

**Integrantes:**

- Cortés Macías Gretel Penélope 317312184
- Velázquez Barrón Marilú Yatzael 318353492
- Peña Nuñez Axel Yael 318279754
- Escalante Castañeda Lenin Alberto 420003193

## 1 Resumen Data Integration:

La integración de big data es una parte crucial de la gestión de enormes conjuntos de datos en múltiples valores. Combina la gestión de datos con operaciones de inteligencia empresarial y abarca múltiples fuentes de datos dentro de la empresa y otras fuentes. Estos datos se pueden integrar en un subsistema único y ser utilizados por las organizaciones para el crecimiento empresarial. La integración de big data también implica el desarrollo y la gobernanza de datos de diferentes fuentes que podrían afectar la capacidad de las organizaciones para manejar estos datos en tiempo real.

La integración de datos en proyectos de big data puede ser crítica y abarca tres etapas: descubrir las fuentes de datos, analizarlas para obtener una comprensión más profunda y perfilar los datos; comprender el valor de los datos y mejorar su calidad; y finalmente, transformar los datos según el entorno de big data.

Los "cinco V" del big data (volumen, velocidad, variedad, veracidad y valor) pueden influir en la integración de datos de muchas maneras. El volumen de datos generado por grandes organizaciones ha aumentado rápidamente, lo que requiere alternativas para integrar estos datos en grandes volúmenes. La velocidad se refiere a la velocidad de transmisión de datos desde la fuente al destino. Los datos pueden ser estructurados o no estructurados, y la integración de datos puede ser más fácil en los datos estructurados. La veracidad se refiere a la confiabilidad y precisión de los datos, y el valor se relaciona con el beneficio que los datos pueden aportar a la organización.

Las organizaciones a menudo implementan metodologías de big data utilizando métodos tradicionales de extracción, transformación y carga (ETL), y Hadoop se utiliza como un nuevo marco de big data para procesar grandes conjuntos de datos de diferentes fuentes.

Las principales desafíos en la integración de datos en un entorno de big data incluyen la gestión del volumen de datos, la inconsistencia de datos, la optimización de consultas, la falta de recursos y la escalabilidad. Además, se necesitan sistemas de soporte para manejar errores y cambios en los requisitos.

Se proporcionan ejemplos de casos de estudio en los que herramientas de integración de datos como Talend y CloverETL se utilizaron con éxito para abordar problemas específicos de integración de datos en empresas.

En conclusión, la integración de datos en el mundo de la tecnología de la información actual es fundamental debido al crecimiento constante de los datos. Los científicos de datos siguen buscando soluciones para simplificar la integración de datos, y se espera que el desarrollo de nuevas herramientas de integración de datos aborde los desafíos de la integración de big data de manera más efectiva en el futuro.