

# ESTRATEGIA DE DATOS EMPRESARIAL:

UNA GUÍA RÁPIDA PARA CREAR  
CAPACIDADES DE DATOS

KARIM TOUMA

# Estrategia de datos empresarial

Karim Touma

Este libro está a la venta en <http://leanpub.com/estrategiadedatos>

Esta versión se publicó en 2020-07-03

ISBN 978-956-401-937-6



Éste es un libro de [Leanpub](#). Leanpub anima a los autores y publicadoras con el proceso de publicación. [Lean Publishing](#) es el acto de publicar un libro en progreso usando herramientas sencillas y muchas iteraciones para obtener retroalimentación del lector hasta conseguir el libro adecuado.

© 2020 Karim Touma

# Índice general

|  |           |
|--|-----------|
| <b>1. Introducción</b>   | <b>1</b>  |
| <b>2. Evaluación</b>   | <b>3</b>  |
| <b>3. Definición de Estrategia</b>   | <b>9</b>  |
| <b>4. Estrategia Táctico-Fundacional</b>                                     | <b>11</b> |
| 4.1. Equipo táctico: la fábrica de casos de uso                              | 13        |
| 4.2. Equipo fundacional: las bases tecnológicas y la gobernanza de los datos | 15        |
| 4.3. Un modelo de evolución  | 17        |
| <b>5. Gestión del conocimiento</b>   | <b>19</b> |
| 5.1. Empresarial: gobierno de datos  | 20        |
| 5.2. Equipos: ciencia de datos   | 22        |
| <b>6. Gestión de los equipos</b>   | <b>23</b> |
| 6.1. Configuración de Equipos  | 24        |
| 6.2. Evolución y desarrollo de plan de carrera                               | 26        |
| 6.3. Agilidad en equipos de datos  | 27        |
| <b>7. Tecnología</b>   | <b>28</b> |
| 7.1. Diferenciar inversiones en tecnología                                   | 29        |
| 7.2. Estrategia de adopción  | 31        |
| <b>8. Principales puntos de dolor</b>  | <b>32</b> |
| 8.1. Calidad de datos  | 33        |
| 8.2. Como encarar un problema de ciencia de datos                            | 35        |
| 8.3. Cultura   | 37        |
| <b>9. Conclusiones y cómo retroalimentar la estrategia</b>                   | <b>38</b> |

# 1. Introducción

Este ensayo intentará proveer una guía práctica para encarar los diferentes problemas por los que pasan las empresas en sus travesías por el mundo de los datos, desde la concepción de una estrategia adecuada hasta cómo embeber sofisticadas tecnologías de inteligencia artificial en sus procesos. Todo tratado desde un punto de vista pragmático y con espíritu de divulgación. Por lo general, los temas que involucran tecnologías emergentes son tratados con bastante sutileza por el mundo académico y con exacerbado cripticismo por expertos, en esta oportunidad intentaremos lo contrario, por ende se involucrará conceptos conocidos y analogías cuando estas sean necesarias para representar definiciones, conjeturas o experiencias.

En la actualidad, la dimensión de datos en las empresas está tomando cada vez más relevancia. Muchas veces, comienza como resultado de evaluaciones posconsultoría o, simplemente, como lineamientos de directorios, los cuales conceptualizan esta nueva dimensión del negocio como una mera lista de requisitos (algo que se considera deseable como una nueva capacidad) o como la bala de plata para enfrentar problemas de toda índole. En estas concepciones, muchas veces se sobreestima la capacidad de evolucionar digitalmente una compañía, así como la carga de sistemas legados o antiguos que cada compañía pueda poseer. Evolucionar digitalmente una start-up o un consorcio no tradicional de varias décadas, requiere de estrategias de ataque para solucionar problemas completamente diferentes.

Para que este viaje tenga sentido, es necesario demostrar el valor futuro de las iniciativas desde el primer momento de su implementación. Por esta razón, es que en el presente texto se utilizará un enfoque pragmático en función a la “distancia” entre los datos y la cadena de valor. La máxima de la construcción se basa en entregas incrementales y alineadas con los objetivos de la compañía para construir la nueva capacidad en forma que sustente valor a mediano y largo plazo.

Las siguientes ideas se encuentran distribuidas en nueve capítulos, los cuales abordan desde lo macro a lo micro en todo lo relativo a la implantación de las capacidades de datos en cualquier empresa, con principal énfasis en compañías no digitales (podríamos decir que para compañías digitales este trabajo se hace un tanto más fácil). El primer capítulo tiene por objetivo la revisión de diferentes metodologías y claves para lograr levantar la realidad empresarial desde un punto de vista holístico. Por su parte, los dos siguientes capítulos tienen por finalidad representar la etapa de planificación estratégica y diseño de las iniciativas, nociones sobre cómo empezar y dar sustento a las inversiones tanto de capital como de recurso humano que deben involucrarse en el desarrollo del proyecto. Del cuarto al sexto capítulo se detallan temas de gestión en sus diferentes dimensiones centradas en el mundo de los datos, y finalizamos con posibles problemas y conclusiones, en los últimos tres capítulos.

Los temas se abordarán de forma pragmática, considerando las expresiones y opiniones como una guía general, asimismo, se ahondará en los temas independientes transversales a la estrategia de datos, muchas veces privilegiando mantenerse fuera del contexto de la empresa, dado que se asume

que el lector poseerá muchísima más experiencia que el escritor en su determinado dominio de negocio.

## 2. Evaluación

Como dijo Alfred Korzybski: “El mapa no es el territorio”. Esto aplicado al mundo de las empresas, la complejidad tras el quehacer diario es extensa y variada. Para navegar la organización es necesario entender que la variabilidad aguas abajo en los equipos que ejecutan se traduce en mayor abstracción aguas arriba. Esto es especialmente confuso en empresas grandes, pero, concretamente en las compañías, los procesos son complejos, poco eficientes y, por lo general, poco o mal documentados. Las grandes excepciones a esta regla tienen que ver con empresas altamente reguladas, en las que esa misma regulación, obliga a configurar procesos más simples o a sobreinvertir en herramientas o equipo para regulación.

Considerando a los datos en el centro, podríamos dividir, a grosso modo, las diferentes compañías en seis grandes grupos:

- industriales (fábricas, minería, energía, petróleo, gas, etc.)
- venta al por menor (retail de todo tipo)
- tecnología, medios de comunicación y telecomunicaciones
- instituciones financieras
- cuidado de la salud
- gobierno y fuerzas armadas.

La anterior es una taxonomía simple de los diferentes niveles de industrias, agrupados por rubro. Posteriormente, podríamos decir que la competitividad y el nivel de digitalización de cada empresa dentro del clúster también proveerá luces sobre cómo realizar la evaluación sobre la madurez de datos en la compañía. La lista exhaustiva de cada característica sería la siguiente:

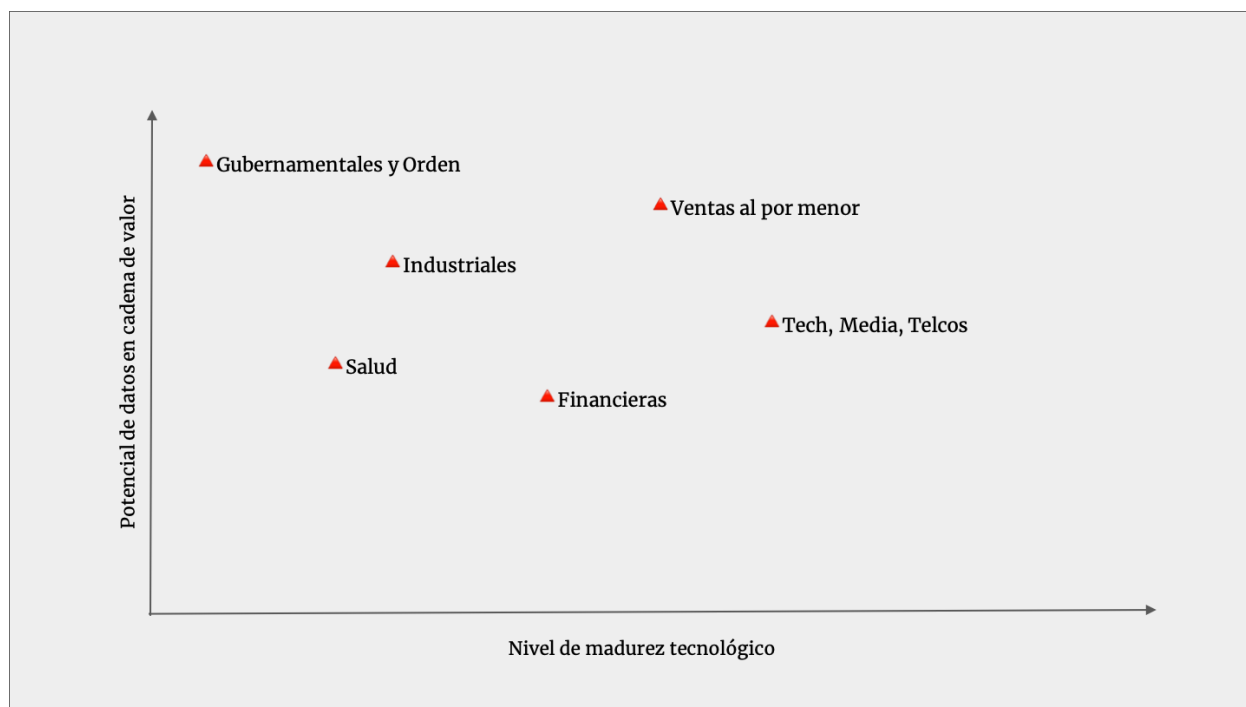


Figura 1. Madurez tecnológica y Potencial de datos en la cadena de valor.

1. Industriales, Energía, Petróleo, Gas: las compañías industriales son altamente demandantes en procesos productivos, por ende, los principales aspectos de evaluación involucran el estado del arte de la sensorización, pues esta será la principal fuente de información para los diferentes casos de uso para sustentar la estrategia de datos. Otra parte importante tiene que ver con todos los activos financieros, los cuales representan oportunidades para evaluar el nivel de evolución digital de las áreas de control de gestión y financieras. Por lo general, los sistemas informáticos en este rubro son sistemas cerrados que tienen bastantes dificultades para integrar y entregar datos. Las áreas de tecnología, al estar bastante alejadas del núcleo del negocio, encuentran dificultades para implantar metodologías disruptivas o de modernización.

2. Venta al por menor: las ventas al por menor son un caso interesante. Existe muchísima variedad, desde las más digitales hasta las más parecidas a empresas industriales. Así también, el mundo de la cadena de abastecimiento ocupa más o menos preponderancia según cuanta integración vertical se vea en la empresa. Las áreas de tecnología, por lo general, se encuentran divididas y existen áreas dedicadas a la innovación, en donde es frecuente encontrar pequeños silos digitales que impulsan el carro de la transformación digital. Lo importante en el levantamiento en este tipo de compañías es identificar tanto las áreas de innovación como de sistemas. Lo importante es lograr identificar los sistemas críticos (de venta, de facturación, proveedores, vendedores), así como dominios importantes (comercio online y todo lo relativo a la función de marketing digital) y los actores del área de innovación. Al ser industrias altamente competitivas, es común observar iniciativas de transformación, esas serán las principales aliadas para consolidar la estrategia.

3. Tecnología, Medios de comunicación y Telecomunicaciones: industrias en donde la tecnología está

en el núcleo del negocio, pero no necesariamente democratizada en todas las áreas de la compañía. Lo interesante al observar este tipo de organizaciones es el nivel de evolución tecnológica en la principal área tecnológica: redes en las telcos, ingeniería en las empresas tecnológicas y marketing digital en los medios. Luego de identificar esa área principal y su nivel tecnológico, se hace necesario levantar el estado de los procesos de cara a cliente y las iniciativas de marketing. Es importante relevar el estado de madurez tecnológico de los sistemas con principal énfasis en identificar porcentaje de sistemas cerrados y nivel de aislamiento de las áreas de gestión y tecnológicas.

4. Instituciones Financieras: gracias a la regulación, las empresas financieras requieren tener algún grado de orden en sus sistemas de información. Este orden no es ampliamente compartido por todas las unidades, pero por lo menos en las principales áreas de venta, marketing y riesgo, existirá trazabilidad y claridad sobre la información. Es posible que los sistemas núcleo sean bastante anticuados, por ende, los principales problemas en este tipo de empresas es la logística de datos.

5. Cuidado de salud: las empresas del rubro de la salud tienen complejidades en el manejo de información de clientes, los cuales, desde el punto de vista de los datos, se encuentran resguardados en un esquema de privacidad extrema. Por eso, existe una clara separación entre los mundos de procedimientos y clientes, por sobre el esquema de manejo tradicional de marketing, operaciones, entre otros.

6. Gobierno y Fuerzas armadas: este tipo de organizaciones son, posiblemente, de las más burocráticas y reguladas. El gran desafío en instituciones que se mueven lento y que no gozan de requerimientos de eficiencia o especialización técnica es la transformación digital. Lo ideal en este tipo de empresas es enfocar la misión y visión de datos en construir un intraemprendimiento en algún subconjunto o nueva misión de la organización, como la creación de empresas adjuntas o nuevas unidades que cuenten con respaldo a nivel de proyecto para funcionar independientemente.

