

Modelo Hub&Spoke Adaptativo

El reto de una organización eficiente para la gestión de productos de datos



AUTORES:

Miguel Reyes

Director Estrategia Informacional en CaixaBank

Antonio Rodríguez

Gerente de Estrategia Informacional en CaixaBank

EN COLABORACIÓN CON:

Rafael Arana

Solutions Architect en Databricks

Iain Humphries

Strategy & Value Consultant en Databrick



PREFACIO:

Optimizando los equipos de datos para desbloquear el valor de los proyectos de datos e IA



Que los datos y la inteligencia artificial ofrecen un significativo potencial valor de negocio en organizaciones de casi todas las industrias es hoy un hecho reconocido.

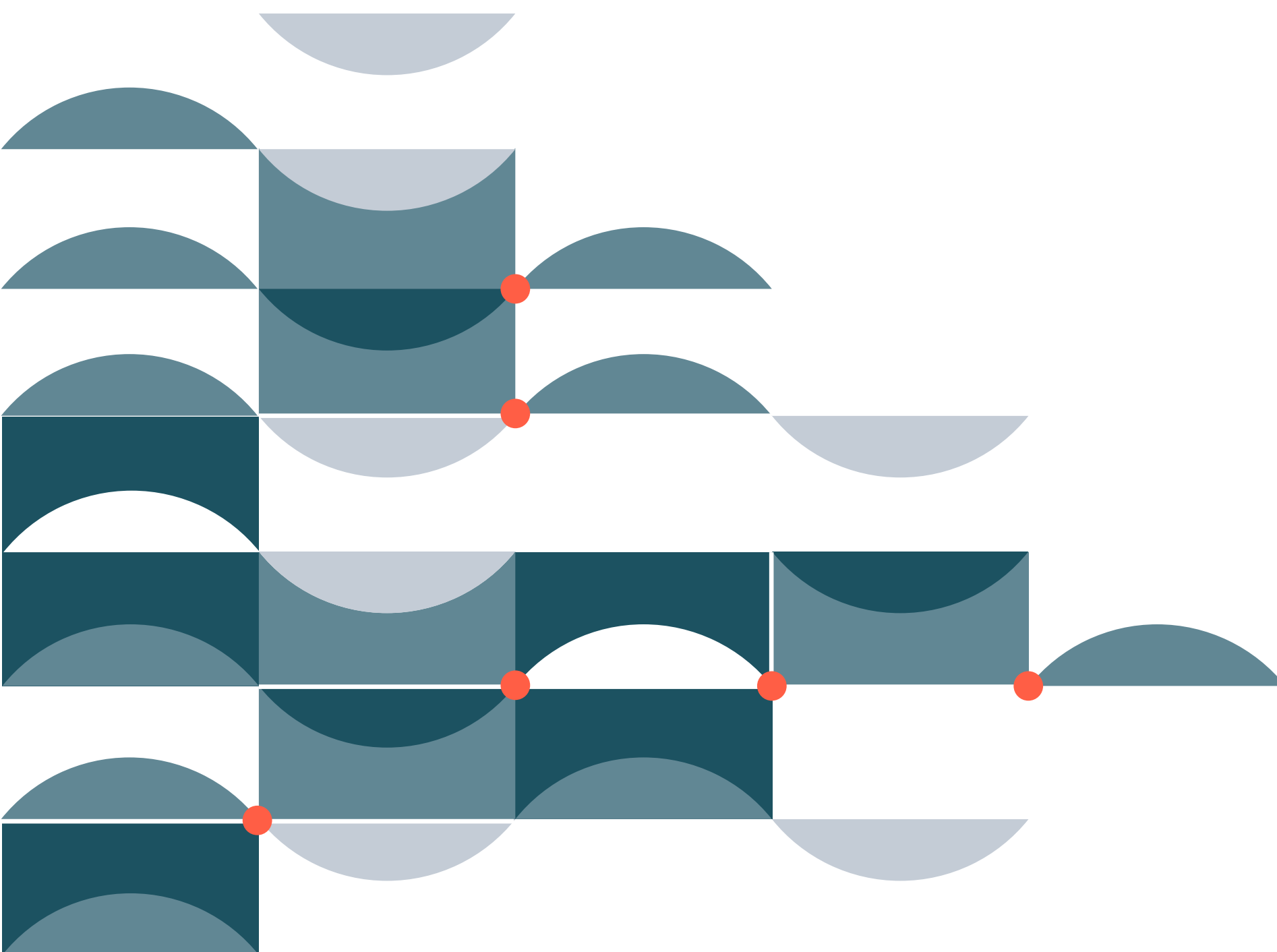
Sin embargo, diversos estudios (Gartner 2017, VentureBeat 2019) han destacado que el porcentaje de éxito de los proyectos de datos apenas llega al 15% sobre el total. Algunos de los factores clave destacados por estos estudios relacionan estas bajas cifras con los recursos, tanto con su ausencia (o falta de experiencia) como por su organización (modelo operativo analítico, AOM).

