A. Requirements of a Cloud Data Solution

1. Discuss the functional requirements that must be considered when deciding on a cloud data storage solution for the provided scenario and why these requirements are important.

1. Scalability
    a. Requirement: The solution must support rapid scaling to accommodate daily data generation of approximately 10 TB.
    b. Importance: Alliah Company is experiencing growth, and the system must handle increasing transaction volumes without performance degradation.
2. High Availability and Reliability
    a. Requirement: The storage solution should ensure minimal downtime and high fault tolerance.
    b. Importance: Continuous uptime is crucial for real-time transaction processing and analytics in e-commerce.
3. Real-Time Data Processing and Analytics
    a. Requirement: The system must allow for real-time data ingestion and processing.
    b. Importance: Business decisions regarding marketing, inventory, and customer interactions rely on real-time data insights.
4. Data Security and Compliance
    a. Requirement: The storage solution should include encryption, access control, and compliance with GDPR and other legal frameworks. (*General Data Protection Regulation (GDPR) – Legal Text*, 2024)
    b. Importance: Ensuring customer data privacy and securing sensitive transaction records is critical to maintaining trust and avoiding legal penalties.
5. Integration with Business Intelligence and Analytics Tools
    a. Requirement: The storage system must integrate seamlessly with BI tools and data visualization platforms.
    b. Importance: Stakeholders need easy access to analytics for decision-making.
6. Efficient Backup and Disaster Recovery
    a. Requirement: Automated backup solutions and failover strategies should be in place.
    b. Importance: Business continuity and data recovery are essential to prevent data loss due to failures or cyber threats.
7. Multi-Source Data Integration
    a. Requirement: The cloud solution should support data ingestion from multiple sources (e-commerce platform, customer service, vendors, etc.)
    b. Importance: Ensures a unified data pipeline for analysis and reporting.
8. Structured and Semi-Structured Data Storage
    a. Requirement: Must support both relational (transaction data) and non-relational (JSON-based shopping cart details) data.
    b. Importance: The hybrid nature of transaction data requires flexible storage formats for efficient querying and processing.

2. Discuss the non-functional requirements that must be considered when deciding on a cloud data storage solution for the provided scenario and why these requirements are important.

1. Performance
    a. Requirement: The system should provide low-latency access to data and support high-throughput transaction processing.
    b. Importance: With approximately 10 TB of transaction data generated daily, fast read/write speeds ensure smooth customer experiences and efficient analytics.
2. Availability
    a. Requirement: The cloud solution must have a high availability SLA (99.9% or higher).
    b. Importance: Downtime can lead to lost revenue and poor customer experience, making high availability essential for continuous operations.
3. Reliability
    a. Requirement: The system must ensure consistent and accurate data storage with redundancy and failover mechanisms.
    b. Importance: Data corruption or loss could impact critical business functions such as financial reporting and inventory management.
4. Security
    a. Requirement: Robust security measures, including encryption, role-based access control (RBAC), and multi-factor authentication (MFA), must be implemented.
    b. Importance: Protecting sensitive customer and financial data is crucial for regulatory compliance and brand trust.
5. Compliance & Regulatory Requirements
    a. Requirement: The solution must comply with GDPR, PCI-DSS, and other applicable data privacy regulations. (*PCI DSS Compliance Software*, n.d.)
    b. Importance: Legal non-compliance can result in hefty fines and reputational damage.
6. Scalability
    a. Requirement: The system must support vertical and horizontal scaling to handle increased data volumes.
    b. Importance: Alliah Company is rapidly growing, requiring a scalable solution to prevent performance bottlenecks.
7. Maintainability
    a. Requirement: The storage solution should be easy to monitor, update, and troubleshoot.
    b. Importance: A well-maintained system reduces downtime, optimizes performance, and minimizes operational cost.
8. Cost Efficiency
    a. Requirement: The system should provide a balance between performance and cost, with options like tiered storage and pay-as-you-go pricing.
    b. Importance: Managing cloud costs effectively ensures profitability without unnecessary expenses.
9. Disaster Recovery & Business Continuity
    a. Requirement: The cloud provider should offer automated backups, geo-redundancy, and a disaster recovery plan.
    b. Importance: Ensures that Alliah Company can recover quickly from outages or cyberattacks.
10. Interoperability & Integration
    a. Requirement: The storage system should integrate seamlessly with existing BI Tools, CRM, and third-party vendors.
    b. Importance: Smooth data flow between platforms enhances operational efficiency.

