SOW - Real-time analysis of user clickstream data from client’s website



**MothersonSumi INfotech & Designs Ltd.**

(A CMMi Level 5 Company)

C-26, Sector-62, NOIDA-201309, India

Website: [www.mind-infotech.com](http://www.mind-infotech.com)

**Bishan Chauhan**

+91 – 120 4365 124

Bishan.Chauhan@mind-infotech.com

The information in this document shall not be duplicated, used, or disclosed in whole or in part outside the customer. If a contract is awarded to MothersonSumi INfotech & Designs Ltd. (MIND) as a result of or in connection with the submission of this document, MIND shall have the right to duplicate, use, or disclose the information within its organization to the extent provided by the contract between the customer and MIND. This restriction does not limit the customer’s right to use information contained in this document if it is obtained from another source without restriction.

Contents

[1. Introduction 3](#_Toc58598916)

[2. Purpose of Real-time analysis of user clickstream data from client’s website 3](#_Toc58598917)

[3. Scope of Work 4](#_Toc58598918)

[3.1 Scope 4](#_Toc58598919)

[3.2 Architecture Diagram 5](#_Toc58598921)

[4. Delivery Timelines 5](#_Toc58598922)

[5 Assumptions 6](#_Toc58598923)

[6 Applicable Terms 6](#_Toc58598924)

# Introduction

MothersonSumi INfotech & Designs Ltd. (MIND) is a part of Joint venture between **Samvardhana Motherson Group** (SMG) of India and **Sumitomo Wiring Systems** of Japan (SWS).

MIND is a provider of end-to-end software and engineering design solutions to companies around the globe. MIND started as an IT arm of the group in the year 2000 to support the IT needs of Samvardhana Motherson Group and Sumitomo Wiring Systems worldwide. MIND has further ventured into European and American Market to customers who are non-SWS and SMG to expand our services.

MIND's headquarters and development centers are located in Noida (near New Delhi), India. MIND is a CMMi Level 5, an ISO 9001:2008 and ISO 27001 certified company. Since its inception in 2000, MIND has emerged as a strong world class IT Company with projects across the globe. MIND has multi-lingual software development capabilities including Japanese and German.

MIND is a Microsoft Gold Certified Partner, AWS, Azure & Google Cloud Service Provider, Oracle GOLD OPN partner and also partner with other big IT brands.

MIND has Data Center (Level 3) services, Security Consulting Services Enterprise IT Helpdesk (Multi-lingual), Remote Application Management, Performance Management & Capacity Planning, Network Management Services and Application Hosting. MIND has defined Business Continuity (BC) and Disaster Recovery (DR) plans to mitigate risk of business disruption for its customers.

# Purpose of Centralized logging and analysis of multi-regional AWS WAF logs

**Business Problem**

Clickstream events are small pieces of data that are generated continuously with high speed and volume. In this business problem, they are generated by user actions and it is very useful to analyze them for data-driven decision making, influenced by the user behavior.

But capturing and processing data clickstream events in real time is difficult due to its large volume. The customer wanted to analyze the effectiveness of its new application features, website layout and marketing campaigns, in real time so that he can take actions faster.

For doing this the biggest challenge is to store metadata coming from collegedunia’s website, mainly the metadata about which links are clicked most, which link is seen for most of the time duration etc.

To cater to this need, the metadata is extracted from their website and is restructured to a specific format which ensures seamless data flow in the pipeline. Troubleshooting and root-cause analysis further identifies anomalies.