La distribución de los recursos de los equipos de datos en la organización y la asignación de las funciones y responsabilidades individuales y colectivas para respaldar los proyectos de datos e inteligencia artificial son claves para acelerar el éxito empresarial.

El modelo operativo analítico de cada organización debe buscar maximizar la capacidad de desbloquear el valor para el negocio de sus activos de datos y este documento se enfoca principalmente en un modelo que alinea habilidades y modelo operativo de proyectos para optimizar los recursos del equipo de datos. También abordaremos la importancia de invertir en una arquitectura de datos y una plataforma de análisis que sea simple, abierta y colaborativa para aumentar aún más las garantías de éxito del proyecto en plazos de tiempo menores.

Índice

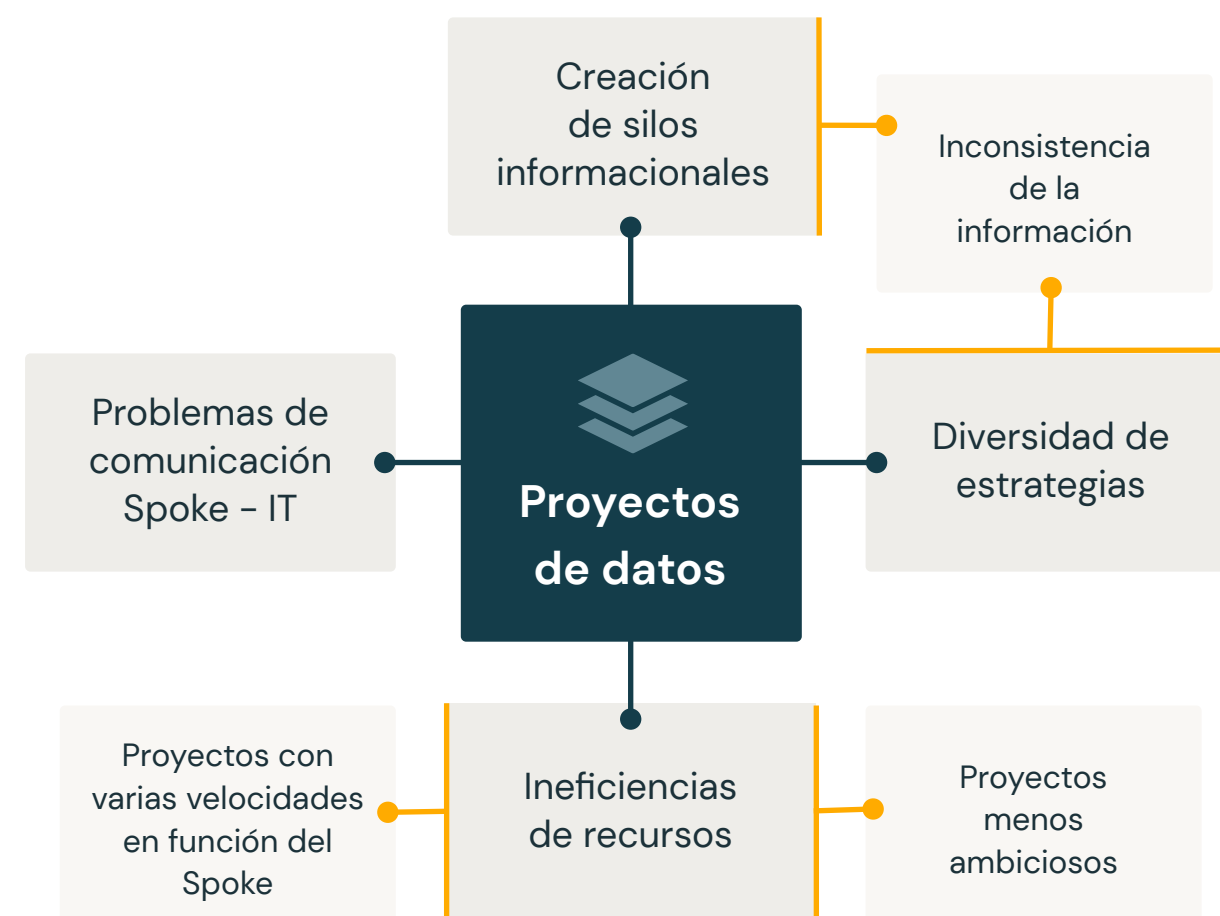
INTRODUCCIÓN AL MODELO HUB&SPOKE	04
Fases de participación del Hub	06
Gobierno	07
Análisis	07
Construcción	08
Mantenimiento	08
Enfoques de participación	09
CUESTIONES CLAVE	10
Modelo adecuado	11
Selección del enfoque	13
Variables relevantes	15
CASOS DE NEGOCIO	21
CONSIDERACIONES TECNOLÓGICAS	25
CONCLUSIÓN	28



CAPÍTULO 1

Introducción al modelo Hub&Spoke

El valor crucial que tiene la información en las empresas hoy en día ha convertido el dato en uno de los pilares claves de toda entidad. Según se han extendido los proyectos informacionales y los modelos analíticos por las entidades, los problemas asociados a los mismos se han vuelto más relevantes y visibles. Como resultado, la organización de personas y tecnologías dentro de un ecosistema informacional es cada día más importante.



