

FRONT SHEET
Individual Coursework

CANDIDATE NUMBER (C-NUMBER)	C2109911
MODULE NAME	Data Design Management
WORD COUNT	2690
SUBMISSION DATE	30/01/2025

DECLARATION

I certify that this assessment submission is entirely my work and I have fully referenced and correctly cited the work of others, where required. I also confirm the contents of my submission have not been generated by a third party, or through an Artificial Intelligence generative system*.

I have read the Student Discipline Regulations ([Student Discipline Regulations](#)) and understand any Assessment Related Offence/ Academic Misconduct may result in penalties being applied.

By submitting this assessment submission, I am confirming that I am fit to sit according to the Assessment Regulations.

I declare that:

- This is my own unaided work.
Yes No
- The word count stated by me is correct.
Yes No
- I'm happy for my work to be retained on the Elite repository and made available to staff and future students**
Yes No

*Please note that all the assignments are submitted to Turnitin.

**Please note personal information (such as names) will be deleted.

Instructions to candidates:

1. Please complete this cover sheet by entering your Candidate Number, Module Name, Word Count, and Submission Date.
2. You must NOT use your NAME on this cover sheet or on any part of your coursework.



DATA DESIGN MANAGEMENT

Introduction to Data Design Management

Importance of Databases in Modern Applications

- Fundamental to e-commerce operations
- Supports customer registration, product tracking, and order processing

Role of Structured Data Management

- Ensures organized and efficient data handling
- Facilitates seamless data flow across business functions

Database Lifecycle Management

- Optimizes data storage, retrieval, and security
- Enhances business scalability and operational efficiency

Databases are the backbone of modern applications, especially in e-commerce, where businesses manage vast amounts of customer and transaction data. Structured data management ensures smooth customer registration, real-time product tracking, and efficient order processing, contributing to operational success (Connolly & Begg, 2015). Effective **database lifecycle management** plays a crucial role in maintaining data integrity, optimizing storage, and securing sensitive information. By streamlining the flow of data, businesses can improve decision-making, enhance customer experience, and scale operations efficiently (Kumar & Banyal, 2020).

CURRENT SITUATION ANALYSIS – TECHNOLOGICAL ADVANCEMENTS IN E-COMMERCE

Real-Time Data Analysis

- Pricing is a major factor influencing sales in e-commerce.
- Companies like **Amazon** use data (sales ranks, prices, algorithms) to **reverse engineer demand** (Ghose & Sundararajan, 2006).
- **Dynamic pricing** adjusts prices in real time based on customer purchase behaviour.
- Optimized pricing boosts **sales and competitiveness**.

Customization Strategies

- **Big data** enables personalized product recommendations and tailored offerings.
- Retailers analyze **clickstream activity, social media, and mobile app usage** to understand customer behaviour.
- The **Consumer-to-Business (C2B) model** offers customized products, improving satisfaction and loyalty (Chen, 2021).
- **Machine learning & real-time recommendation systems** provide targeted suggestions during browsing.
- **Big data platforms** improve scalability and eliminate database bottlenecks, ensuring smooth operations.

Strategic Impact

- **Enhances customer experience** with tailored services.
- **Improves operational efficiency** and decision-making.
- **Ensures competitiveness** in a rapidly evolving e-commerce landscape.

Current Situation Analysis – Technological Advancements in E-Commerce

Data is being used widely by businesses and invariably in e-commerce industry for a myriad of ways.

Real-time data analysis:

Finding an optimal pricing strategy for products is a constant challenge for businesses as it is one of the major driving factors of sales. Companies like Amazon use large amount of data such as sales ranks, prices, and algorithms to “reverse engineer” demand based on sales ranks. This enables them to perform real-time data analysis which in turn enables them to perform dynamic pricing (Ghose & Sundararajan, 2006). These price adjustments reflect the purchase behaviour of customers based on prices and sales ranks which help them in optimizing product prices dynamically.

Customization strategies:

In the context of e-commerce, big data plays a critical role in customization and personalization, allowing stores to tailor their offerings to individual consumer preferences. By analysing clickstream data, social media activity, and mobile app usage, retailers can gain valuable insights into customer behaviour and preferences. The Consumer-to-Business (C2B) model, highlighted in the study, leverages these insights to offer personalized products that cater to specific customer needs, thereby enhancing customer satisfaction and loyalty. Advanced technologies such as data integration and preprocessing platforms, combined with machine learning algorithms, support the creation of real-time recommendation systems that deliver targeted product suggestions during browsing sessions. Additionally, big data platforms address the limitations of traditional databases, such as query bottlenecks, by enabling real-time data processing and ensuring scalability for large datasets. These technological advancements empower online retail stores to refine their strategies, improve operational efficiency, and remain competitive in a dynamic market environment (Chen, 2021).

Data is widely utilized by businesses, especially in the e-commerce industry, for various strategic purposes. Real-time data analysis and customization strategies are two key technological advancements shaping the strategies of online retail stores.

Real-Time Data Analysis:

Finding an optimal pricing strategy is one of the major challenges for e-commerce businesses, as pricing significantly influences sales. Companies like Amazon leverage vast amounts of data, including sales ranks, prices, and algorithms, to "reverse engineer" demand based on sales ranks. This process enables them to perform **real-time data analysis**, allowing for **dynamic pricing** that adjusts prices in response to customer purchase behaviours. These adjustments are informed by the relationship between sales ranks and price, optimizing product prices dynamically to boost sales and competitiveness (Ghose & Sundararajan, 2006).

