

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 de Grandes Datos (Big Data), no solo tiene que ver con su volumen, variedad y velocidad sino también con otras variables como la veracidad y el valor que se puede obtener de ellos actualmente.

La Ciencia de Datos (Data Science, Analytics and Visualization) tiene que ver con el conjunto de habilidades que permiten extraer conocimiento a partir de los datos, mediante el desarrollo de modelos analíticos, predictivos, de aprendizaje automático y prescriptivos, con el objetivo de enriquecer el análisis descriptivo tradicional para la toma de decisiones de manera ágil y efectiva.

2. Notas Introductorias

Los negocios buscan desarrollar sus capacidades alrededor de Grandes Datos y Ciencia de Datos con el deseo de encontrar y actuar sobre nuevas oportunidades estratégicas. Esta área de conocimiento promueve la innovación haciendo que grandes conjuntos de datos estén disponibles para su explotación.

3. Metas

- Aprovechar las relaciones entre los datos corporativos y los procesos de negocio para obtener resultados favorables
- Proveer mecanismos para la explotación de datos y generación de insights de forma escalable e iterativa en la organización.
- Publicar datos utilizando técnicas de visualización adecuadas, confiables y éticas.

4. Preguntas Principales

1

¿Se maneja alguna metodología para el desarrollo de iniciativas de ciencia de datos (CRISP-DM, KDD, entre otras)?

2

¿Considera que el estado actual de los datos permite resolver las preguntas de negocio a través de diferentes tipos de análisis?

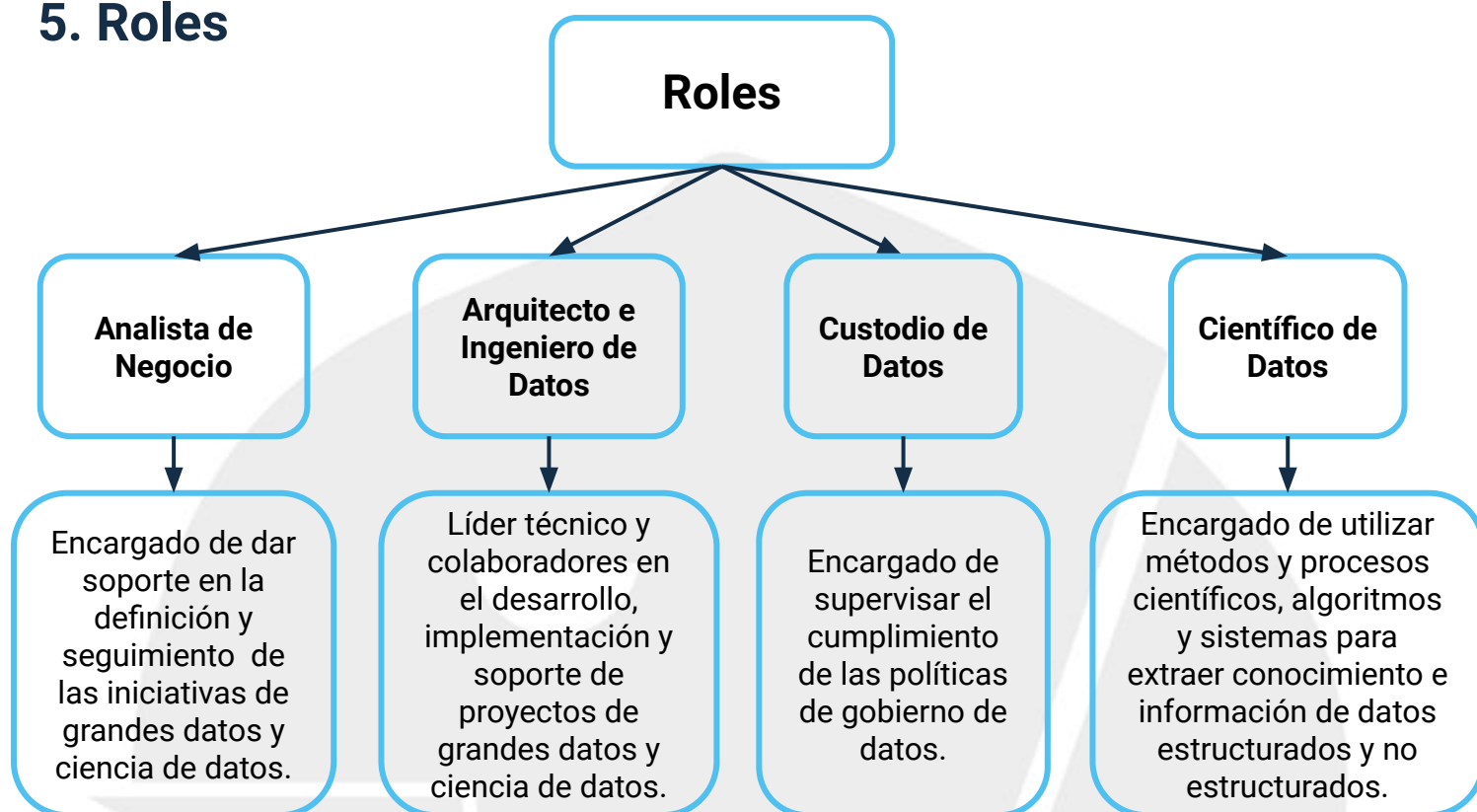
3

¿Considera que los procesos de análisis de datos (reportes, modelos, tableros, etc) implementados en la organización permiten tomar decisiones de negocio en el momento que se requiere?