Uno de los enfoques más comunes hoy en día para dar respuesta a este desafío es el modelo Hub&Spoke (H&S). En esta estrategia, el hub o nodo central, con conocimientos tanto del negocio de la entidad como analíticos y de IT, sirve como Gobierno de la Información y/o como Centro de Excelencia para la ejecución de proyectos de datos.

Las claves de éste modelo son las siguientes:

Estrategia del dato

El modelo Hub&Spoke favorece una estructura centralizada que asegura una visión y estrategia del dato transversal en la entidad. Maximiza el valor del dato como activo de la compañía, bajo un enfoque basado en la colaboración, estandarización, escalabilidad y flexibilidad. Pone el Gobierno como eje central en el uso del dato

Generación de valor

El dato aporta valor directo al negocio gracias a la confianza y a la certidumbre de la información sobre la que éste toma las decisiones.

Eficiencia

Se reduce el consumo de recursos en los proyectos mediante la estandarización de tareas y la eliminación de procesos duplicados. "Build once, use many". Se persigue la reusabilidad de la información. Al centralizar recursos, también se pueden abordar proyectos más amplios que los que podría abordar un spoke por sí solo.

Time-to-market

Mediante la utilización de economías de escala se consigue que el producto final sea liberado con mayor celeridad. Y por tanto los usuarios puedan tomar decisiones de negocio en un menor tiempo. La agilidad en dar respuesta al usuario aumenta la satisfacción de éste.

Democratización del dato

Evita la existencia de dos velocidades en el uso de la información, poniendo a disposición de todas las áreas la realización de proyectos de datos, ya cuenten con perfiles técnicos o no. Así mismo, se fomenta la proactividad en las iniciativas de datos, promoviendo desde el Hub el uso de información/modelos entre Spokes.

Comunicación Negocio - IT

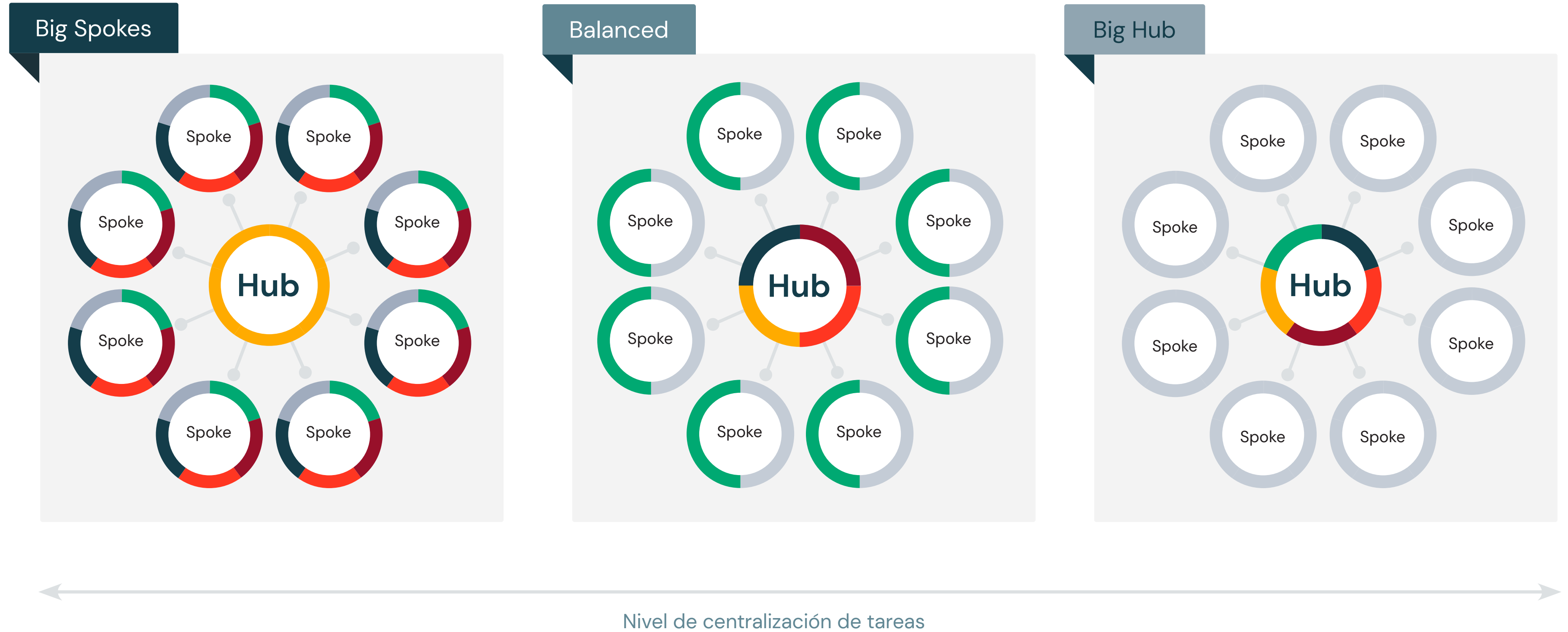
Se evitan los problemas de entendimiento entre las áreas funcionales y las áreas de IT, consiguiendo un mejor alineamiento y una mayor satisfacción del usuario con el producto generado.