Las descripciones anteriores solo representan una vaga opinión superficial que puede o no servir de guía, simplemente trato de enumerar algunos conceptos generales sobre una guía de estructura lógica que describe los posibles problemas entre los niveles de digitalización y madurez en sistemas de información. En lo referente a industrias se asume que el lector tendrá una mayor comprensión sobre su determinado rubro.

Continuando con el tema, existen cuatro componentes estratégicos que describen el estado de madurez de cualquier empresa y su ecosistema de datos:



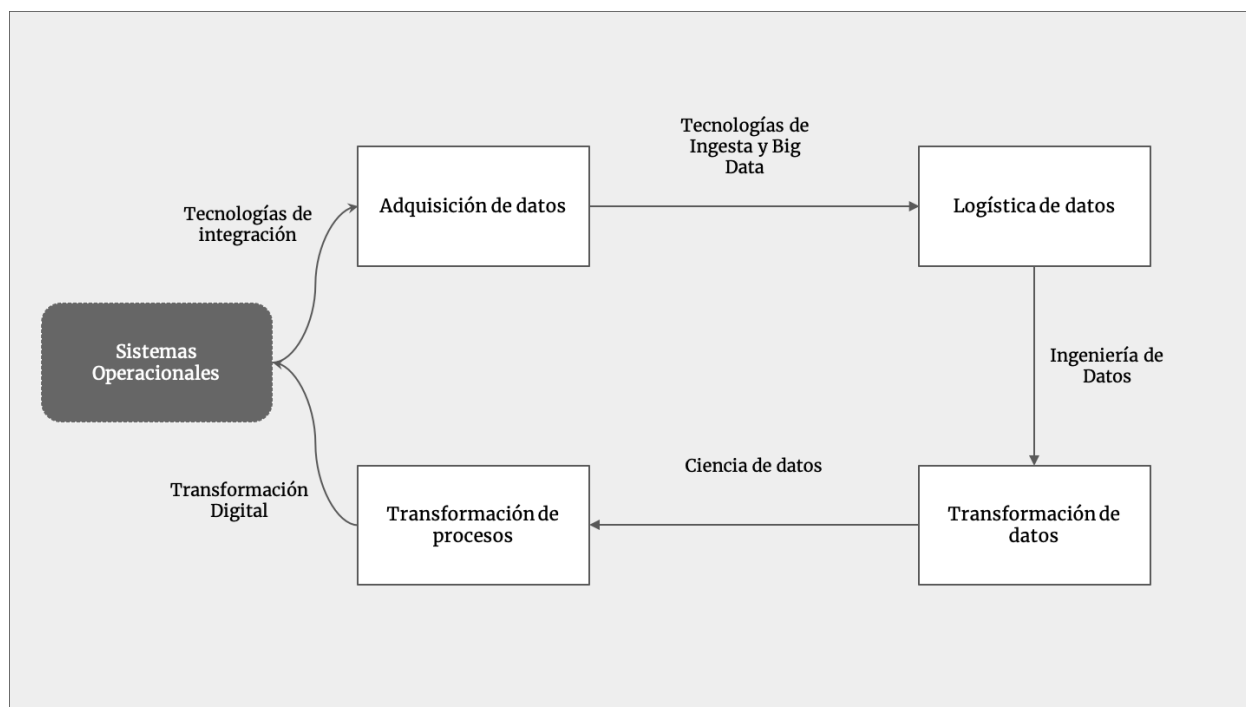


Figura 2. Componentes estratégicas y su interrelación tecnológica.

- **Adquisición de datos:** la capacidad de acceder, obtener, capturar o redireccionar los datos que generan los procesos de la compañía, ya sea desde sistemas de información digitalizados, manuales o sensorizados. Es importante mencionar que las capacidades de adquisición dependen de la complejidad de la información. Una compañía que procesa data volumétrica, de alta frecuencia de actualización o de una alta variedad puede requerir tecnologías avanzadas para acceder a la información o para liberar la información de sus sistemas.

Otra barrera de adquisición relevante es la regulación, la que puede mitigar cualquier posibilidad de mejora hasta sortear los requerimientos regulatorios, legales, éticos o morales para acceder a los datos. Por ende, para el levantamiento de esta componente es necesario cubrir las principales tipologías de datos con su par en habilitador tecnológico, para esto ocuparemos la más simple estructura de velocidad de generación, volumen de generación, variedad de generación y regulación con sus consecuentes habilitadores tecnológicos. Aplicando la Regla de Pareto, lo más probable es que las principales fuentes de información en el sentido estricto vivan en el 20% de los sistemas de la compañía, aquellos serán los primeros que deberían habilitar en su sentido de acceso.

- **Logística de datos:** movilizar la data toma en consideración que la explotación de la información no ocurre en el contexto de la generación, es decir, que el sistema que resuelve las necesidades de negocio soporta operatoria de datos contextual al proceso, más no se encarga de relacionar esos datos y que ese interrelacionamiento con información de otros contextos es lo que enriquece la toma de decisiones de la empresa como un todo. Es por esto que se debe mover

la información a unidades supercontextuales donde se puedan analizar; estas pueden ser almacenes de datos, lagos de datos, lagos de lagos de datos y así sucesivamente (las plataformas de procesamiento y explotación las veremos en el apartado de tecnología). Movilizar estos datos, en tiempo y forma, requiere de capacidades tecnológicas en directa relación al mundo de generación. Muchas capacidades de movilización crecen en complejidad en función al posible uso o problemáticas en su logística, como pueden ser temas de conectividad, formatos, velocidades o volumen de información. Es por esto que muchas empresas optan por soluciones ad hoc para el tema de la logística de datos, y lo consideran como una función más dentro de la compañía, la cual se encuentra circunscrita en la primera fase de la “Ingeniería de Datos”. Los principales puntos en la etapa de descubrimiento tienen que ver con las capacidades de la compañía para movilizar los datos, por ende, lo relevante son las capacidades de conectividad, la cantidad de “silos” de información, así como la antigüedad promedio de los sistemas.

- Transformación de datos (modelamiento): en la transformación es donde ocurre la magia. Existen una infinidad de procesos dentro de las transformaciones de datos e incluyen acciones tan variadas- desde limpieza de datos incongruentes hasta modelos analíticos. En la fase de levantamiento es importante recabar información del estado de documentación sobre las fuentes y transformaciones de datos, el nivel regulatorio actual y el nivel de documentación de los procesos de la compañía que utilizan datos. La gestión de conocimiento de la compañía es la base fundacional de todas las transformaciones de datos y representan la mejor interpretación o reglas de facto para la toma actual de decisiones. Entre mayor el nivel de documentación de los datos en la compañía, mayores serán las capacidades de escalar el conocimiento en analítica avanzada.
- Transformación de procesos de Negocio (acción con datos): una vez terminados los procesos de transformación, es relevante contar con capacidades para accionar estos nuevos activos. Entre mayor el nivel de digitalización en la punta de la acción, mayor será la capacidad de los equipos de data de tomar relevancia y cobertura sobre los procesos de la compañía, por contrario, si se abusa de integraciones o procesos manuales, se coarta la capacidad de esta nueva información para mejorar los nuevos procesos.

Cabe mencionar que si estos componentes se encuentran encapsulados en algún sistema que contiene plataformas de procesamiento de datos en el interior del proceso de negocio, coarta las posibilidades exponenciales para acceder a los insumos de la información. Por ejemplo, en una plataforma de telefonía, que realiza la tasación del tráfico móvil, puedes obtener señales sobre cuándo se acaba el tráfico para cada cliente, por ende, otro sistema ocupa esta señal para cortar el plan del cliente.

Esto en particular es una plataforma operacional dedicada que tiene un fin de cara al servicio de cada cliente, en cambio, si se lograra intervenir el ciclo de procesamiento para esta plataforma, y, dado el caso, realizar un análisis dedicado a los consumos de cada cliente, sería posible anticiparse a los consumos o patrones de tráfico, y alertar al usuario sobre consumos fuera de su patrón habitual, lograríamos evolucionar por completo una plataforma de tasación que cumple una función de “bien primario” hacia una plataforma de optimización de experiencia cliente. Estas consideraciones deben estar presentes al definir qué sistema deberá intervenir para extraer los insumos o subproductos de datos que este genera. Actualmente, muchos sistemas que solucionan problemas estandarizados

en las empresas cuentan con capacidades para obtener los datos y procesarlos para mecanismos de diferenciación o innovación, lo cual justifica también, inversiones en modernización de tales plataformas. Ahora, es decisión del caso de negocio y de la voluntad de la empresa decidir cuáles serán las primeras batallas, dado que en la planificación de desarrollo sobre estas plataformas, también puede existir el desarrollo de este tipos de casos de uso avanzados, que reciben como insumo información de la compañía para mejorar la experiencia de tráfico.

Como verán, existe una eterna batalla entre plataformas de operación en obtener capacidades de diferenciación, así como de compañías que quieren tener control de sus capacidades de operación para escalar más rápido la diferenciación. La clave para elegir este camino no es otra más que la capacidad que tiene una empresa para atacar problemas de evolución tecnológica. En un sentido pragmático, es decidir si se quiere invertir en talento digital o en grandes inversiones en soluciones que cuenten con características que suplan estas necesidades, la competitividad del sector, así como la capacidad real de generar capacidades técnicas, son unas de las principales palancas para tomar esta importante decisión, que muy probablemente la mayoría de las organizaciones de América Latina ni siquiera se han empezado a cuestionar.

### 3. Definición de Estrategia

La estrategia de datos considera la definición, abordaje, así como consideraciones técnicas y funcionales. Para simplificar este despliegue, es necesario diseñar la evolución, creación o consolidación de las capacidades de datos para la empresa, lo que abordaremos en este capítulo. La primera gran pregunta -¿cuál es el dolor que va a atacar la estrategia?- divide a las empresas en dos grandes agrupaciones: empresas con estrategias de datos defensivas y ofensivas. Posterior a la evaluación, sería posible encasillar el modo en función a las características del negocio, nivel de competitividad y nivel de madurez tecnológico, incluso la evaluación de contextos dentro de mundos complejos. Luego, a esta definición es necesario definir los roles de liderazgo necesarios para empujar cada estrategia, en donde las capacidades de datos se deben mezclar con el conocimiento al interior de la compañía.

Para simplificar la infinita variedad de posibles diseños, los englobamos en tres grandes grupos. Se definirán sus principales características, para que el lector defina cuál se adecua mejor a su realidad y posteriormente, a través de los siguientes capítulos, definir en qué partes relajará o altera su ejecución y profundidad. Las estrategias no son mutuamente excluyentes ni exhaustivamente complementarias, es decir, debe considerarse que la estrategia es el pilar que sostiene la evolución, pero no el camino que traza el rumbo, las competencias son complementarias.

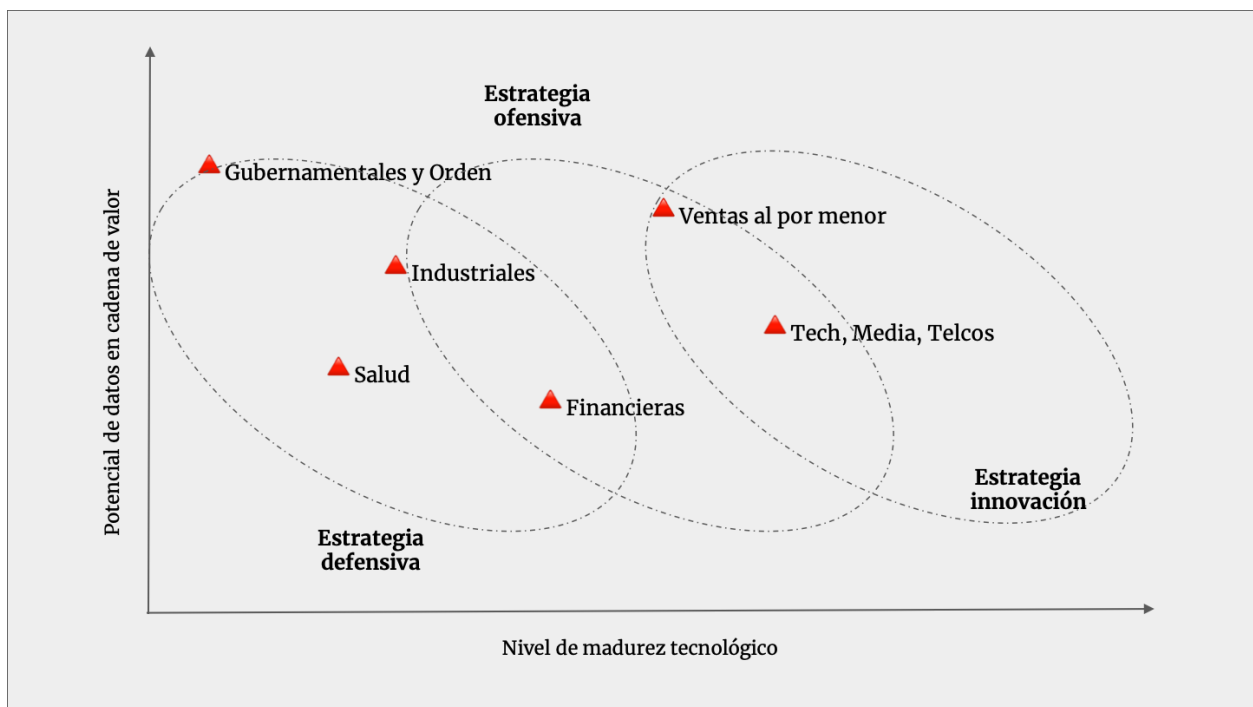


Figura 3. Diferentes estrategias y su adaptación al contexto industria.

