# Analysis of Digikala

# Online Shopping System

**Hamed Soleimani**

**May 2024**

**Table of Contents**

# 1. Introduction to Data Management

In today's digital age, data has emerged as one of the most invaluable assets for organizations spanning diverse industries. Effective data management plays a pivotal role in enabling businesses to extract actionable insights, make well-informed decisions, and gain a competitive advantage in the market. This encompasses a comprehensive framework of practices, processes, and technologies aimed at organizing, storing, safeguarding, and analysing data throughout its lifecycle. From initial collection and secure storage to thorough analysis and strategic utilization, data management ensures that information is handled efficiently, securely, and in full compliance with regulatory requirements.

Data is universally acknowledged as the cornerstone of business success, with its efficient collection, utilization, and comprehension being integral not only to a company's prosperity but also to its survival within the fiercely competitive landscape of the 21st century. The paramount significance of data to businesses is underscored by leading scholars in the field. As Redman (2008) aptly articulates, all employees routinely interact with data multiple times daily. This highlights the imperative for companies to establish a cohesive and robust data management system, necessitating consistent and coordinated efforts from virtually every facet of the organisation.

In essence, data management involves the systematic handling of data to ensure its availability, integrity, and usability for stakeholders within an organization. It encompasses a wide range of activities, including data governance, data quality management, data integration, and data security. Through robust data management practices, organizations can unlock the full potential of their data assets, driving innovation, enhancing operational efficiency, and fostering growth.

# 2. Data Management Practices

Below are some data management practices:

## 2.1. Data Integration

Data integration is the process of combining and reforming data from multiple sources into user-friendly format for data analysis and decision-making purposes (Dell, 2021). Today organisations rely on data gathered from various sources such as internal data from customer relationship management (CRM) systems, sales and marketing data, operational data, or external date sources like social media data, public data sources and more. Often when using multiple sources, it leads to data silos and inconsistencies, data integration provides the solution to such challenges alike which permits better informed decision making, streamlined operations within organisations.

Since the 1970s, the industry standard approach to data integration has been extraction, transform and load (ETL) (Chang, 2022). In this type of data integration, the extracted data is transformed in separate staging area before it is loaded into the data storage. The alternative method of data integration to ETL and much more modern approach to data integration is extract, load, transform (ELT), here data can be loaded to the data storage before transformation. Chang (2022) argues that the rise of ELT as an improved alternative to ETL can be attributed to the emergence of cloud computing, furthermore, (Chang, 2022) asserts that in most cases ELT is the better option due to its exploitation of modern data storage systems. ELT offers more flexibility as it can manage both structured and unstructured data whereas ETL was originally intended for relational, structured data. Given the nature of ELT is cloud based software as a service it provides greater scalability and accessibility to broader range of users compared to ETL. Khambra et al, 2014, identifies three major issues in data integration with ETL these are schema integration, data redundancy and detection.

## 2.2. Data Interoperability

In general, interoperability is defined as the seamless exchange and integration of different systems, organisations, or individuals to achieve a common goal (Berryman, et al., 2013), additionally, Reif, et al. (2022), contends that the definition of interoperability largely depends on the context and often its definition can be a great indicator of the most pressing issues in that ecosystem. In the healthcare sector, the European governance initiative (eHGI, 2012) defined interoperability by three key characteristics, first is the legal interoperability; this pertains to laws, policies and other legally binding agreement required to exchange data between organisations or regions. Second is the semantic interoperability which is associated with ensuring accurate interpretation of exchanged information between systems, services, and users. Lastly is the technical interoperability which alludes the seamless exchange of data between multiple application with little to zero manual intervention.

Chojka (2020), argues that because most interoperability models are based on the levels of conceptual interoperability model (LCIM) which consists of 4 levels of interoperability maturity, the layer of semantic and syntactic interoperability appears in most interoperability frameworks across different sectors such healthcare, military, and defence, and more. Therefore, asserting that both issues should be considered when building any framework to achieve an effective data interoperability strategy.

## 2.3. Streamlined Data Management Process

The introduction and ongoing adoption of a streamlined data management system are paramount for any company striving for success in the modern era. This system is characterized by efficient and optimized data handling, encompassing storage, retrieval, and maintenance within the organization. It relies on standardized procedures, along with technologies and tools, to effectively manage data from creation to disposal. The aim is to minimize redundancy, enhance data quality, improve accessibility, and ensure compliance with government regulations and industry standards (Peddireddy, 2023).

However, despite the myriad advantages of a streamlined data management system, its implementation and maintenance present various challenges. Firstly, the quality of available data significantly impacts its effective utilization within the streamlined system, with poor-quality data leading to errors and misinformation. Legacy systems and outdated technologies pose another hurdle, as many organizations grapple with integrating old data into modern data management practices. Additionally, data silos within organizations impede seamless data integration and sharing, resulting in duplicated efforts, inconsistencies, and difficulties in data management and application across departments (Peddireddy, 2023).