Especialización de tareas

Se busca maximizar el valor aportado por cada miembro del equipo, esto es, posibilitar que los spokes, o áreas de negocio dediquen un mayor porcentaje de tiempo a actividades donde su generación de valor es mayor, como es el análisis de la información y su aplicación para la toma de decisiones.

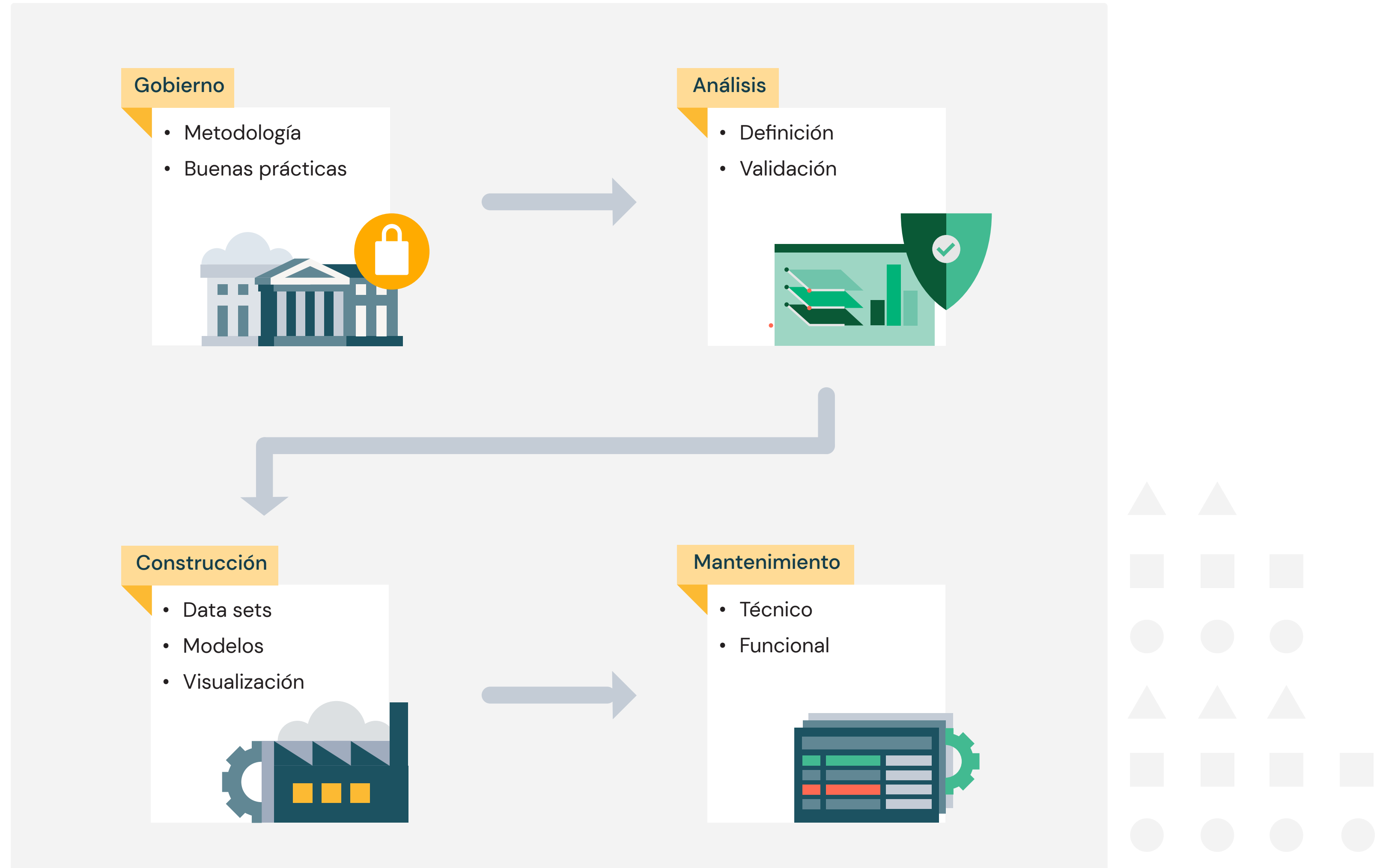
Desde el surgimiento de éste nuevo paradigma, se han establecido diferentes tipologías o enfoques de modelos en función del nivel de centralización seguido:

- Aplicaciones de visualización
- Análisis ad-hoc
- Analítica avanzada
- Procesos y tecnología
- Gobierno del dato
- Definición requerimientos y puesta en valor de la Información

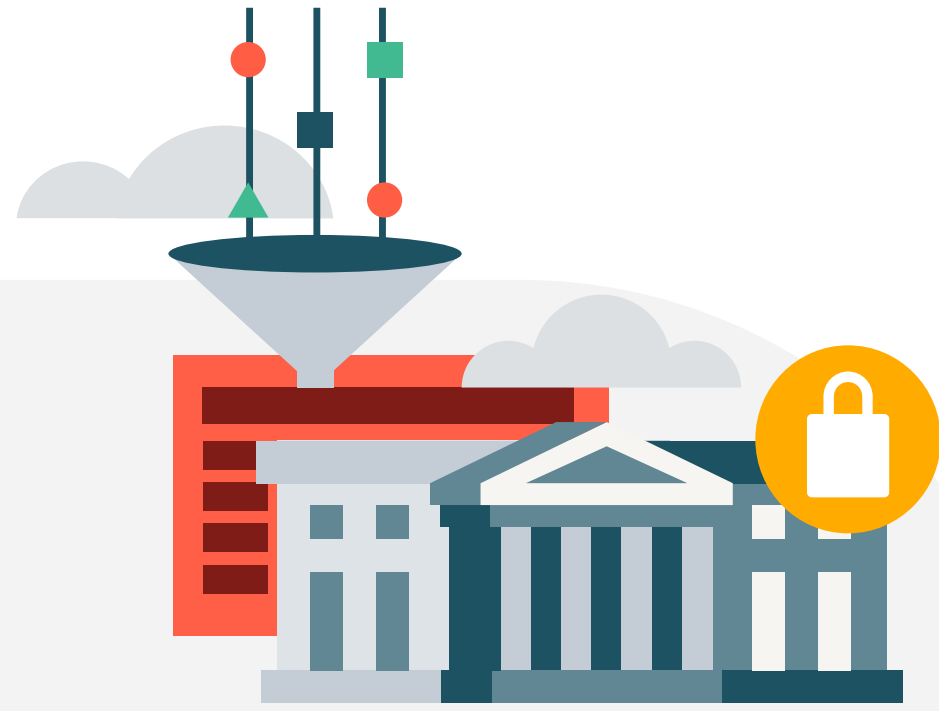


1.1. Fases de participación del Hub

Para entender mejor estos enfoques, vamos a tratar de explicar brevemente las tareas asociadas a un proyecto de datos y su posible centralización o descentralización. Para ello, agruparemos todas las tareas en 3+1 bloques principales.



Gobierno



Definición y seguimiento de una metodología para garantizar la correcta usabilidad de la información. Implica, entre otras subtareas, la documentación de los datos utilizados durante todo el ciclo de vida de cada proyecto informacional así como de las lógicas seguidas en los modelos analíticos y usos resultantes implementados.

Persigue, por tanto, asegurar que se cumplen buenas prácticas en la gestión y uso de los datos de la entidad, evitando duplicidades y garantizando la coherencia entre los usos. Las tareas de Gobierno posibilitan una correcta trazabilidad de la información que facilita y reduce la carga en las fase de aprovisionamiento o ingesta de los proyectos y además permite la reusabilidad de modelos y datos.

Esta tarea no es descentralizable y siempre ha de ser ejecutada por el Hub.

Análisis

Esta fase comprende las tareas necesarias para la definición de los requerimientos de negocio del producto (necesidades del usuario), así como los trabajos de validación del mismo.

Esta tarea siempre es responsabilidad del Spoke, que es quien tiene el poder de decisión sobre la definición funcional, si bien su ejecución se puede delegar en el Hub.

Debido a que tiene tareas tanto técnicas como funcionales se puede separar en dos subfases:



A. Fase técnica

Como soporte al componente funcional, tendríamos en la fase de Análisis una serie de tareas más técnicas que tienen una doble función:

Análisis técnico previo: Da soporte al análisis funcional permitiendo refinar los requerimientos a un nivel de detalle suficientemente preciso para la correcta interpretación por equipos técnicos. Para ello se realizan análisis exploratorios de variables, catálogos, fuentes de información, etc.

Validaciones de usuario: Se refiere a la facilitación de las pruebas necesarias (y definidas por el Spoke) para certificar la correcta implementación del producto y su utilidad para el negocio.

