



Rich it

Grandes Datos y Ciencia de Datos

1. Propósito
2. Notas Introductorias
3. Metas
4. Preguntas Principales
5. Roles
6. Técnicas, Herramientas y Métricas
7. Áreas de Proceso Relacionadas
8. Productos de Trabajo
9. Niveles de Madurez
10. Identificación de Riesgos



Rich it

Niveles de Madurez RDM²

Grandes Datos y Ciencia de Datos

1. Propósito

La recopilación (Big Data) y el análisis (Data Science, Analytics and Visualization) de muchos tipos de datos para encontrar respuestas e ideas para preguntas que no se conocen al inicio del análisis.

2. Notas Introductorias

- ◆ Definir la estrategia de grandes datos, ciencia de datos (analítica) y las necesidades del negocio.
- ◆ Adquirir e ingestar las fuentes de datos para su análisis.
- ◆ Seleccionar correctamente las fuentes de datos apropiadas.
- ◆ Desarrollar una hipótesis y los posibles métodos para llegar a una respuesta.
- ◆ Integrar / alinear los datos para su análisis.
- ◆ Explorar datos usando modelos

3. Metas

- Descubrir relaciones entre datos y negocios
- Soporte a la integración interactiva de fuentes de datos en la empresa.
- Descubrir y analizar nuevos factores que podrían afectar al negocio.
- Publicar datos utilizando técnicas de visualización adecuadas, confiables y éticas.

4. Preguntas Principales

1

¿Se maneja alguna metodología (Ciclo de Vida) para la ciencia de datos (CRISP-DM o KDD)?

2

¿Los modelos analíticos y los datos resuelven las preguntas del negocio?

3

¿Cómo se evalúan las métricas de salida (outputs) de los modelos analíticos?

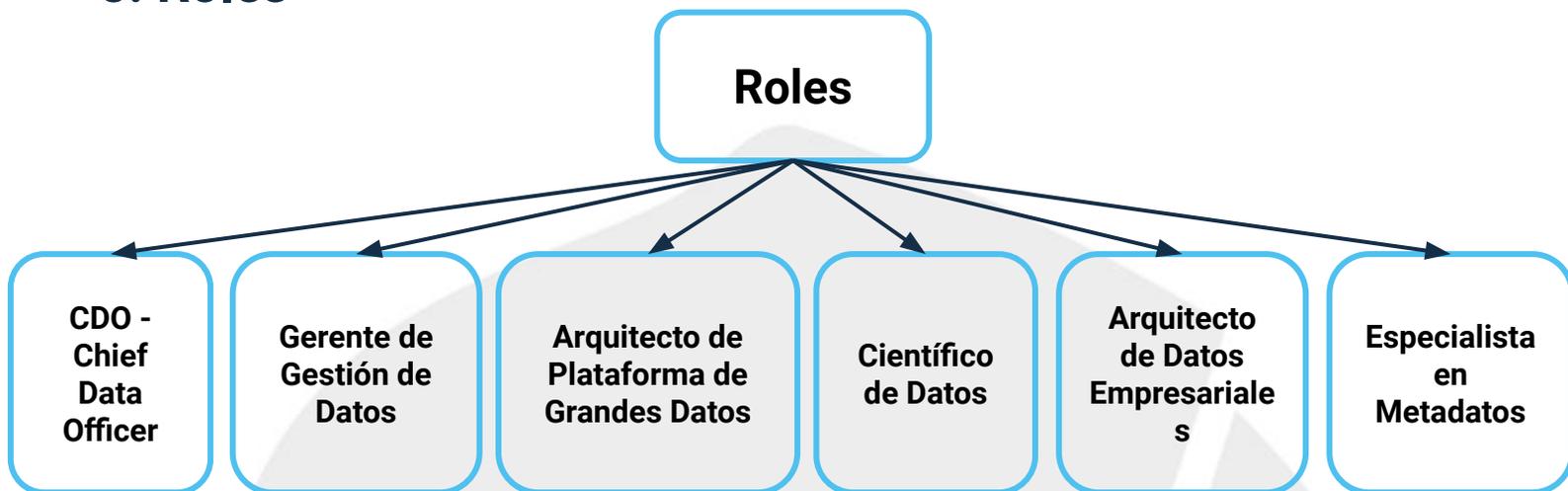
4

¿Cada cuanto tiempo se reentrenan los modelos analíticos?

5

¿Cómo se maneja la seguridad y calidad de los datos?