B. Cloud Data Solution

1. Propose a cloud data storage solution that addresses the provided scenario.

a. Identify the preferred cloud vendor and explain why this vendor is preferred.

b. Describe the major features (e.g., class, size) of your proposed solution.

Preferred Cloud Vendor: Amazon Web Services (AWS)
AWS is the preferred cloud provider for Alliah Company due to its robust storage solutions, scalability, security features, and seamless integration with business intelligence tools. Key reasons include:

a. Scalability: AWS provides auto-scaling features that can handle Alliah Company's 10 TB of daily transaction data.
b. High Availability & Reliability: AWS offers multi-region replication and a 99.99% uptime SLA.
c. Security & Compliance: AWS complies with GDPR, PCI-DSS, and SOC 2, ensuring customer data protection.
d. Cost Efficiency: Tiered storage pricing and pay-as-you-go models help manage costs effectively.
e. Integration Capabilities: AWS integrates with Amazon Redshift, QuickSight, and third-party analytics platforms for real-time insights.

Major features of the proposed solution:

1. Primary Storage: Amazon S3 (Simple Storage Service) (*Amazon Simple Storage Service Documentation*, n.d.)
    a. Class: S3 Standard (for frequently accessed transaction data) & S3 Glacier (for archival data).
    b. Size: Scales dynamically based on daily data growth.
    c. Features
        i. High durability (99.99%) availability
        ii. Supports encryption at rest and in transit
        iii. Lifecycle policies to transition older data to Glacier for cost savings.
2. Database storage: Amazon RDS (Relational Database Service) - PostgreSQL
    a. Purpose**:** Stores structured transaction data for business operations.
    b. Features
        i. **Multi-AZ deployment** for failover protection.
        ii. **Read replicas** for load balancing high-traffic queries.
        iii. **Automated backups** with point-in-time recovery.
3. Data Warehouse: Amazon Redshift
    a. Purpose: Processes and analyzes large transaction datasets.
    b. Features:
        i. Columnar storage for high-performance analytics
        ii. Integration with AWS Glue for ETL operations
        iii. Supports complex SQL queries for business intelligence.
4. Real-Time Data Processing: AWS Kinesis
    a. Purpose: Ingests and processes transaction data in real time.
    b. Features
        i. Handles streaming data for instant analytics
        ii. Scalable to support high transaction volumes.
5. Security & Compliance: AWS IAM & KMS

    a. **AWS IAM (Identity and Access Management):** Implements role-based access control (RBAC).
    b. **AWS KMS (Key Management Service):** Provides encryption for sensitive transaction data.
6. Disaster Recovery & Backup: AWS Backup & S3 Cross-Region Replication
    a. **Automated backups** ensure data integrity.
    b. **Geo-redundancy** protects against regional failures.

The AWS-based cloud storage solution ensures scalability, security, and real-time processing while keeping costs manageable. With Amazon S3, RDS, Redshift, and Kinesis, Alliah Company can efficiently store, manage, and analyze its growing transaction data.