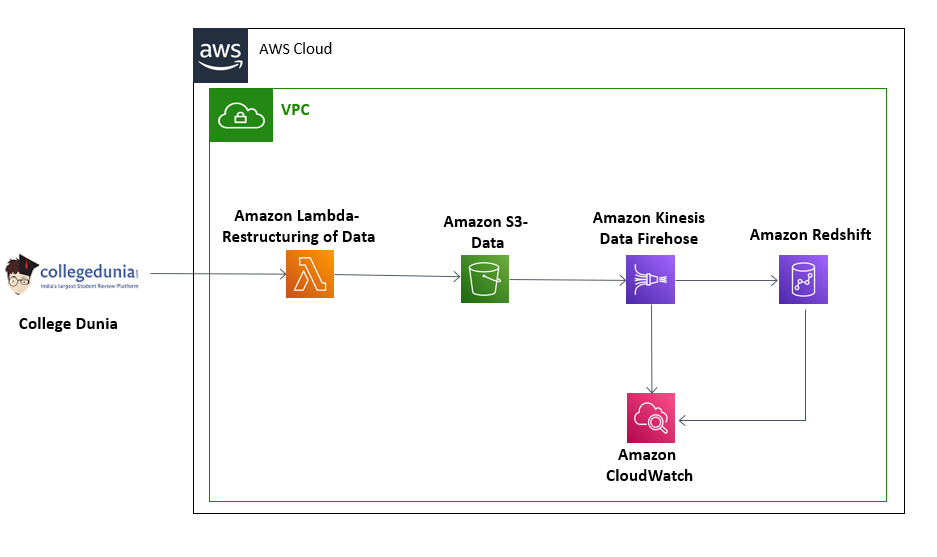
# Scope of Work

### 3.1 Scope-

MIND discussed the problem with the customer and after analyzing the business problem, following steps were taken to deliver an efficient solution-

* This solution detects the user behavior on collegedunia website by analyzing the sequence of clicks that the user is making, the duration of time the user spends, where it usually begins to navigate, and how the user ends the session.
* By tracking this user behavior in real time, this solution helps to update recommendations, perform advanced A/B testing, push relevant notifications based on duration of session, and a lot more.
* As the number of users and web and mobile assets you have increases, so does the volume of data which is why this solution uses Amazon Kinesis which provides you with the capabilities necessary to ingest this data in real time and generate useful statistics immediately so to take necessary action.
* AWS Lambda service is used to fetch metadata from college dunia website and then process this data so that the output is restructured data, AWS S3 is used to store this restructured data.
* Lambdas are no doubt a powerful service but using them to directly ingest data is not the best approach so we have paired it with some other AWS services which also lowers down the cost of the pipeline.
* Firehose’s job is simply receiving data and pouring it at a target destination. There are options to choose between S3, RedShift, Lambdas, and Elasticsearch. Firehose, however, is an excellent fit for this scenario as its job is simply to receive the data and pour it at the destination location.
* Amazon Firehose streams the restructured data coming from S3 to Amazon Redshift.
* Then Amazon Redshift is used to run high performance queries on this streamed data with the customer then creates powerful reports and dashboards for quick insights.

### 3.2 Architecture Diagram



### 4. Delivery Timelines

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Scope Schedule** | | | | | | |
| **Activity** | **Wk1** | **Wk2** | **Wk3** | **Wk4** | **Wk5** | **Wk6** |
| Working on textract console |  |  |  |  |  |  |
| Building Pre-Processing Lambda |  |  |  |  |  |  |
| Building of Post-Processing Lambda |  |  |  |  |  |  |
| Building of lambda function which saved data to DynamoDB |  |  |  |  |  |  |
| Building of the Front-end |  |  |  |  |  |  |

# Assumptions

* Logs are near real time and without any delay.

6. Non-Functional Requirements

1. No specific workload related DR requirements are specified (unless specifically mentioned)
2. Cloud Watch metrices -i.e. workload operational health can be checked using CloudWatch, not applicable for this workload.
3. MIND ensures that capacity matches but does not exceed what is needed for workload, including by using a demand-based, buffer-based, or time-based approach. This will ensure smallest possible Payload size

MIND solutions are designed to keep bandwidth and message costs down by eliminating unnecessary messaging. MIND Architect takes care for optimized data transfer

# 6 Applicable Terms

1. MIND will deliver work as per SoW defined above, and as part of scope mentioned in support, MIND will manage and operate the workload on mutual agreement.
2. MIND will hand over the workload on acceptance of SoW work, hand off responsibilities include (if support is not applicable as part of support scope)
3. Customer will manage the workload account and will be responsible for maintenance of it including security / permissions of the workload