted resources?
Relevance	Do you really need this information?
Timeliness	How up- to-date is information? Can it be used for real-time reporting?

The background is a complex, layered illustration in shades of brown and tan. It features a central globe, various data charts (line graphs, bar charts, pie charts), and two robotic hands at the bottom holding a glowing green data stream. The overall theme is data science and technology.

Data Analytics Process

Overview of the Analytics Process Model



Business Analytics Process Model

Src: <https://blogs.sas.com/content/sgf/2019/05/14/big-data-in-business-analytics-talking-about-the-analytics-process-model/>

Different Kinds of Data Models

Relational DBs are the
most successful technology
for the last 50 years

Relational

IBM Research Defines the Relational Database

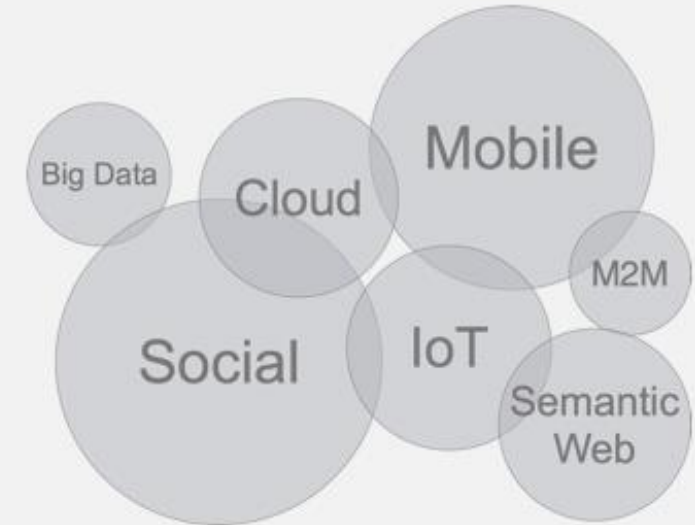
Until the mid-1970s, computers sorted information using rigid, one-off database programs. Predecessor systems like IBM's IMS and VSAM on the mainframe could store megabytes of data, but it had to be entered and retrieved in the same structured way every time. IBM researcher E. F. "Ted" Codd wanted to improve the way data was sorted and handled. He sought to create a generalized description of how to store, update and extract data with accuracy, and query responses so any changes to data produced consistent results. In 1970, Codd completed his definition of the relational database, which became the foundation for IBM DB2 products.



NoSQL

A NoSQL database provides a mechanism for **storage** and **retrieval** of data that is modeled in means **other than the tabular relations** used in **relational databases**. Motivations for this approach include: simplicity of design, "horizontal" scaling, which is a problem for relational databases, and finer control over availability

Structured Data	→	Unstructured Data
Small Datasets	→	Large Volume
Few Relationships	→	Connected Data
Waterfall Approach	→	Agile Approach
Scale Up	→	Scale Out
CIO	→	Developers