OLAP (Online Analytical Processing) is one of the applications commonly associated with streamlined data management. OLAP technology enables users to analyse multidimensional data interactively from multiple perspectives (Plattner, 2009). It allows for complex analytical and ad-hoc queries to be performed swiftly, providing insights into various aspects of business operations.

OLAP plays a crucial role in streamlined data management by facilitating efficient and interactive data analysis. It enables organizations to quickly access and analyse large volumes of data, helping them make informed decisions and identify trends or patterns in their data (Plattner, 2009).

# 3. Digikala: A Historical Overview

Digikala is a prominent online shopping platform which was established in 2015. That has garnered significant attention for its diverse range of products and exceptional customer service. Founded with the mission to provide customers with convenient access to a wide array of items, Digikala has emerged as a leader in the e-commerce industry. With a seamless registration process, customers can effortlessly navigate through the platform, search for products, place orders, and complete purchases, all within a user-friendly interface. The company's commitment to delivering high-quality products at competitive prices has contributed to its exponential growth in sales and popularity among consumers.

# 4. Current System, Identified Problem and proposed solution

In the increasingly competitive landscape of e-commerce, Digikala has emerged as a dominant player, leveraging data-driven strategies to drive growth and enhance customer experience. However, like many successful companies, Digikala faces critical challenges in managing surges in demand for certain products, particularly during peak times or promotional events. These fluctuations often surpass the company's capacity to replenish inventory in a timely manner, posing risks of stock shortages and subsequent customer dissatisfaction. Such bottlenecks not only hinder Digikala's ability to fulfill orders promptly but also jeopardize its reputation for reliability and customer service excellence.

Recognizing the urgency of addressing this issue, Digikala is poised to implement advanced data management solutions, including an OLAP system, as part of its streamlined data processes. By automating data pipelines and enabling seamless collection, cleansing, and real-time analysis of information, Digikala aims to gain actionable insights into customer behaviours, preferences, and purchasing patterns. Through the utilization of real-time data analytics, Digikala will be equipped to promptly identify surges in demand for specific product categories and proactively adjust inventory levels. Moreover, the integration of predictive modelling and machine learning algorithms will enable Digikala to forecast demand accurately and optimize its supply chain operations.

The implementation of these streamlined data processes underscores Digikala's commitment to making swift, data-driven decisions to ensure optimal inventory management and enhance customer satisfaction.

## 4.1. Context Diagram (As is)

A context diagram, within the realm of systems analysis and design, is a visual representation that illustrates the external entities interacting with a system and the flow of data between them. It serves as a high-level overview, depicting the boundaries of the system and its relationships with external entities without delving into internal details. By presenting this holistic perspective, a context diagram aids stakeholders in understanding the system's scope, its interfaces with external entities, and the flow of information across these interfaces.

Below is the context diagram for the Digikala's online shopping system:

At the core of Digikala's operations is its online shopping system, serving as the central entity that facilitates transactions and interactions within its digital ecosystem. This system acts as the nexus through which customers, payment processors, and inventory systems interact. Customers engage with Digikala's online shopping system via multiple channels, including the mobile application and the website.

# Entities

Externally, the customer, payment processor, and inventory systems are considered as Digikala's key external entities. These entities interact with the online shopping system to fulfill various functions:

Customer: Customers initiate purchases, browse products, and complete transactions through the online shopping system.

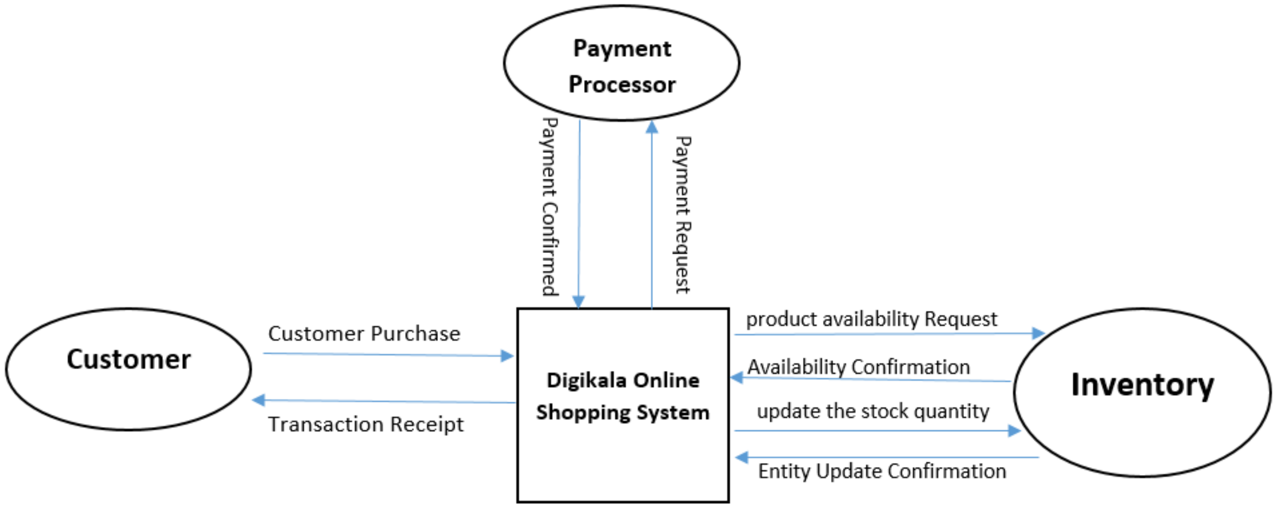


*Diagram1:* Context Diagram Digikala shopping system – As Is

Payment Processor Interaction: A third-party service that handles online payment transactions securely from the customer's bank account.

