****

**CST3440: Business Intelligence**

**Coursework 1 Report**

**Submission: Friday 1st December 2023**

**Student ID:M00810926**

**Name: Abdullahi Mohamed**

**CAMPUS: HENDON**

**Contents**

**Introduction……………………………………………………………………………………………………………………. 3**

**Data requirements………………………………………………………………………………………………………….. 4**

**Data Storage …………………………………………………………………………………………………………………… 4**

**Conclusion………………………………………………………………………………………………………………………. 11**

**References……………………………………………………………………………………………………………………….12**

**Case Study: Improve Amazon data driven decision to enhance customer experience and operations.**

**Introduction**

Amazon is an online retailer that provide customers with electronics, clothing, groceries, and books. They also offer services and products, this highlights the depth they already go to satisfying customer needs, however the scenario I have decided is to enhance the customer experience and optimize the current operations of the world largest e-commerce. Data can improve the decision making and to achieve any objective, in this case study Amazon can improve the customer satisfaction, supply chain, product recommendation to stay competitive within the e-commerce industry. The goal of this case study is to improve the customer experience, operations of Amazon and to exceed sales expectations. Amazon distributes for a range of demographics, maintaining and providing for all associations is mandatory, and I will be delivering data driven solutions including data requirements, storing, and warehousing them to execute and satisfy the needs of all customers. This topic outlines the operational segment of Amazon, it includes the physical warehouse where products are distributed, sorted, and dispatched to consumers. Faulty products are common in this process, for example a spillage of drink packages due to poor handling of staff leads to a delay of customers receiving their order. Therefore, I will be providing data driven solutions that can enhance the productivity of Amazon by understanding customer needs and optimizing the operatives of Amazon.

**Data Requirements**

Internal Data

**Sales data:** Amazon can collect data by using customer relationship management because it is used for customer preferences based on purchase history, the benefits are as followed, buying patterns of customers become evident to which helps cater to them and then can personalize customers experience by providing recommendations based on their order history.

**Supply chain and coordination:** data on shipping time, order tracking, successful deliveries and return data can help enhance and progress the sales of Amazon. For example, having a stored table of data daily can identify the causes of delayed shipment or returned orders and find a way of refining the logistic and customer reputation.

**Customer service data:** this set of data can consists of what the customers want such as restocking a product, this allows for the voices of the customers to be heard. For example, the customer service data can provide Amazon an understanding of what their consumer preferences are and how they can implement it within the business.

External data

**Competitor Analysis**: The use of web scraping tools can help gather the updated information of competitors statistics to which can include customer reviews and prices. With this external data collection Amazon can improve the customer experience through understanding what are the negatives comments left from customers of the web scraped data, to which it provides a detailed insight on what Amazon can do to ensure their customers are satisfied through the data being used effectively.

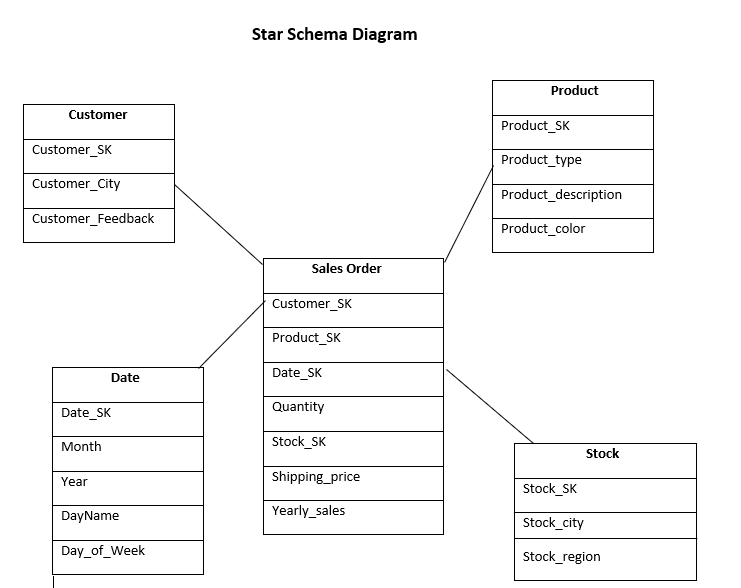
**Social media data:** Amazon gathering data on social media helps to identify trends within their customer, for example many customers online complaining about the timing of delivery can lead to Amazon to find a solution to speed up the process. This also gathers the understanding of the current reputation of the business and what methods they can apply to ensure the customers perspective are considered within the organisations.