2. Justify your proposed solution based on the functional and non-functional requirements you described in A1 and A2.

1. Functional Requirements Alignment

Scalability	**Amazon S3 & Redshift** provide auto-scaling to handle **10 TB of daily transaction data**. S3 lifecycle policies optimize storage costs.
High Availability & Reliability	AWS ensures **99.99% uptime** with **multi-AZ deployment** in **Amazon RDS** and **cross-region replication** in **S3**.
Real-Time Data Processing	**AWS Kinesis** enables real-time data ingestion, ensuring timely analytics for **marketing, inventory, and CRM**.
Security & Compliance	AWS complies with **GDPR, PCI-DSS**, and **SOC 2**. **IAM & KMS** manage **role-based access control** and encryption.
Integration with BI & Analytics	**Amazon Redshift + AWS Glue** allow for seamless integration with **business intelligence (BI) tools**.
Efficient Backup & Disaster Recovery	**AWS Backup & S3 Cross-Region Replication** ensure automated backups and rapid data recovery.
Multi-Source Data Integration	**AWS Glue** supports ETL pipelines from various sources (e-commerce, customer service, vendors).
Structured & Semi-Structured Data Storage	**Amazon RDS (PostgreSQL) stores structured data**, while **S3** handles semi-structured JSON-based shopping cart details.

2. Non-Functional Requirements Alignment

Performance	**Amazon RDS & Redshift** provide **fast read/write speeds**, optimizing e-commerce transaction processing.
Availability	AWS offers **multi-region replication** and **99.99% uptime**, preventing service disruptions.
Reliability	**Automated failover in RDS**, **multi-AZ deployment**, and **data durability** in S3 ensure reliable data storage.
Security	**IAM, KMS, encryption at rest & transit**, and **network security groups** protect sensitive customer data.
Compliance	AWS adheres to **GDPR, PCI-DSS**, and **data residency regulations**, reducing legal risks.
Scalability	AWS solutions **scale horizontally and vertically** to match increasing transaction volumes.
Maintainability	**AWS Management Console & CloudWatch** provide easy monitoring, maintenance, and automated updates.
Cost Efficiency	**Pay-as-you-go pricing**, **S3 tiered storage**, and **Redshift compression** optimize cloud costs.
Disaster Recovery & Business Continuity	**S3 Cross-Region Replication & AWS Backup** provide geo-redundancy and **point-in-time recovery**.
Interoperability & Integration	**AWS Glue & Lambda** facilitate integration with **BI tools, CRM, and third-party services**.

Conclusion

The AWS cloud storage solution is the best fit for Alliah Company because it ensures:
1. Scalability for growing data needs
2. Security & compliance for customer protection
3. High availability & disaster recovery to prevent downtime
4. Real-time processing & analytics for data-driven decision-making
5. Cost-efficient storage & performance optimization

This future-proof cloud strategy enables seamless expansion, operational efficiency, and competitive advantage for Alliah Company.

C. Strategies and Requirements

1. Describe *all* security and legal requirements that must be considered in migrating the data to the cloud.

Migrating Alliah Company's 10 TB of daily transaction data to the cloud involves strict security and legal compliance measures to protect customer information, financial records, and business transactions.

1. Security Requirements

Requirement	Justification	AWS Implementation
Data Encryption	Protects sensitive customer data (PII, financial transactions) from unauthorized access.	**AWS KMS (Key Management Service)** for encrypting data at **rest** and **in transit** (TLS 1.2).
Identity & Access Management (IAM)	Ensures that only authorized users can access specific data.	**AWS IAM with RBAC (Role-Based Access Control)** and **Multi-Factor Authentication (MFA)**.
Network Security	Prevents unauthorized external access.	**VPC (Virtual Private Cloud) & Security Groups** restrict access to authorized IPs.
Data Integrity & Versioning	Ensures data is not altered or corrupted.	**S3 Object Versioning & Checksums** maintain data integrity.
Monitoring & Logging	Detects and prevents unauthorized access attempts.	**AWS CloudTrail & AWS GuardDuty** for real-time security audits and threat detection.
Backup & Disaster Recovery	Ensures business continuity in case of failures.	**S3 Cross-Region Replication & AWS Backup** for redundant data storage.

2. Legal & Compliance Requirements

Requirement	Justification	AWS Compliance Support
General Data Protection Regulation (GDPR)	Protects EU customer data with privacy laws.	**AWS GDPR compliance features** (data residency, encryption, access control).
Payment Card Industry Data Security Standard (PCI-DSS)	Protects credit card transactions.	**AWS PCI-DSS certification** ensures secure payment processing.
California Consumer Privacy Act (CCPA)	Protects personal data of California residents.	**Data anonymization & access logs** ensure compliance.
Health Insurance Portability and Accountability Act (HIPAA)	Protects healthcare-related customer data if applicable.	**AWS HIPAA-compliant services** for secure storage of health-related transactions.
Data Residency Requirements	Ensures customer data stays in the designated region.	**AWS Regional Data Control** enables compliance with location-specific regulations.

