**1 ARCHITECTURE DESIGN** 

Architecture Design **Amazon Sales Data Analysis**

|  |  |
| --- | --- |
| **Written By** | Siri S |
| **Document Version** | 0.3 |
| **Last Revised Date** |  |

**Amazon Sales Data Analysis** 1

**2 ARCHITECTURE DESIGN** 

**DOCUMENT CONTROL**

**Change Record:**

|  |  |  |  |
| --- | --- | --- | --- |
| **VERSION** | **DATE** | **AUTHOR** | **COMMENTS** |
| 0.1 | 29- April -  2024 | Siri S | Introduction and architecture defined |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

**Reviews:**

|  |  |  |  |
| --- | --- | --- | --- |
| **VERSION** | **DATE** | **REVIEWER** | **COMMENTS** |
|  |  |  |  |

**Approval Status:**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **VERSION** | **REVIEW**  **DATE** | **REVIEWED BY** |  | **APPROVED BY** | **COMMENTS** |
|  |  |  |  |  |  |

**Amazon Sales Data Analysis** 2

**3 ARCHITECTURE DESIGN** 

**Contents**

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

**2. System Architecture………………………………………………………………………………………….4**

**2.1 Overview…………………………………………………………………………………………………….4**

**2.2 Components………………………………………………………………………………………………..4**

**2.3 Interactions…………………………………………………………………………………………………4**

**3. Technologies and Tools…………………………………………………………………………………….5**

**3.1 Programming Languages………………………………………………………………………………5**

**3.2 Libraries and Frameworks……………………………………………………………………………5**

**4. Data Flow…………………………………………………………………………………………………………6**

**4.1 Data Extraction…………………………………………………………………………………………….6**

**4.2 Data Pre-processing……………………………………………………………………………………..6**

**4.3 Analysis………………………………………………………………………………………………………….6**

**4.4 Visualization……………………………………………………………………………………………………6**

**5. Scalability and Performance………………………………………………………………………………….7**

**5.1 Scalability………………………………………………………………………………………………………..7**

**5.2 Performance…………………………………………………………………………………………………….7**

**6. Security and Privacy………………………………………………………………………………………………8**

**7. Deployment……………………………………………………………………………………………………………8**

**7.1 Deployment Environment…………………………………………………………………………………8**

**7.2 Deployment Process………………………………………………………………………………………..8**

**8. Maintenance and Monitoring………………………………………………………………………………..9**

**8.1 Maintenance…………………………………………………………………………………………………….9**

**8.2 Monitoring………………………………………………………………………………………………………9**

**Amazon Sales Data Analysis** 3

**4 ARCHITECTURE DESIGN** 

1. **Introduction**

The architecture document serves as a blueprint for the design and implementation of a robust system tailored for analyzing Amazon sales data. With the ever-increasing importance of data-driven decision-making in the e-commerce domain, this document outlines a comprehensive approach to harnessing Amazon sales data effectively. By leveraging modern technologies and best practices, the proposed architecture aims to provide insights into sales trends, customer behaviour, and operational efficiency.

In today's competitive landscape, understanding and interpreting sales data is crucial for businesses to stay ahead. The architecture detailed herein offers a structured framework comprising various components, each contributing to the overall process of data analysis and visualization. Through systematic extraction, pre-processing, analysis, and visualization of Amazon sales data, organizations can derive actionable insights to optimize strategies, improve customer satisfaction, and drive business growth.

This document delves into the intricacies of the system architecture, elucidating the roles and interactions of each component. Furthermore, it highlights the technologies and tools utilized, emphasizing their relevance in ensuring scalability, performance, security, and privacy. From data extraction to deployment and maintenance, the architecture document provides a comprehensive guide for implementing a robust and efficient system tailored for analyzing Amazon sales data effectively**.**

**Amazon Sales Data Analysis** 4

**5 ARCHITECTURE DESIGN** 

**2. System Architecture**

**2.1 Overview**

* The system architecture follows a modular and scalable design to facilitate data analysis and visualization of Amazon sales data.
* It consists of multiple components, each responsible for specific tasks such as data extraction, pre-processing, analysis, and visualization.