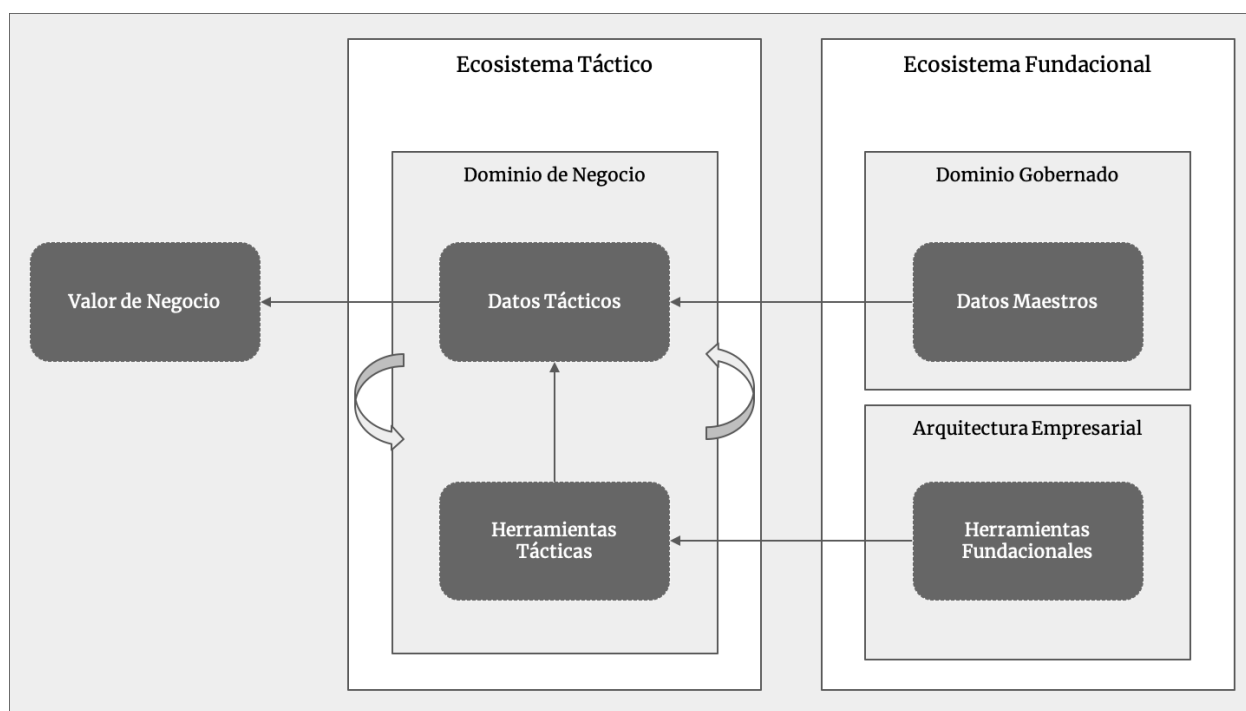
- **Estrategia Defensiva:** en esta estrategia, lo primordial es utilizar los datos para asegurar la sostenibilidad del negocio, en modelos que dependen fuertemente de regulación, normativa o seguridad de la información. El foco en esta estrategia es construir capacidades de seguimiento exhaustivo de activos de datos a través de mecanismos de arquitectura, gobernanza y sistemas de información, es una metodología basada en procesos, con poco espacio para innovación o diferenciación. Por lo general esta estrategia se ocupa para empresas con alto riesgo de operación como alta industria, minería, gas, y alta intensidad de gastos de capital. En empresas en donde los datos tienen implicancias regulatorias esta estrategia puede ser la manera de ejecutar gobierno, pero no quiere decir que se obvian las innovaciones o analítica avanzada, existen problemáticas defensivas que si requieren modelos de machine learning, es decir, donde la madurez de las soluciones no recoge todas las implicancias. Esto quiere decir que muchas veces las dificultades pueden ser abordadas mediante soluciones analíticas cuando una metodología existente no contiene soluciones para problemas propios de gestión, manejo y consistencia de datos.
- **Estrategia Ofensiva:** considera una orientación pragmática donde el valor de los datos se construye mediante iteraciones de aplicaciones de datos en el negocio. La mejora operacional, optimización de mercadeo, reducción de gastos, mejora en las inversiones, son algunas de las verticales de ataque, en las que todos los objetivos de desarrollo están pragmáticamente orientados a mejorar los resultados de negocio de la compañía. Puede resultar compleja esta estrategia si no existen, o están en construcción, las capacidades para monetizar los nuevos activos de datos que los equipos de datos construyan, por lo tanto, en una estrategia ofensiva es necesario considerar mayores prerequisites en su establecimiento, dado que los focos de ataque pueden ser tan variados o complejos como los focos de la matriz. Por eso, muchas veces es necesario compensar o sobreinvertir en la aceleración de la compañía como un todo.
- **Estrategia de Innovación:** esta estrategia resulta ser útil para organizaciones grandes que poseen dificultades para formar nuevas capacidades dada su inercia, tamaño y extrema complejidad. Se encuentra frecuentemente en adquisiciones de empresas especializadas, creación de innovación interempresa, formación de centros de competencias o centros de investigación y desarrollo. Las formas de ejecución pueden ser variadas, pero se caracterizan por no seguir los lineamientos de la corporación y trazar, definir y especificar claramente en un inicio del proyecto las problemáticas en donde trabajar. Es una estrategia muy útil para modernizar la operación de la compañía como un todo, e incluso, probar soluciones que puedan tener un riesgo reputacional o regulatorio para la compañía madre. Es un concepto en el que los equipos de datos se mueven como una start-up, con el presupuesto de una gran compañía y en el que la orientación puede o no tener relación con los objetivos de la compañía. Por lo general, hay una mayor probabilidad de que sobreviva si los objetivos se alinean, pero no es un requerimiento en su formación.

Las estrategias deben considerarse centroides y no limitaciones, pues se pueden enmarcar mediante un fin, o mediante las restricciones organizacionales, pero son altamente variables en su configuración, tanto en el transcurso como en la vida de los equipos, en el estado del mercado o en el estado general de la organización. Lograr definir un foco y luego construir las capacidades para ganar flexibilidad es crítico considerando el estado de madurez de las capacidades y tecnologías ligadas al mundo de los datos, tópicos que esperamos aclarar en el transcurso de los siguientes capítulos.

## 4. Estrategia Táctico-Fundacional

La estrategia táctico-fundacional corresponde a una heurística para afrontar problemáticas tecnológicas y de entendimiento en la ejecución de casos de negocio. Se denomina táctico fundacional por la existencia de equipos de pioneros, así como fundadores y planificadores completamente diferenciados. Esta estrategia está destinada a ser un catalizador, principalmente en estrategias ofensivas, las cuales, con base en la realidad de Latinoamérica, concentran la mayoría de los grandes proyectos de modernización de análisis de datos. Esta estrategia se soporta sobre dos grandes equipos: uno con orientación a proyectos y el segundo con orientación a tecnología. En este último hay dos grandes vasos comunicantes: los equipos de gobernanza y los equipos de análisis de negocio.

A grandes rasgos el esquema estratégico sigue la figura 4, en el que ambos equipos o roles funcionales se retroalimentan en función de los objetivos de negocio. Se distinguen funcionalmente tres grandes dominios, los cuales se clarifican antes de desglosar los roles o funciones que laboran en cada uno de los contextos.



*Figura 4. Esquema táctico-fundacional.*

- Ecosistema táctico: corresponde a la traducción inmediata entre valor de negocio y activos de datos, todas las construcciones están orientadas en la evaluación priorizada de hipótesis de datos que puedan tener un impacto positivo en un KPI concreto de la compañía. Por ejemplo,

si se trabaja con metodologías ágiles, corresponde considerar una historia de usuario basada en beneficios concretos y no en activos de información, datos, reportes o modelos de aprendizaje automático. Esto quiere decir que no necesariamente se pesa la complejidad de la solución en esta etapa, sino más bien se itera progresivamente en aprender sobre el problema y encontrar las palancas concretas para resolverlo. En esta etapa existe mucha interacción entre el estado del arte para los problemas y la correcta traducción a la realidad empresa. Los elementos mínimos para generar valor desde el punto de vista táctico, son los datos y herramientas para manipularlos.

\* Data táctica: corresponde a esquemas de información en fases de exploración. Pueden contener variadas fuentes de información en el sentido de volumen, velocidad y variedad. Se utilizan principalmente integraciones rápidas, privilegiando el acceso ante la robustez. En esta etapa táctica es necesario determinar el valor de la fuente de información para priorizar el desarrollo de las integraciones sistémicas, documentación y cualquier otro tipo de esfuerzo técnico o regulatorio para acceder a esta información.

\* Herramientas tácticas: en esta etapa se privilegian las tecnologías de autoatención, así como las herramientas que no requieran fases extensas de desarrollo. El sentido de estas herramientas es simplificar las integraciones y no incurrir en gastos extensos sin antes justificar el valor de la conexión entre los sistemas y los prototipos funcionales.

- Ecosistema Fundacional: el ecosistema fundacional es el principal soporte a mediano y largo plazo para los activos de datos. Priorizados por los equipos tácticos, se logra que las tecnologías se construyan en función al valor y no como innovaciones sin sentido práctico. Este enfoque asegura que los datos por su parte consideren la creación de consolidaciones parciales y depuración de las principales fuentes de datos que requieran los proyectos, con el fin de disminuir la cantidad de errores, funciones operativas y problemas de gestión de conocimiento para el área. En este enfoque, además de los elementos mínimos para encarar las problemáticas, se debe velar por la mantenibilidad de las soluciones, así, tanto los datos como las herramientas deben soportar algún nivel de evolución y sostenibilidad para múltiples proyectos y áreas:

\* Datos maestros: la información ideal para explotar tiene como característica la resiliencia, robustez, y aseguramiento de calidad, regulación y trazabilidad. Como en muchas organizaciones estas máximas no se encuentran aseguradas, en la estrategia táctico-fundacional, los datos maestros se construyen priorizados en función a los proyectos estratégicos.

\* Herramientas fundacionales: contienen características de robustez, escalabilidad, sinergias, ahorros de costo e integraciones sistémicas que consideran un programa de evolución orgánica, en el que se aseguren las características en la medida en que soportan las funcionalidades de las herramientas tácticas priorizadas por los casos de uso. Existen varias formas de priorizar las herramientas fundacionales, desde las recomendaciones del mundo de arquitectura corporativa a los desarrollos incrementales de ingeniería de software. Es importante destacar que el valor de las herramientas fundacionales tiene sentido en el mediano plazo, por ende la elección de tecnologías y su estrategia tendrá su apartado particular en el capítulo siete.

## 4.1. Equipo táctico: la fábrica de casos de uso

Una de las principales problemáticas al encarar el enfoque táctico-fundacional, tiene que ver con la persecución orientada del valor. La identificación de las áreas claves y de los casos de uso particulares en los que la tecnología sumada a los nuevos datos, pueden ser diferenciadores. En esta sección, exploramos la fábrica de casos de uso como una composición pragmática de un equipo táctico.

Los roles de una fábrica de casos de uso consideran con una proporción mínima toda la cadena de valor del dato, es decir, considerando las piezas suficientes para acceder, movilizar, manipular, transformar y extraer el valor final de los datos con su consecuente aplicación al problema de negocio. Los roles se especifican de una manera conceptual y se dan luces de cómo se denomina ese concepto frecuentemente en el mercado, en donde existen cinco grandes grupos de roles en la cadena de valor:

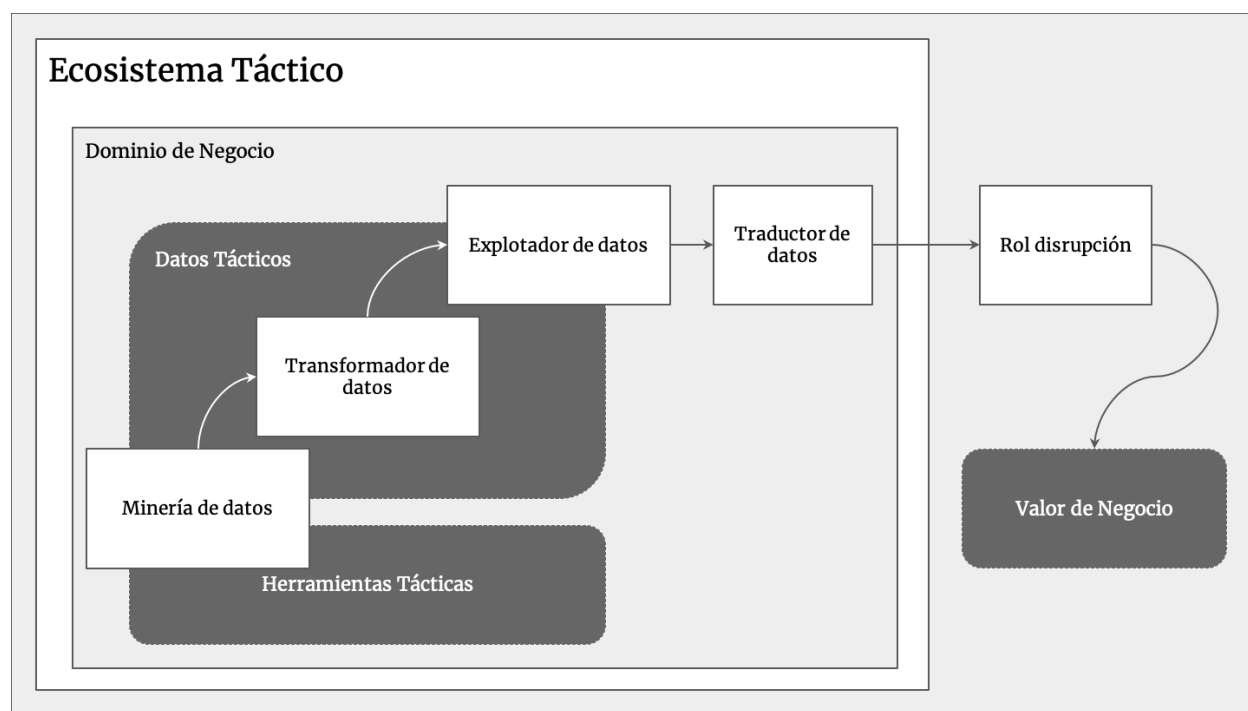


Figura 5. Principales roles en la fábrica de casos de uso.

