

Informe Técnico: clase 13 de abril de 2024

Juan Esteban Pulido Lancheros¹, Jorge Hernando Chávez Camargo²,

Andrés Felipe Martínez³

Universidad Central

Maestría en analítica de datos

Automatización e integración de datos

Bogotá, Colombia

¹jpulido2@ucentral.edu.co, ²jchavezc1@ucentral.edu.co, ³amartinezh3@ucentral.edu.co

Abril 19, 2024

1 Resumen ejecutivo

La clase del día 13 de abril de 2024 inició con la sesión del observatorio de noticias sobre tecnología, en este espacio se socializaron cuatro noticias.

La primera estaba relacionada con un sistema llamado Udio que permite a los usuarios crear música a partir de texto, incluyendo letras, y ofrece una amplia gama de opciones de personalización.

La segunda contextualizó que chat GPT-4 logró recuperar su posición como líder entre los chatbots, después de haber sido superado por Claude 3. OpenAI ha lanzado GPT-4 Turbo, que incluye datos más actualizados y un precio más accesible, buscando mantener su dominio en el campo de la inteligencia artificial.

En tercer lugar, se socializó la noticia que indicó que en marzo de este año se dio a conocer el robot humanoide más cercano a lo ahora llamado como embodiment o “encarnación”, Su nombre es Figure 01 y tiene la capacidad de distinguir formas, objetos y el uso de los objetos; para su fabricación participó OpenAI, NVIDIA y Jeff Bezos dueño Amazon.

La última noticia trató sobre Whisper, que según el New York Times (NYT), OpenAI desarrolló esta aplicación para transcribir videos de YouTube y entrenar GPT-4 con la transcripción de un millón de horas de videos.

La segunda parte de la clase consistió en la presentación de las exposiciones sobre la plataforma de servicios cloud Amazon Web Services (AWS) y el sistema de gestión de bases de datos NOSQL Apache Cassandra.

La clase finalizó con la socialización de los avances de los proyectos de la clase de los grupos: 1) Turing, con el tema “Segmentación de clientes para empresa de comercialización de automóviles”; 2) Monitor, con el tema “modelo predictivo para la acción anticipatoria y alerta temprana en eventos climáticos asociados a inundaciones”; y una introducción del grupo 3) Infinity DataFlow IA, con el tema “fraude en tarjetas de crédito” quienes terminarán de mostrar su proyecto en la clase del 20 de abril.

2 Introducción

Los avances continuos de la tecnología han permitido desarrollar sistemas capaces de obtener información desde prácticamente cualquier objeto que tenga conexión a internet. Todos estos datos se han convertido en un insumo vital para las sociedades y las compañías al ser uno de los mecanismos que nos permiten comprender los comportamientos y las necesidades de los diferentes grupos de la sociedad de acuerdo con los intereses y objetivos de las empresas y los estados. Ante esta situación la innovación en programas de captura, integración, procesamiento y análisis de información ha ido en aumento exponencial en las últimas décadas. Por esto, la clase de automatización e integración de datos es el espacio en el que los estudiantes de la maestría en analítica de datos podremos conocer y comprender como se estructura de manera correcta e integral un flujo de datos, los componentes necesarios en cada etapa del flujo y explorar los diferentes softwares que existen para las necesidades de cada proyecto.

En esta clase se abordaron las herramientas Amazon Web Services (AWS), el sistema de gestión de bases de datos NOSQL Apache Cassandra y Zapier.

3 Contexto y antecedentes

En esta sección veremos los antecedentes principales y el contexto de las tres herramientas expuestas en clase en primer lugar la plataforma de servicios cloud Amazon Web Services (AWS), en seguida el sistema de gestión de bases de datos NOSQL Apache Cassandra y finalmente sobre Zapier.

Amazon Web Services (AWS), lanzado por Amazon en 2006, revolucionó la industria de la tecnología al ofrecer una plataforma de servicios en la nube que se convirtió en líder del mercado. Originado por la necesidad interna de Amazon de escalabilidad y flexibilidad en su infraestructura de TI, AWS se convirtió rápidamente en un proveedor de servicios en la nube confiable y escalable para empresas de todos los tamaños. Su oferta inicial incluyó servicios básicos como almacenamiento (Amazon S3) y capacidad informática (Amazon EC2), y desde entonces ha crecido para abarcar una amplia gama de servicios, desde bases de datos y análisis hasta inteligencia artificial y aprendizaje automático.