2. Discuss the proposed backup strategies in the cloud. The following points must be explicitly addressed:

- **Which types of backups are appropriate?**
- **Which types of backups are most cost-effective?**
- **How often is data backed up?**
- **How does the frequency with which the data is backed up impact the proposed architecture?**
- **How does the frequency with which the data is backed up impact cost?**

1. **Types of Backups**

Backup Type	Description	Appropriate for
Full Backup	A complete copy of all data.	Periodic comprehensive backups (e.g., weekly/monthly).
Incremental Backup	Backs up only data changed since the last backup.	Daily transaction data to minimize storage costs.
Differential Backup	Backs up all changes since the last full backup.	Balances speed and data restoration.
Snapshot Backup	Captures system state at a given time.	Quick restores for critical transaction data.

Best Approach: Hybrid Strategy (Full + Incremental + Snapshots)

- Full backup (weekly/monthly)
- Incremental backup (hourly/daily)
- Snapshots (real-time for critical data)

2. Most Cost-Effective Backup Types

Backup Type	Cost Consideration
Incremental Backup	Cost-effective as it stores only changes, reducing storage usage.
S3 Intelligent-Tiering	Automatically moves older backups to lower-cost storage classes.
S3 Glacier	Cheapest long-term storage for archival backups.
Automated Lifecycle Policies	Moves data from S3 Standard → S3 Glacier based on retention rules.

3. Backup Frequency

Data Type	Backup Frequency	Backup Type
Transaction Data (active)	Hourly	Incremental
Customer & Product Data	Daily	Differential
Full Database	Weekly	Full Backup
Long-Term Archives	Monthly	Glacier Backup

4. Impact on Proposed Architecture

- Frequent incremental backups reduce performance impact by avoiding large-scale data transfers.
- Snapshots allow real-time restoration without disrupting services.
- Multi-region replication ensures disaster recovery, protecting against data center failures.
- Automated lifecycle management optimizes storage costs, transitioning old backups to Glacier.

5. Impact on Cost

Backup Frequency	Cost Impact
More Frequent Backups	Higher S3 Storage and compute costs for processing
Less frequent backups	Lower storage costs but longer recovery times
Hybrid approach (smart tiering)	Balances cost and performance using incremental backups and glacier storage archives

3. Discuss how failover and disaster recovery planning are accommodated in the proposed design.

Failover Planning

Failover ensures that services remain operational even if a component fails. For Alliah Company, the proposed AWS design includes automatic failover mechanisms across key infrastructure components:

- Amazon RDS (PostgreSQL) uses Multi-AZ Deployment, which means a standby database is maintained in a different availability zone. If the primary database fails, AWS automatically switches to the standby instance with minimal downtime.
- Amazon S3 enables Cross-Region Replication (CRR) to ensure that transaction data is available in another AWS region in case of failure.
- Amazon Redshift includes automated snapshots, allowing the system to recover from data corruption or a node failure.
- AWS Kinesis, used for real-time data processing, operates across multiple availability zones, ensuring that incoming data streams are not disrupted if a processing node fails.
- Auto Scaling and Elastic Load Balancing (ELB) are used for compute resources, ensuring that if a server or EC2 instance fails, new ones are provisioned automatically to handle traffic.

With these failover mechanisms in place, Alliah Company's critical data and services can continue running with minimal interruption in case of localized failures.