- Rol de minería de datos: el rol de minería de datos, accede a las fuentes de información, localiza las piezas de datos que tienen mayor valor dentro de las fuentes, las obtiene y moviliza a un lugar consolidado. Frecuentemente, este rol lo ejecutan analistas de sistemas, programadores, expertos en sistemas de información, arquitectos de datos, ingenieros de datos. Entre mayor sea la complejidad sistémica y/o empresarial, mayor será la profundidad de las habilidades necesarias para ejecutar este rol y por ende, al evolucionar en el caso de uso, evolucionan también las interacciones de las partes del equipo que ejecute este rol.



- Rol de transformador de datos: la transformación de datos tiene variadas etapas, desde la refinación, calidad, transformaciones, hasta un modelo de aprendizaje automático. El foco de este concepto es considerar que el dato, como materia prima, necesita procedimientos que definan su valor, o que mediante transformaciones o interrelaciones se logren construir nuevos artefactos de datos según el fin último o intermedio. Los datos son unidades de información que sufren modificaciones orientadas a nuevos objetivos, por ende, según la complejidad de estos objetivos se complejiza el mecanismo de transformación. Los cargos que tienen sentido con transformadores de datos son los analistas, analistas de visualización, científicos de datos, analistas de calidad de datos. Finalmente, a nivel conceptual, un complejo modelo matemático que utiliza datos históricos, así como un filtro de negocio a datos sin sentido, proporciona una transformación, que cumple un nuevo objetivo y lo que puede tener o no sentido organizacional-muchas veces el sentido organizacional o coherencia de negocio no tiene sentido al inicio de un caso de negocio.
- Rol de explotador de datos: la explotación de datos tiene que ver con encontrar el sentido de valor sobre los artefactos transformados. Existen algunos roles de transformación que también ejecutan la explotación, pero se diferencian en el sentido que los explotadores de datos generan conocimiento o sabiduría desde las transformaciones en el punto en que se interrelacionan con información de negocio, ya sea a nivel de indicadores o descubrimientos. Frecuentemente, encontramos este rol a analistas de negocio, analistas de inteligencia de negocios, analistas de relaciones o comunicaciones, ingenieros de campañas y científicos de datos.
- Rol de traductor de datos: el traductor de datos es el encargado de transformar los hallazgos o transformaciones en elementos tangibles, ya sea mediante la intervención de procesos de una compañía, la creación de nuevos procesos, así como la monetización directa o la creación de nuevas fuentes de valor desde los datos, desde luego en asociación con las demás capacidades de la compañía. El rol de traductor también posee retroalimentación de problemas de negocio a problemas de datos. Esto es de vital importancia, ya que es el camino más corto para optimizar un proceso real de la compañía y enlazar la optimización matemática con una mejora en ingresos-tema que se verá en profundidad en la sección 8.3. Los cargos comúnmente encontrados en este rol son los gerentes de producto, gerentes de proyecto, jefes de proyecto, analistas de negocio, ingenieros y consultores.
- Rol de disrupción con datos: la disrupción con datos equivale a ejecutar la labor planificada por el traductor. Cuanto mayor sea el desacople del ser humano en la toma de decisiones, mayor será la intervención del área de disrupción. En algunas oportunidades, el valor estará autocontenido por los traductores, en caso contrario, se requiere de modernización o desarrollos de interfaces en los sistemas de la compañía. Los encargados de esta función son Desarrolladores de Software, Arquitectos de Datos, Ingenieros de Transformación Digital, Gerentes de Desarrollo, Ingenieros de Aprendizaje Automático, Ingenieros de Datos.

Evidentemente, la configuración de los equipos y las planificaciones pueden adaptarse a la realidad de la ejecución de proyectos en la compañía, utilizando agilidad o gestión de proyectos tradicionales. Lo importante es la configuración de los roles en el equipo, incluso compartiendo roles en la misma persona si la realidad así lo amerita.

## 4.2. Equipo fundacional: las bases tecnológicas y la gobernanza de los datos

Los equipos fundacionales se retroalimentan con las herramientas utilizadas en la fábrica de datos y se encargan de escalarlas, para esto deben rediseñarlas, construirlas, darles mantenibilidad y un ciclo de vida. Las decisiones de la ingeniería de detalle se verán en mayor profundidad en la sección 7.1, aun cuando las características profundas de la componente tecnológica y su relevancia en el futuro de cualquier organización alcanzarían a completar varias veces este escrito.

Otro aspecto relevante de los equipos fundacionales son las áreas de gobernanza de información, las cuales son el garante del valor futuro y el conocimiento exhaustivo de los datos y sus usos, en donde las áreas de gobierno, no solamente mantienen catálogos, diccionarios, sino que también, construyen las representaciones técnicas del conocimiento de la organización.

Los equipos fundacionales son menos estandarizables, dado que no existe configuración mínima que contenga todas las posibles competencias requeridas, dado que se podrían compatibilizar roles que administren herramientas o construyan nueva tecnología, por ende, se describirán a grandes rasgos sus focos y diferencias y la construcción detallada se abordará en el capítulo 6.

- Equipo orientado a la tecnología: independientemente de las decisiones de compra, arriendo o construcción, la complejidad de este equipo compondrá el principal músculo para dar evolución tecnológica a los equipos de data. Dado que, en el estado actual de las diferentes plataformas, no existe un equipo de explotación de datos que pueda hacer uso exclusivo de plataformas de registros, o al menos con un alto grado de madurez, y si se consideran los casos de uso que tienen el mayor potencial de disrupción en cualquier negocio, estos tendrán una alta dificultad técnica difícil de adquirir y de adaptar a la realidad empresarial.

Es por esta razón que el foco es construir tecnología evolutiva, mentalidad de productos mínimos viables en el sentido que, aún hoy día en lo referente a los datos, no existe un nivel de madurez en industria que pueda considerarse como el estándar de facto, sino más bien, como la interrelación de componentes en un ecosistema de soluciones. Esto no quiere decir que para organizaciones que estén comenzando sus procesos de digitalización las inversiones sea relativamente altas, es más, la complejidad emerge cuando la estrategia de evolución tecnológica de datos tiene que sostener y sobreinvertir por decisiones del pasado que se hicieron sin pensar que los datos tendrían que ser capturados, transformados y modelados, por ende, este concepto de “deuda técnica” es la principal razón de las altas inversiones o sobreingeniería.

En una empresa joven, en donde las plataformas transaccionales (que es el lugar que alberga la mayor cantidad de los datos) son modernas, o consumidas en modalidad software como servicio, es posible disminuir notablemente los costos y tiempos de implementación. Los roles dentro de este equipo consideran una fábrica de software orientada a soluciones que interactúen con datos, es decir, desarrolladores, arquitectos, ingenieros de datos, ingenieros de aprendizaje automático, entre muchos otros.

- Equipo orientado a gobierno: el gobierno de datos tiene dos principales vertientes: entre los sistemas de información hacia los repositorios de datos y desde estos repositorios hacia los procesos de la compañía y por ende hacia las personas que trabajan en estos. Entiéndase por estas relaciones al ciclo de vida de acceso, transformación y traducción final, proceso cuyo su diseño, desde el punto de vista de la eficiencia, se apoya en la arquitectura de datos y en su ejecución y mantenibilidad se soporta en mecanismos de gobernanza. Los equipos de arquitectura de datos son los responsables de diseñar las integraciones desde los sistemas de información. En esa integración, existen responsabilidades en las que custodios de datos se aseguran que la representatividad establecida en esta integración se mantenga en el tiempo, o de planificar una correcta actualización de criterios o nuevos datos desde aquel sistema transaccional, el flujo continúa con el aseguramiento de calidad y la gestión del conocimiento, en donde existen flujos transversales de legislación, consideraciones de privacidad, acceso y seguridad.
- Equipo de Operaciones de Datos: existe una tercera vertiente indispensable en los equipos fundacionales, que se desempeña en el aspecto de operación. El ideal de una fábrica de software es que la operación se delegue a piezas no customizables del sistema como puede ser el caso de utilización de software como servicio o aplicaciones con soporte, sea cual sea el caso, las transformaciones de datos, el modelo operativo de gobernanza, y los artefactos de software requieren documentación, mantenibilidad y ciclos de desarrollo que aseguren la solución de errores, la corrección evolutiva, entre otros.

### 4.3. Un modelo de evolución

La evolución de esta estrategia considera tres grandes estados y sus correspondientes evoluciones: la etapa de descubrimiento, la etapa de asentamiento, y la etapa de consolidación. En los que tanto la configuración de los roles como los diferentes tipos de actitudes y aptitudes dentro de los equipos, varían diametralmente. En esta sección analizaremos las principales características de estos tres diferentes estados y sus consecuencias en el plan de largo plazo de las competencias de datos al interior de cualquier compañía.

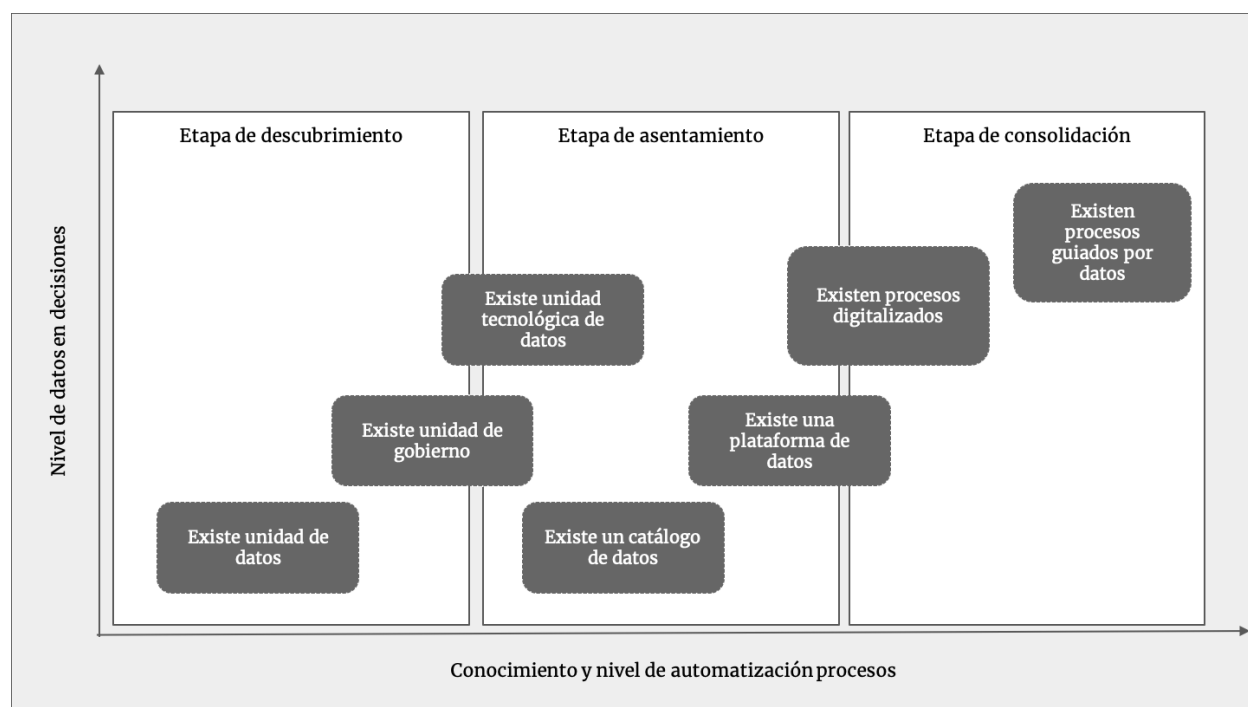


Figura 6. Principales etapas de madurez y algunos hitos referenciales.