Inventory System Interaction: Inventory systems manage and track the availability of products in Digikala's inventory, ensuring that orders can be fulfilled promptly.

## Processes

Product browse and selection: Here the customer will browse the system to select their desired products. They begin by adding details typing names or description of the product.

Order Processing: This process begins when the customer selects products and adds it to cart to purchase. It captures the customer's order details, including the selected products, quantities, and shipping information. When customer initiates a purchase, the online shopping system processes the order request.

Payment authorisation: Upon receiving the order information from the order processing system, this process handles the payment transaction securely through the payment processor. It ensures that the payment details provided by the customer are valid and processes the payment. After successful payment, it informs the order processing system to proceed with order fulfilment. If the payment successfully processed, a payment confirmation is sent to the online shopping system.

Order fulfilment: Once the order has been through payment authorisation, the product is prepared ready for delivery to the customer. During this stage the customer will be notified on the progress of their order, ''pending', processing', 'shipped'.

Inventory Management: This process monitors the inventory levels of products in real-time. Concurrently, during the purchase process, the online shopping system communicates with the inventory system to check the availability of the purchased product.

If the product is in stock, the inventory system confirms its availability to fulfill the customer's order.

Upon confirmation of product availability and after the customer makes the payment, the online shopping system sends a request to the inventory system to update the stock quantity, indicating that the product has been sold. This ensured accurate stock management.

The inventory system responds by updating its records to reflect the reduced stock quantity, ensuring accurate inventory management.

## 4.2. DFD Level 0 (As is)

A Data Flow Diagram (DFD) is a graphical representation of the flow of data within a system or process. It shows how data moves from one process to another, how it is stored, and how it is processed. DFDs are used to model the data flow and transformation within a system, helping to understand the system's functionality and identify potential areas for improvement or optimization.

### Data Store

Customer Account: Stores information about customer orders, including order details (e.g., products, quantities, prices), customer details, and order status.

Product Catalogue: This database contains details of available products, including their unique identifiers or primary keys, names, product descriptions, prices, and quantities available in stock. This database is updated simultaneously as products are added, sold, or restocked.