A. Fase funcional

Esta subfase del análisis se enfoca en la lectura e interpretación de regulaciones o normativas, tendencias del mercado, estrategias comerciales,... que permiten definir desde un punto de vista teórico el objetivo deseado del producto.

Construcción

Son implementaciones técnicas de procesos que permiten la generación de un producto informacional en cualquiera de sus vertientes y tecnologías: cuadro de mando, data mining, plantillas regulatorias, modelos de autoconsumo, etc.

Su ejecución comprende varias tareas que dependerán de la metodología de industrialización de cada entidad, como pueden ser el modelado de la información, elaboración de diseños técnicos, la construcción de los procesos (de ingesta, enriquecimiento, controles de calidad, etc.), pruebas unitarias, pruebas integradas, etc.

Esta fase de los proyectos de datos suele estar bajo la responsabilidad de las áreas de IT. El Hub debe velar por la correcta implementación de todos los procesos que deban ser implementados en el contexto del proyecto en cuestión.



Mantenimiento

En una fase posterior a la puesta en marcha de los procesos y/o modelos en un entorno productivo es importante tener en cuenta **el mantenimiento o BAU** que estos llevan asociados (especialmente en los modelos analíticos).

Para evitar su degradación, es necesario un esfuerzo adicional tanto **funcional** como **técnico**, mediante una constante monitorización, recalibrado y corrección de los mismos.

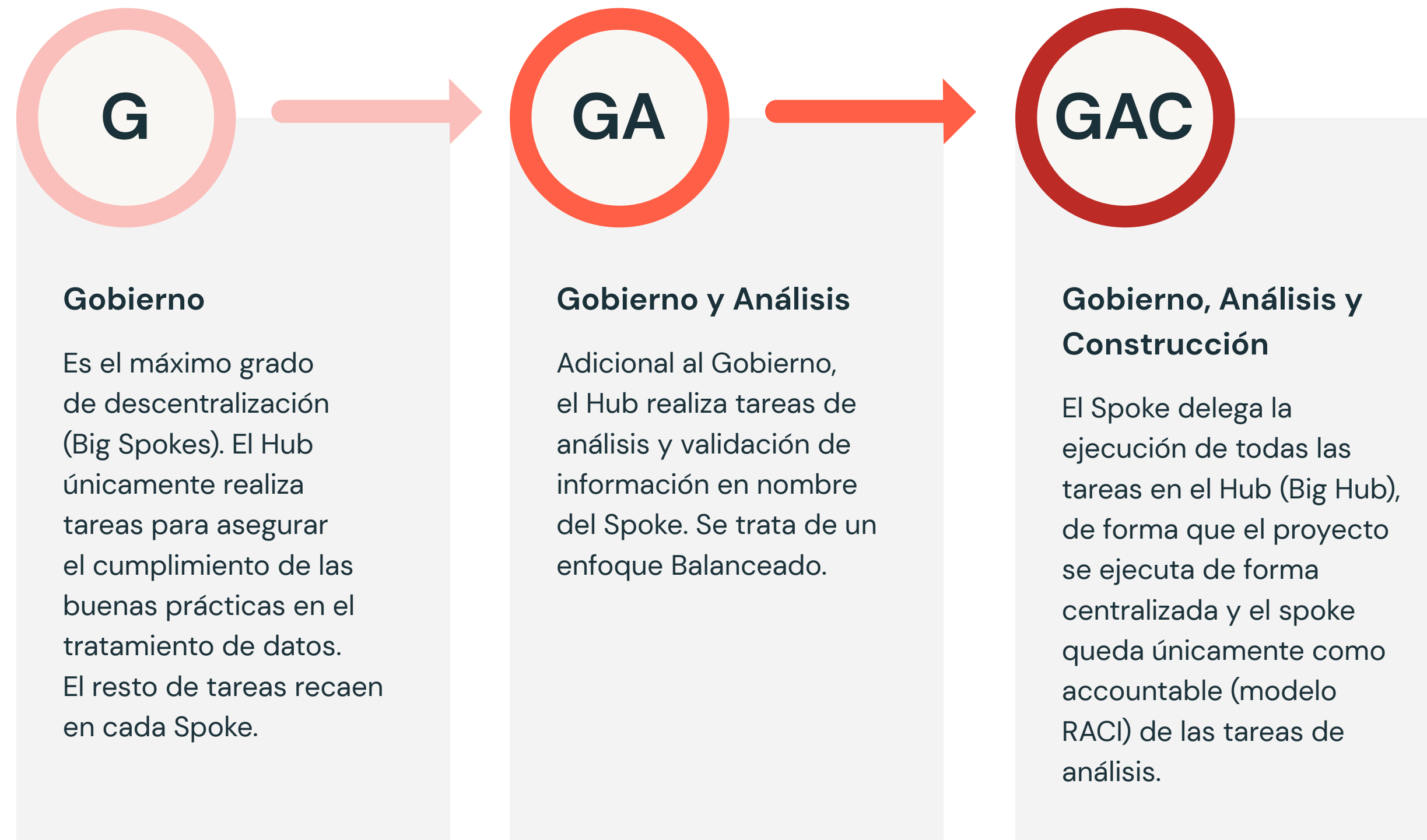
Este esfuerzo, por simplicidad, se puede distribuir de forma proporcional entre las fases de Análisis y Construcción de forma que su responsabilidad, y el esfuerzo dedicado, recaen en el equipo que desempeñe estas tareas. Sin embargo, por su relevancia en los proyectos, se ha considerado importante destacarlo en esta sección.