Disaster Recovery Planning

Disaster recovery ensures that Alliah Company can restore services and data quickly after a major outage, data corruption, or cyberattack. The proposed AWS design includes multiple disaster recovery strategies:

- For database failures, RDS provides automated failover, but if a failure affects both the primary and standby instances, point-in-time recovery (PITR) can restore data from automated backups.
- For accidental deletions or data corruption, S3 Object Versioning and RDS Snapshots allow restoration of previous versions of data.
- For storage failures, S3 Cross-Region Replication ensures that a second copy of data is always available in a geographically distant AWS region.
- For compute failures, Auto Scaling and Elastic Load Balancing ensure that new instances are provisioned automatically.
- For full AWS region outages, Route 53 DNS Failover redirects traffic to a backup region, minimizing service disruptions.

With these disaster recovery measures, Alliah Company can restore business operations within minutes to hours, depending on the severity of the failure.

Recovery Time and Data Loss Considerations

Failover and disaster recovery strategies must balance speed of recovery (Recovery Time Objective - RTO) and data loss tolerance (Recovery Point Objective - RPO).

- Mission-critical services, such as the transaction database, require fast recovery with near-zero data loss. RDS Multi-AZ failover ensures minimal downtime, while incremental backups provide up-to-date transaction records.
- Archived data can tolerate longer recovery times and is stored in cost-effective solutions like S3 Glacier, which can take hours to restore.

- For severe failures, such as region-wide outages, Route 53 DNS Failover and S3 Cross-Region Replication ensure that a backup environment is available within minutes to hours.

By layering these strategies, Alliah Company ensures that critical systems recover instantly, while non-essential data is restored in a cost-effective manner.

Cost Considerations

Frequent backups, multi-region replication, and automatic failover mechanisms increase cloud costs, but they ensure business continuity.

- More frequent backups and multi-region failover reduce downtime but require higher storage and compute expenses.
- S3 Glacier and lifecycle management help reduce long-term storage costs by archiving less frequently accessed backups.
- Using a combination of real-time failover for mission-critical systems and slower backups for non-essential data balances cost and recovery speed.

The proposed strategy optimizes cost by using a tiered approach, ensuring that the most critical data is always available, while archived data is stored cost-effectively.

Conclusion

The AWS-based failover and disaster recovery plan for Alliah Company ensures high availability, data integrity, and business continuity.

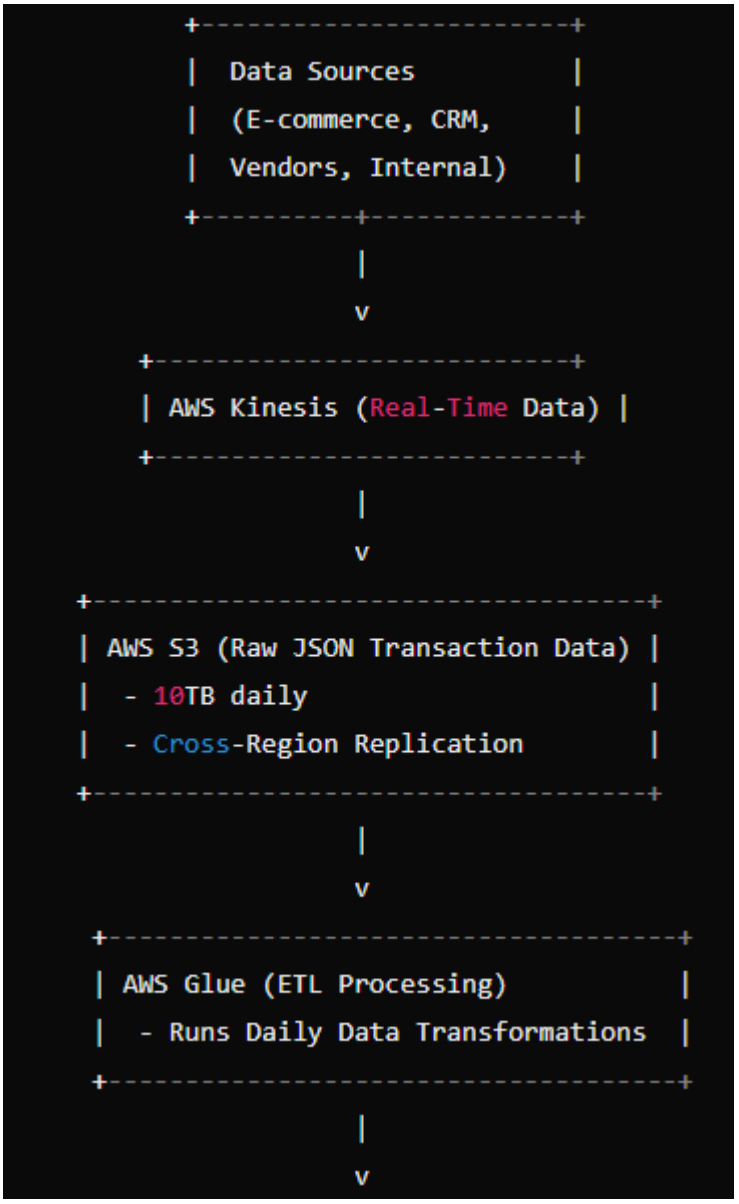
- Automatic failover mechanisms in RDS, S3, Redshift, and Kinesis keep services running with minimal disruption.
- Disaster recovery solutions such as backups, snapshots, and cross-region replication ensure that data is restored quickly after failures.
- A multi-tiered backup and recovery approach balances cost and performance, ensuring the company can continue operating smoothly in any failure scenario.