Customization Strategies:

Big data plays a critical role in enabling online retail stores to offer tailored products and personalized recommendations. By analysing data such as clickstream activity, social media interactions, and mobile app usage, businesses gain insights into customer preferences and behaviour. The **Consumer-to-Business (C2B) model**, as described by Chen (2021), leverages these insights to provide personalized products that meet specific customer needs, thereby increasing satisfaction and loyalty. Advanced technologies, such as **data integration and preprocessing platforms**, combined with

machine learning algorithms, support real-time recommendation systems. These systems deliver targeted product suggestions during browsing sessions, enhancing the shopping experience. Additionally, **big data platforms** address the limitations of traditional databases, such as query bottlenecks, by enabling real-time processing and scalability for large datasets. These advancements empower online retail stores to refine their strategies, improve operational efficiency, and remain competitive in a dynamic market environment (Chen, 2021).

EVALUATION OF KEY RDBMS TECHNOLOGIES AND ADVANCEMENTS

CLOUD-BASED RDBMS

- ENHANCES SCALABILITY, PERFORMANCE, AND AVAILABILITY
- EXAMPLE: **AMAZON RDS** AUTOMATES DATABASE MANAGEMENT FOR HIGH TRANSACTION VOLUMES (CARPENTER & FOWLER, 2021)

DISTRIBUTED DATABASES

- ENSURES FAULT TOLERANCE AND LOW-LATENCY ACCESS
- EXAMPLE: **APACHE CASSANDRA (USED BY EBAY)** SUPPORTS GLOBAL DATA DISTRIBUTION (CARPENTER & FOWLER, 2021)

ROLE OF INNOVATION IN DATABASE PERFORMANCE

OPTIMIZED QUERY PROCESSING & DATA REPLICATION

- ENSURES HIGH AVAILABILITY AND FASTER RESPONSE TIMES (ELMASRI & NAVATHE, 2020)

INDUSTRY EXAMPLES:

- **AMAZON RDS:** AUTOMATES SCALING, REDUCING DOWNTIME AND IMPROVING CUSTOMER EXPERIENCE (ELMASRI & NAVATHE, 2020)
- **SHOPIFY & GOOGLE CLOUD SPANNER:** SUPPORTS SEAMLESS GLOBAL TRANSACTION MANAGEMENT (ELMASRI & NAVATHE, 2020)

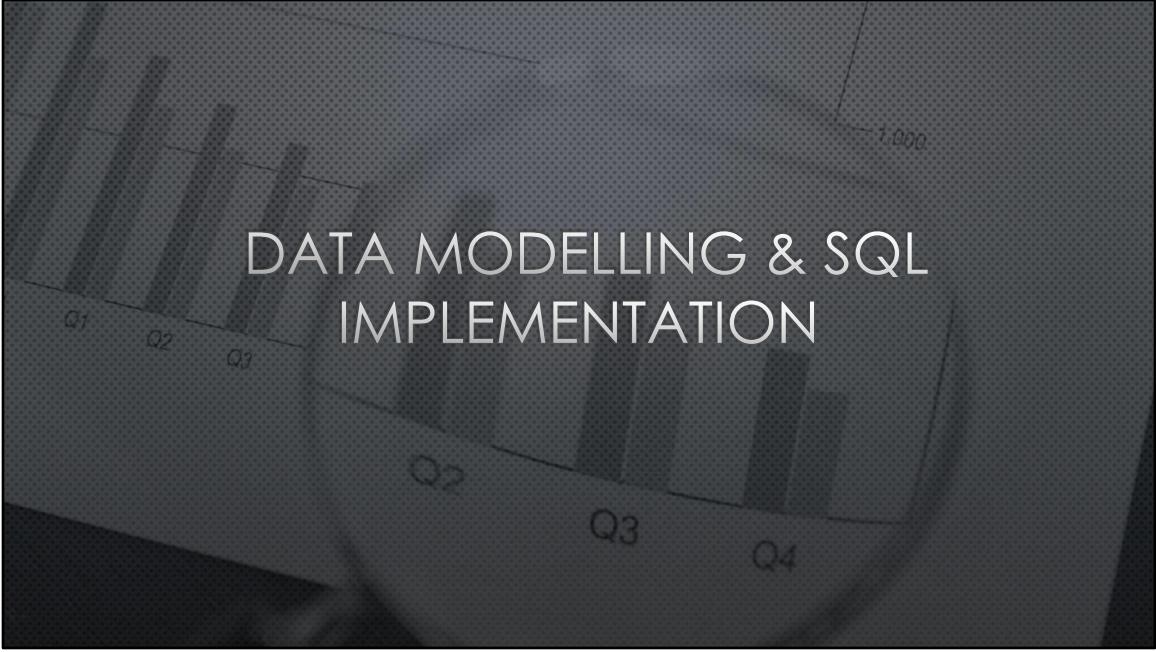


ROLE OF INNOVATION IN DATABASE PERFORMANCE

Modern RDBMS technologies enhance **scalability, availability, and fault tolerance**. Cloud-based RDBMS solutions, such as **Amazon RDS**, streamline database management for high-traffic e-commerce sites. **Amazon RDS automates administrative tasks like patching and backups**, enabling businesses to focus on core operations (Carpenter & Fowler, 2021).

Distributed databases, like **Apache Cassandra**, used by **eBay**, provide low-latency access across multiple regions, ensuring uninterrupted availability even during server failures (Carpenter & Fowler, 2021).

Innovations in database technology, including **automated scaling and optimized query processing**, improve performance and reliability. **Amazon RDS** reduces downtime through automation, improving customer experience. **Google Cloud Spanner**, leveraged by **Shopify**, ensures transactional consistency across a global customer base, supporting real-time analytics and business growth (Elmasri & Navathe, 2020).