Cabe destacar la necesidad de planificar adecuadamente este esfuerzo puesto que no suele tenerse en cuenta en la realización de un proyecto pero sí consume recursos para mantener los procesos generados.

Enfoque de centralización

El enfoque de centralización se define como el grado de concentración en el Hub de tareas que se sigue, dentro de un modelo Hub&Spoke, para un proyecto y un spoke específicos.

De acuerdo a las tareas comentadas previamente (y considerando el bloque de Mantenimiento repartido entre los de Análisis y Construcción) podemos encontrar 3 enfoques distintos:



Cuestiones clave

Una vez entendido el Modelo Hub&Spoke y las diferentes opciones que permite con respecto a su implementación surgen varias cuestiones claves que vamos a tratar de responder en este documento:



¿Cuál es el modelo adecuado para una empresa?



¿Cómo seleccionamos el enfoque adecuado?



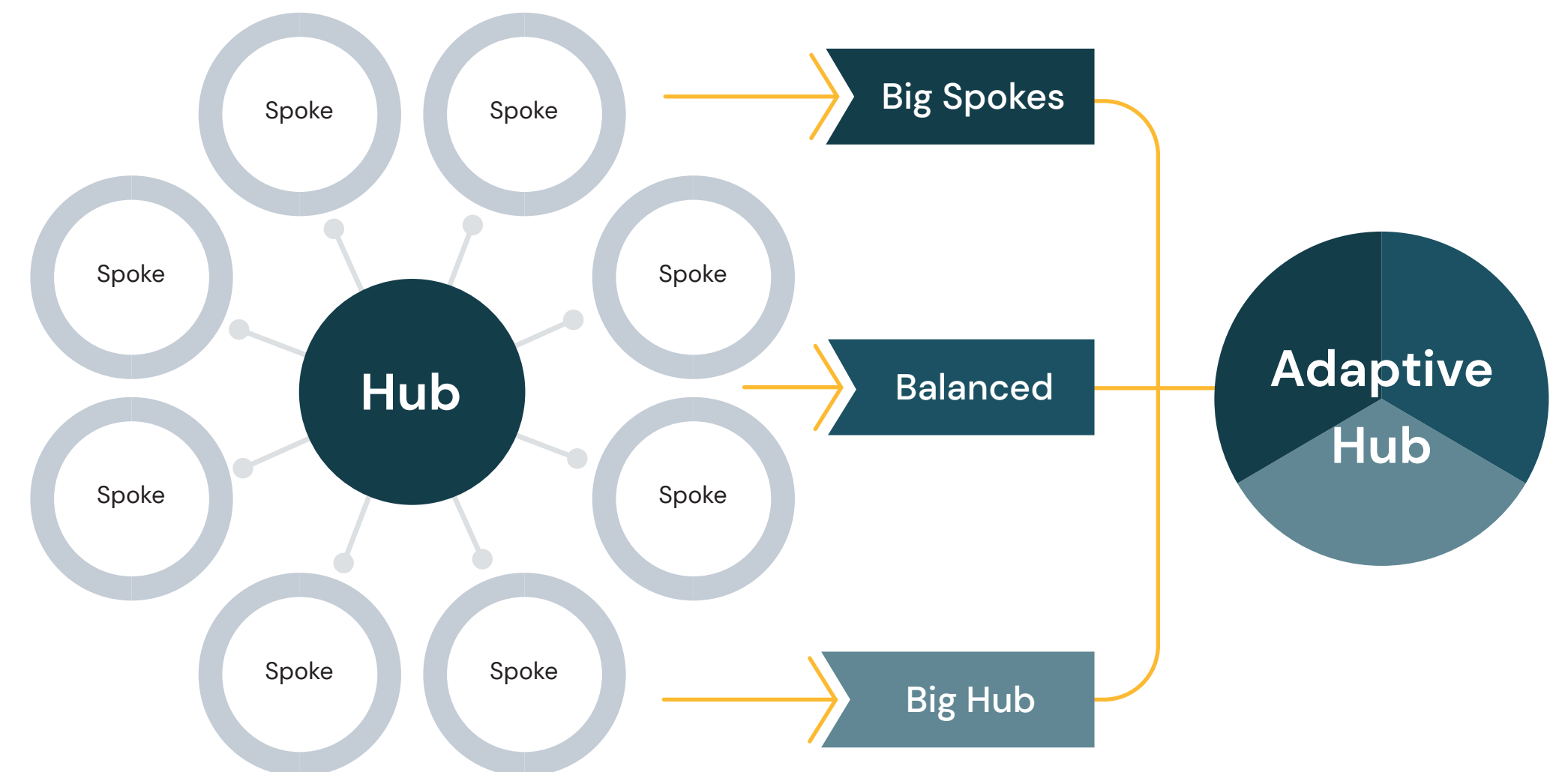
¿Qué variables son relevantes en la decisión?