La popularidad de AWS se debe a varios factores clave, incluida su capacidad para escalar recursos informáticos según sea necesario, su modelo de negocio de pago por uso, que elimina la necesidad de inversiones iniciales significativas en hardware, y su continua innovación y actualizaciones de servicios. Utilizado por millones de clientes en todo el mundo, AWS ha transformado la forma en que las empresas gestionan y escalan sus recursos informáticos, permitiendo a las organizaciones adaptarse rápidamente a las cambiantes demandas del mercado y mantener su competitividad en la era digital.

Apache Cassandra es un sistema de base de datos distribuida, altamente escalable y de código abierto, diseñado para manejar grandes volúmenes de datos en entornos distribuidos y de alta disponibilidad. Fue creado originalmente por Facebook en 2008 y luego se convirtió en un proyecto de código abierto de la Apache Software Foundation en 2009.

Cassandra está diseñado para ofrecer una alta disponibilidad sin un único punto de fallo, utilizando una arquitectura distribuida y descentralizada que distribuye los datos a través de

múltiples nodos en un clúster. Esto permite que Cassandra mantenga el rendimiento y la disponibilidad incluso en situaciones de fallo de hardware o red.

Zapier es una plataforma de automatización en línea que conecta más de 3.000 aplicaciones web, permitiendo a los usuarios crear flujos de trabajo automatizados llamados "Zaps". Estos Zaps pueden realizar acciones automáticamente en una aplicación cuando ocurre un evento en otra, eliminando la necesidad de tareas manuales repetitivas. Desde integraciones con herramientas de productividad como Google Workspace y Microsoft Office hasta aplicaciones de redes sociales y comercio electrónico, Zapier ofrece una amplia gama de posibilidades para mejorar la eficiencia y la productividad al simplificar los flujos de trabajo y permitir una mayor automatización sin necesidad de conocimientos de programación.

Una de las características distintivas de Cassandra es su modelo de datos sin esquema, que permite a los usuarios almacenar y recuperar datos de manera flexible sin necesidad de definir un esquema rígido de antemano. Esto hace que Cassandra sea especialmente adecuado para aplicaciones con requerimientos de escalabilidad y flexibilidad de datos, como aplicaciones web y móviles a gran escala.

4 Metodología

La clase está estructurada para abarcar todos los temas de interés asociados a la automatización e integración de datos. Cada una de las sesiones permite generar un conocimiento amplio y sólido sobre el contexto actual de la tecnología utilizada en analítica de datos y además se abarcan todos los temas objeto de estudio de la clase mientras que a su vez se monitorean los proyectos y se retroalimentan continuamente tomando como referencia los temas vistos en la cátedra.

5 Desarrollo

La clase integra la participación activa de los estudiantes en cada una de las secciones de la clase y permite que en cada uno de los proyectos de los compañeros se pueda ver como se pueden aplicar las herramientas vistas en clase.

6 Conclusiones

En la clase del 13 de abril de 2024 se observó una vez más como los avances continuos en tecnología nos muestran innovaciones en diversos campos como el surgimiento de Udio para la creación de música a partir de texto, el desarrollo de humanoides y la integración de tecnologías actuales para fortalecer los avances en inteligencia artificial y robótica. Además, se exploraron herramientas fundamentales como AWS, Apache Cassandra y Zapier, destacando su papel en la gestión de datos y la automatización de procesos.

7 Referencias

Amazon Web Services. (s.f.). Computaci3n en la nube con AWS. Recuperado el 18 de abril de 2024, de <https://aws.amazon.com/es/what-is-aws/?nc2=hqlent>.

Apache Cassandra (2009-2024). Apache Cassandra's documentation. Recuperado el 18 de abril de 2024, de <https://cassandra.apache.org/doc/latest/>.

Zapier (2024). Make your work flow with Zapier. Recuperado el 18 de abril de 2024, de <https://zapier.com>