

Big Data

BIG DATA

- Conjuntos extensos de datos, principalmente en las características de volumen, variedad, velocidad y/o variabilidad, que requieren una arquitectura escalable para su almacenamiento, manipulación y análisis eficientes.
- Conjunto de datos cuyo tamaño supera la capacidad de las herramientas típicas de SW de base de datos para capturar, almacenar, administrar y analizar.

Las 5 'V'

- Volumen → grandes volúmenes de datos.
- Variedad → diversos formatos de datos procedentes de diferentes fuentes de información.
- Velocidad → altas velocidades de acumulación y procesamiento de datos.
- Veracidad → precisión e integridad en la generación y procesamiento de los datos.
- Valor → valor propio de los datos recolectados y analizados para el negocio.

Usos

- UX (Experiencia de Usuario).
- Desarrollo de Productos.
- Fraude y conformidad.
- Eficiencia operacional.
- Mantenimiento predictivo.

Desafíos

- Lidiar con el crecimiento de los datos.
- Generar conocimiento en forma oportuna.
- Reclutar y retener talento de *Big Data*.
- Integrar diferentes fuentes de datos.
- Validación de datos.
- Seguridad.

Tecnologías y Productos

- | | |
|----------------------------|----------------------------------|
| • Inteligencia Artificial. | • R. |
| • Blockchain. | • Data Lakes. |
| • Bases de Datos NoSQL. | • Predictive Analytics. |
| • Python. | • Big Data Governance Solutions. |
| • Hadoop. | • Big Data Security Solutions. |
| • Spark. | • Prescriptive Analytics. |

Seguridad

SEGURIDAD EN BIG DATA

Conjunto de acciones de protección de datos y de procesos de análisis frente a factores que podrían comprometer su confidencialidad e integridad.

- Requiere combinaciones de herramientas de seguridad y procesos inteligentes para monitorear la seguridad a lo largo de la vida de la plataforma.
- Se busca que los datos se enruten a través de un circuito establecido no vulnerable.
- Opera sobre datos de entrada, datos almacenados y datos de salida.
- Tecnologías:
 - Cifrado → protección de datos en tránsito y en reposo.
 - Control de Acceso a Usuarios → configuración de acceso con roles y usuarios.
 - Detección y Prevención de Intrusiones.
 - Seguridad Física → debe considerarse siempre.
 - Gestión Centralizada de Claves → registro de uso, entrega de claves bajo demanda, abstracción de la administración de claves respecto de su uso, etc.
- Los responsables de la seguridad son los DBAs, los programadores, las áreas de calidad y las áreas de *compliance*.

Data Engineering, Data Science, Data Analytics¹

DATA ENGINEERING

Ingeniería que se dedica a superar los cuellos de botella en el procesamiento de datos y los problemas de manejo de datos para aplicaciones que usan *Big Data*.

DATA SCIENCE

Campo multidisciplinario centrado en encontrar información procesable a partir de grandes conjuntos de datos tanto sin procesar como estructurados.

- Focalizada en encontrar respuestas a las cosas que no sabemos que no sabemos.
- Se usan varias técnicas para obtener respuestas, mediante análisis predictivo, ciencias de la computación, estadísticas y machine learning para analizar conjuntos masivos de datos en un esfuerzo por establecer soluciones a problemas que aún no fueron pensados.

DATA ANALYTICS

Centrado en procesar y realizar análisis estadísticos en conjuntos de datos existentes, con el objetivo de mostrar dicha información con una herramienta adecuada.

- Se busca crear métodos para capturar, procesar y organizar datos que permitan tanto descubrir información procesable sobre problemas actuales como establecer la mejor manera de presentar estos datos.
- Busca resolver problemas disparados por preguntas cuyas respuestas aún no conocemos, pudiendo generar resultados que pueden conducir a mejoras inmediatas.

¹ Data Engineering nutre a Data Science y Data Analytics, quienes permiten la toma de decisiones.

Blockchain

BLOCKCHAIN

- Red P2P que no depende de entidades centralizadas para llegar a un consenso.
- Tecnología que permite tener un registro distribuido donde cada par tiene su propia copia de la información, cuya validez y veracidad se establece por consenso entre los pares.