**Demographic data:** a set of data analysing the age, gender, location, salary can help Amazon to improve customer experience and optimisation because they can target the least sold products of a demographic then offer discounted items through mailing them which can help lure potential customers to become repeat purchasers.

**Data Storage**

**Star Schema**

The "Sales Order” is the transactional data such as client, product, date, and stock details, is the centrepiece of the star schema. The surrogate keys, the dimension tables "Product," "Customer," "Date," and "Stock" are linked to this is the fact table. Entries for customer and product relationships, along with dates, stock, quantity, shipping charges, and annual sales, are stored in the "Sales Order" database. Customer city, and feedback are all included in the "Customer" field, whereas the product type, description, and colour are all included in the "Product" table. With the use of this dimensional structure, extensive analysis over a wide range of business aspects is made possible, providing insights into consumer behaviour, sales, and product performance and can add value to customer experience at Amazon.



Advantages

* Simplistic: easy to use and structured design
* Flexible: allows for new dimensions or measures without existing structures
* Efficient query: optimizes query performances.

Disadvantages

* Data denormalization: this may result in data overload and storage needs.
* Complexity in maintenance: altering in business requirements may result in complex schemas.
* Limited flexibility: when using query requirements and specialised data models is preferred.

**ETL**

Extract: this is the phase that involves data from a variety source and bring into a central location for further processing. Amazon can use data extracted from data sources such as transaction database, customer interactions and websites data.

Web scraping: tool is available from online sources such as customer reviews, competitor websites and market trends, used for a genuine insight on competitor analysis and can help decision making within Amazon.

Transform: Amazon has a vast amount of data that they contain and has effective techniques for analysis, to which the data is aggregated meaning that is set into weeks, months, and years for Amazon to improve their operations through having quality data. Amazon can personalise customer experiences, increasing customer happiness and engagement by customising content and recommendations based on insights from aggregated data. To optimise operations Amazon can ensure effective operations, aggregated data is used to estimate demand, simplify coordination, and enhance inventory control.

Load: The transformed data is kept in a data warehouse in the third stage, where it is made available for reporting and analysis. This includes loading large volumes of data speeds the process and minimise waiting for a long time and is suitable for Amazon because this process concludes with a data visualization and allows for an informed decision to be finalized.

Advantages

* Integrates diverse data sources ensuring consistency and accuracy.
* Data is cleaned, valued, and enhances of the overall quality and enhances reliability.
* Historical Data: captured and stored data and helps with future marketing trends.

Disadvantages

* Builds and maintained ETL processes and is time-consuming.
* Data latency: data available can impact the real time decision of Amazon.
* Data losses or inconsistencies during transformation phases leading to inaccurate results.

ETL for Amazon.

Amazon can improve customer experiences by using ETL because it provides a cohesive display of consumer behaviour. ETL transformed data leads to a more focused marketing efforts and better search functions. By ensuring correctness, data cleansing promotes improved decision-making. ETL help to customise customer care and quickly address issues and improves customer loyalty by streamlining operations and providing a more tailored and effective shopping experience.

**Online Analytical Processing (OLAP) for Amazon**

Amazon use OLAP (Online Analytical Processing) to manage a variety of cloud databases to help the organizations store and perform OLAP operations, Amazon analyst can use the database to build and align with company requirements. Amazon can use OLAP to analyse customer preferences for books and devices (laptops, smartphones) and create customised strategies to improve customer experiences and streamline operations by slicing, dicing, and aggregating data to obtain insights into the product preferences of different consumer categories.