4

¿Cuentan con planes de mejora continua, mantenimiento y soporte para la actualización frecuente de tableros o reentrenamiento de modelos analíticos?

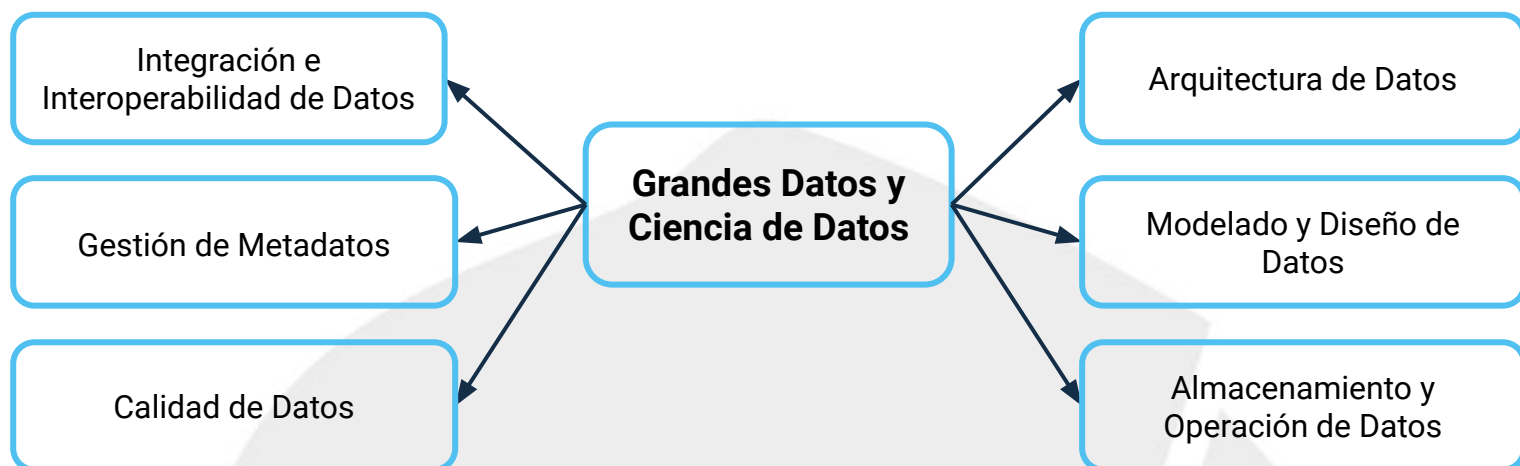
5. Roles



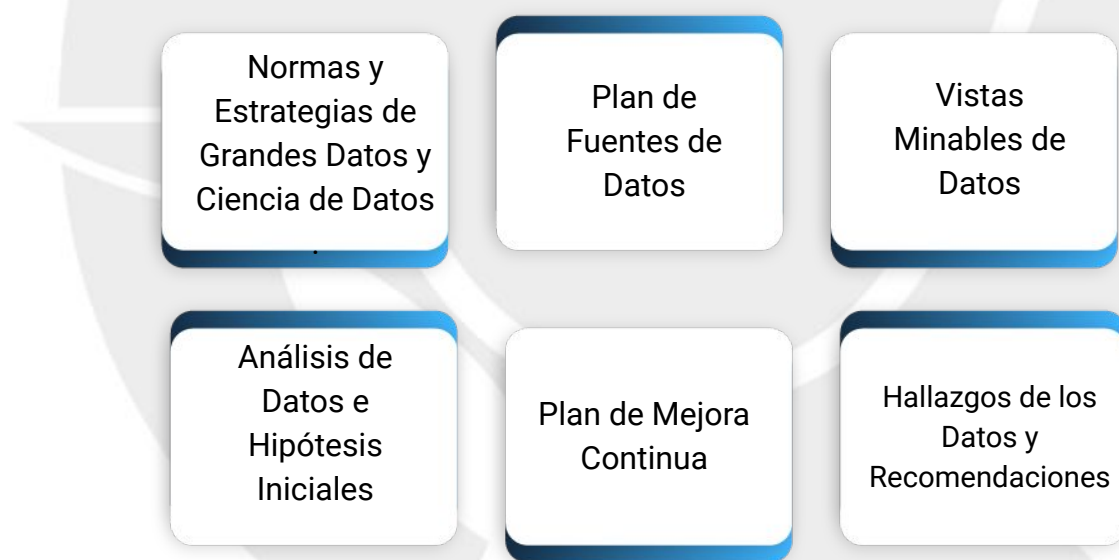
6. Técnicas / Herramientas / Métricas

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 aprendizaje de máquina automática (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)
	Algoritmos en la base de datos (SQL server, postgresQL, mongoDB)	Métricas de confiabilidad y precisión de los modelos
	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



Brindar resultados erróneos debido al análisis de conjuntos de datos sesgados o mal procesados.



Falta de disponibilidad de herramientas o recursos para la obtención, análisis y desarrollo de modelos analíticos.



Aumento de costos en términos de tiempo y recursos en el desarrollo de iniciativas que no responden a las preguntas claves del negocio.

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.