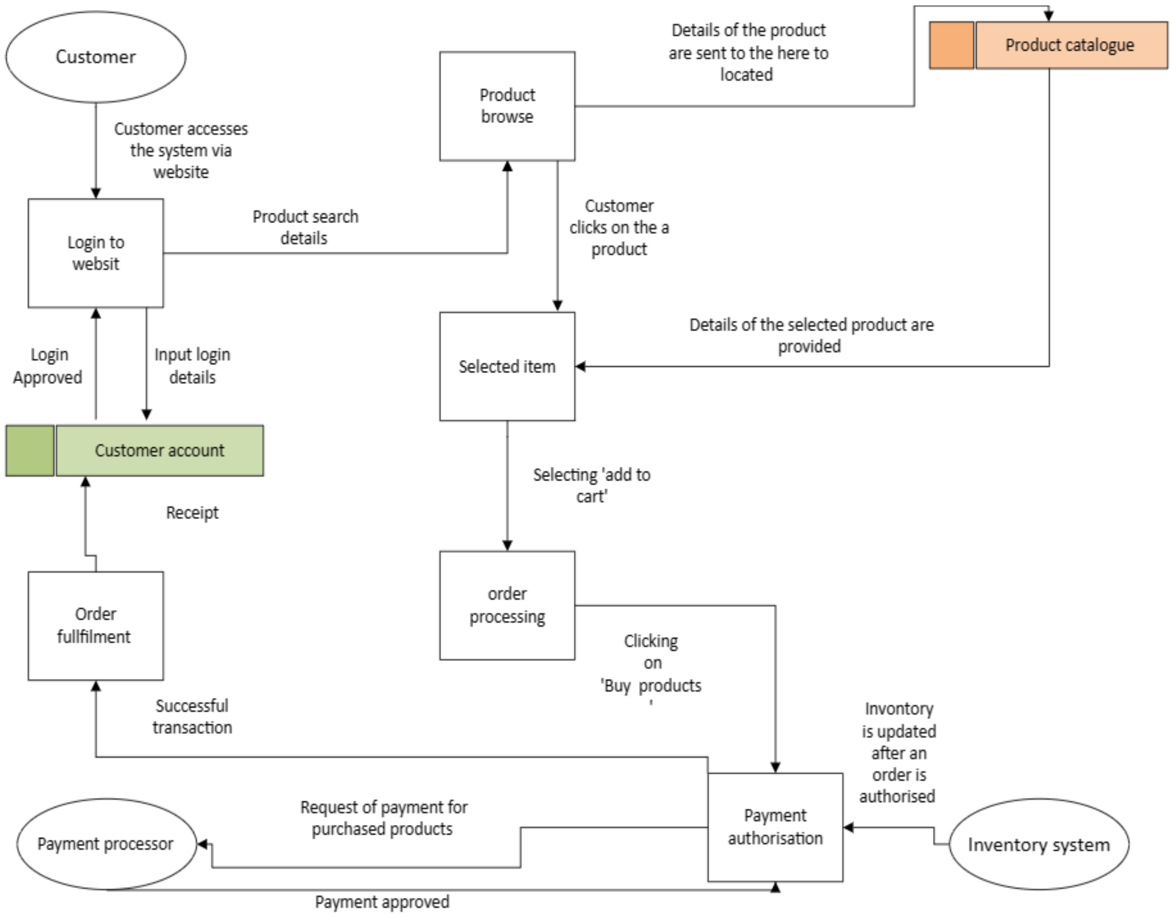


*Diagram2:* DFD Level 0 Digikala shopping system – As Is

## 4.3. Context Diagram (To be)

Digikala intends to implement an OLAP system as a crucial component of its streamlined data processes, aimed at promptly identifying surges in demand for specific product categories and proactively adjusting inventory levels. This involves the continuous transmission of information regarding searched products, sales products, and current inventory status to the OLAP system. Once received, the OLAP system processes this data, leveraging advanced analytics and real-time processing capabilities to generate actionable insights.

Subsequently, the OLAP system actively sends suggestions regarding restocking products to the inventory management system. These suggestions are based on the analysis of customer behaviors, sales trends, and inventory levels, enabling Digikala to optimize its inventory management practices in response to changing market demands.

By integrating the OLAP system into its data processes, Digikala aims to enhance its ability to anticipate and respond to fluctuations in customer demand effectively. This proactive approach to inventory management will not only improve operational efficiency but also contribute to maximizing customer satisfaction by ensuring the availability of popular products when needed.