**2.2 Components**

* Data Extraction Component: Responsible for extracting data from the provided dataset and loading it into the system.
* Data Pre-processing Component: Handles data cleaning, pre-processing, and transformation tasks to prepare the data for analysis.
* Analysis Component: Conducts exploratory data analysis (EDA) to identify sales trends, key metrics, and factors influencing sales performance.
* Visualization Component: Generates visualizations and interactive dashboards to present analysis results in an understandable and actionable format.

**2.3 Interactions**

* The components interact with each other in a sequential manner, with data flowing from the extraction component to preprocessing, analysis, and finally visualization.
* Each component receives input data from the preceding component, performs its tasks, and passes the processed data to the next component in the pipeline.

**Amazon Sales Data Analysis** 5

**6 ARCHITECTURE DESIGN** 

**3. Technologies and Tools**

**3.1 Programming Languages**

**Python:**

Used for data manipulation, analysis, and preprocessing tasks.

**SQL:**

Employed for querying and manipulating data stored in databases, if applicable.

**3.2 Libraries and Frameworks**

**Pandas:**

Used for data manipulation and preprocessing tasks.

**Matplotlib, Seaborn, Plotly:**

Utilized for data visualization.

**Scikit-learn:**

Employed for advanced analysis tasks such as clustering or predictive modeling, if applicable.

**Amazon Sales Data Analysis** 6

**7 ARCHITECTURE DESIGN**



**4. Data Flow**

**4.1 Data Extraction**

* The data extraction component reads the provided dataset using pandas' read\_csv() function and loads it into memory.

**4.2 Data Preprocessing**

* The pre-processing component handles missing values, duplicates, data type conversions, and outlier detection using pandas and NumPy libraries.

**4.3 Analysis**

* The analysis component conducts exploratory data analysis (EDA) to identify sales trends, distribution of key variables, and calculate key metrics such as average resolution time, requester wait time, and satisfaction scores.

**4.4 Visualization**

* The visualization component generates visualizations and interactive dashboards using Matplotlib, Seaborn, Plotly.

**Amazon Sales Data Analysis** 7

**8 ARCHITECTURE DESIGN** 

**5. Scalability and Performance**

**5.1 Scalability**

* The system architecture is designed to be scalable, allowing for the processing of large volumes of data efficiently.
* Components can be scaled horizontally by deploying multiple instances of the system in parallel to handle increased workload.

**5.2 Performance**

* Performance optimizations such as parallel processing, efficient data structures, and caching mechanisms are implemented to ensure optimal performance.

**Amazon Sales Data Analysis** 8

**9 ARCHITECTURE DESIGN** 

**6. Security and Privacy**

* Data security measures are implemented to protect sensitive information and comply with privacy regulations such as GDPR.
* Access controls, encryption, and data anonymization techniques are employed to safeguard data privacy.

**7. Deployment**

**7.1 Deployment Environment**

* The system can be deployed on-premises or in the cloud, depending on organizational requirements and infrastructure capabilities.
* Cloud platforms such as AWS, Azure, or Google Cloud are preferred for their scalability, flexibility, and managed services.

**7.2 Deployment Process**

* Deployment involves packaging the system components into containers using Docker and orchestrating them using Kubernetes or a similar container orchestration platform.
* Continuous integration and continuous deployment (CI/CD) pipelines are established to automate the deployment process and ensure smooth updates and releases.

**Amazon Sales Data Analysis** 9

**10 ARCHITECTURE DESIGN** 



**8. Maintenance and Monitoring**

**8.1 Maintenance**

* Regular maintenance activities include monitoring system health, updating dependencies, and optimizing performance.
* Automated monitoring tools and alerts are set up to detect and respond to issues proactively.
  1. **Monitoring**
* Key performance indicators (KPIs) such as system uptime, response time, and resource utilization are monitored using monitoring tools like Prometheus or Datadog.
* Logs and metrics are collected and analyzed to identify performance bottlenecks and areas.