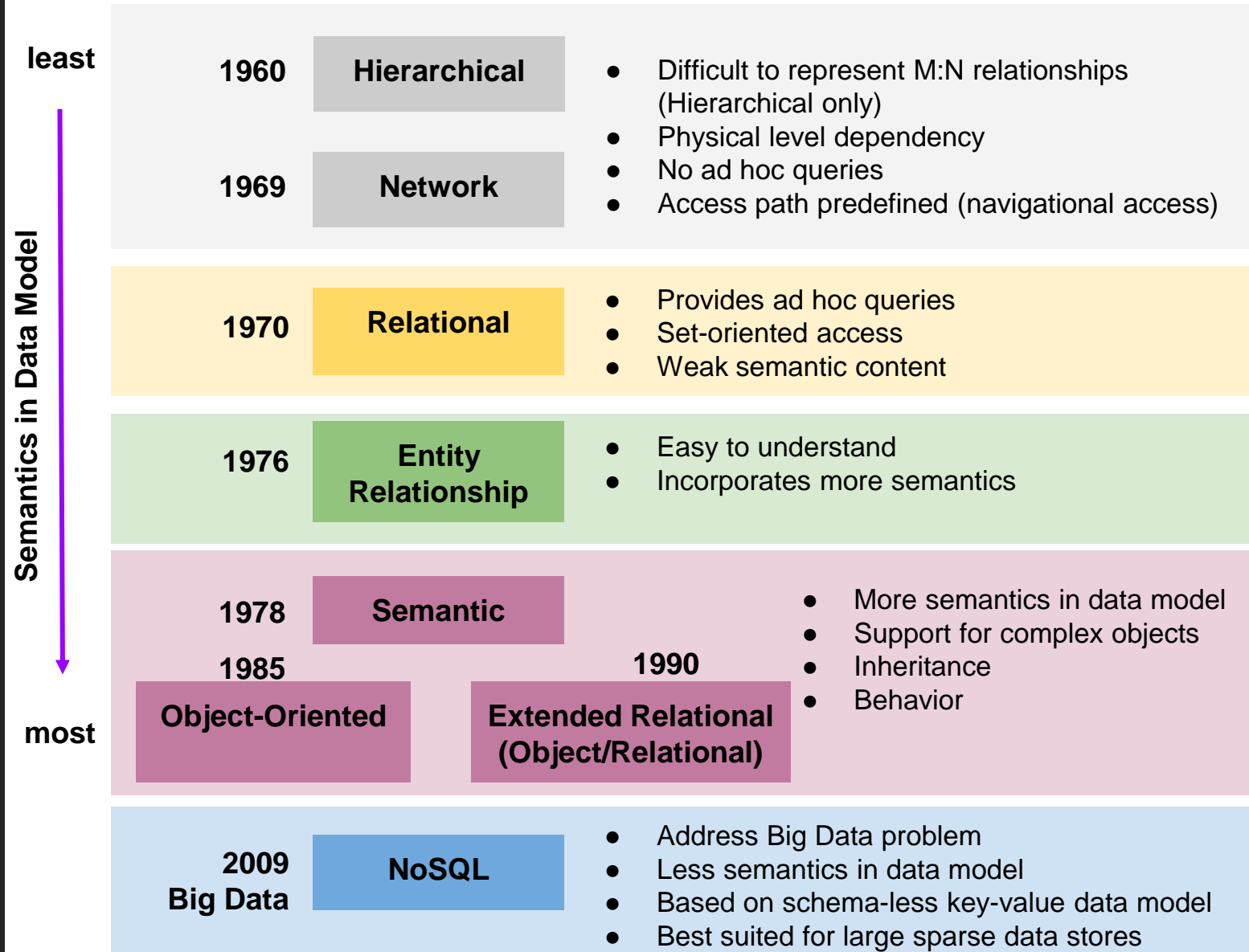


1970

2009

What's Next?

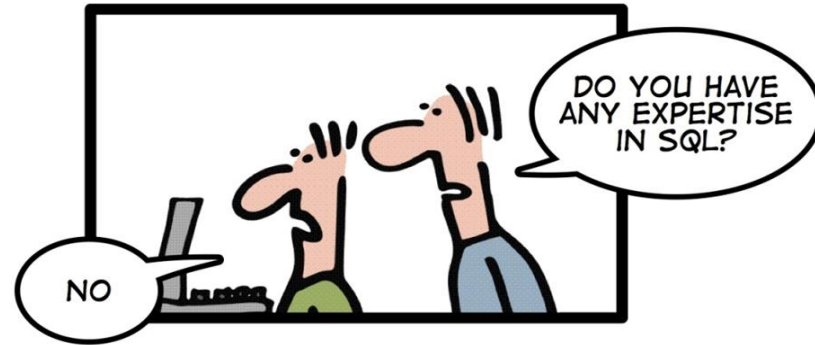
Evolution of data models



A decorative background featuring a network diagram. It consists of numerous nodes, represented by small circles, some of which are solid grey and others are hollow with a grey outline. These nodes are interconnected by thin, light-grey lines, forming a complex web-like structure that is more dense on the left and right sides of the image, leaving the center where the text is located relatively clear.

NoSQL Database Concepts

HOW TO WRITE A CV



Leverage the NoSQL boom

A Little Humor...



