Amita Worlikar

AMITA.WORLIKAR@CITI.COM

Hackathon - Bigdata

AdventureWorks Case Study

**Objective:**

As a Data Engineer, you will assist AdventureWorks in the solution design and implementation to meet the Business, functional and technical requirements that the company has set forth to be successful for growth, expansion and innovation strategies. You will execute this is a way that minimizes operational costs and can be monitored for effectiveness.

**Data Analysis:**

* **Existing DWH solution**

**Current State:**

* Single on-premise database which is configured as Historical DWH database – **AdventureWorksDW** used for Reporting and descriptive analysis
* This Server has been struggling to process the reports timely manner

**Future State:**

* Data warehouse to be migrated from on-premise to the new platform
* You should ensure that access to the data is restricted
* Also wants predictive analytics capabilities as new feature
* New platform or analytics engine required which is resilient and performant

**Expected Output:**

POC of Sqoop, Hive and Other Big Data tools to explore the capabilities of the product

**Suggested Use Cases:**

|  |  |  |
| --- | --- | --- |
| **S#** | **Use case** | **Suggested Solutions** |
| U001 | Data warehouse to be moved from on-premise to new platform which is accessible globally | Azure Cloud solution to be adopted |
| U002 | Predictive Analytics Capabilities, analytics engine which is required to be resilient and performant | Hadoop, Hive Database, Sqoop |

* **Customer Service /Pre-sales**

**Current state:**

* Customer service and pre-sale experiencing high call volumes
* Fraudulent cases where customers asking for support of bikes not in warranty or not even purchased from AdvenureWorks
* Currently dependent on the agents to identify fraudulent cases

**Future State**

* Introduction of **chatbot** for Customer service and pre-sale which would provide
  + Through a set of questions customer recommendations are given, order is placed or it is redirected to sales specialist
  + Retrieve status of the order and estimated delivery time
  + Find suitable bicycle parts suitable for existing bicycle.
  + Bicycle owners can upload the picture of their bicycle or take a serial number of their bicycle which will assist further to recommend bicycle parts
* Mechanism to identify the bike indeed purchased from AdventureWork and may need to capture warranty period, serial number, start and end date where it will be tracked systematically to help agents
* Chatbot need to respond to Cust service/pre-sales in near real-time regardless of where customer is located. Chatbot should support multiple languages such as Dutch, German, French, English, Spanish and Japanese which will be done by AI engineers but they need platform to store conversation history

**Suggested Use Cases:**

|  |  |  |
| --- | --- | --- |
| **S#** | **Use case** | **Suggested Solutions** |
| U003 | Using the SalesOrderID, retrieve delivery date or ship date | User the SalesOrderheader database to search based on specific SalesOrderID |
| U004 | Find suitable bicycle parts suitable for existing bike | Spare parts table needs to be created based on the bicycle model id  Select parts based on the bicycle model id |

* **Social Media Analysis**

Capability is required to track hashtags during the campaign at any time. – **This requirement is not applicable.**

* **Connected bicycles**

**Current State:**

This feature does not exist.

**Future State:**

* Bicycles will be equipped with an innovative built-in computer which consist of automatic locking features of the bicycle as well as operational status. Information captured includes:
* Bicycle model, serial number and registered owner
* Bicycle location (latitude longitude)
* Current status (stationary, in motion)
* Current speed in km per hours
* Bicycle locked /unlocked
* Bicycle parks and components information (on electrical bicycles)
* Access

First Party and 3rd party applications can have access the information of the bicycles computer that must be secure and for the integration into mobile applications and real time display of location and bike ride sharing information.

* Furthermore, daily summary data can be saved to flat files that include Bicycle model, serial number, registered owner and a summary of the total miles cycled per day and the average speed

(New data base to maintain access rights, summary of data)

**Suggested Use Cases:**

|  |  |  |
| --- | --- | --- |
| **S#** | **Use case** | **Suggested Solutions** |
| U005 | New database to capture the data receive from build-in computer | Create new database to capture  Bicycle model, serial number, registered owner  Bicycle location (latitude longitude)  Current status (stationary, in motion)  Current speed in km per hours  Bicycle locked, unlocked  Bicycle Type Electric, manual  Bicycle parts information (for electrical bicycles) |
| U006 | Find suitable bicycle parts suitable for existing bike | Spare parts table needs to be created based on the bicycle model id  Select parts based on the bicycle model id |

* **Bicycle Maintenace Services**
* Existing bicycle owners can opt in to getting notifications on when the bicycle needs repair, based on
* Telemetry from electrical bicycle based on sensor data
* Bicycle usage information coming from the built in computers based on average mileage / wear and tear

This service can be offered as paid service.

**Design and Architecture:**

**Inject and Data Storage**

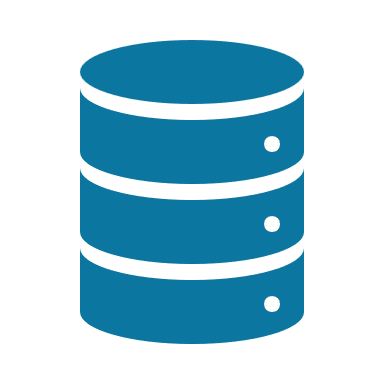
**A close-up of a logo

Description automatically generated**

**Data Sources**

**Data Processing**

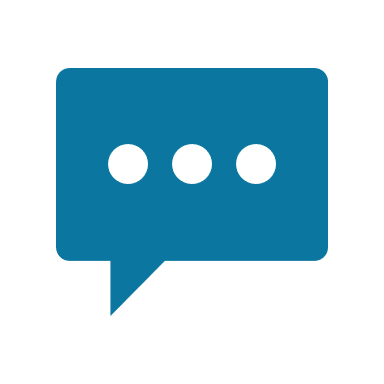
**Data Usage**

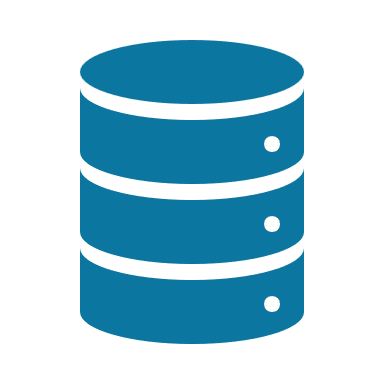
****

**SQL DB**

**Data visualization, Reports**

**AdventureWorks Web site**

****

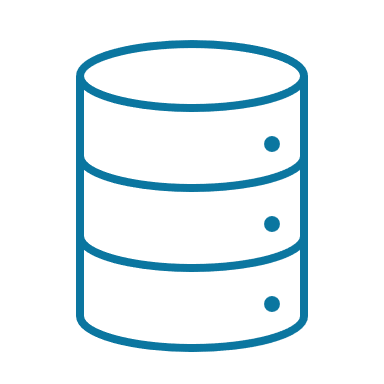
****

**ChatBot**

****

**Hive / HDFS/ Hadoop**

**Call centre**

****

**AdventureWorksDW**

**A black and white logo

Description automatically generated**

**For Streaming**

**Microsoft Azure Cloud Services**

**Streaming from Sensors on bicyle**

**POC results for SQOOP, HDFS:**

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A close up of a text

Description automatically generated

**Conclusion of the POC:**

The SQL database and tables were listed however sqoop import had issues due to SQL driver JAVA version mismatch with the Cloudera installed on RPS cloud. It was concluded to receive CSV files to cover next steps for this project.

**Data Source:**

[AdventureWorks/ at main · Kiran-255666/AdventureWorks · GitHub](https://github.com/Kiran-255666/AdventureWorks/tree/main)