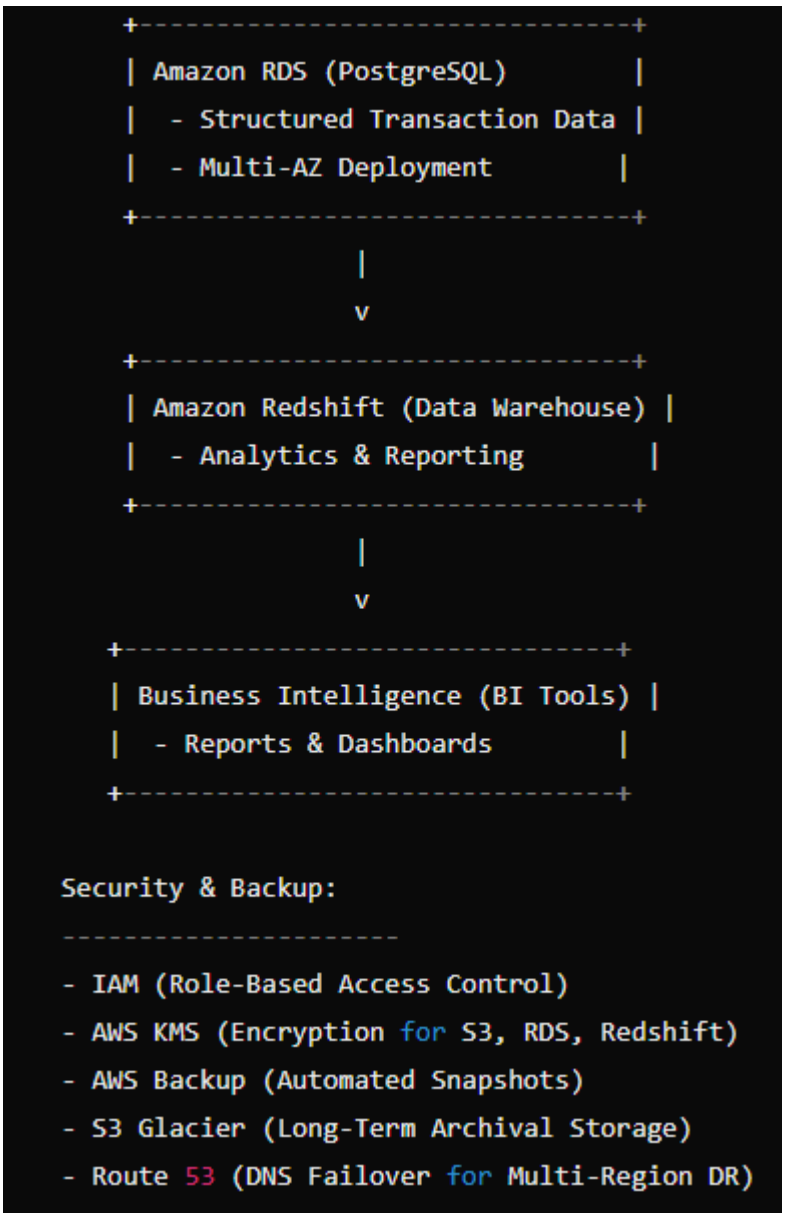
By implementing failover, automated backups, and cross-region replication, Alliah Company can protect its transaction data and recover swiftly from any disaster.