3 DATABASE ADMINS



WALKED INTO



A NOSQL BAR ...



A LITTLE WHILE LATER



THEY WALKED OUT BECAUSE



THEY COULDN'T FIND A TABLE

NO SQL



CONCEPTS AND CHARACTERISTICS

NoSQL Origin

Generally newer databases solving new and different problems;

Not only SQL;

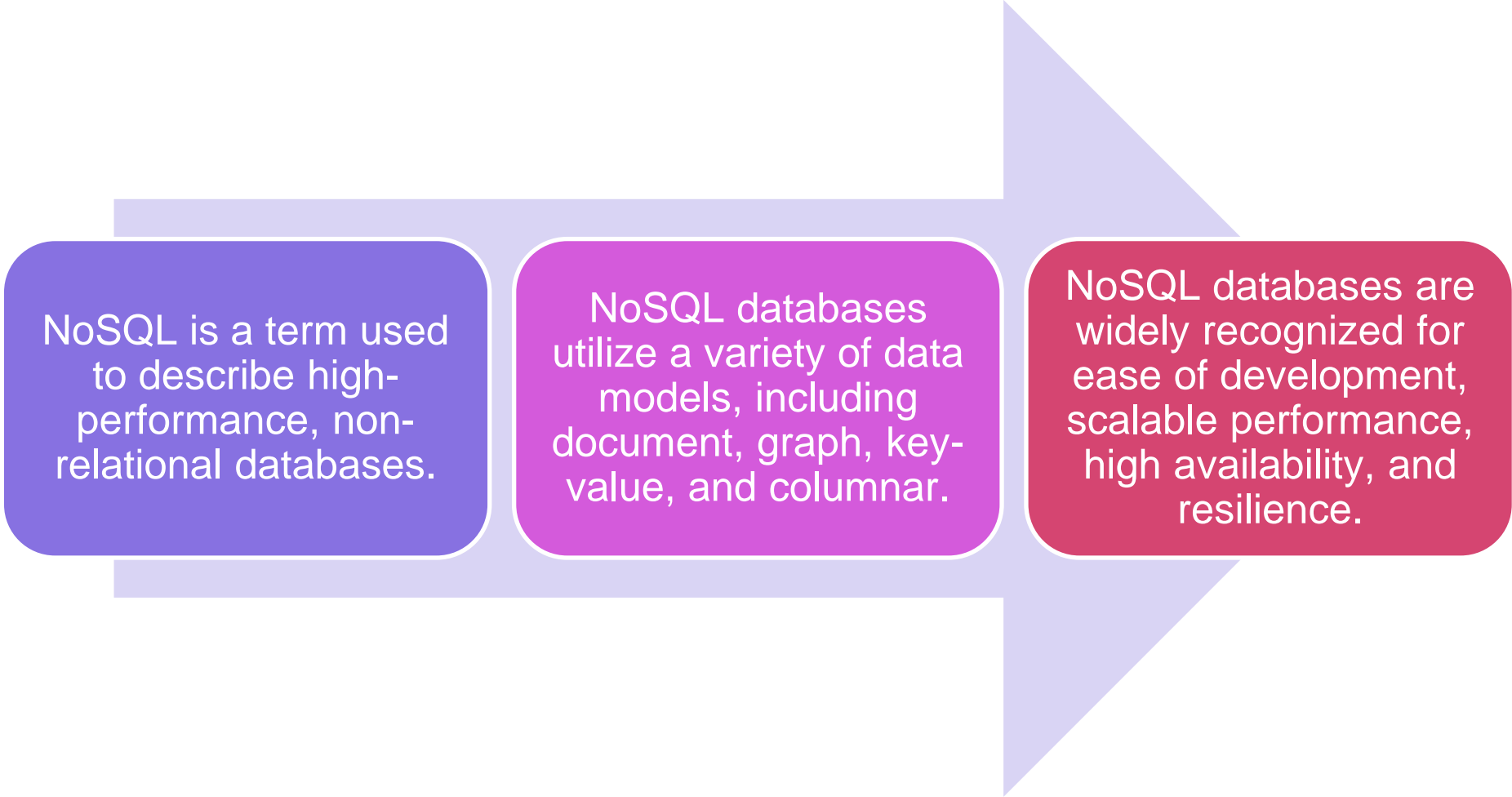
Problems not solved by RDBMSs;

Limitation of RDBMSs, not SQL;

NO SQL

NoSQL is a database technology designed to support the requirements of cloud applications and architected to overcome the scale, performance, data model, and data distribution limitations of relational databases (RDBMS's).