**Roll Up**

* Reduces data from levels of finer to coarser texture.
* Trend analysis of a range of time periods (weeks, months, years).
* Offers a more comprehensive picture for making strategic decisions.

Right Below is an example of a roll up for Amazon to which it reduces the data and offers a detailed picture of the phone being sold per quarter in each city.

A white cube with black text

Description automatically generated

**Roll Down**

* Enables more in-depth analysis of time periods or segments by Amazon.
* Enables focused enhancements or adjustments founded on detailed understandings.

Below is a roll down for Amazon to which in depth displays a period time that these TV categories were watched and in the countries to.

A cube with different names

Description automatically generated with medium confidence

**Slice**

* A single-dimensional data filter.
* Assists in the examination of data subsets for in-depth analysis.
* Allows Amazon to narrow its focus on specific factors or standards, improving its decision-making procedures.

Below is a diagram of the slice method, to which it narrows down in June the clothing categories that were common amongst specific countries.

A diagram of clothing items

Description automatically generated

**Dice**

* Involves choosing and analysing multidimensional data concurrently along several dimensions.
* Enables Amazon to concentrate on combinations or intersections in the data.
* Facilitates in-depth analysis by reducing the data set to parameters along several dimensions, supporting more focused insights and decision-making.

Below is a diagram of dice, to which shows whether physical or digital books are common more in the summer or the winner in European countries.

A diagram of a cube

Description automatically generated

Advantages

* Organised analysis: fast responses, provide insight into trends and pattern within dataset.
* Decision-Making: data exploration allows for multidimensional analysis for decision-making.
* Users can navigate by drilling down, slicing, and dicing to provide detailed information.

Disadvantages

* Complexity: systems is hard to use, specialised knowledge to data modelling and must be skilled in data analysis to use OLAP.
* Internet Connection: Real time update might be challenging and data refreshes and lead to delays due to latency.

**Big Data**

Big data is a complexed dataset that cannot be proceeded nor analysed with a data tool, I will be defining the big data using the 3 Vs:

Volume: it is when data is being generated in sizes usually from terabytes to petabytes and growing exponentially.

Velocity: Big data is generated and collected at an unprecedented speed, the data is in real time and usually is fast to produce.

Variety: data comes in many different formats such as spreadsheets, usually its unstructured data.

Competitor analysis usually contains unstructured data that contains a variety of sources, assists Amazon with gaining insights in marketing, pricing strategies, and making an informed decision.

Examples For Amazon using big data

Customer reviews: Amazons customer reviews is a set of unstructured text data that data warehousing can struggle to process and analyse vast amounts of data.

Social media web-scraping data: platforms such as Instagram comments, tweets is unstructured data to which is typically sizeable, real time and is usually an example of big data.

Customer service: this includes transcripts, email messages, phone calls are all unstructured data that contain valuable information, and the big data can be used for improvements in customer service.

Supply Chain data: can use big data to optimize operations and can be unstructured data.

Why not suitable for storing in data warehouse.

Customer reviews usually consists of unstructured data and are not maximised to analyse a few customers that makes it difficult to manage within a database.

Social media data has a variety of posts, comments, tweets, reposts, reply that are all unstructured data in real time. Traditional data warehouses lack the efficiency and speed to analyse data in real time which is beyond the capabilities of databases.

Customer service contains plenty of unorganized text data such as emails, chat bots and transcripts. storing a high volume of customer data is challenging due to the variety and complexity of it and it requires efficient methods such as big data.

Supply chain data includes variety of unorganized elements such as communication, shipping tracking and warehouse stock report and requires real time analysis which traditional data warehouse are not designed to oversee.

Why Data Warehouse cannot manage big data.

