

Arquitectura de eventos para la ingesta de datos en tiempo real

Laura Milena Ramos Bermudez
Maestría en Informática
Escuela Colombiana de ingeniería Julio Garavito
Bogota, Colombia
laura.ramos-b@mail.escuelaing.edu.co

Resumen—Hoy en día las aplicaciones modernas utilizan diferentes fuentes de información como redes sociales, bases de datos relacionales, métricas de uso de aplicación, etc, esto genera altos volúmenes de datos que deben ser procesados con el objetivo de unificar su almacenamiento y explotar la información para generar acciones en tiempo real que ayuden al negocio.

I. PROBLEMA

Actualmente estoy trabajando en el desarrollo del proyecto de grado que tiene como objetivo la implementación de un sistema Big Data que permita alertar en tiempo real posibles irregularidades en la contratación de gasto publico y licitaciones publicas. Una parte muy importante para la implementación de la arquitectura del proyecto es el almacenamiento y limpieza previa de los datos que son necesarias para realizar las alertas o banderas rojas de corrupción sobre los contratos.

Para esto se plantea una arquitectura orientada a eventos con el objetivo de procesar esta información en tiempo real y almacenarla en un único punto para su futura explotación.

A demás, de acuerdo a la investigación realizada previamente en el proyecto de grado se requiere los procesos de contratación y los contratos licitados que son publicados en www.datos.gov.co

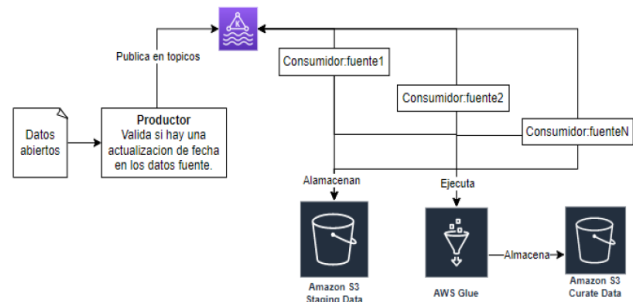


Figura 2. Arquitectura propuesta

COP I y II, en dado caso de que la fecha sea actualizada publicara un evento en un tópico diferente por cada fuente de información.

- **Consumidor:** Una aplicación en Scala que consuma los eventos publicados, descargue, almacene los datos en Amazon S3 en la zona curated para su posterior limpieza con AWS Glue y guarde su schema de datos por cada fuente en catalogo de AWS Glue.

Consulte la información de los Procesos de Contratación de SECOP publicados en datos.gov.co

- ▼ SECOP I
 - Plan Anual de Adquisiciones
 - Proponentes por Proceso
 - Multas y sanciones
 - Adiciones
 - Procesos de Contratación
- SECOP II
- SECOP I - II - Integrado
- TVEC

Figura 1. Datos a descargar

A continuación se propone la arquitectura:

- **Productor:** Mediante una aplicación en Scala se validara la fecha de actualización de los datos publicados en SE-