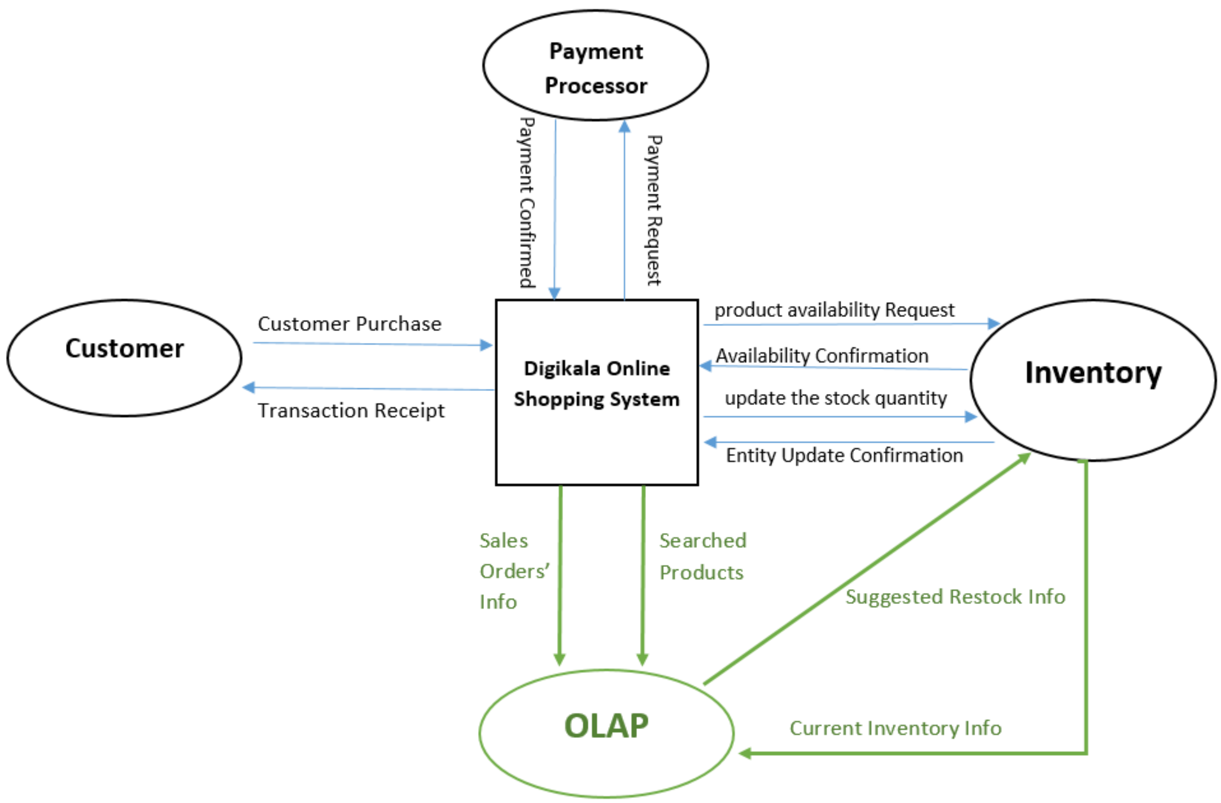


*Diagram3:* Context Diagram Digikala shopping system – To Be

## 4.4. DFD Level 0 (To Be)

In order to improve the Digikala system an online analytical processing (OLAP) system was added. OLAP is a technology that enables multi-dimensional and multi-level analysis on a large volume of data (Quieroz-Sousa & Salgado, 2020). The primary use of this tool in the Digikala system is an integrated management of inventory, the OLAP system would analyse the combined data sources using them to provide extremely accurate recommendations for the exact quantities and timings of product restocking required throughout the supply chain of the company.

This system provides several advantages to the company, the real time data reporting provided by the OLAP system allows for inventory managers to access up to date information and recommendations allowing them to make better and faster decisions, allowing higher levels of responsiveness in a dynamic and changing market. Secondly, OLAP systems are able to identify seasonal trends, changes in demand, and correlations between different product categories, enabling more accurate inventory planning and procurement strategies.

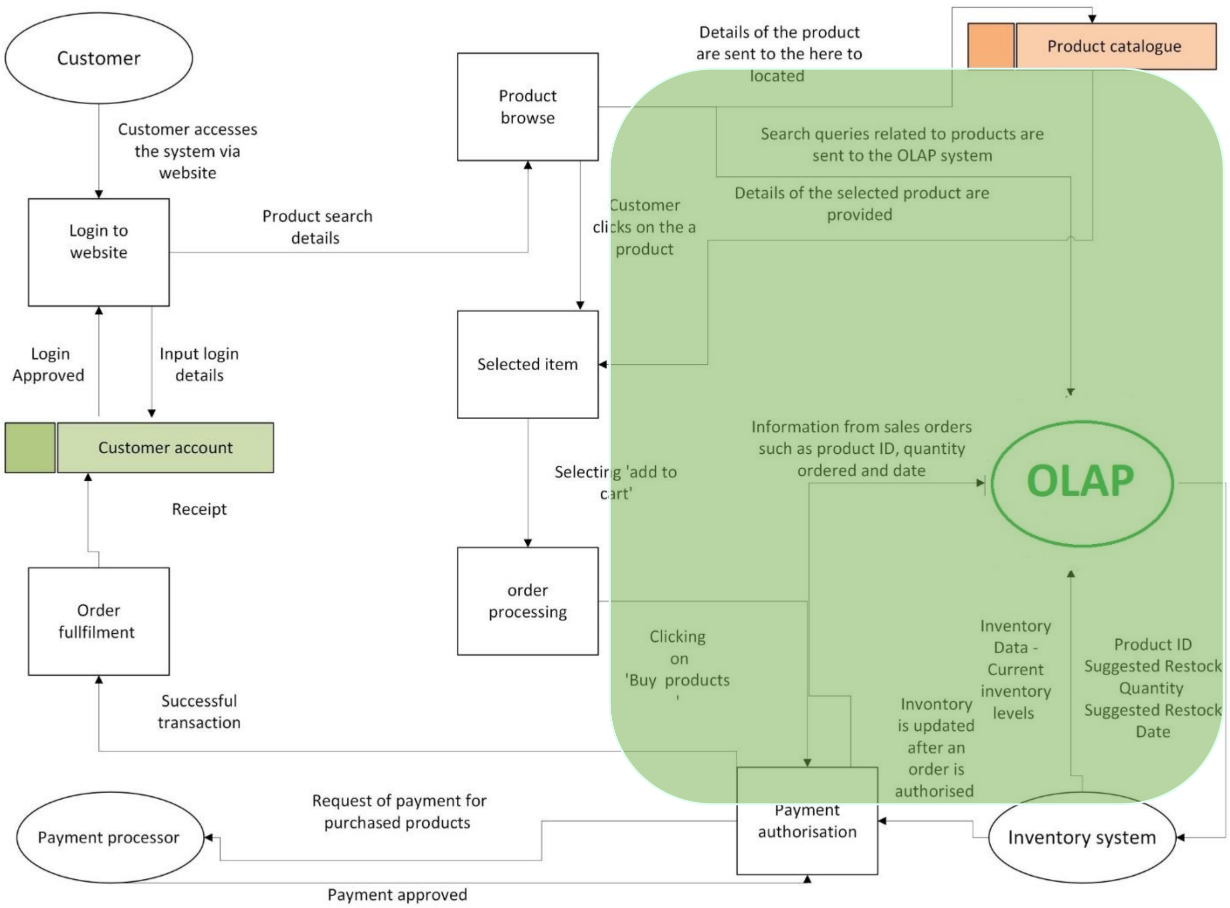


*Diagram4:* DFD Level 0 Digikala shopping system – To Be

## 5. Entity-Relationship Diagram (ERD)

An Entity-Relationship (ER) diagram serves as a visual representation of data, facilitating our understanding of the system and the connections among its components. The four components of the ER diagram are entity, attributes, relations, and cardinality. An entity is a distinct object or concept in the real world that can be represented in a database to store data about. An attribute is a characteristic or property of an entity in a database. A relation refers to the connection between two or more entities and cardinality indicates how many instances of one entity can be linked to instances of another entity. Cardinality can be:

One-to-One (1:1), One-to-Many (1: N) or Many-to-Many (M: N) (Dedhia, Jain and Deulkar, 2015). As the relationship between entities is bidirectional, cardinalities are defined in both ways. Because transferring the primary key of one entity to another would result in a multivalued attribute issue, representing a Many-to-Many relationship solely with two base tables is insufficient. Therefore, creating a new table to keep the relationship is inevitable. This intermediary entity converts a Many-to-Many relationship into two separate One-to-Many relationships (Li and Chen, 2009). According to Shoval, Danoch and Balabam (2004), the ER diagram is one of the most common data models in the database field.

Li and Chen (2009) indicated that with the normalisation method, data attributes are classified into categories to make sure entities are non-redundant, flexible, stable, and adaptable. Normalisation has multiple types:

1NF, 2NF, 3NF, 4NF, 5NF, BCNF, ONF, DKNF. 1NF, 2NF, and 3NF are the most common among them.

1NF: Eliminating the repeating groups and giving each a primary key.

2NF: Removing redundant data by extracting attributes that depend only on part of a composite key and placing them into a separate table.

3NF: Removing the columns that are not dependent on the primary key and placing them in a different table.

The discussion initially revolves around the entities and their attributes, followed by the assignment of cardinalities between these entities. Detailed Entity-Relationship Diagram (ERD) tables are provided, offering more granularity in comparison to the data stores in the Data Flow Diagram (DFD).

While the focus of the DFD is on presenting a conceptual model of how the business operates and how it can be enhanced through data management practices, the ERD delves deeper into the structure of the database. For instance, instead of simply representing a generic "product Catalogue" as a data store in the DFD, separate tables for "product" and "product category" are created in the ERD. This decision allows for a more detailed representation of the data and facilitates its use in queries and other database operations.

SalesOrder: It contains details of each sales order, including unique identifier (SalesOrderID), order date, shipping date, sales order number, and the corresponding customer who placed the order.

SalesOrderDetail: It contains detailed information about products in a sales order. Attributes include a unique identifier (SalesOrderDetailID), sales order ID, product ID, order quantity, unit price, and any discounts applied.

Product: It represents characteristics of a product with attributes such as unique identifier (ProductID), name, product number, color, cost, original price, product category, and current stock level.

Product Category: It describes different categories of products, identified by a unique ID (ProductCategoryID) and a name.

Customer entity: It contains information about customers, including unique identifier (CustomerID), title, first name, last name, address, city, postal code, email address, phone number, gender, marital status, birth date, and annual income.

City: It represents cities with their unique identifier (CityID), name, and the corresponding stateID it belongs to.

State: It represents states by a unique ID (StateID) and the name of the state.

The attributes of each entity is shown in tables below:

| **SalesOrder Detail** | **SalesOrder** | **Product** | **Product Category** | **Customer** | **City** | **State** |
|---|---|---|---|---|---|---|
| SalesOrder DetailID (PK) | SalesOrderID (PK) | ProductID (PK) | Product CategoryID (PK) | CustomerID (PK) | CityID (PK) | StateID (PK) |
| SalesOrderID (FK) | CustomerID (FK) | Product CategoryID (FK) | Name | CityID (FK) | StateID (FK) | State Name |
| ProductID (FK) | ShipDate | Name | | Title | City Name | |
| OrderQty | SalesOrder Number | ProductNumber | | FirstName | | |
| UnitPrice | OrderDate | Color | | LastName | | |
| UnitPrice Discount | | Cost | | Address Line | | |
| | | OriginalPrice | | PostalCode | | |
| | | Stock | | Email Address | | |
| | | | | Phone Number | | |
| | | | | Gender | | |
| | | | | Marital Status | | |
| | | | | BirthDate | | |
| | | | | Annual Income | | |

In the table below, the entities, relations, and cardinalities of the Digikala are given.

| Entity | SalesOrder Detail | Sales Order | Product | Product Category | Customer | City | State |
|---|---|---|---|---|---|---|---|
| SalesOrder Detail | | N:1<br>Is in | N:1<br>Includes | | | | |
| Sales Order | 1:N<br>Includes | | | | N:1<br>Belongs to | | |
| Product | 1:N<br>Can be in | | | N:1<br>Belongs to | | | |
| Product Category | | | 1:N<br>Includes | | | | |
| Customer | | 1:N<br>Can have | | | | N:1<br>Is in | |
| City | | | | | 1:N<br>Can have | | N:1<br>Is in |
| State | | | | | | 1:N<br>Has | |

The Entity-Relationship diagram of the Digikala is represented below:

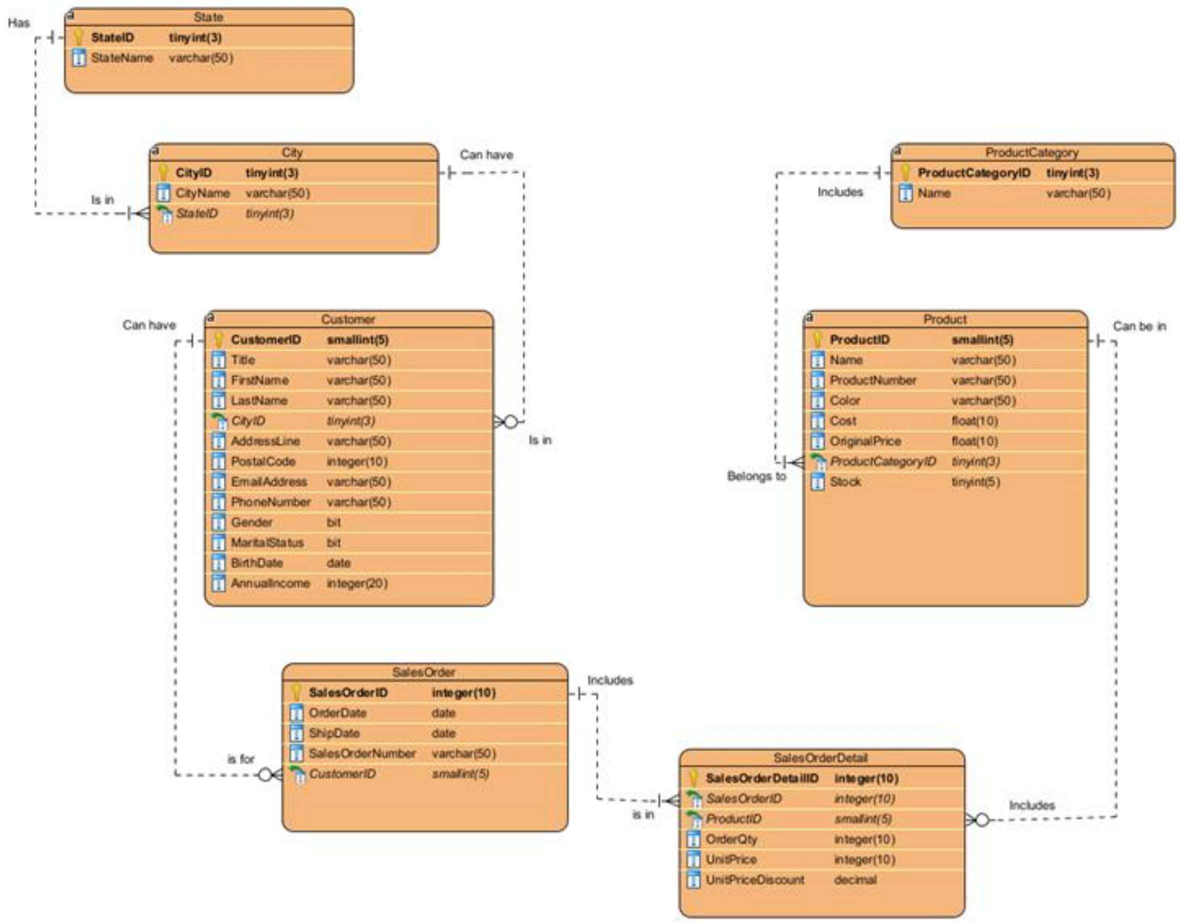


*Diagram5:* ERD Digikala shopping system

# References

Ahamed, B, Ramkumar, T (2016) ‘Data integration- challenges, Techniques and future directions: A comprehensive study’ *Indian journal of science and technology,* vol 9(44). [online] [Accessed 02/04/2024] available at: https://sciresol.s3.us-east-2.amazonaws.com/IJST/Articles/2016/Issue-44/Article73.pdf

Bennett, K. (1995) ‘Legacy systems: coping with success’ *IEEE Software*, 12 (1), pp. 19-23 https://info.lse.ac.uk/staff/services/Policies-and-procedures/Assets/Documents/eleDatRetPur.pdf

Berryman, R, Yost, N, Dunn, N, Edwards, C (2013) ‘Data interoperability and information security in healthcare’ *western Michigan university: centre for health information technology.* Vol.2, No1. [Online] [accessed on 02/04/2024] available at: https://core.ac.uk/download/pdf/144155395.pdf

Chojka A, (2020) ‘semantic and syntactic interoperability issues in the context of SDI’ *Geomatics and environmental engineering*. Vol 14, No3 [Online] [Accessed on 02/04/2024] Accessible at: https://www.gaee.agh.edu.pl/gaee/article/view/54/37

Dedhia, R., Jain, A. and Deulkar, K. (2015) 'Techniques to automatically generate entity relationship diagram', *International Journal of Innovations & Advancement in Computer Science,* 4(10), pp. 68-73.