- Etapa de descubrimiento: para incrementar el impacto de los conquistadores, el equipo táctico se configura con un agregado de expertos de negocio y facilitadores técnicos del interior de la compañía, enriqueciendo con conocimiento de negocio y prioridades a los profesionales del área de datos. En esta etapa también es posible considerar ayuda de recursos externos de consultoría las capacidades avanzadas de modelamiento matemático o tecnología.

Como se desarrolló en secciones anteriores, el equipo táctico es el protagonista de la evolución de la gestión con datos, tomando un contexto completo y mejorándolo a través de información y tecnología. Los focos en la etapa de descubrimiento son configurar las puntas de lanza y empoderar a los equipos con un sentido de urgencia por aprender de las problemáticas y configurar el medio para testear las hipótesis. Los perfiles en esta etapa deben considerar orientación al cambio, empoderamiento interno y redes de contacto con las áreas claves al interior de la compañía.

- Etapa de asentamiento: una vez planificada la captura de valor por la de exploración, es necesario sentar las bases de la transformación. En la etapa de asentamiento, deben ingresar los perfiles de construcción, así como los primeros planificadores, lo que incluye conceptos de tecnología de la información, para diseñar la evolución mediante metodologías de arquitectura empresarial. También es una buena oportunidad para asesorarse con empresas con experiencia en aceleración digital para aprovechar las experiencias.

En esta etapa debe ponerse foco en capturar perfiles con experiencia en el mercado, que servirán como semilleros de la configuración futura de los equipos. Se requiere comenzar a desarrollar las metodologías de contratación y configuración de planes de carrera. El principal foco en la etapa de asentamiento es generar robustez al equipo de datos al interior de la compañía, mediante la contratación de expertos y consolidación de las capacidades tecnológicas.

- Etapa de Consolidación: las capacidades de datos en esta etapa se consideran maduras al interior de la fábrica y el equipo táctico por lo que el principal foco en la consolidación es salir del modo proyecto y dotar a la organización completa con un enfoque de solución de problemas más orientado a datos. En pos de dotar con esta capacidad a la organización, es necesario permear a la capa directiva con conocimientos de base, como los son el método científico, estadística e ingeniería de software. En el sentido de percibir las nuevas capacidades y estar atento en como embeberlas en la organización.

Los perfiles en esta etapa comienzan por interiorizarse en la organización como un todo, desplegando las habilidades de desarrollo desde el mundo de los datos hacia las demás áreas de tecnología, mientras que por el lado de casos de uso, estos adoptan nuevas responsabilidades y pueden o no depender funcionalmente de otras áreas más cercanas a la captura de valor. La evolución natural de este equipo es involucrarse en las capacidades de la compañía. De otra forma, la escalabilidad de la adopción de datos no podrá tomar presencia transversal en la compañía, por ende, estará destinada a vivir en un alcance limitado.

Estas tres diferentes etapas, claramente deben ser adaptadas a la estrategia de negocios y a la estrategia tecnológica del negocio, por lo que pueden sufrir modificaciones, una particularidad relevante son los liderazgos de estas áreas, en donde abordaremos mayor profundidad en el capítulo 6.

## 5. Gestión del conocimiento

La gestión del conocimiento es la mantenibilidad de cualquier proyecto tecnológico. Usualmente es muy complejo considerar una estrategia disruptiva sin hacerse cargo del manejo del conocimiento empresarial, dado que, en función a este, es como se obtienen las mejoras sustanciales a mediano plazo. Si se piensa la empresa como un ser vivo, o un sistema complejo, las capacidades de memoria colectiva, al contrario de un mamífero, las de esta no se encuentran centralizadas. Es muy común observar que los comportamientos empresariales en temas de gestión de conocimiento se encuentran desperdigados en los tomadores de decisiones, las componentes individuales se sienten valoradas por sus conocimientos y no por sus capacidades y cuidan este conocimiento de la organización.

## 5.1. Empresarial: gobierno de datos

Es la gobernanza de datos el subconjunto de la gestión del conocimiento que hace referencia a cómo y por qué se utiliza la información. La información en una empresa tradicional es la sangre que compone un sistema aún más complejo. Es la forma de comunicar y transferir procesos entre áreas, la forma de entender las realidades de los clientes, así como es la forma para comunicar. Desde el punto de vista científico, los datos en una empresa son el lenguaje en el que esta empresa se relaciona con sus partes, por ende, entre mayor claridad de los datos, mayor efectividad en esta comunicación. En este punto, es importante considerar que entre mayores sean las complejidades de las organizaciones, también lo serán sus flujos de datos. En una empresa con problemas de gestión del conocimiento, los procesos complejos incrementan la cantidad de partes relacionadas y esto agravará a largo plazo los problemas de conocimiento que ocasionaron la complejidad, claramente un ciclo de reforzamiento negativo.

Los problemas en la gestión de los datos son la principal oportunidad de punto de inflexión para la cultura de una empresa. Tomar en consideración que las inversiones en documentación no son gastos de tiempo y cambiar la forma de mirar los aprendizajes y priorizar los estados de finalización para los desarrollos considerando la documentación, son los primeros pasos para sentar la estrategia de gobierno.

Existe mucha discusión en el mundo de los datos sobre la utilidad real de los estándares para gobierno de datos, tal como lo fue la generación de estándares o buenas prácticas para gestión de proyectos. En el mundo real, la utilización de un estándar no asegura que la cultura adoptará los procedimientos, pero sí plantea una guía inicial para facilitar la adaptación de esta a la cultura de la organización, para, posteriormente, evolucionar a la organización en el sentido de adoptar esta nueva realidad orientada a la mejora evolutiva, la memoria colectiva y constante eficiencia y adaptabilidad.

Las claves para un equipo de gobierno de datos, alineadas con las estrategias globales, son, sin dudar, el sentido de valor futuro del conocimiento, lo que es muy difícil de estimar desde un primer momento, pero que, gracias a la sostenibilidad de los casos de uso, puede habilitar inversiones y gestiones para disminuir los errores o mejorar las fuentes de información y así, reducir la manualidad de los procesos. Una estructura organizacional de gobierno de datos tiene tres grandes pilares, los cuales son:

- **Arquitectura de datos:** la arquitectura de datos se encarga de definir, desde un punto de vista técnico, tanto en diseño como en políticas, todas las integraciones y requerimientos de tiempo y forma, para el libre movimiento de los datos entre los sistemas.
- **Aseguramiento de calidad:** esta sección toma el control de los custodios de datos desde los sistemas, así como los analistas de calidad de datos, los cuales evalúan constantemente la sintaxis, la forma y la semántica de la información que fluye desde los sistemas.
- **Gestión de Metadatos y Conocimiento:** el área de metadatos y conocimiento es la memoria y el área de comunicaciones corporativas del área de gobierno. Se encarga de comunicar, transparentar y recabar la información necesaria para servir de base para construir nuevo conocimiento y aprender de la información contenida en los repositorios de datos. Sus

principales activos son los diccionarios de datos, los catálogos y los servicios de linaje de datos; los cuales muestran todas las componentes y trazabilidades de cada criterio en construcción de datos más complejos.



## 5.2. Equipos: ciencia de datos

Cuando nos referimos a gestión de conocimiento en equipos de datos, queremos tomar en consideración todos los procesos y subproductos de los proyectos analíticos y cómo estos enriquecen una organización. Esto trae a colación discusiones filosóficas y muy profundas en consideración a la resolución de los problemas que típicamente se atacan con ciencia de datos, los cuales escapan el alcance de esta sección. Por tanto, nos centraremos en discutir cómo se utiliza el método científico en la resolución de problemas de información en la compañía y daremos las bases de las implicancias en la cosmovisión de los problemas en una compañía.

La ciencia de datos se puede definir como un esquema sistematizado para recolectar evidencia para validar o rechazar una hipótesis desde una pregunta de negocio. Los principales aceleradores para la ciencia de datos moderna son: las nuevas capacidades tecnológicas, la gestión de conocimiento, y las capacidades de experimentación, así como la consolidación y completitud de datos históricos.

El objetivo de la ciencia de datos es generar nuevo conocimiento para la organización, además de establecer un sentido de robustez para este nuevo conocimiento. Este nuevo conocimiento claramente puede estar condicionado por temas estructurales (sesgos) o bien depender de una determinada estacionalidad (tiempo), por lo que la clave para mantener este valor es construir la validación científica sobre bases estadísticas robustas o bien, generar un músculo de aprendizaje automático para mantener en el tiempo vivas las hipótesis.

La principal diferencia entre estos dos enfoques, tiene que ver con asumir una constante revisita según tiempos de expiración de cada hipótesis de forma manual, o configurar los sistemas de información para que realicen este aprendizaje de forma automática, claramente la sustitución tecnológica de este avanzado criterio experto requiere de desarrollos tecnológicos avanzados y es ahí donde reside el valor exponencial de un equipo de tecnología de datos.

Los principales facilitadores para la gestión de conocimiento de los equipos de ciencia de datos, son activos que logren materializar, compartir y democratizar las hipótesis, que pueden ser desde catálogos, hasta documentos de carácter científico. Lo importante en la gestión de conocimiento en los equipos de datos es plantear las hipótesis y enumerar los avances y posibles puntos ciegos en sus parámetros.

Los equipos de ciencia de datos han tomado importante preponderancia en los equipos de datos, se les cree campeones o unicornios en pos de solucionar todos los problemas, cuando en la práctica son un eslabón más en un intrincado y complejo juego de equipo, en el que los hallazgos desaparecen si en el tiempo no se toman consideraciones de sustentabilidad- que da equipos de tecnología y gobierno. La capacidad de aprender de la información es importante, pero aplicar este conocimiento, así como transmitirlo en un lenguaje organizacional, es la clave para poder extraer el valor de los científicos de datos de una forma efectiva.

## **6. Gestión de los equipos**

La gestión de equipos corresponde al principal activo de la estrategia de datos, pues es común en cualquier tipo de organizaciones que el conocimiento colectivo sea la mejor fuente de información de calidad. Así también, la integración de nuevas capacidades a las existentes y la convivencia de los actores son fundamentales para lograr ahorrar horas de desarrollo e investigación con conocimiento de negocio, así como evitar problemas operacionales con una buena ingeniería.

## 6.1. Configuración de Equipos

La primera etapa de cualquier proyecto de datos tiene que ver con reconvertir, calibrar o salir a adquirir al mercado las nuevas competencias en datos. Es importante mencionar que las áreas de recursos humanos de las compañías tradicionales, por lo general, no están capacitadas para generar y evolucionar capacidades de alta especificación técnica. Los principales activos a construir en esta etapa son los siguientes:

- Perfil y etapa de construcción de oficina de datos y analítica avanzada (ciencia de datos y análisis). Importante en este punto es la estructura y las diferenciaciones entre una oficina centralizada, federada o distribuida. Las principales diferencias en estas configuraciones tiene que ver con el nivel de madurez y relevancia de los datos en las otras áreas de la compañía. Si se observa que la organización tiene equipos con capacidades, pero falta comunicación, es factible comenzar con estrategias de datos federados para empoderar a las áreas y facilitar la innovación. En el caso de empresas no dedicadas a la tecnología, o que la tecnología no sea núcleo del desarrollo de las capacidades, es posible considerar procesos de consolidación centralizados, para luego evolucionar las capacidades.

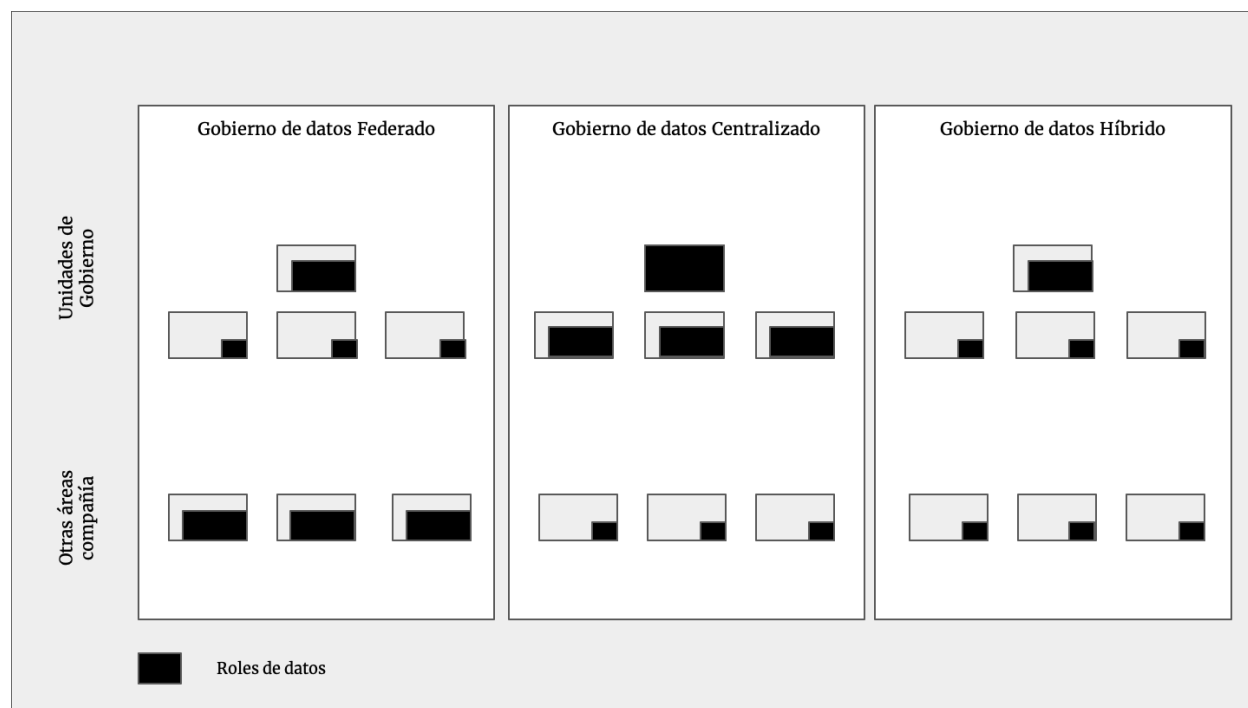


Figura 7. Diferentes cargas de roles de datos en cada área según tipo de gobierno

- Diferentes capacidades necesarias y cómo evaluar los perfiles involucrados, matrices de capacidades y competencias de un equipo ideal, tanto para perfiles tácticos como fundacionales.