1. Volume and Velocity of Big Data Passing Through the ETL Process

Big data provides difficulties to data warehousing, passing data through ETL process is overwhelmed by the amount of volume of data designed. The flow of data is challenging to ETL and is designed to manage data efficiently. Traditional data storing struggles with large amount and speed of big data, Amazon using ETL can slow down the operations.

2. Speed of Analysis Required

Customer behaviour can change rapidly therefore data warehouses are unable to provide an accurate real time timely insight.

3. Variety of Data, Unstructured, Semi-Structured

They lack flexibility to manage and analyse unstructured and semi-structured data to which implementing more advanced big data systems can manage the variety of data.

**Framework**

Hadoop provides a good foundation for managing the substantial amounts of data in Amazon's customer-centric operations. Hadoop makes use of components such as HDFS for distributed storage and MapReduce for processing huge datasets. For example, Amazon can use Hadoop to effectively analyse a variety of data sources, including clickstream data and customer reviews. Beyond the technical aspects, Hadoop, and other scalable frameworks that Amazon uses have a considerable influence on its customer-focused operations.

Utilising Hadoop's distributed processing and storage capabilities, Amazon can manage the ever-growing data environment. This method makes it possible to extract complex information from a variety of sources, including clickstream activity and customer reviews. These observations form the basis of Amazon's unwavering quest to improve customer experiences. Amazon uses Hadoop's scalability and parallel processing to process large datasets and drive advances in personalised marketing, supply chain optimisation, and customization.

**HDFS**

* HDFS splits and spreads out big data sets to be stored on several machines.
* Provides dependable and scalable storage in Hadoop clusters by ensuring fault tolerance through data replication between nodes.

**MapReduce**

* The MapReduce programming model enables distributed computer clusters to process large datasets concurrently.
* Splits work into phases called "map" (data processing) and "reduce" (aggregation), making it easier to compute and analyse enormous amounts of data efficiently.

Providing a scalable storage and effective processing of large data sets, HDFS and MapReduce improve Amazon's customer experiences. While MapReduce's parallel processing powers enable Amazon to extract knowledge from vast amounts of data, enabling suggestions, optimised services, and increased operational efficiency, HDFS guarantees dependable data storage and help optimize operations.

**Cloud Storage**

Cloud computing is storing, managing, and delivering services via remote servers connected to the internet, allowing for scalability and on-demand access and offers flexibility in data warehousing, allowing for effective scalability of storage requirements.

* It can oversee the volume and variety of data for big data systems, but there are expenses and significant latency associated with data transit.
* The ability to store and retrieve data on these platforms facilitates smooth scalability, guaranteeing dependable and high-performing storage solutions for customer-facing apps and services.

Benefits for Amazon

* Saving media files or website material on Amazon guarantees fast and dependable access, which enhances user experience by cutting down on load times and ensuring continuous resource availability.
* Storage's adaptability and dependability eventually improve operational agility by freeing up resources for Amazon to concentrate on innovation and service improvements rather than infrastructure maintenance, leading to a more customer-focused and responsive strategy.

**Conclusion**

The case study concludes by emphasising how data-driven choices can improve Amazon's operational effectiveness and customer experience. External data such as competitor analysis and internal data like sales data are essential. Organised analysis can be facilitated by using a star schema in a data warehouse; however, unstructured sources such as social media and consumer reviews provide big data issues and is advantageous to use Hadoop for scalable processing and cloud storage for flexibility. Personally, after reviewing all data models, using a hybrid strategy that combines Hadoop for unstructured data and data warehousing for structured data is the best solution for Amazon. As a result, data can move between systems with ease, using their respective advantages, to which using a hybrid approach to manage the complexity of its sizeable dataset and offer analysis, informed decision-making, and ongoing development of the organization and improve the customer experiences and optimizing operations.

**Reference**

Imperva (2023), What is web scraping? available from articles content at: <https://imperva.com/learn/application-security/web-scraping-attack/> [ date accessed 11th November 2023]

AWS (2023), What is OLAP? Available from AWS at <https://aws.amazon.com/what-is/olap/> [date accessed 22nd November 2023]