DATA MODELLING & SQL IMPLEMENTATION

NORMALIZATION PROCESS – ENSURING DATA INTEGRITY

1NF – Eliminating Redundancy	Atomic values (no multiple values in one field). No repeating groups; each row is unique.
2NF – Removing Partial Dependencies	Every attribute depends on the entire primary key.
3NF – Eliminating Transitive Dependencies	Product variants and order details stored separately. Attributes depend only on the primary key.
Final Benefits:	
<input checked="" type="checkbox"/> Reduced redundancy & improved efficiency. <input checked="" type="checkbox"/> Enhanced data integrity & security.	

Normalization Process – First Normal Form (1NF)

Normalization reduces redundancy and enhances data integrity through stages (1NF, 2NF, 3NF).

1NF ensures:

- Atomic values (each field holds a single, indivisible value).
- No repeating groups or multivalued attributes.
- A unique identifier (Primary Key) for each row (Chris, 2022).

1NF Application:

- **Customer Registration:** Phone numbers and addresses stored in separate rows.
- **Product Management:** Each product variant (size, colour) stored separately.

- **Order Management:** Orders split so each row represents one product per order.
- **Employee Management:** Emergency contacts stored separately per employee.

Normalization Process – Second Normal Form (2NF)

2NF eliminates partial dependencies by ensuring all attributes depend on the entire primary key.

2NF Application:

- **Customer Registration:** Already in 2NF.
- **Product Management:** Product variants moved to a new table.
- **Order Management:** Product details separated; only product_id referenced in order details.
- **Employee Management:** Already in 2NF.

Establishing Relationships After 2NF

With 2NF, relationships between tables are clearer:

1. **Customer Registration Table** → customer_id is referenced in the **Order Table**.
2. **Product Management Table** → product_id is referenced in **Order Details and Product Variants**.

New relationships established between tables ensure referential integrity. Additional tables include Inventory and Stock Movement for better tracking.

Normalization Process – Third Normal Form (3NF)

3NF removes transitive dependencies, ensuring all attributes depend only on the primary key (Chris, 2022).

3NF Application:

- **Customer Registration:** City-related details moved to a Location Table.
- **Product Management:** Already in 3NF.
- **Order Management:** Payment method moved to Customer Payment Table.
- **Employee Management:** Manager details moved to a separate table.

A Payment Information Table enhances security and flexibility.

A separate Payment Information Table is introduced for better security and flexibility.

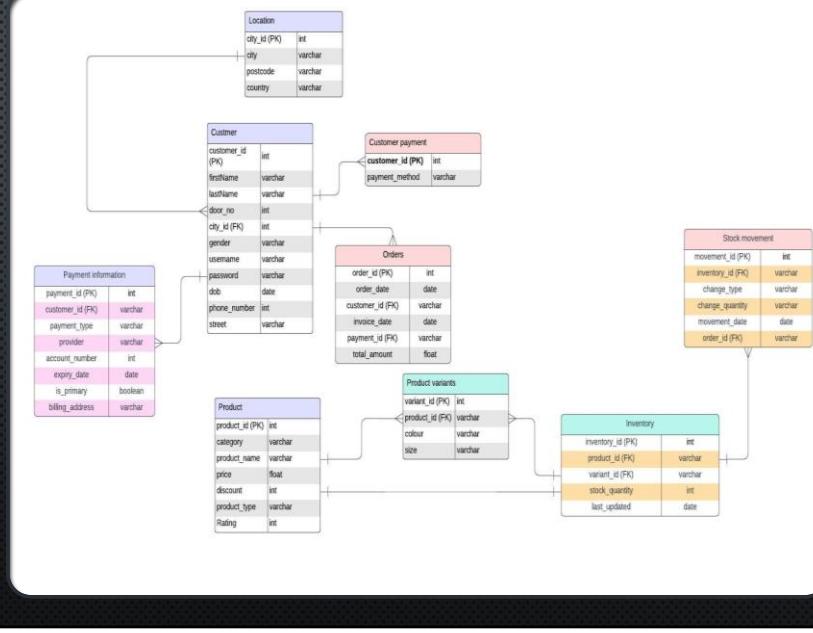
Final Summary:

- **1NF:** Eliminated duplicate data, ensured atomic values.
- **2NF:** Removed partial dependencies.
- **3NF:** Eliminated transitive dependencies for optimal efficiency.

This structured approach ensures minimal redundancy and maximizes data consistency.

ER – DIAGRAM

- **CUSTOMER MANAGEMENT**
- **PRODUCT MANAGEMENT**
- **ORDER MANAGEMENT**



The ER diagrams illustrate a comprehensive design for managing customer interactions and employee operations in a relational database system.

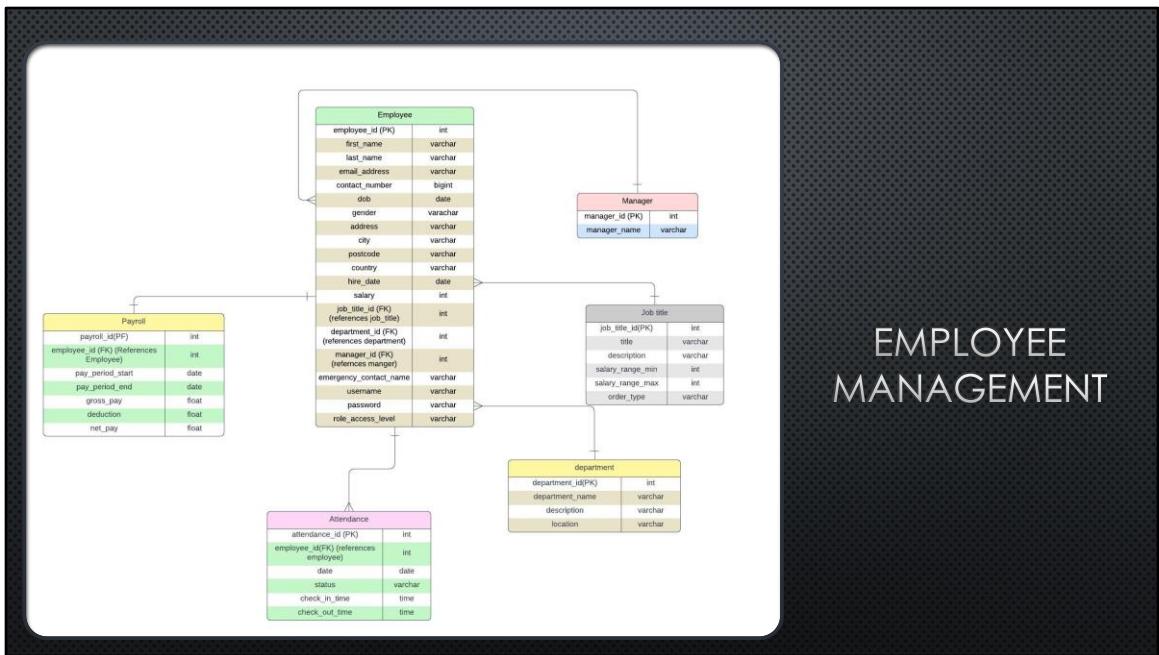
Customer Management , Product management, and Order management Diagram:

This schema organizes customer data into modular tables, ensuring normalization and reducing redundancy. Key entities include:

- **Customer:** Stores personal details like name, address, and contact information. The city is linked to the **Location** table via **city_id** for better data integrity.
- **Orders:** Captures order details, linked to the **Product Variants** and **Inventory** tables for tracking product specifics like size and colour. The **Stock Movement** table records inventory changes.
- **Payment Information:** Enables secure payment processing by segregating payment data into its table, linked to **Customer Payment** and **Orders** via

payment_id.

This structure supports scalability, ensuring smooth customer data handling, order management, and stock control.



Employee Management Diagram:

The employee-focused schema integrates workforce data efficiently. Key entities include:

- **Employee:** Central table containing personal and professional details, linked to the **Department**, **Manager**, and **Job Title** tables for hierarchical management.
- **Payroll:** Tracks salary details and deductions for each employee using **employee_id**.
- **Attendance:** Monitors daily employee activities with attributes like check-in/out times and status.

The design incorporates robust relationships, maintaining referential integrity across tables. It enables accurate reporting and seamless HR operations while ensuring adaptability to organizational needs.

Together, these diagrams demonstrate a robust relational database design optimized for operational efficiency and data integrity.

SQL Scripts

```
CREATE TABLE Customer
(customer_id INT AUTO_INCREMENT PRIMARY KEY, -- Primary key with auto-increment
firstName VARCHAR(50) NOT NULL,
lastName VARCHAR(50) NOT NULL, street VARCHAR(100),
door_no INT,
city_id INT, -- Foreign key to Location table gender ENUM('Male', 'Female', 'Other') NOT
NULL,
username VARCHAR(50) UNIQUE NOT NULL, -- Ensure username is unique
password VARCHAR(255) NOT NULL, dob DATE,
phone_number BIGINT, -- Use BIGINT for phone numbers to accommodate larger values
FOREIGN KEY (city_id) REFERENCES Location(city_id)
ON DELETE SET NULL ON UPDATE CASCADE -- Cascade updates, set null on delete);

CREATE TABLE Location
(city_id INT AUTO_INCREMENT PRIMARY KEY, -- Primary key
city VARCHAR(50) NOT NULL,
postcode VARCHAR(10) NOT NULL,
country VARCHAR(50) NOT NULL
);
```

```

CREATE TABLE Payment_Information (
    payment_id INT AUTO_INCREMENT PRIMARY KEY, -- Primary key
    customer_id INT NOT NULL, -- Foreign key to Customer table
    payment_type ENUM('Credit Card', 'Debit Card', 'PayPal', 'Other') NOT NULL,
    provider VARCHAR(50) NOT NULL,
    account_number BIGINT NOT NULL,
    expiry_date DATE NOT NULL,
    is_primary BOOLEAN NOT NULL DEFAULT TRUE,
    billing_address VARCHAR(255) NOT NULL,
    FOREIGN KEY (customer_id) REFERENCES Customer(customer_id)
    ON DELETE CASCADE ON UPDATE CASCADE -- Cascade updates and deletes);

```

```

CREATE TABLE Orders (
    order_id INT AUTO_INCREMENT PRIMARY KEY, -- Primary key
    order_date DATE NOT NULL,
    customer_id INT NOT NULL, -- Foreign key to Customer table
    invoice_date DATE,
    payment_id INT, -- Foreign key to Payment_Information table
    total_amount FLOAT NOT NULL CHECK (total_amount >= 0), -- Ensure non-negative totals
    FOREIGN KEY (customer_id) REFERENCES Customer(customer_id)
    ON DELETE CASCADE ON UPDATE CASCADE,
    FOREIGN KEY (payment_id) REFERENCES Payment_Information(payment_id)
    ON DELETE SET NULL ON UPDATE CASCADE
);

```

```

create table product (
    product_id int auto_increment primary key, category varchar(50) not null,
    product_name varchar(100) not null,
    price float not null check (price >= 0), -- Ensure price is non-negative
    discount int default 0 check (discount between 0 and 100),
    product_type varchar(50),
    rating int check (rating between 0 and 5) -- Ensure rating is between 0 and 5
);

```

```

create table product_variants (
    variant_id int auto_increment primary key,
    product_id int not null
    colour varchar(30).size varchar(30),
    foreign key (product_id) references product(product_id)
    on delete cascade on update cascade
);