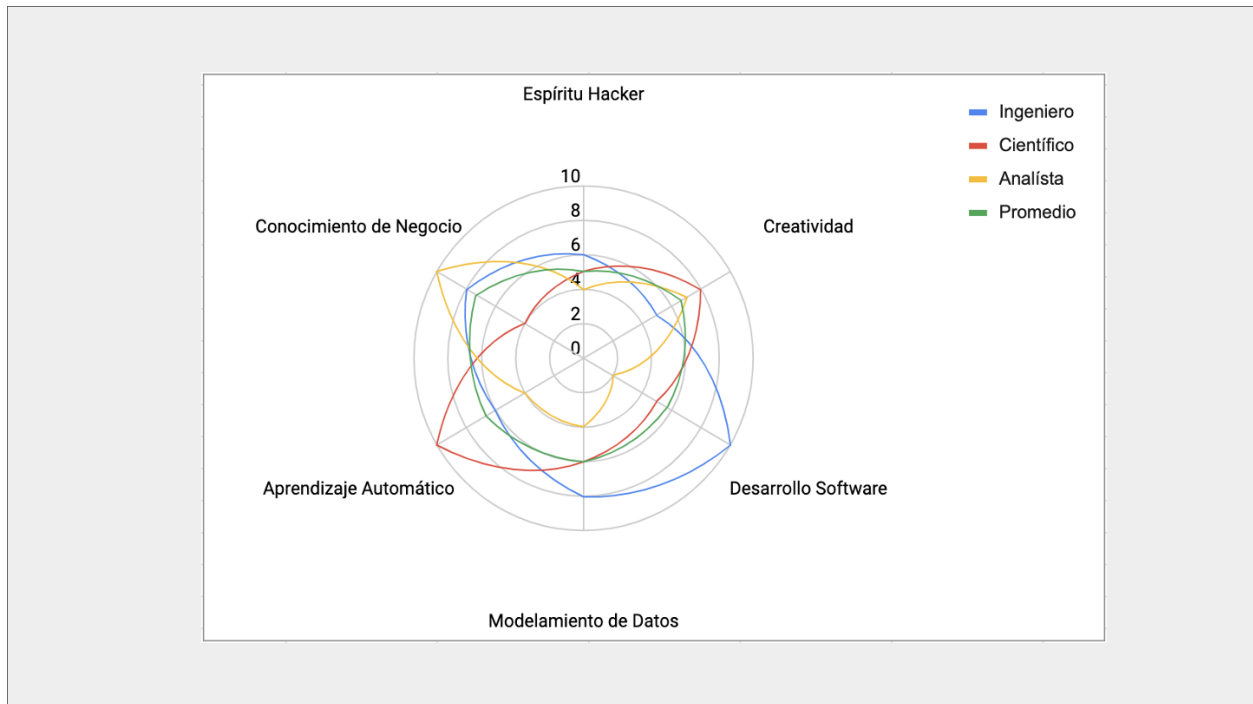


Figura 8. Esquematización de capacidades y establecimiento de requerimientos por cargo y evaluación del equipo.

Es importante mencionar que los sistemas de evaluación deben considerar a lo menos dimensiones blandas y técnicas. En el estado de madurez actual de la tecnología es importante diseñar ambientes colaborativos

## 6.2. Evolución y desarrollo de plan de carrera

Existe poca claridad en empresas tradicionales sobre el actual estado de madurez para los roles de datos, es por esto que decisiones sobre la pertenencia organizacional de estos equipos entra en conflicto con las actuales estructuras. En Latinoamérica, históricamente, las áreas de tecnología se encuentran lejos de los sustentos de valor, por ende, son vistas como áreas de apoyo y de aseguramiento de continuidad operacional, existen muy pocas empresas y sectores en donde las áreas de negocio viven cerca de la tecnología, por lo que carecen de las capacidades para convertirse en áreas de diferenciación y menos innovación. No es raro hoy en día encontrar áreas de datos más cerca de unidades de negocios, en el sentido en que la evolución de una carrera en datos debe estar aplicada a la creación de valor de la compañía. Incluso no es difícil encontrar áreas de datos dependiendo directamente de la gerencia general.

Aparte de la pertenencia organizacional, en un inicio es conveniente dar cierta autonomía para las contrataciones y los planes de carrera, privilegiando la flexibilidad para contratar a los primeros roles y dar espacio en la organización para mejorar el ambiente de trabajo y las condiciones que fortalezcan ambientes de innovación, creatividad y multidisciplinariedad.

[ilustracion\_9.png)

*Figura 9. Esquematización de Principales roles de datos y sus evoluciones orgánicas dentro de la organización*

### 6.3. Agilidad en equipos de datos

Una buena alternativa para la gestión de proyectos de datos es involucrar metodologías ágiles, tomando principalmente las máximas de empoderamiento en la toma de decisiones y la conformación de equipos interrelacionados, en los cuales las diferentes áreas enriquezcan, mediante el interrelacionamiento de conocimiento, la operación y creación de nuevo conocimiento-fin máximo de los equipos de datos.

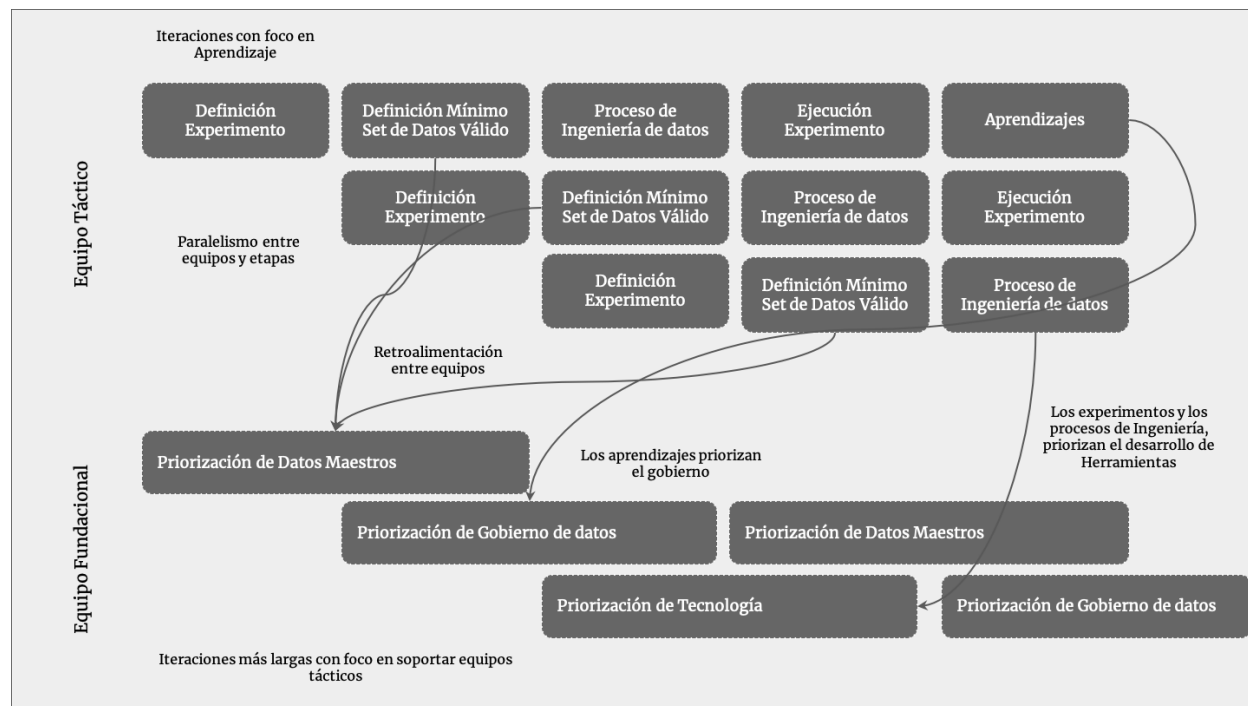


Figura 10. Esquemmatización de los diferentes ciclos de trabajo, tanto en los equipos tácticos y cómo estos realimentan la estrategia fundacional.

La agilidad en equipos de datos no es construir más rápido, sino aprender cada vez más rápido, en el sentido en que el aprendizaje de forma -robusta- sea logrado cada vez con menos iteraciones de validación o incertezas, tomando la experiencia de la creación de conocimiento científico y la epistemología, mientras se construyan bases firmes de conocimiento, cada vez será más sencillo construir nuevas teorías sobre las nuevas hipótesis. Una construcción colaborativa de hipótesis validadas no es un compendio de axiomas, por ende, los mecanismos de revisita de este conocimiento adquirido debe considerarse una nueva capacidad. Esta cualidad de la ciencia de datos puede parecer contraintuitiva, pero por el contrario, enfoca el fin último de los equipos y la importancia de estos, dado que el objetivo no es construir una batería de teorías, sino más bien, un músculo (equipo) aceitado (con tecnología) que esté orientado a encontrar y explotar oportunidades, mientras constantemente se revisan las hipótesis que estén dejando de reportar valor de negocio.

## 7. Tecnología

Luego de los equipos -unión personas y cultura-, el principal catalizador de la estrategia de datos en cualquier compañía es la tecnología. La estrategia tecnológica no puede estar desacoplada de la estrategia de datos, pues es esta última la que soporta gran parte de las nuevas capacidades, logrando una simbiosis que, de escalarse, puede lograr un real diferenciador entre las diferentes empresas, mediante la construcción de plataformas de diferenciación e innovación.

La tecnología en datos se puede considerar como un diferenciador de procesos de negocio, mientras que los casos de negocio pueden considerarse desde el punto de vista tecnológico como plataformas de innovación, dado que el posible valor futuro de la aplicación de esta tecnología está aún en discusión y análisis, por lo tanto, los equipos especificados en la sección de estrategia táctica fundacional tienen menos ancho de banda de ingeniería de software en los equipos que no han extraído necesariamente el valor de los casos de uso, mientras que el equipo fundacional tiene toda la fuerza de las capacidades de construcción, aun así, analizaremos cada macrocomponente tecnológica y luego, describiremos si es necesario invertir en capacidades de construcción, o si la diferenciación puede no valer la pena por sobre soluciones de mercado o adaptaciones de estas soluciones.

En términos organizacionales, la tecnología no reemplazará el rol de los ejecutivos, sino, más bien, los ejecutivos que logren explotar adecuadamente la tecnología reemplazarán a los demás. Como a nivel empresarial, las empresas que logren adaptar, crear o adoptar la tecnología en tiempo y forma adecuado para sus procesos internos, sobrepasaron a las empresas que no lo hagan, aun cuando las curvas de valor futuro sean diferentes en función de si la empresa construye capacidades de desarrollo, o adapta un world class en sus procesos. Lo importante es tener conciencia de dónde generar las capacidades y de dónde adaptar herramientas.

Un error común en empresas digitales es sobre generar equipos de tecnología para favorecer al mercado o a fondos de inversión con características que brindan un carácter tecnológico emergente a la estructura de la compañía. Esta sobreingeniería puede dotar de flexibilidad en un principio y ser diferenciador a largo plazo, pero requiere configuraciones organizacionales complejas, desde el punto de vista de ambiente de trabajo como planes de desarrollo.

Las compañías digitales no son digitales por tener muchos desarrolladores, sino que configuran un ambiente en donde los desarrolladores se sienten desafiados y motivados para innovar. Los extremos siempre son complicados de gestionar y, en este punto, lo ideal es encontrar un equilibrio en la especialización de los equipos, y determinar en dónde están los puntos críticos de dolor para mejorar los procesos y en esos, invertir en flexibilización. Por esta razón es importante saber dedicar los esfuerzos de innovación y diferenciación en los hitos tecnológicos que estén alineados con la estrategia de mediano plazo para la organización.