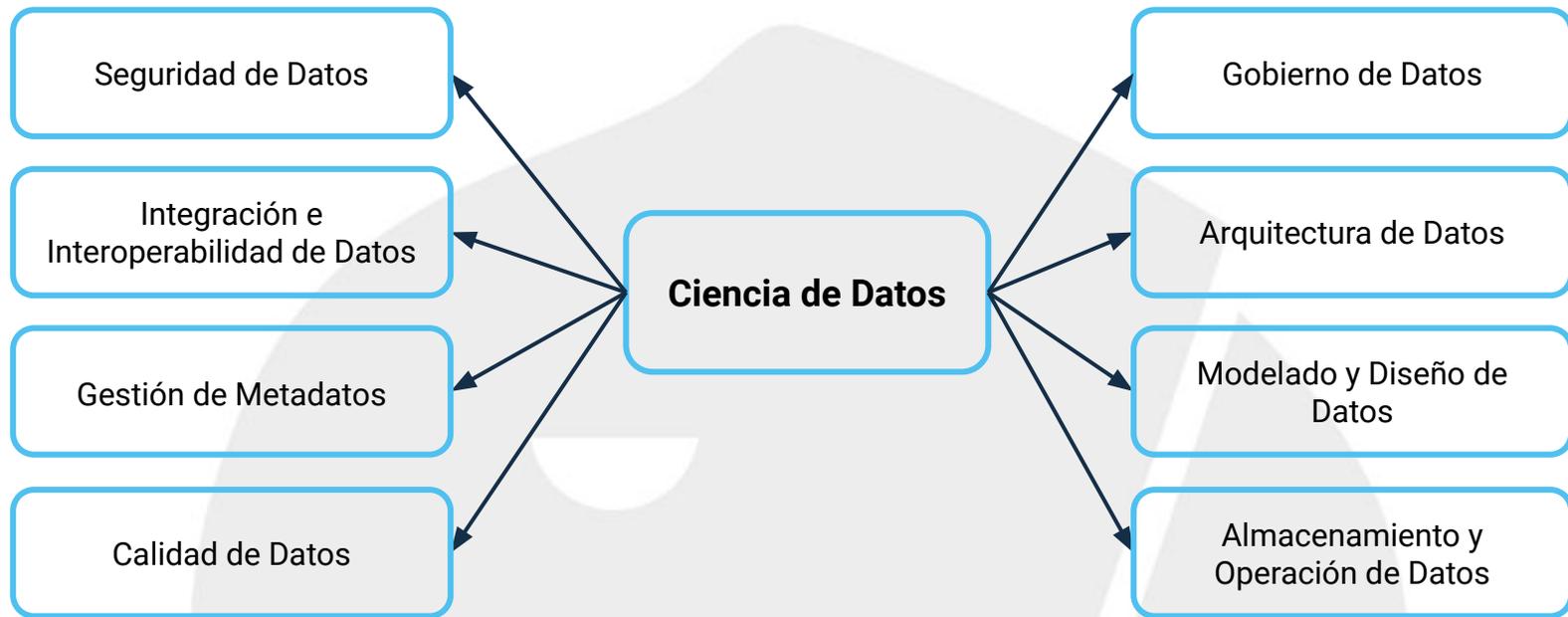
5. Roles



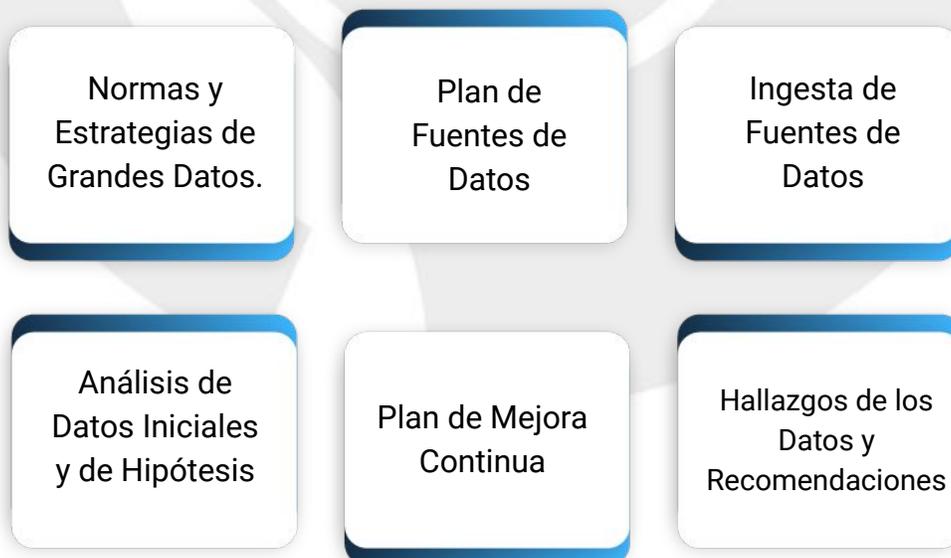
6. Técnicas / Herramientas / Métricas

Técnicas	Herramientas	Métricas
Mezcla y/o Combinación de Fuentes de Datos (Integración de un Datawarehouse)	Soluciones Distribuidas Basadas en Archivos (Hadoop)	Métricas de Uso de Datos
Técnicas de Machine Learning (Machine Learning)	Compresión Columnar (Hadoop, Data Lake)	Métricas de Respuesta y Desempeño
Aprendizaje Supervisado Avanzado (Deep Learning, Data Mining, Text Mining)	MPP Arquitecturas Shared-Nothing	Métricas de Carga de Datos y Escaneo (Exploración)
	Computación en Memoria y Bases de Datos	Historias y Aprendizaje
	Algoritmos en la Base de Datos (SQL Server, PostgreSQL, MongoDB)	
	Conjuntos de Herramientas de Visualización de Datos (Kibana, Grafana, Power BI,	