D. Architecture Diagram

1. Create an architecture diagram for your proposed cloud data storage solution that includes the following:

- **labels for data, including volume and type**
- **labels for target databases, including type and schema**
- **labeled processes, including cloud components used and frequency of operations**
- **labels for security groups**
- **labels for locations and availability zones**





2. Identify which database objects need to be included.
a. Create a logical data diagram to store the data in the proposed solution, including the objects identified in D2.

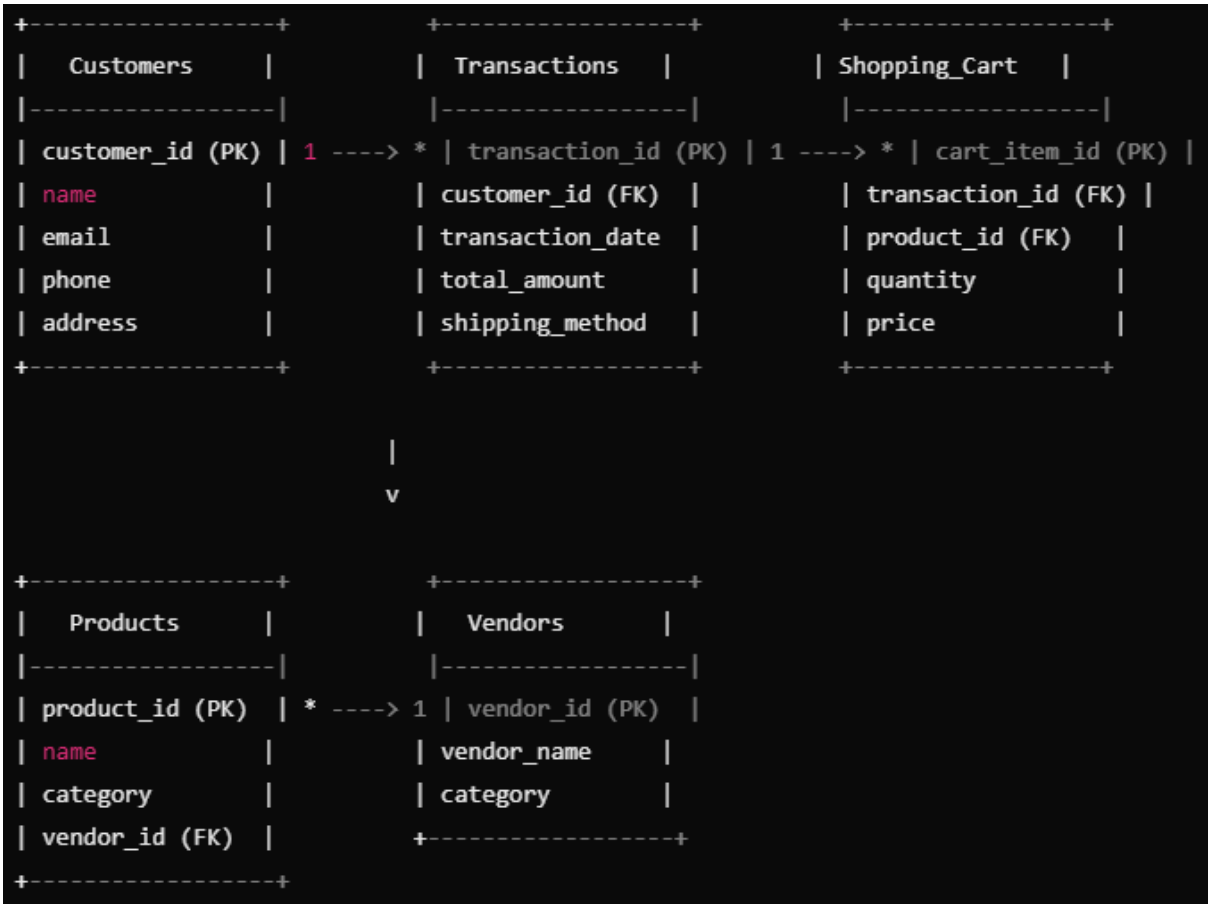
1. Database Objects to Include

Primary Tables:

- **Customers**: Stores customer information (ID, name, email, phone, address).
- **Transactions**: Stores purchase details (ID, date, total amount, shipping method, customer reference).
- **Shopping_Cart_Items**: Stores individual products purchased in a transaction (ID, transaction reference, product reference, quantity, price).
- **Products**: Stores product details (ID, name, category, vendor).
- **Vendors**: Stores vendor information (ID, name, category).

Supporting Objects:

- **Indexes** on frequently queried columns like transaction date, product category, and customer email for fast lookups.
- **Foreign Keys** for referential integrity between transactions, customers, and products.
- **Views** for simplified reporting on **total sales per customer, top-selling products, and revenue by vendor**.
- **Stored Procedures** for **automating order processing and generating reports**.



Explanation of Relationships:

- **Customers (1) → Transactions (Many):** A customer can have multiple transactions.
- **Transactions (1) → Shopping_Cart_Items (Many):** A transaction can include multiple products.
- **Products (1) → Shopping_Cart_Items (Many):** Each cart item references a product.
- **Vendors (1) → Products (Many):** A vendor supplies multiple products.

Primary and Foreign Keys:

- **Primary Keys (PK):** Unique identifiers for each table.
- **Foreign Keys (FK):** Establish relationships between tables.

3. Explain why *each* of the database objects identified in part D2 needs to be included.