## 7.1. Diferenciar inversiones en tecnología

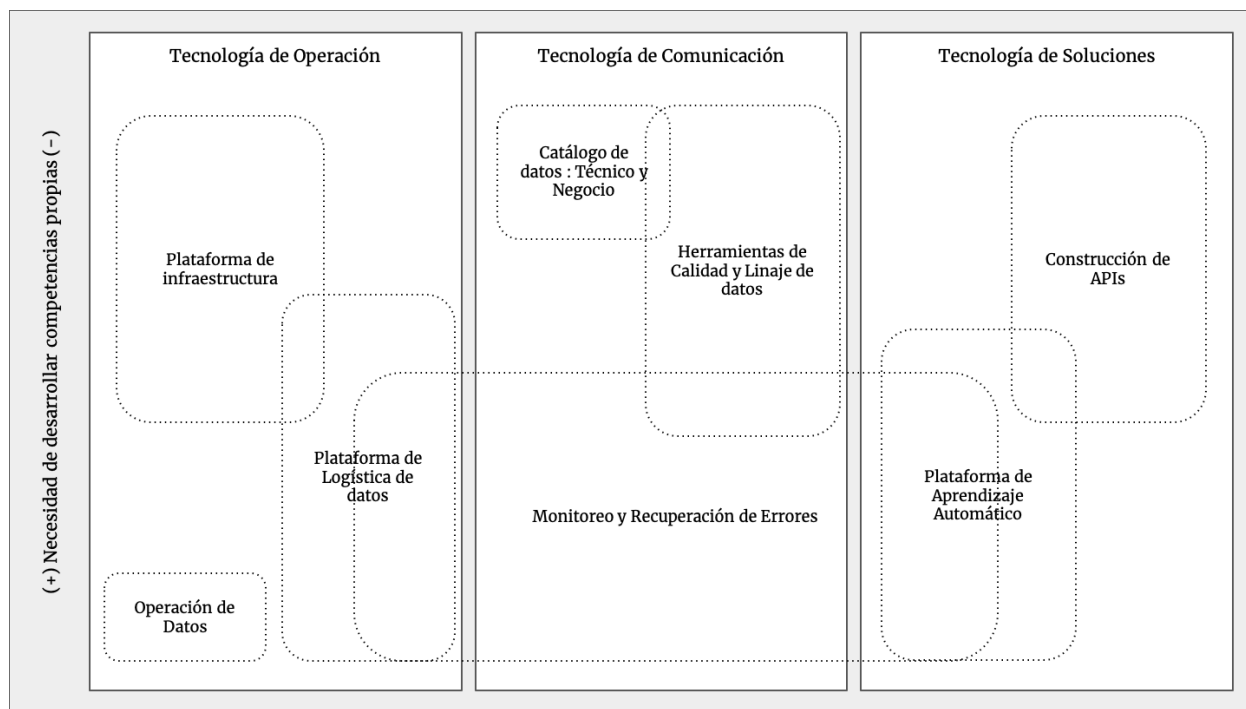


Figura 11. Principales herramientas dentro de las diferentes componentes tecnológicas y su priorización de desarrollo interno (diferenciación).

Las principales componentes de intensidad tecnológica se pueden definir en función a su orientación. Tenemos, por una parte, las tecnologías con orientación a cargas de trabajo, lo que puede incurrir en orquestación de procesamiento, acceso, integraciones, operación o ingreso de documentación. Existe otra componente de comunicación, que incluye la reportería, las herramientas de catálogo, entre otras cosas, y así también existen las componentes tecnológicas de soluciones, las cuales son la integración directa de un caso de uso aplicado y sirven para capturar la punta de valor.

- Componente tecnológica de operación: se definen como componentes de operación, las tecnologías de acceso a datos, las tecnologías de ingeniería de datos y transformación de datos, incluyendo mecanismos de analítica avanzada, como ciencia de datos o aprendizaje automático. La primera gran decisión estratégica tiene que ver con definir si será una estrategia monolítica con base en soluciones operadas en centros de datos propios o se construirá sobre capacidades de nube pública. Por lo general, si existen condiciones regulatorias o de volumen de información puede adoptarse una estrategia dual, pero, en el mejor de los casos, debiese ser una única estrategia en función a las características de la organización y de sus expectativas.
- Componente tecnológica de comunicación: la comunicación es clave para la eficiencia de los procesos en los equipos de datos. Uno de los principales activos de la estrategia de datos es comunicar los nuevos hallazgos, así como consolidar los canales de comunicación desde



las áreas operacionales (de sistemas), así como hacia las áreas usuarias o de negocio. Otra parte importante de la comunicación tiene que ver con la especificación de requerimientos de los equipos, así como el monitoreo, calidad de datos y manejo de fallas, las cuales durante la transición de digitalización pueden o no estar automatizadas, por esta razón, puede que tome un tiempo transformar capacidades análogas de comunicación a procesos automáticos de recuperación ante fallas.

- Componente tecnológica de soluciones: una pieza integral de las integraciones desde los activos de datos y los procesos de la compañía. Para efectos de un enfoque holístico, se denominan soluciones, dado que la complejidad de acceder a procesos no regularizados puede incluir un trabajo ad hoc de dos vías, considerando los desarrollos necesarios en las plataformas para tomar datos y enriquecer la gestión propia de su dominio como puede ser el caso de plataformas de marketing digital, así como construir vasos comunicantes desde las plataformas de datos hacia procesos mediante patrones de comunicación entre aplicaciones. La diferencia está en el diseño, planteamiento de arquitectura y estrategia de plataformas, dado que la decisión en este punto tiene que ver con enriquecer un dominio de negocio con datos o avanzar desde los datos en soluciones de negocio integrales.

## 7.2. Estrategia de adopción

Estas componentes tecnológicas sustentan la plataforma de explotación de datos. Para comenzar la estrategia de adopción comenzaremos con algunas definiciones:

- Almacén de datos (Data Warehouse): un almacén de datos corresponde al repositorio centralizado de información para fines analíticos, es decir, mediante procesos de transformación y desnormalización, se toman los datos desde los sistemas transaccionales y luego se depositan a un repositorio común, cabe mencionar que los almacenes de datos tienen como finalidad almacenar datos para ser consultados, por ende, la estructura de estos almacenes tiene finalidad de análisis o de operación de diferenciación básica, como campañas de contactabilidad o repositorios para uso de consultoría.

Un error común en áreas de datos es utilizar los almacenes de datos para fines operacionales. Cuando los equipos toman control de deficiencias sistémicas mediante herramientas de explotación de datos se observa el fenómeno de la informática paralela. Es por esta razón que las plataformas de información no pueden delegarse por completo a control de usuario, pero más grave aún es hacer la vista gorda con los procesos quebrados y no fortalecer los desarrollos de sistemas para que realmente se hagan cargo de las deficiencias apalancadas por procesos.

- Lago de datos (Data Lake): los lagos de datos son la evolución de los almacenes en cuanto a capacidades. Comenzaron como solución para procesar grandes volúmenes, grandes velocidades o grandes variedades de datos que vinieron principalmente de la sensorización, las transacciones de internet o la capacidad para obtener información de imágenes, videos o textos y en el corto tiempo, fueron evolucionando sus capacidades, contextos y requerimientos de orden en el sentido que esta nueva masividad compleja de información requirió nueva tecnología para procesar esta información. Esto hizo que la industria de los almacenes evolucionara, unido con la vertiente de soluciones de código libre y otras iniciativas de empresas digitales lograran en el transcurso de unos años democratizar las herramientas para procesar información. Un lago de datos puede ser muy simple o muy complejo en su arquitectura; todo va a depender de las elecciones de arquitectura y de los casos de uso que soporten su construcción.

Según la figura 10, el valor futuro y el nivel de madurez son claves para tomar decisiones estratégicas de inversión de tiempo o recursos en una estrategia de datos, las capacidades internas y la subcontratación, también deben ser parte de la evolución tecnológica, las recomendaciones anteriores tienen el contexto actual tecnológico, por lo que las componentes deben actualizarse según el nivel de madurez tecnológico en los tiempos de evaluación.

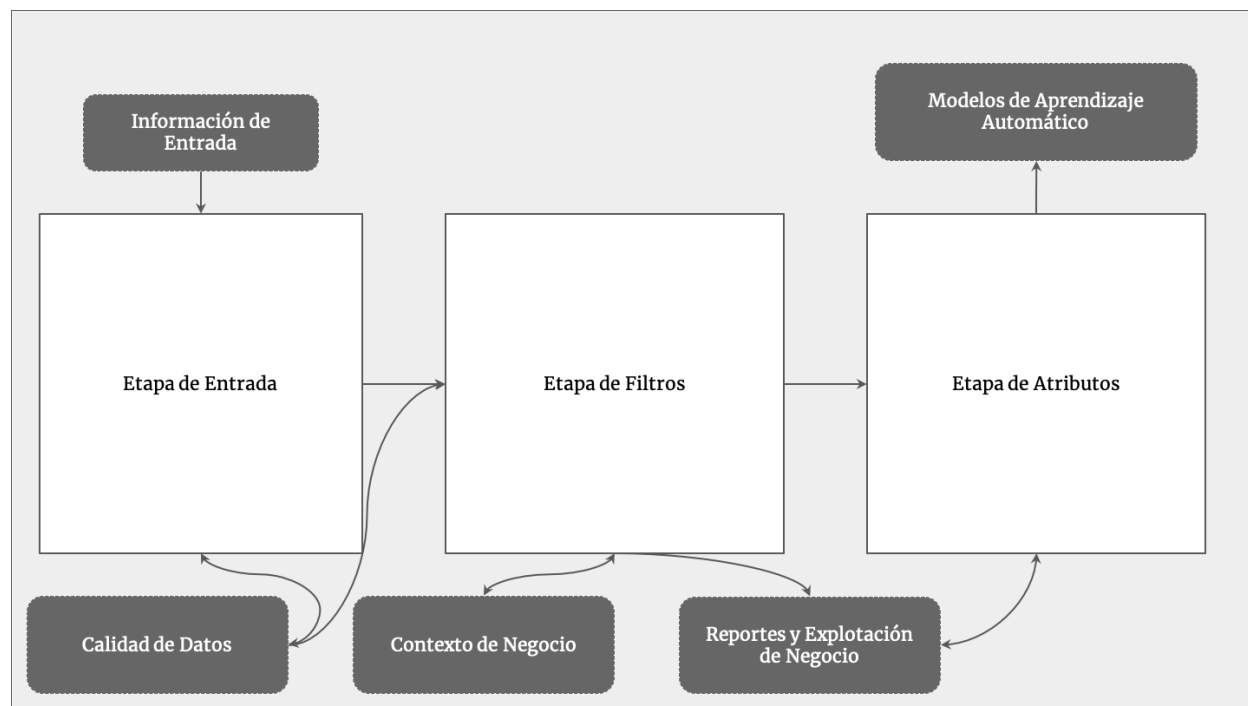
La tendencia actual es aprovechar las evoluciones en temas de datos basadas en tecnologías de nubes privadas o públicas, es por esto que la adecuada elección de los socios tecnológicos debe hacerse desde el punto de vista de la estrategia de evolución del negocio, más que en la estrategia de evolución de costos en sistemas de información. Este es el principal diferenciador al elegir las componentes tecnológicas que integrarán el ecosistema.

## 8. Principales puntos de dolor

Muchas veces, cuando las empresas se mueven, el seguimiento del plan se vuelve una máxima, pero, tarde o temprano, existirán problemas que no pueden planificarse como tales. En esta sección, enumeramos los más comunes, simplemente para agregarlos al plan desde un principio, dado que en base a la experiencia estos temas son los menos considerados y los más complejos de solucionar. Claramente, la complejidad para solucionar estos problemas tiene que ver con el establecimiento de capacidades simples, que no son difíciles de construir, pero que, muchas veces, requieren de una disciplina no presente en muchas organizaciones, principalmente en organizaciones con excesiva presión competitiva o donde el éxito oculte las propias ineficiencias.

## 8.1. Calidad de datos

Como vimos en la sección de gobierno de datos, la calidad desde los sistemas, así como la calidad del conocimiento de negocio son problemas persistentes y diseminados a través de la organización. La calidad de las soluciones está estrechamente vinculada a la calidad de los insumos en las “suposiciones” de transformación. Se asumen suposiciones y no verdades en el sentido que la calidad de datos en las empresas no existe, dada la complejidad observada en la gestión de conocimiento, las métricas de calidad, completitud, correctitud, deben ser variables y para no incurrir en problemas de mantenibilidad, es necesario construir “filtros” con estas reglas antes de utilizar la información. Puede parecer un sobreproceso, pero las siguientes recomendaciones, en el mediano plazo, disminuirán tiempo de reprocesamiento, posibilidad de errores, así como la transparencia de las reglas asumidas por los equipos que explotan los datos.



*Figura 12. Principales herramientas dentro de las diferentes componentes tecnológicas y su priorización de desarrollo interno (diferenciación).*

Desacople de transformaciones orientado a la calidad: como se observa en la figura 12, las transformaciones de los datos pueden orientarse en tres grupos, en donde cada uno posee un fin para su estructura. Existen múltiples mecanismos en la literatura para transformar datos, este se recomienda en contextos empresariales, en donde no existe un buen nivel de documentación o las reglas de datos son variables en el tiempo. Cada etapa tiene sus consideraciones de calidad de datos, la cual está contenida en su estructura y es consecuencia y requerimiento mandatorio para su construcción. Fuera de las transformaciones está la información de entrada, los inputs corresponden a la manipulación estándar de la información de entrada -entiéndase por manipulación

estándar a todo procedimiento que tiene como fin estandarizar el formato de la información, es decir, transformaciones sintácticas que no alteran el conocimiento, pero si optimizan su manipulación futura y homologación del sistema y sus métricas.