What is NoSQL?



NoSQL is a term used to describe high-performance, non-relational databases.

NoSQL databases utilize a variety of data models, including document, graph, key-value, and columnar.

NoSQL databases are widely recognized for ease of development, scalable performance, high availability, and resilience.

Schema-less Database: what is?

In Relational DB (schemaful DB), there are limitations:

- Cannot add a record which does not fit a schema
- Needs to add NULL values to unused data attribute in a record
- Strong datatyping
- Composite attributes and multivalued attributes are not allowed!!

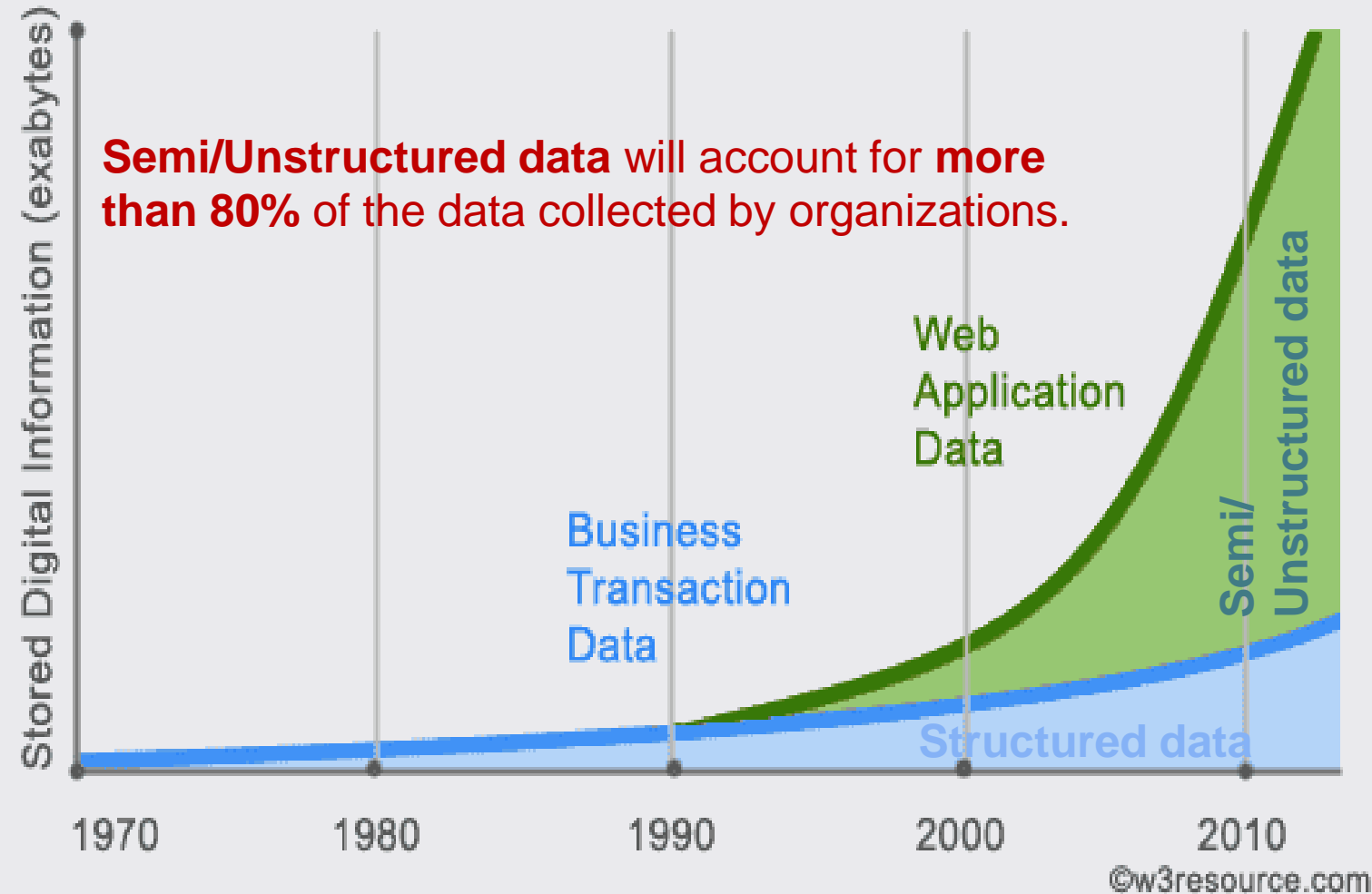
Schema-less Database: what is?