1. Customers Table

Why is it needed?

- Stores customer information such as name, email, phone number, and address.
- Ensures that transactions can be linked to a specific customer.
- Supports customer service, order history tracking, and targeted marketing.

2. Transactions Table

Why is it needed?

- Stores details of each purchase, including transaction ID, customer reference, transaction date, total amount, and shipping method.
- Serves as the central link between customers and purchased products.
- Enables tracking of sales trends, revenue, and customer spending patterns.

3. Shopping_Cart_Items Table

Why is it needed?

- Stores each product purchased in a transaction, linking transactions to products.
- Allows for multiple items per transaction, tracking quantity and price per item.
- Ensures detailed purchase records for reporting and customer service.

4. Products Table

Why is it needed?

- Stores product details such as name, category, and vendor reference.
- Supports inventory management, product recommendations, and reporting on best-selling products.
- Links to Shopping_Cart_Items to associate each purchase with a product.

5. Vendors Table

Why is it needed?

- Stores vendor details such as vendor name and product category.
- Allows tracking of which suppliers provide specific products.
- Supports vendor performance analysis, contract management, and supply chain optimization.

Sources:

General Data Protection Regulation (GDPR) – legal text. (2024, April 22). General Data Protection Regulation (GDPR). https://gdpr-info.eu/

PCI DSS Compliance Software. (n.d.). https://try.netwrix.com/change-tracker-pci-dss?cID=701Qk000007op0IIAQ&nwm=1&sID=1225956491281265&aID=dat-2328422030664645:loc-190&creative_id=&placement_id=&location_id=111138&msclkid=b054bc601c0c17b22cb8dba66f9a3440&utm_source=bing&utm_medium=cpc&utm_campaign=NAM_US_English_Search_Desktop_Change%20Tracker_Generic_P&utm_term=https://try.netwrix.com/change-tracker-pci-dss&utm_content=Dynamic

Amazon Simple Storage Service documentation. (n.d.). Retrieved March 3, 2025, from https://docs.aws.amazon.com/s3/?icmpid=docs_homepage_featuredsvcs