La siguiente etapa contiene los filtros, los cuales tienen en su estructura un subconjunto del input, el cual tiene contexto y calidad en su estructura, dominio definido y estandarización semántica, o sea que esta etapa se puede considerar como la resiliencia ante cambios y como única capa de explotación válida para las siguientes etapas. Esta consideración vuelve al activo “filtro” el más relevante para realizar calidad de contexto de negocio y construcción mantenible de información. Los atributos por otra parte se consideran construcciones orientadas al análisis, capas de explotación o capas semánticas, el sentido de los atributos es mantener el conocimiento de negocio.

## 8.2. Como encarar un problema de ciencia de datos

Para encarar un problema de ciencia de datos, existe bastante metodología, principalmente, desde el mundo de la minería de datos, documentación y tecnología moderna, que aliviará el proceso desde el set de datos hacia ajustar matemáticamente un modelo. Por lo que el principal problema que abordaremos en esta sección tiene que ver en cómo encaramos la construcción de los datos, que representan el dominio del problema, en forma que el objetivo de tal representación, represente la solución al determinado problema con la consiguiente adecuada comunicación de los hallazgos. Todo esto en problemas de ciencia de datos estructurados, es decir, con información tabulada. Para problemas no estructurados el enfoque de solución y planteamiento es diferente, ya que la estructuración de la información no estructurada es uno de los problemas que solucionan las herramientas modernas como redes neuronales o aprendizaje profundo.

La primera gran incómoda verdad del aprendizaje automático es que no existe solución óptima para un determinado problema (al menos demostrable por la matemática más avanzada de esta época). El enfoque de estos, en cambio, es encontrar la mejor respuesta útil a un problema sin solución conocida en un tiempo prudente, y para esto se ocupa una gran batería de diferentes técnicas y alcances que pueden transformar un espacio de características, o información de un determinado contexto en información que no está presente en el dominio, la cual se ajusta matemáticamente en función de la información que representa el problema en sí.

Claro está, existen muchas limitaciones estadísticas en la creación de este espacio de características, pero lo primordial, en un inicio y para evitar complejidad innecesaria a posteriori, es lograr representar, mediante dominios mutuamente excluyentes y completamente exhaustivos, el dominio de un problema, para luego lograr definir la información que se necesita para obtener la información deseada. Luego de definir la información deseada, se requiere representar esta información en los datos históricos. Esta debe ser gestionable e interpretable en el sentido de poder validar los hallazgos. En un principio, los modelos de aprendizaje automático deben privilegiar la explicabilidad, y lograr con esto, información robusta sobre los dominios que representan el problema. Esto ayudará a definir las líneas de investigación e incluso, a retroalimentar la gestión interna de la compañía. Luego de este primer paso de descubrimiento, la idea es optimizar el espacio de características en función de la importancia de cada dominio en un flujo de aprendizaje y priorización de dominios continuo.

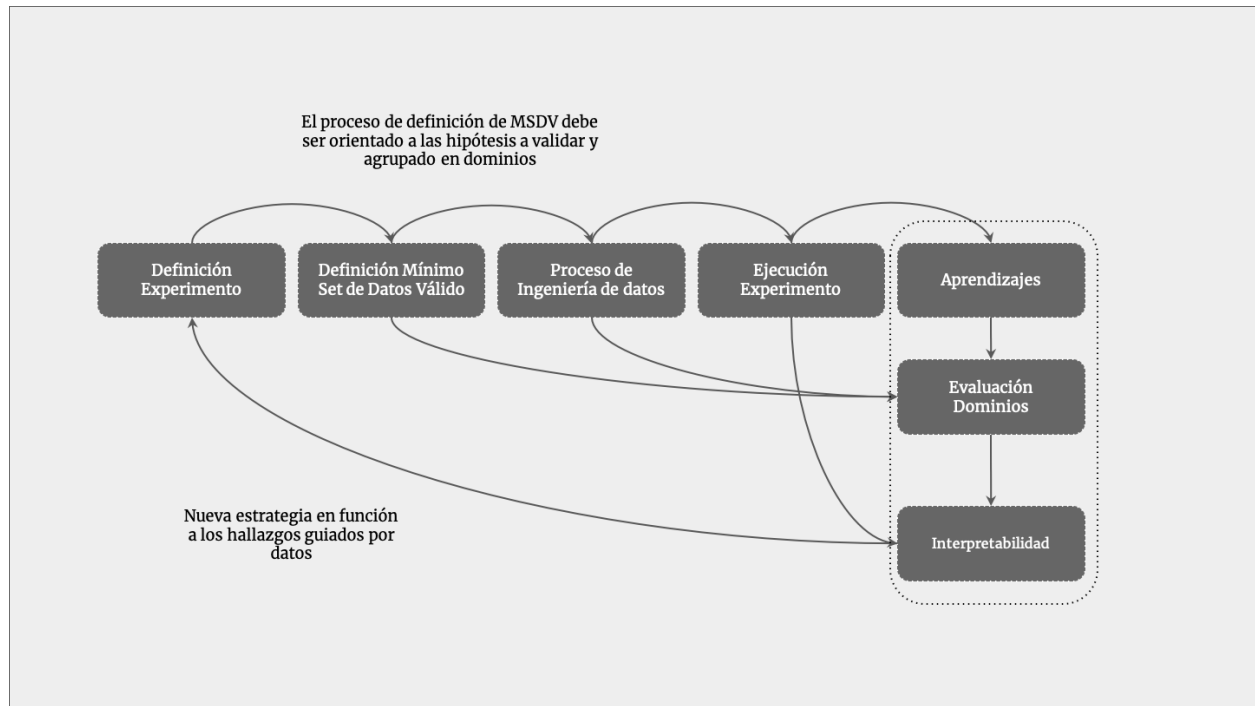


Figura 13. Esquema del proceso de indagación analítica para profundizar dominios de análisis para los problemas de datos.

## 8.3. Cultura

La cultura es el principal cuello de botella para la transformación de una empresa en todo sentido. Desde establecer y manejar las expectativas hasta lidiar con problemas de comunicación en culturas que no confían en cajas negras y que requieren de validación al perder el control de algunas decisiones. Los principales dolores culturales se dan en los dos extremos de la cadena de valor de los datos: en su inicio con las áreas que tienen en su responsabilidad los sistemas operacionales que soportan los insumos de datos, y las áreas que toman decisiones. Ambas ven amenazadas sus condiciones de privilegio y, en consecuencia, la colaboración y evolución de ambos equipos es la prioridad en la transformación cultural. Para solventar esta situación, es clave crear embajadores de la estrategia de datos en ambas áreas, en donde los roles actuales deben evolucionar para tomar control de sus datos y responsabilidad sobre los cambios de estos.

Para el caso de los responsables de los sistemas, generalmente orientados a resolver problemáticas de operación, la cultura de cambio tiene que ver con denotar la importancia específica de los datos como activo. Aquí surge el rol de Custodio de datos, el cual tiene la gorra de responsabilidad sobre los datos que surgen de la operación de los sistemas, su correcta evolución y disponibilización a las áreas usuarias. En las áreas de gestión, existe una metodología de evangelización similar, tiene que ver con empoderar la función estableciendo propietarios de datos, los cuales son responsables de la data como activo en el área de negocio.

En la medida en que estos activos representan valor para la compañía, ya sea para monitoreo de los indicadores de negocio o como soporte a decisiones o acciones, el responsable de los datos de cara a la continuidad de negocio reporta al propietario, el cual, desde la utilización, es finalmente el sujeto más interesado con el correcto gobierno, interpretabilidad, mantenibilidad y seguridad de la información, mientras que el rol de las áreas de negocio que no guardan relación con los datos debería enfocarse en donde los datos no están: tareas como innovación, creatividad usando más información, estudiando a la competencia, clientes o mercado.

Muchas organizaciones cuando plantean la necesidad de generar competencias de datos, se encuentran en paralelo cambiando su mentalidad para promover nuevos esquemas organizacionales para la toma de decisiones. Estos procesos interfieren con algunos tipos de equipos y con los desarrollos necesarios en las capacidades de soporte (como recursos humanos o tecnología), de forma que al principio se corre en el riesgo de “quemar” talento transformador digital. Es por esto que los equipos deben tener fusibles o blindajes organizacionales en su formación, así como para minimizar roces con equipos no acostumbrados a la innovación o al proceso creativo que demanda trabajar con tecnologías de emergentes.



## 9. Conclusiones y cómo retroalimentar la estrategia

El propósito de este ensayo fue dar una orientación práctica para pararse frente a una organización que comienza su viaje de datos. Aun cuando gran parte de la gestión tradicional de las empresas se realiza con datos, escalar este enfoque tiene que ver con lograr explotar el conocimiento de la organización de forma incremental y sustentable que nos proponen las nuevas dinámicas digitales. A continuación se enumeran los principales argumentos para ejecutar una estrategia de datos, los cuales son el fin último de toda aplicación tecnológica o de procesos:

- Los datos pueden ser catalizadores del negocio, al ofrecer una alternativa de mejora continua: lo observamos a lo largo de las iniciativas de negocio como los planteamientos de enriquecimiento de procesos utilizando aprendizaje automático, los quehaceres de una empresa pueden ser optimizados con datos, así como son claves para la diferenciación de sus procesos o iniciativas de innovación.
- Generar nuevos modelos de negocios como evolución de propuesta de valor vertical u horizontal: es claro que la información ha generado las mayores iniciativas de negocio, no es casualidad que la justificación de actores digitales la apalanquen estrategias de obsesividad con datos. Con datos un vendedor puede hacerse cargo del total de la experiencia de cliente, sin consideraciones subjetivas. Con datos de los clientes se pueden traducir procesos en experiencias, y así, priorizar su desarrollo. Es por estas razones que actualmente empresas pueden realizar movimientos entre las líneas de valor o en líneas de valor paralelas mediante información, para optimizar audiencias, la nueva economía del arbitraje o iniciativas que propongan rentabilizar este conocimiento de mercado, clientes, entre otros. En donde el nuevo activo es la información, como cualquier nuevo activo, este dependerá de cuán refinado se pueda obtener de los procesos de la compañía y cuán sofisticados están los consumidores para aprovecharlo.
- Vasos comunicantes entre la estrategia digital y la actual cadena de valor: lo que tienen en común todas las organizaciones son los datos, por ende, el único vaso comunicante se sustenta en validaciones estables (método científico) sobre información. La estrategia digital debe evolucionar mediante los preceptos de valor al cliente mediante análisis de historias de usuario y diferenciadores, cuyo insumo principal es la información. También es claro que para priorizar las entidades a atacar mediante la transformación digital, los datos plantean una metodología única para mejorar las priorizaciones.
- Rediseñar los procesos para mejorar la eficiencia de la organización: el rediseño de procesos requiere estándares de lenta evolución, o de experimentación. Los datos pueden acelerar esta mejora continua en situaciones en las que no es trivial realizar las mejoras. Estas mejoras pueden ser de ámbitos desconocidos por la organización o en donde se desconoce que realmente

hay datos involucrados. La estrategia en este punto es simple: donde existe un criterio, puede ser mejorado con datos.

- Construir un modelo consistente de mejora en la toma de decisiones: la toma de decisiones dentro de una organización se sustenta en el conocimiento colectivo, por ende, el planteamiento de utilizar datos para esta mejora tiene que ver con especificar estos factores y lograr objetivar las discusiones y los supuestos. El proceso no es simple, pues las estructuras y gobiernos corporativos asumen una experticia sobreestimada a la hora de tomar decisiones estratégicas, principalmente por razones culturales en los liderazgos. Sin considerar a los datos como representantes de activos de información, sino como interpretaciones de una realidad que, platónicamente, se describe a través de datos más interpretación. Esta diferencia sin orientación científica vuelve caótico el proceso de decisión, incluso considerando insumos para que seres humanos “aprieten el botón”.

El principal fundamento para comenzar el desarrollo de una estrategia de datos es la propia autoconservación de la empresa, en el sentido de ofrecer una alternativa clara para la mejora continua. Esta alternativa se aprecia claramente cuando el negocio tiene componentes digitales, pues es en el mundo digital en donde más velozmente se puede apreciar cómo los datos optimizan el flujo de valor, para otros procesos más análogos, los ciclos de iteración son más largos, es ahí donde la tecnología y nuevas técnicas de manejo de información pueden, por una parte, acelerar, y por otra, sostener esta nueva digitalización, en donde cada desafío puede transformarse en una oportunidad de diferenciación, mediante la digitalización, sensorización o innovación.

La retroalimentación de la estrategia, tiene que ver con plantearse ciclos de revisita a las hipótesis de valor, para reorientar el trabajo de los equipos de descubrimiento, en donde, a medida que transcurre el tiempo, se disminuye la cantidad de saltos de fe y se cambia la búsqueda de información por articulación de nuevo conocimiento. Este nuevo conocimiento se involucra en la organización y, en el mejor de los casos, ocasiona una espiral de transformación, la cual requiere nuevas capacidades para ejecutarse y abre un nuevo abanico de preguntas y plantea nuevos requerimientos de ajuste en la estrategia. Son las anteriores razones las que plantean la necesidad de una mentalidad de eterna innovación, y cuestionamiento de las estrategias, mentalidad preparada para aprovechar el valor rápido, para evaluar las siguientes batallas, y cómo éstas se deberán librar con estas nuevas capacidades.