```

<pre>create table inventory (inventory_id int auto_increment primary key, product_id int not null, variant_id int, stock_quantity int not null check (stock_quantity >= 0), -- Non-negative stock last_updated date not null, foreign key (product_id) references product(product_id) on delete cascade on update cascade, foreign key (variant_id) references product_variants(variant_id) on delete set null on update cascade);</pre>	<pre>create table customer_payment (customer_id int not null, -- foreign key to customer table payment_method varchar(50) not null, -- payment method (e.g., Credit card, Paypal) primary key (customer_id, payment_method), -- Composite primary key to ensure uniqueness foreign key (customer_id) references customer(customer_id) on delete cascade on update cascade -- Ensure referential integrity);</pre>
<pre>create table stock_movement (movement_id int auto_increment primary key, inventory_id int not null, change_type enum('increase', 'decrease') not null, change_quantity int not null check (change_quantity > 0), movement_date date not null, order_id int, foreign key (inventory_id) references inventory(inventory_id) on delete cascade on update cascade, foreign key (order_id) references orders(order_id) on delete set null on update cascade);</pre>	<pre>create table manager (manager_id int auto_increment primary key, manager_name varchar(100) not null);</pre>

The SQL scripts implement a structured database schema with well-defined data types, constraints, and referential integrity rules. Here's a critical breakdown:

Data Types:

- **INT & BIGINT:** customer_id, order_id, and product_id use INT AUTO_INCREMENT for unique primary keys. BIGINT is used for phone_number and account_number to handle large values.
- **VARCHAR:** Text fields like firstName, lastName, city, and username use VARCHAR with reasonable length limits.
- **ENUM:** Fields such as gender, payment_type, and change_type restrict values to

predefined options.

- **FLOAT & INT:** price, total_amount, and discount ensure numerical accuracy, with constraints enforcing non-negative values.
- **DATE:** Used for dob, order_date, and expiry_date, ensuring correct date formats.

Constraints & Rules:

- **Primary Keys:** Each table has a unique identifier (AUTO_INCREMENT PRIMARY KEY).
- **Foreign Keys:** Relationships enforce data integrity (e.g., customer_id in Orders references Customer).
- **Cascade Rules:**
 - ON DELETE CASCADE ensures dependent records are removed when a parent is deleted (e.g., Orders linked to Customer).
 - ON DELETE SET NULL retains data integrity for nullable references (e.g., payment_id in Orders).
- **CHECK Constraints:** Ensure logical values (e.g., total_amount \geq 0, discount BETWEEN 0 AND 100).
- **Unique Constraints:** username is unique to prevent duplicates.

- Extract all the customers from a specific city

```
SELECT * FROM Customer c
JOIN Location l ON c.city_id = l.city_id
WHERE l.city = 'London';
```

- Count how many customers are from a specific city

```
SELECT COUNT(*) AS customer_count
FROM Customer c
JOIN Location l ON c.city_id = l.city_id
WHERE l.city = 'New York';
```

- Calculate the average of the unit price.

```
SELECT AVG(price) AS average_price
FROM Product;
```

- Extract all current date

```
SELECT *
FROM Orders
WHERE invoice_date IS NULL;
```

- Extract all payments with credit cards

```
SELECT * FROM customer_payment
WHERE payment_method = 'Credit Card';
```

- Count how many employees handle male clothing order

```
SELECT COUNT(*) AS male_clothing_employees FROM
Employee e
JOIN Job_title jt ON e.job_title_id = jt.job_title_id
WHERE jt.order_type = 'Male Clothing';
```

- Count the employees with the first name starting with the letter S (S capital)

```
SELECT COUNT(*) AS employees_with_s
FROM Employee
WHERE first_name LIKE 'S%';
```

SQL statement for queries:

1. Extract all customers from a specific city

This query retrieves customer details by joining the Customer and Location tables on city_id, filtering for customers in 'London'. The JOIN ensures location data is included.

2. Count customers from a specific city

It counts the number of customers from 'New York' using COUNT(*), leveraging an inner join between Customer and Location on city_id.

3. Calculate the average unit price

This query uses AVG(price) on the Product table to determine the average price of all products, summarizing numeric data.

4. Extract all orders with a missing invoice date

It filters the Orders table where invoice_date IS NULL, retrieving orders that haven't been invoiced.

5. Extract all payments made with credit cards

A simple SELECT * FROM customer_payment WHERE payment_method = 'Credit Card' filters all credit card transactions.

6. Count employees handling male clothing orders

This query joins Employee and Job_title, counting employees handling 'Male Clothing' orders.

7. Count employees with first names starting with 'S'

Using LIKE 'S%', this query counts employees whose first name starts with 'S'.

Gartner Hype Cycle Overview (Gartner, 2024)

Represents the maturity and adoption of merging technologies.

Five Key phases:

1. Innovation Trigger – Early breakthrough, limited viability.
2. Peak of Inflated Expectations – Hype leads to unrealistic expectations.
3. Trough of Disillusionment – Adoption slows due to challenges.
4. Slope of Enlightenment – Practical applications emerge.
5. Plateau of Productivity – Widespread adoption and value realization.

EVOLUTION OF DATA PROCESSING TECHNOLOGIES (GARTNER, 2023; IBM, 2024)

DATA FABRIC – IN TROUGH OF DISILLUSIONMENT, BUT EXPECTED TO GROW.

AI-POWERED DATA MANAGEMENT – ADVANCING TOWARD MAINSTREAM ADOPTION.

DATA MESH – DECENTRALIZED APPROACH PROMOTING DOMAIN OWNERSHIP.

GRAPH DATABASES & AUGMENTED DATA CATALOGS – METADATA – DRIVEN AUTOMATION GAINING TRACTION.

EDGE COMPUTING & HYBRID CLOUD – ENHANCING REAL-TIME DATA PROCESSING.

Key Takeaway:

The future of data processing is AI-driven, automated, and decentralize, aligning with business agility and innovation (Gartner, 2024).