Dr Tolk, A, Maguira, J.A, (2003) ‘the levels of conceptual interoperability model’ *Virginia modelling analysis and simulation centre college of engineering and technology*. VA 23529. [online] [accessed on 02/04/2024. Available at: https://www.mscoe.org/content/uploads/2017/12/Tolk-Muguira-The-Levels-of-Conceptual-Interoperability-Models.pdf

ETSI, world class standards (unknown date) ‘solve the challenge of interoperability: interoperability best practices’ ETSI.org. [online] [accessed on: 02/04/2024] available at: https://www.etsi.org/images/files/Events/interoperability_best_practices_handbook.pdf

European union (2012) ‘discussion paper: on semantic and technical interoperability’ eHealth governance initiative (eHGVI). [online] [accessed on 02/04/2024] available on: https://health.ec.europa.eu/system/files/2016-11/ev_20121107_wd02_en_0.pdf

Fujitsu (2021) ‘The importance of modern-day data archiving’ *Freeform dynamics* Available at: https://sp.ts.fujitsu.com/dmsp/Publications/public/wp-analyst-importance-of-data-archiving-ww-en.pdf. [online] [accessed: 01/01/2024].

IBM, ‘What is data integration’. IBM solution, [online] [Accessed on 02/04/2024] Available at: https://www.ibm.com/topics/data-integration

Khambra, G, Richariya, P (2014) ‘Survey on data integration in data warehouse’ Binary Journal of Data mining and networking. vol 4. [online] [accessed on 04/04/2024] available at: https://core.ac.uk/download/pdf/290043388.pdf

Lenzerini, M (2002) ‘Data integration: A theoretical perspective’ *Research gate* DOI: 10.1145/543613.543644. [online] [Accessed on 02/04/2024] available at: https://www.researchgate.net/publication/220266329_Data_Integration_A_Theoretical_Perspective

Liu, J. Li, J. Li, W & Wu, J. (2016). 'Rethinking big data: A review on the data quality and usage issues.' *ISPRS Journal of Photogrammetry and Remote Sensing*, 115(May), Pp.134-142

Li, Q. and Chen, Y.-L. (2009) *Entity-Relationship Diagram.* p. 125-139.

Shoval, P., Danoch, R. and Balabam, M. (2004) 'Hierarchical entity-relationship diagrams: the model, method of creation and experimental evaluation', *Requirements Engineering,* 9(4), pp. 217-228.

Nwori, J.C, Nwosu, O (2021) 'Best practices for efficient and effective database archiving' *ResearchGate*. DOI: 10.14662. [online] [Accessed: 01/01/2024] Available at: https://www.researchgate.net/publication/357285029_Best_practices_for_efficient_and_effective_database_archiving

Patel, J. (2019) 'Bridging Data Silos Using Big Data Integration' *International Journal of Database Management Systems.* 11(2/3). Pp.1-6.

Plattner, H., (2009), June. A common database approach for OLTP and OLAP using an in-memory column database. In *Proceedings of the 2009 ACM SIGMOD International Conference on Management of data* (pp. 1-2).

Peddireddy, K., (2023), May. Streamlining Enterprise Data Processing, Reporting and Realtime Alerting using Apache Kafka. In *2023 11th International Symposium on Digital Forensics and Security (ISDFS)* (pp. 1-4). IEEE.

Queiroz-Sousa, P.O. and Salgado, A.C., (2019). 'A review on OLAP technologies applied to information networks'. *ACM Transactions on Knowledge Discovery from Data (TKDD*), 14(1), pp.1-25

Redman, T.C., (2008). 'Data driven: profiting from your most important business asset'. *Harvard Business Press*. 22(Sept).

Reif, V, Maeus, L (202220 'smart metering interoperability issues and solutions; Taking inspirations from other ecosystems and sectors' *Science Direct*, vol 76, No101360 [online] [accessed on 02/04/2024. Available at: https://www.sciencedirect.com/science/article/pii/S095717872200025X

Szaefman, A, Levine, J.G, Tonning, J.M (2022) 'recommendations for archiving interoperable and shareable medical data in the USA'. Vol 2, PMID: 35865358. [online] accessed on 02/04/2024] available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9293957/

Sayiram, B (2021) 'importance of data integration in the data decade' *Dell Technologies*. [online] [Accessed on 04/04/2022] Available at: [https://education.dell.com/content/dam/dell-emc/documents/en-us/2021KS_Sayiram-Data_Integration.pdf](https://education.dell.com/content/dam/dell-emc/documents/en-us/2021KS_Sayiram-Data_Integration.pdf)

The London school of economics and political science (Unknown date) 'Electronic data retention, archiving and purging poli' [online] [accessed: 01/01/2024] Available at:

Valle, L.D (2017) 'Data integration' *University of Plymouth research outputs*. [Online] [accessed on 02/04/2024] available at: https://pearl.plymouth.ac.uk/bitstream/handle/10026.1/9294/Data%20Integration_stat08014.pdf?sequence=1&isAllowed=y