In Schema-less DB

- No fixed, rigid Schema
- No NULL constraint/enforcement
- No datatyping

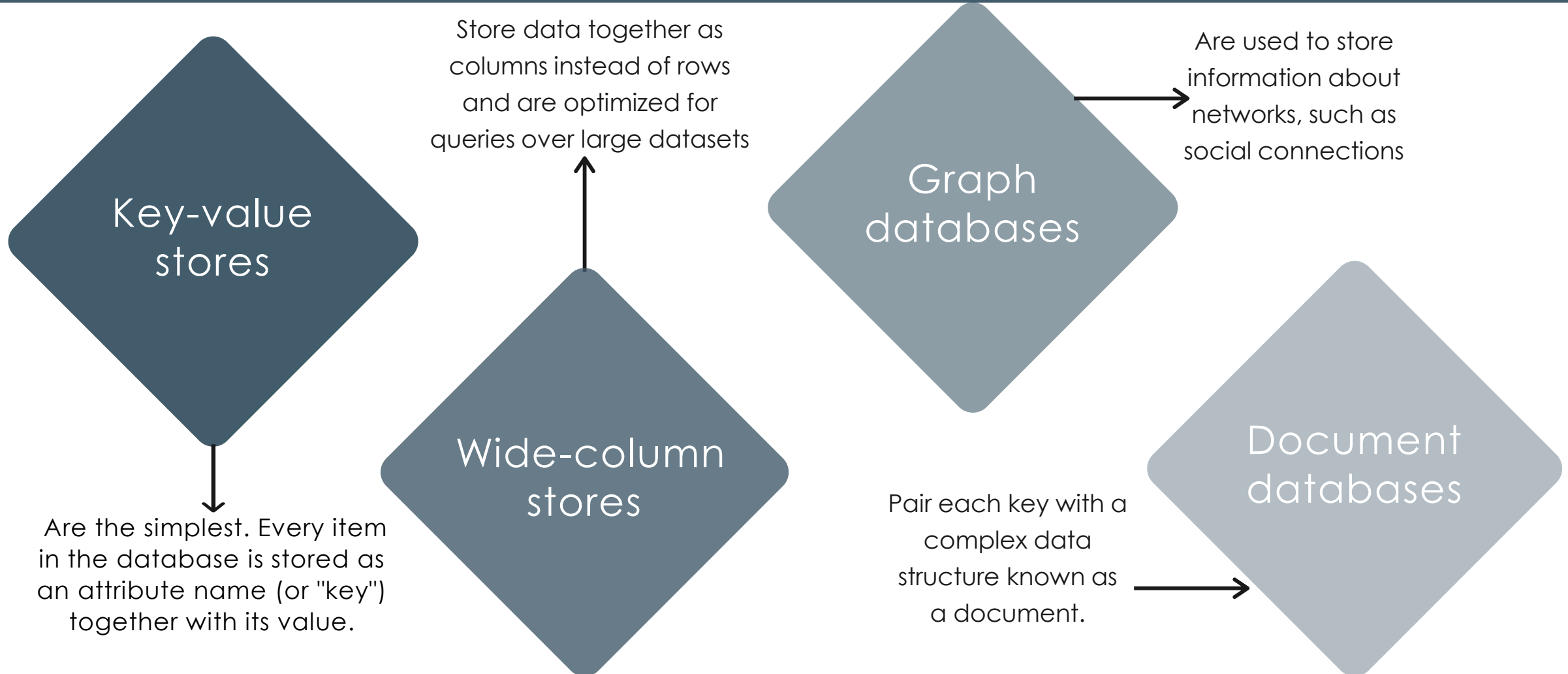
This is Schema-less Database!

