HIGH-LEVEL DESIGN DOCUMENT

**Designing a Streaming Data Pipeline with Dataflow and visualization dashboard using Looker**

Created Date: 10/04/2023

Revision History

|  |  |  |  |
| --- | --- | --- | --- |
| Version | Date | Created By | Reviewed By |
| 1.0 | 10.04.2023 | Saikumar Chappa | Pasha |
| 1.1 |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

TABLE OF CONTENTS

Contents

[1. Introduction 4](#_Toc129867130)

[1.1 Purpose 4](#_Toc129867131)

[1.2 Scope 4](#_Toc129867132)

[1.3 Document Overview 4](#_Toc129867133)

[2. System High-Level Design Overview 4](#_Toc129867134)

[2.1 System High-Level Design Goals, Scope, and Objectives 4](#_Toc129867135)

[2.2 High Level Architecture 4](#_Toc129867136)

[3. System Design Consideration or Approaches 4](#_Toc129867137)

[4. Deployment / Reporting 4](#_Toc129867138)

# Introduction

## Purpose:

This document provides the information related to the Designing a Streaming Data Pipeline with Dataflow and visualization dashboard using Looker. Complete end to end functionality and the High Level Architecture was documented. This Document is supported by the Detailed document of end to end Technical Steps in Power Point Presentation.

## Scope:

The scope of a streaming data pipeline involves the processes and technologies required to collect, process, and analyze real-time data streams. Visualizing the streaming data using Looker Studio which includes the scenarios related to activity details.

The dashboard provides single point of view to analyze the retail business and build business strategy and future decisions.

# System High-Level Design Overview

## System High-Level Design Goals, Scope, and Objectives

**Scope:**

The scope of a streaming data pipeline refers to the various stages and components involved in processing and analyzing real-time data streams. This includes the Ingestion of data from various sources, processing and transforming the data, storing the data, monitoring and analyzing the data for insights. The dashboard provides single point of view to analyze the retail business and build business strategy and future decisions.

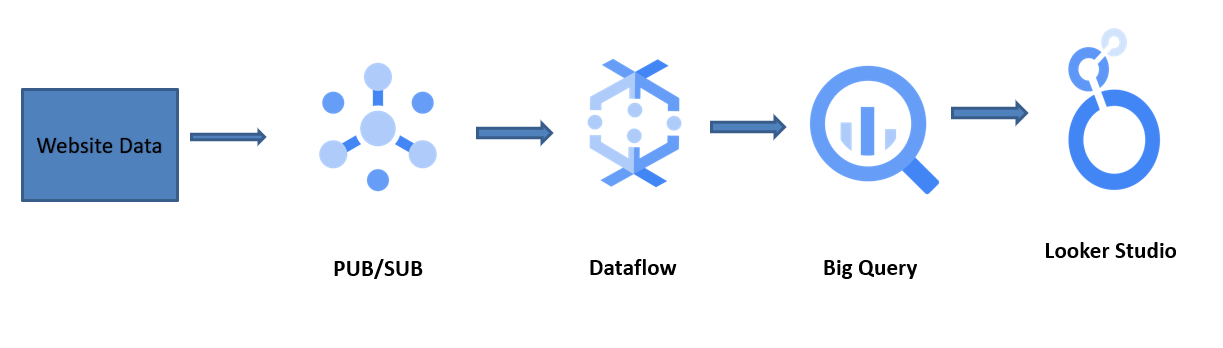
**Objective:**

Designing a Streaming Data Pipeline with Dataflow and visualization dashboard using Looker.

**Goal:**

Our goal is to load the steaming data to Bigquery using Dataflow and visualize real time Insights using Looker Studio.

## High Level Architecture



## System Design Consideration or Approaches:

Building a streaming data pipeline requires careful consideration of various system design factors to ensure that the pipeline can handle the large volume of data in real-time and provide reliable insights.

To Build this we uses Google Cloud Platform, Dataflow, BigQuery, Pub/Sub, Looker Studio.

**Cloud Dataflow** is a fully managed, cloud-based service provided by Google Cloud Platform (GCP) for building and running large-scale data processing pipelines. It is a serverless data processing service that allows users to build batch and streaming data processing pipelines using Apache Beam programming model, which provides a unified programming model for both batch and stream processing.

**Cloud Pub/Sub** is a fully-managed, scalable messaging service provided by Google Cloud Platform (GCP). It provides reliable, real-time messaging that enables asynchronous communication between distributed systems and applications. Cloud Pub/Sub supports both pub-sub and streaming messaging patterns and can handle millions of messages per second.

**BigQuery** is a fully-managed, cloud-based data warehousing and analytics platform provided by Google Cloud Platform (GCP). It enables users to analyze massive datasets using SQL queries and provides scalable, high-performance data processing capabilities.

**Looker** is a business intelligence and data analytics platform that allows users to visualize and analyze their data using a web-based interface. Looker Studio is a feature within Looker that provides a collaborative workspace for data teams to develop, manage, and share their data models and analytics.

## Deployment / Reporting:

Over all data showing the eCommerce website related activity details. The visualization report also provides the insights related to the particular -activity.

