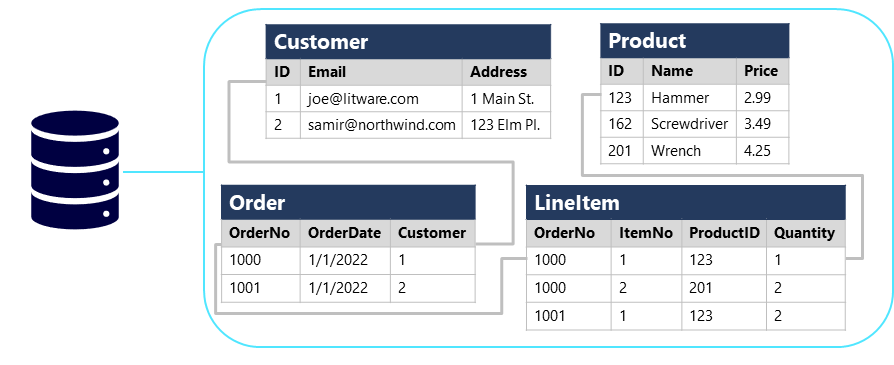
**Explore databases**

A database is used to define a central system in which data can be stored and queried. In a simplistic sense, the file system on which files are stored is a kind of database; but when we use the term in a professional data context, we usually mean a dedicated system for managing data records rather than files.

**Relational databases**

Relational databases are commonly used to store and query structured data. The data is stored in tables that represent entities, such as customers, products, or sales orders. Each instance of an entity is assigned a *primary key* that uniquely identifies it; and these keys are used to reference the entity instance in other tables. For example, a customer's primary key can be referenced in a sales order record to indicate which customer placed the order. This use of keys to reference data entities enables a relational database to be *normalized*; which in part means the elimination of duplicate data values so that, for example, the details of an individual customer are stored only once; not for each sales order the customer places. The tables are managed and queried using Structured Query Language (SQL), which is based on an ANSI standard, so it's similar across multiple database systems.

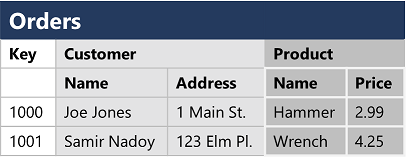


**Non-relational databases**

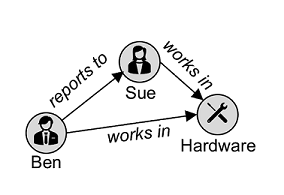
Non-relational databases are data management systems that don’t apply a relational schema to the data. Non-relational databases are often referred to as NoSQL database, even though some support a variant of the SQL language.

There are four common types of Non-relational database commonly in use.

* **Key-value databases** in which each record consists of a unique key and an associated value, which can be in any format.
* **Document databases**, which are a specific form of key-value database in which the value is a JSON document (which the system is optimized to parse and query)
* **Column family databases**, which store tabular data comprising rows and columns, but you can divide the columns into groups known as column-families. Each column family holds a set of columns that are logically related together.



* **Graph databases**, which store entities as nodes with links to define relationships between them.



**Khám phá Cơ sở dữ liệu**

Một cơ sở dữ liệu được sử dụng để định nghĩa một hệ thống trung tâm nơi dữ liệu có thể được lưu trữ và truy vấn. Theo nghĩa đơn giản, hệ thống file nơi các file được lưu trữ là một loại cơ sở dữ liệu; nhưng khi chúng ta sử dụng thuật ngữ này trong ngữ cảnh dữ liệu chuyên nghiệp, chúng ta thường muốn nói đến một hệ thống chuyên dụng để quản lý các bản ghi dữ liệu thay vì các file.

**Cơ sở dữ liệu quan hệ (Relational databases)**

Cơ sở dữ liệu quan hệ được sử dụng thông thường để lưu trữ và truy vấn dữ liệu có cấu trúc. Dữ liệu được lưu trữ trong các bảng đại diện cho các thực thể, chẳng hạn như khách hàng, sản phẩm hoặc đơn đặt hàng. Mỗi thể hiện của một thực thể được gán một **khóa chính (primary key)** duy nhất xác định nó; và các khóa này được sử dụng để tham chiếu đến thể hiện thực thể trong các bảng khác. Ví dụ: khóa chính của khách hàng có thể được tham chiếu trong bản ghi đơn đặt hàng để chỉ ra khách hàng nào đã đặt hàng. Việc sử dụng các khóa để tham chiếu các thực thể dữ liệu này cho phép cơ sở dữ liệu quan hệ được **chuẩn hóa (normalized)**; một phần nghĩa là loại bỏ các giá trị dữ liệu trùng lặp để chẳng hạn, chi tiết của từng khách hàng chỉ được lưu trữ một lần; không phải cho mỗi đơn đặt hàng mà khách hàng đó đặt. Các bảng được quản lý và truy vấn bằng Ngôn ngữ truy vấn cấu trúc (Structured Query Language - SQL), dựa trên tiêu chuẩn ANSI, vì vậy nó tương tự nhau trên nhiều hệ thống cơ sở dữ liệu.

**Cơ sở dữ liệu phi quan hệ (Non-relational databases)**

Cơ sở dữ liệu phi quan hệ là các hệ thống quản lý dữ liệu không áp dụng lược đồ quan hệ cho dữ liệu. Cơ sở dữ liệu phi quan hệ thường được gọi là cơ sở dữ liệu NoSQL, mặc dù một số cơ sở dữ liệu này hỗ trợ một biến thể của ngôn ngữ SQL.

Có bốn loại cơ sở dữ liệu phi quan hệ phổ biến thường được sử dụng.

* **Cơ sở dữ liệu key-value:** Trong đó mỗi bản ghi bao gồm một khóa duy nhất và một giá trị được liên kết, giá trị này có thể ở bất kỳ định dạng nào.
* **Cơ sở dữ liệu tài liệu:** Là một dạng cụ thể của cơ sở dữ liệu key-value trong đó giá trị là một tài liệu JSON (mà hệ thống được tối ưu hóa để phân tích cú pháp và truy vấn).
* **Cơ sở dữ liệu họ cột (column family databases):** Lưu trữ dữ liệu dạng bảng gồm các hàng và cột, nhưng bạn có thể chia các cột thành các nhóm được gọi là họ cột (column-families). Mỗi họ cột chứa một tập hợp các cột có liên quan logic với nhau.
* **Cơ sở dữ liệu đồ thị (graph databases):** Lưu trữ các thực thể dưới dạng các nút với các liên kết để xác định mối quan hệ giữa chúng.