Gartner Hype Cycle Overview

The Gartner Hype Cycle is a methodology that represents the maturity, adoption, and social perception of emerging technologies. It follows five key phases (Gartner, 2024):

1. Innovation Trigger – A breakthrough generates early interest but lacks proven viability.
2. Peak of Inflated Expectations – Hype and unrealistic expectations drive enthusiasm, often leading to failures.
3. Trough of Disillusionment – Interest wanes as early adopters face challenges in implementation.

4. Slope of Enlightenment – Organizations refine best practices, and real-world applications gain traction.
5. Plateau of Productivity – The technology reaches mainstream adoption with clear value realization.

Evolution of Data Processing Technologies According to Gartner

Gartner's predictions highlight the continuous evolution of data processing technologies, emphasizing automation, AI-driven analytics, and decentralized data architectures (Gartner, 2024).

- Data Fabric, positioned in the Trough of Disillusionment, is expected to gain traction as organizations overcome integration complexities (Gartner, 2023).
- AI-powered Data Management tools are advancing toward mainstream adoption, enhancing real-time analytics and decision-making (IBM, 2024).
- Data Mesh is emerging as an alternative to centralized data platforms, promoting domain-oriented data ownership (Gartner, 2024).
- Graph Databases and Augmented Data Catalogs are shifting towards greater adoption as metadata-driven automation becomes essential for enterprise data strategies (Gartner, 2023).
- Edge Computing and Hybrid Cloud Architectures continue to gain momentum, optimizing real-time data processing across distributed environments (IBM, 2024).

These advancements align with the broader industry shift toward scalable, automated,

and AI-driven data ecosystems, driving business agility and innovation (Gartner, 2024).

Data Fabric & Event Stream Processing (ESP) in the Gartner Hype Cycle

⌚ Gartner Hype Cycle Positions

- **Data Fabric:** Trough of Disillusionment – Adoption challenges but high potential (Gartner, 2024).
- **ESP:** Peak of Inflated Expectations – Rising adoption for real-time analytics (AWS, 2024).

📌 Key Features & Benefits

✓ Data Fabric (IBM, 2024; Pollard, 2024)

- Breaks data silos across RDBMS, lakes & clouds.
- AI-driven insights for automation & decision-making.
- Scales efficiently with integrated governance.

✓ ESP (AWS, 2024; Lawton, 2021)

- Processes real-time streams for instant insights.
- Scalable with Kafka & AWS Kinesis.
- Enhances user experience with fast decision-making.

📘 Practical Examples

- **Data Fabric:** Retailers track supplier risks & logistics in real time (Pollard, 2024).
- **ESP:** Zillow's Zestimate uses Apache Spark for real-time property valuations (AWS, 2024).

🚀 Future Implications

- **Data Fabric:** AI & self-service analytics will boost scalability & governance (IBM, 2024).
- **ESP:** AI/ML will enhance anomaly detection & automation (AWS, 2024).

Data Fabric:

Definition & Gartner Hype Cycle Position:

Data Fabric is an innovative data management design that simplifies access to dispersed data through flexible, reusable, and automated integration. It leverages metadata, machine learning, and governance to unify structured and unstructured data, enabling seamless connectivity across relational databases, data lakes, and cloud storage. Positioned in the Trough of Disillusionment in the Gartner Hype Cycle for Data Management 2023, Data Fabric faces adoption challenges but is recognized for its transformative potential (Gartner, 2024; IBM, 2024).

Features & Benefits

As illustrated in Gartner's analysis, Data Fabric connects multiple storage environments to:

- Break data silos by unifying systems such as RDBMS and cloud warehouses.

- Provide AI-powered insights and recommendations for decision-making.
- Scale efficiently by integrating diverse data types.
Organizations benefit from enhanced data governance, automation, and faster insights (Pollard, 2024).

Practical Example

For example, a retailer using Data Fabric can integrate sales, logistics, and customer data to track supplier risks and production delays in real time, enabling efficient decision-making and improved customer experiences (Pollard, 2024).

IBM Cloud Pak for Data

IBM Cloud Pak for Data exemplifies Data Fabric's potential. This unified AI and data platform integrates siloed on-premises and cloud data without moving it. It automates data discovery, enforces governance policies, and ensures data privacy while supporting AI model development. By simplifying data access and enabling faster ROI, IBM Cloud Pak empowers organizations to transform their data management strategies and deliver trusted insights (IBM, 2024).

Future Implications

As Data Fabric evolves, its integration with AI and self-service analytics will enhance scalability, agility, and enterprise-wide governance, solidifying its role as a cornerstone of modern data strategies (IBM, 2024)

Event Stream Processing

Definition & Gartner Hype Cycle Position

Event Stream Processing (ESP) involves ingesting and processing a continuous stream of data from multiple sources in real time. It enables organizations to analyze data on the

fly and make immediate decisions, addressing use cases like fraud detection, customer journey mapping, and predictive analytics. In the Gartner Hype Cycle for Data Management 2023, ESP is positioned at the Peak of Inflated Expectations, as businesses embrace its potential for real-time analytics despite the challenges of widespread implementation (Amazon Web Services, 2024; Lawton, 2021).

Features & Benefits

ESP simplifies real-time data management with components like stream ingestion, storage, and processing. Key benefits include:

- **Real-Time Insights:** Analyze logs, clickstreams, and IoT data to make proactive decisions.
- **Scalable Architecture:** Solutions like Apache Kafka and Amazon Kinesis ensure scalability and reliability.
- **Enhanced User Experience:** Applications benefit from faster decision-making and improved personalization.