Web Apps Driving Data Growth

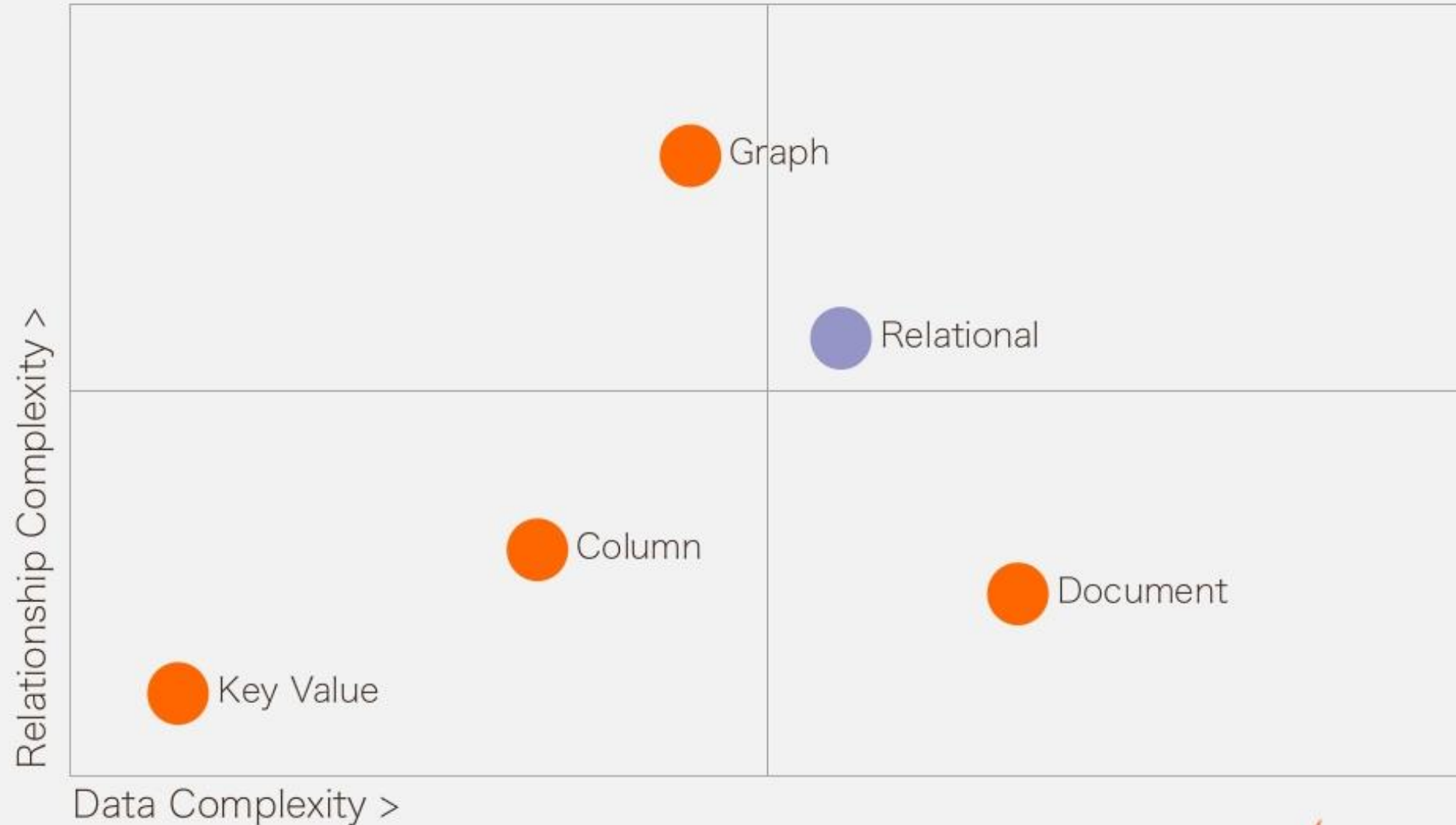


Major Categories of NOSQL Data Models

MAJOR CATEGORIES OF NOSQL DATA MODELS



DBMS Quadrant



One size fits all DB
may not exist!