Conceptos Técnicos

- P2P → protocolo de red de comunicación entre pares.
- Algoritmo de Hash → función que transforma cualquier bloque arbitrario de datos en una nueva serie de caracteres con una longitud fija.
- Criptografía asimétrica → sistema que utiliza dos claves (una clave pública para cifrar y una clave privada para descifrar) para el envío de datos.
- Proof of Work o Prueba de Trabajo → algoritmo de consenso basado en operaciones matemáticas complejas y en fuerza bruta, utilizado para la confirmación de las transacciones y la generación de nuevos bloques.
- Consenso → que toda la red esté de acuerdo con el resultado de una prueba.

Principios de Blockchain

- Integridad en la Red → la integridad está cifrada y distribuida en todas las etapas del proceso y no depende de cada individuo.
- Poder Distribuido → el sistema distribuye el poder por una red de iguales sin que haya ningún punto de control → las partes no pueden apagar el sistema por sí solas.
- El Valor como Incentivo → el sistema alinea los incentivos de los *stakeholders* y sus intereses.
- Seguridad → se garantiza la confidencialidad y la autenticidad de todas las actividades.
- Privacidad → se elimina la necesidad de conocer la verdadera identidad de los pares.
- Preservación de Derechos → los derechos de propiedad están garantizados.
- Inclusión → la economía funciona mejor cuando funciona para todos.

Pasos

1. Un usuario solicita una transacción.
2. Se crea un bloque que representa esa transacción.
3. El bloque se difunde a todos los nodos de la red.
4. Dichos nodos validan el bloque y, así, también validan la transacción.
5. El bloque se añade a la cadena.
6. La transacción se verifica y se ejecuta → la transacción ya no se podrá modificar.

Usos

- Cualquiera donde se quiera asegurar que la información no se altere.
- Contratos.
- Libros contables y libros de registros.
- *Tokenización* de activos.
- Registros automotores.
- Historias clínicas.
- Escrituras de propiedades.

Ventajas

- Naturaleza distribuida → no hay ningún nodo privilegiado, ningún nodo puede apagar la red, no tiene un punto único de falla.
- Estabilidad → es muy poco probable que los bloques confirmados sean revertidos.
- Sistema *Trustless* → no requiere de confianza entre terceros.
- Incorruptible → es casi imposible alterar la información contenida en los bloques.
- Transparencia → cualquiera puede consultar las transacciones en el registro y verificarlas.
- Trazabilidad → la información se puede rastrear de manera sencilla y su historial se comprueba constantemente.
- Libre de Errores → los resultados siempre son comprobados y correctos.

Desventajas

- Ataques del 51% → si un número suficiente de nodos se hace con el control de más del 50%, será la mayoría, pudiendo así crear *nuevas verdades* y *una nueva realidad*.
- Claves privadas → perder la clave privada implica no tener acceso a la información.
- Ineficiente → las *blockchains*, en especial las que usan *proof of work*, usan mucha energía debido al gran procesamiento por la fuerza bruta empleada.
- Almacenamiento → la cantidad de información de cada copia de la *blockchain* es inmensa.
- Apertura → cualquiera puede consultar datos presentes/históricos de cualquier otro.

Posibles Riesgos o Ataques²

- Doble gasto → una misma moneda digital puede gastarse más de una vez.
- Redes fantasmas → red impostora donde se realicen operaciones falsas.

² El protocolo de *blockchain* soluciona ambos ataques.

BaaS

BaaS • BLOCKCHAIN AS A SERVICE

- *Blockchain* pensado como un servicio.
 - Redes basadas en la nube por parte de terceros para empresas en el negocio de la creación de aplicaciones *blockchain*.
- Permiten la integración con contratos inteligentes.
 - Permiten la integración con plataformas que aseguran identidad.
 - Permiten poder trabajar con distintas implementaciones de *blockchain*.
 - Ofrecen mecanismos de consenso basados en la identidad.

BFA • Blockchain Federal Argentina

BFA es una plataforma multiservicios abierta y participativa pensada para integrar servicios y aplicaciones sobre *blockchain* dentro de Argentina.

Características:

- Sin criptomoneda → la plataforma no está pensada para criptomonedas.
- Modelo liviano → no usa fuerza bruta (el minado) para obtener recompensas.
- Permissionada → el consenso se logra porque los integrantes se conocen (no son anónimos).
- Transacciones gratuitas → las transferencias no tienen costo.
- Almacenamiento en off-chain → no se almacena información per se dentro de la *blockchain*, sino los hashes de esa información.
- Software libre → el SW se basa en una implementación abierta.