Practical Example

Zillow leverages Amazon Kinesis and Apache Spark to deliver real-time property valuations through their Zestimate tool. By integrating event streams from diverse data sources, Zillow calculates updated valuations in seconds, enhancing accuracy and customer trust. This use case highlights ESP's role in transforming industries through actionable insights (AWS, 2024).

Future Implications

As ESP evolves, its integration with AI/ML will unlock advanced capabilities such as anomaly detection, trend forecasting, and automation. With its ability to process vast data volumes in real time, ESP will remain central to modern data strategies (Amazon Web Services, 2024).

POWER BI FOR DATA ANALYSIS & DECISION- MAKING

WHY POWER BI?

• POWER BI ENABLES **REAL-TIME DATA ANALYSIS**, AI-DRIVEN INSIGHTS, AND INTERACTIVE **DASHBOARDS**, SUPPORTING DATA-DRIVEN DECISION-MAKING (MICROSOFT, 2023; GARTNER, 2023).

SUITABILITY FOR THE USE CASE

• **REAL-TIME PROCESSING** – CONNECTS TO SQL, AZURE, DATA LAKES (MICROSOFT, 2023). AI & PREDICTIVE ANALYTICS – USES **MACHINE LEARNING** FOR TREND FORECASTING (MCKINSEY, 2023).

• CUSTOM DASHBOARDS & REPORTS – **VISUALIZES KEY TRENDS** FOR BETTER DECISIONS (GARTNER, 2023).

• **SEAMLESS INTEGRATION** – WORKS WITH EXCEL, APIs, AND CLOUD PLATFORMS (FORRESTER, 2023).

JUSTIFICATION & CAPABILITIES

• **SCALABILITY & AUTOMATION** – MANAGES STRUCTURED/UNSTRUCTURED DATA (MICROSOFT, 2023).

• AI-POWERED INSIGHTS – ENHANCES **FORECASTING & RISK MITIGATION** (MCKINSEY, 2023).

• **USER-FRIENDLY** – NO-CODE & LOW-CODE OPTIONS FOR ACCESSIBILITY (FORRESTER, 2023).

◆ POWER BI DRIVES SMARTER DECISIONS, OPERATIONAL EFFICIENCY & STRATEGIC GROWTH (GARTNER, 2023).

POWER BI FOR DATA ANALYSIS & DECISION-MAKING

Why Power BI?

Power BI is a leading business intelligence tool that enables organizations to **analyse, visualize, and share insights** from large datasets in real time (Microsoft, 2023). It supports **interactive dashboards, AI-driven analytics, and seamless data integration**, making it ideal for data-driven decision-making (Gartner, 2023).

Suitability for the Use Case

For analysing online retail sales, Power BI offers:

Real-Time Data Processing – Connects to multiple sources, including **SQL, Azure, and Data Lakes** (Microsoft, 2023).

AI & Advanced Analytics – Uses **machine learning models** to predict trends and automate insights (McKinsey, 2023).

Custom Dashboards & Reports – Interactive **visualizations** to help decision-makers interpret key trends (Gartner, 2023).

Seamless Integration – Works with **Excel, Google Analytics, and APIs** to unify data (Forrester, 2023).

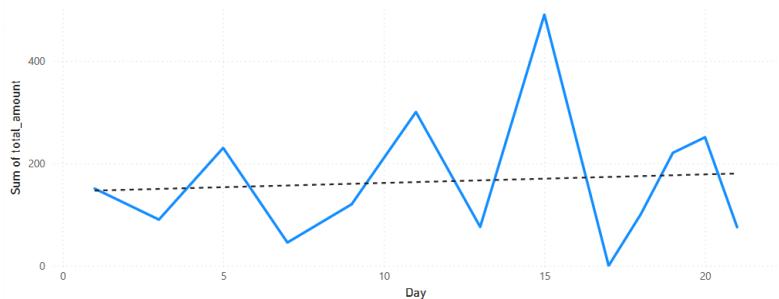
Justification & Capabilities

- **Scalability & Automation** – Handles structured/unstructured data across **on-premise and cloud platforms** (Microsoft, 2023).
- **Enhanced Decision-Making** – AI-powered insights improve **forecasting and risk mitigation** (McKinsey, 2023).
- **User-Friendly Interface** – No-code and low-code capabilities support a **wide range of users** (Forrester, 2023).

By leveraging Power BI, businesses can **uncover actionable insights, optimize operations, and enhance strategic planning**, making it a powerful tool for modern data analysis (Gartner, 2023)