Polyglot Persistence



When storing data, it is best to use multiple data storage technologies, chosen based upon the way data is being used by individual applications or components of a single application.



Different kinds of data are best dealt with different data stores.



In short, it means picking the right tool for the right use case.

Polyglot Persistence example

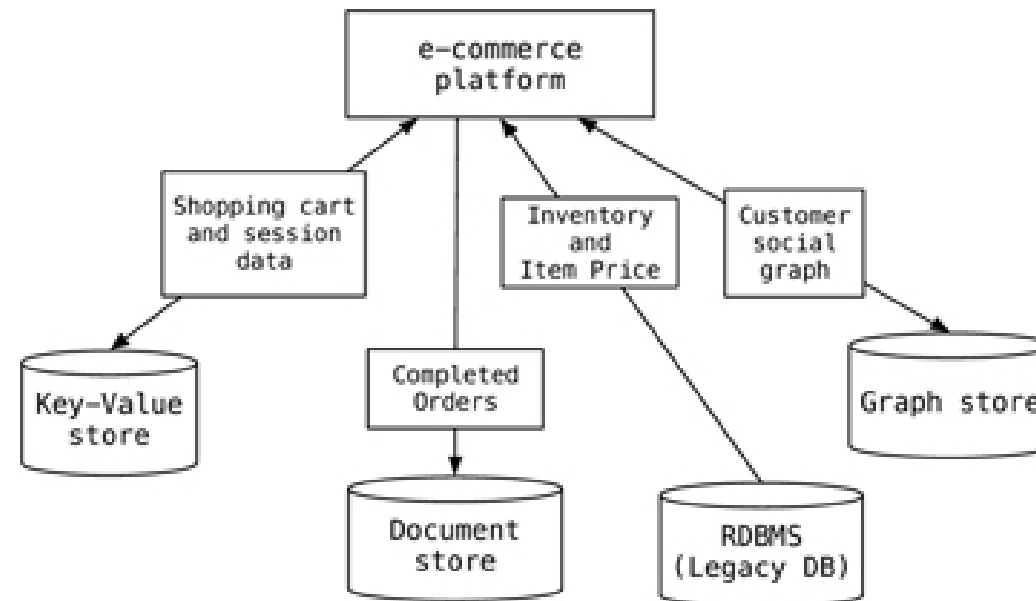


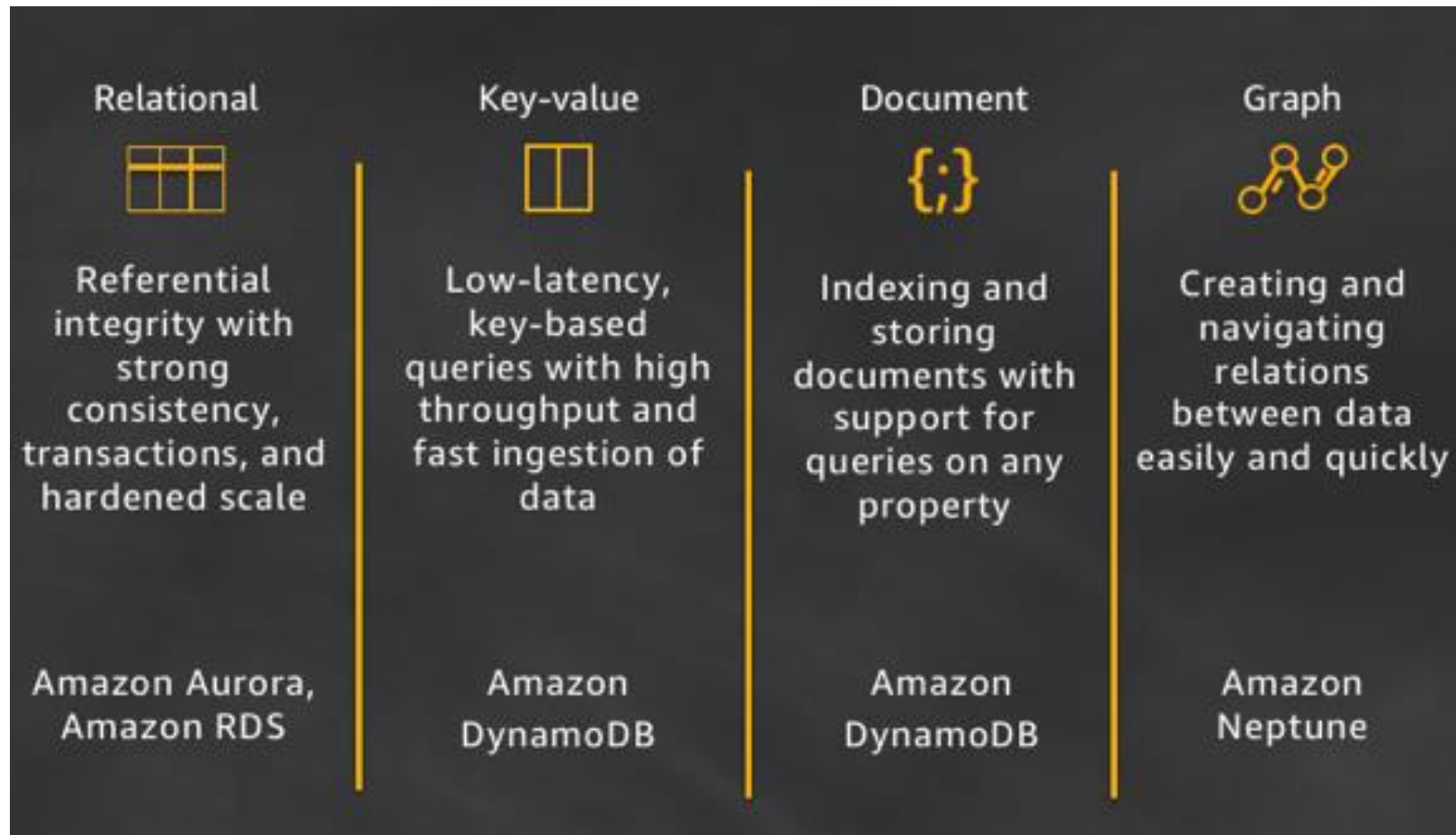
Figure 13.3. *Example implementation of polyglot persistence*

An e-commerce platform will deal with many types of data (i.e. shopping cart, inventory, completed orders, etc) using a mixture of RDBMS solutions with NoSQL solutions

Different Data Stores are suitable with different requirements and usecases

Functionality	Considerations	Database Type
User Sessions	Rapid Access for reads and writes. No need to be durable.	Key-Value
Financial Data	Needs transactional updates. Tabular structure fits data.	RDBMS
POS Data	Depending on size and rate of ingest. Lots of writes, infrequent reads mostly for analytics.	RDBMS (if modest), Key Value or Document (if ingest very high) or Column if analytics is key.
Shopping Cart	High availability across multiple locations. Can merge inconsistent writes.	Document, (Key Value maybe)
Recommendations	Rapidly traverse links between friends, product purchases, and ratings.	Graph, (Column if simple)
Product Catalog	Lots of reads, infrequent writes. Products make natural aggregates.	Document
Reporting	SQL interfaces well with reporting tools	RDBMS, Column
Analytics	Large scale analytics on large cluster	Column
User activity logs, CSR logs, Social Media analysis	High volume of writes on multiple nodes	Key Value or Document

Example: Multiple AWS services



Summary: Different Database and Data Modeling Technologies

Databases are built for a purpose and matching the use case with the database will enable developers to write high-performance, scalable, and more functional applications faster.

Developers also are no longer using a single database for all use cases in an application—they are using many databases.



THANK YOU