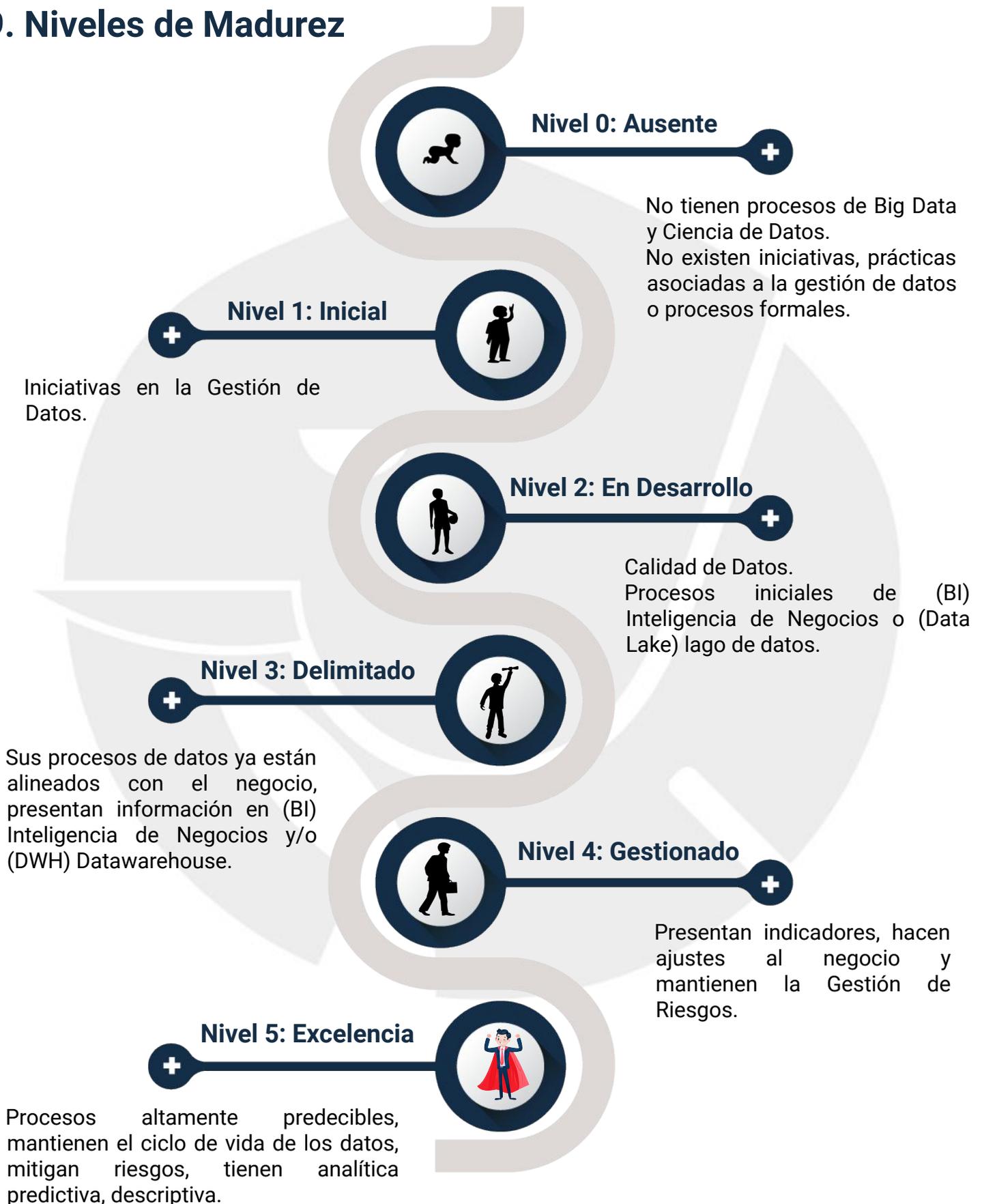
7. Áreas de Proceso Relacionadas



8. Productos de Trabajo



9. Niveles de Madurez



10. Identificación de Riesgos



El manejo de diferentes identificadores para una misma identidad en distintos sistemas impide contar con la trazabilidad de todas las operaciones de dicha identidad en la organización.



Esto incurre en la pérdida de oportunidades de brindar un mejor servicio, operar de forma más eficiente y nuevas oportunidades de negocio.



Al no contar con datos maestros gestionados, las áreas de negocio duplican esfuerzos y actividades debido a la multiplicidad de datos y fuentes.

Acerca de RICH IT

RICH IT es un fabricante de plataforma analítica con 9 años de experiencia en el procesamiento y extracción de valor de los datos.

Cuenta con uno de los equipos más robustos de ingeniería y ciencia de datos en América Latina.

Soporta proyectos de misión crítica y seguridad nacional.

Las metodologías y plataformas de RICH IT acortan hasta un 80% los tiempos de desarrollo de productos analíticos.



Rich it

¿Te gustaría agendar una videollamada para hablar sobre tus proyectos y ver cómo podemos ayudarte?

Agendar una llamada

contacto@richit.ai

richit.ai