Sales Analysis

Total sales per day



Key KPIs

Total Sales

2.15K

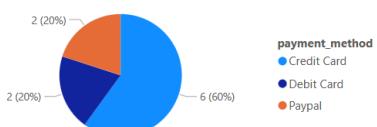
Total Orders

14

Average Order Value

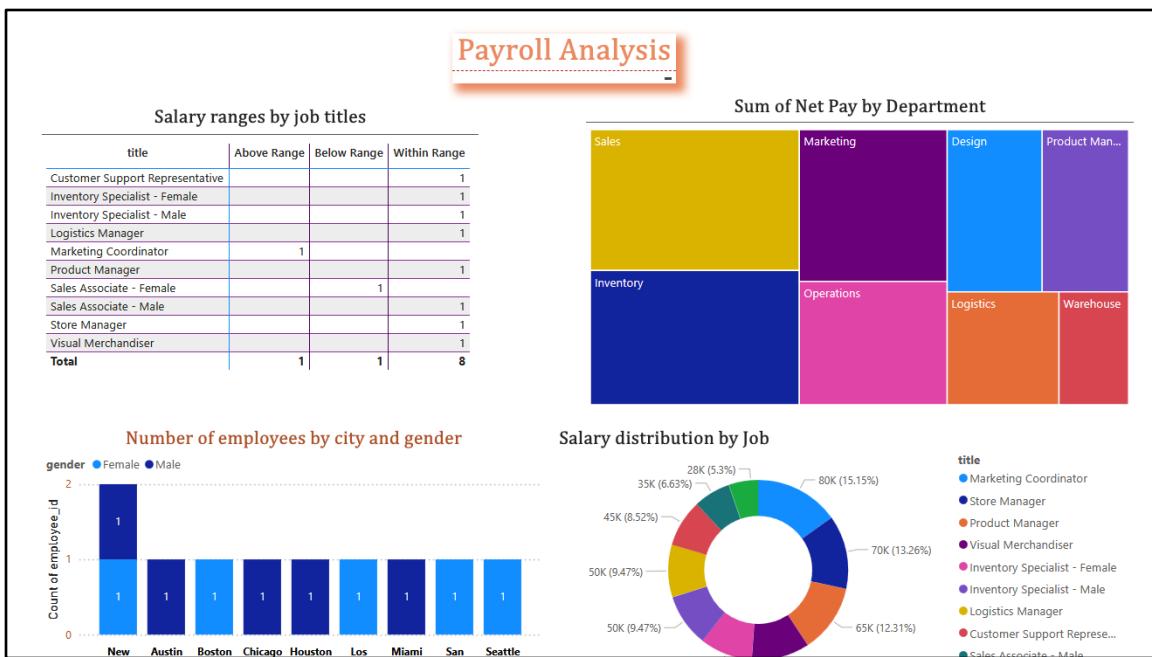
153.46

Payment Method Analysis



Customers distribution by city





Role of Data Visualization in Enhancing Reports and Deriving Insights

Data visualization simplifies complex data, making reports more engaging and insightful. It helps identify trends, patterns, and anomalies, enabling data-driven decision-making. In business reporting, clear visual representation of key metrics ensures stakeholders can quickly grasp performance indicators and take action.

Use Case Context: Sales Performance Analysis

A company tracking sales transactions, customer demographics, product categories, and sales regions aims to optimize stock levels and refine marketing strategies. Sales trends, peak purchasing times, and regional performance help adjust inventory and promotions effectively.

Example Reports

1. **Sales Analysis Report:** Highlights KPIs, daily sales trends, preferred payment methods, and customer distribution by city.

2. **Payroll Analysis Report:** Provides salary insights, total net pay per department, employee distribution by city and gender, and salary ranges per job title.

References

Books and Journals

- Carpenter, B. & Fowler, M., 2021. *NoSQL Distilled: A Brief Guide to the Emerging World of Polyglot Persistence*. Addison-Wesley.
- Chen, S., 2021. Analysis of customization strategy for e-commerce operation based on big data. *Wireless Communications and Mobile Computing*, Volume 2021, p. 11.
- Connolly, T. & Begg, C., 2015. *Database Systems: A Practical Approach to Design, Implementation, and Management*. Pearson, p. 52.
- Elmasri, R. & Navathe, S., 2020. *Fundamentals of Database Systems*. Pearson.
- Ghose, A. & Sundararajan, A., 2006. Evaluating pricing strategy using e-commerce data: evidence and estimation challenges. *Institute of Mathematical Sciences*, 21(2), pp. 131-142.
- Kumar, R. & Banyal, R. K., 2020. Data lifecycle management in big data analytics. *Procedia Computer Science*, Volume 173, pp. 364-371.

Online Sources

- Amazon Web Services, 2024. AWS. Available at: <https://aws.amazon.com/what-is/real-time-data-streaming/> [Accessed 28 Jan. 2025].
- AWS, 2024. AWS Case Studies: Zillow Zestimate. Available at: <https://aws.amazon.com/solutions/case-studies/zillow-zestimate/> [Accessed 28 Jan. 2025].
- Chris, K., 2022. Database Normalization (1NF, 2NF, 3NF) Table Examples. FreeCodeCamp. Available at: <https://www.freecodecamp.org/news/database-normalization-1nf-2nf-3nf-table-examples/> [Accessed 30 Jan. 2025].
- Forrester, 2023. The Forrester Wave: Augmented BI Platforms, Q2 2023. Available at: <https://www.forrester.com> [Accessed 21 Jan. 2025].
- Gartner, 2023. Magic Quadrant for Analytics and Business Intelligence Platforms. Available at: <https://www.gartner.com> [Accessed 21 Jan. 2025].
- Gartner, 2024. Data Fabric Overview. Available at: <https://www.gartner.com/en/data-analytics/topics/data-fabric> [Accessed 28 Jan. 2025].
- IBM, 2024. What is Data Fabric?. Available at: <https://www.ibm.com/think/topics/data-fabric> [Accessed 28 Jan. 2025].
- IBM, 2024. IBM Data Fabric Solutions. Available at: https://www.ibm.com/data-fabric?p1=Search&p4=43700081258725525&p5=p&p9=58700008825615518&gad_source=1&gclsrc=ds [Accessed 28 Jan. 2025].

- Lawton, G., 2021. *Understanding Stream Processing*. TechTarget. Available at: <https://www.techtarget.com/searchdatamanagement/definition/stream-processing> [Accessed 28 Jan. 2025].
- McKinsey & Company, 2023. *The Future of AI-Driven Business Intelligence*. Available at: <https://www.mckinsey.com> [Accessed 22 Jan. 2025].
- Microsoft, 2023. *Power BI: Business Intelligence for All*. Available at: <https://powerbi.microsoft.com> [Accessed 22 Jan. 2025].
- Pollard, J., 2024. *Data Fabric in Retail: The Go-To Solution to Boost Customer Experience and Personalization*. Retail Technology Review. Available at: <https://www.retailtechnologyreview.com/articles/2024/07/22/data-fabric-in-retail-the-go-to-solution-to-boost-customer-experience-and-personalization/> [Accessed 28 Jan. 2025].