¿Cuál es el modelo adecuado para una empresa?



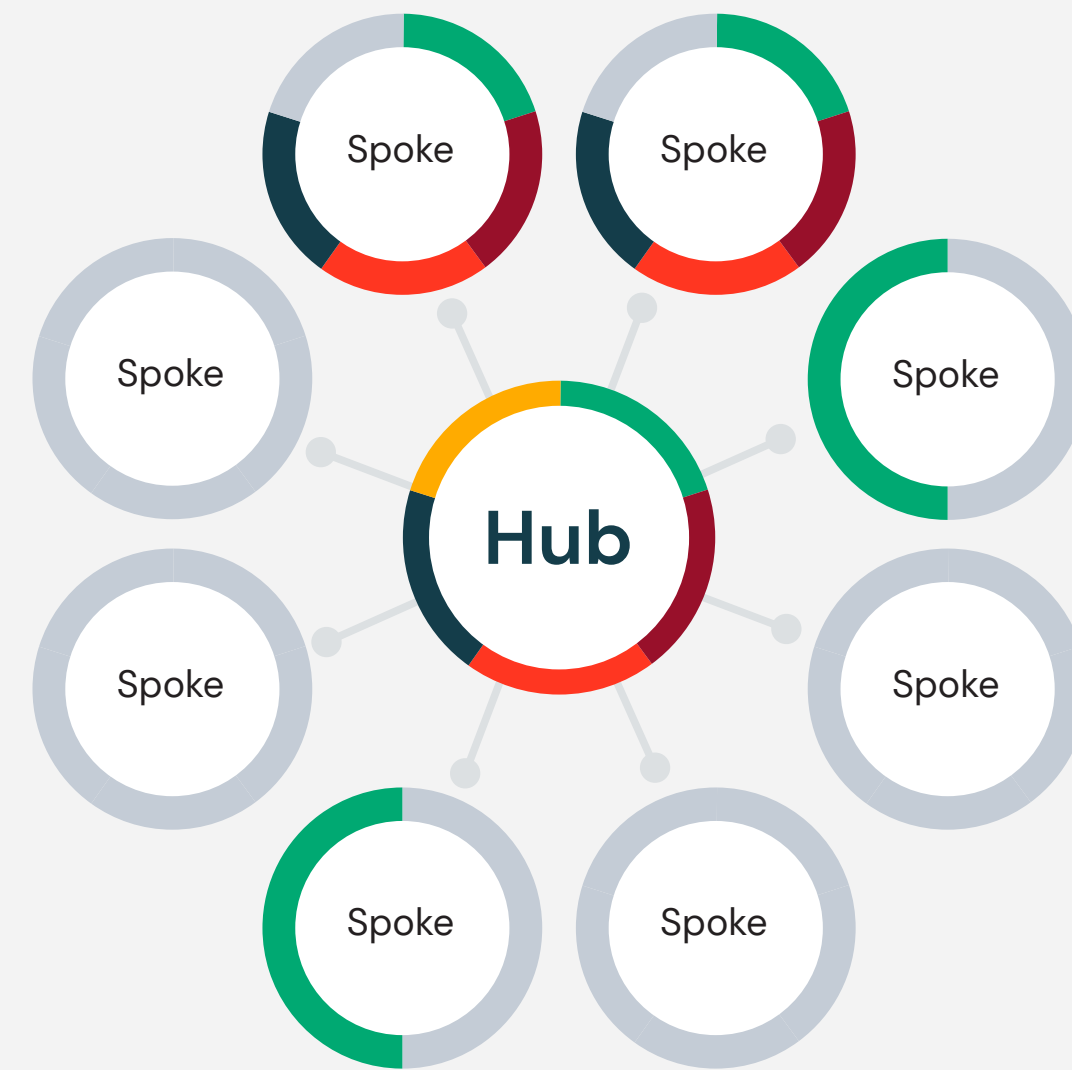
Las empresas son entornos dinámicos, con equipos muy diversos y por tanto con competencias diferentes. A este entorno variable, se suma la heterogeneidad de los proyectos de datos y la tecnología asociada a ellos, que supone una derivada adicional en la complejidad del proyecto.

Como resultado, el mejor enfoque para una entidad es aquel en el que Hub se adapta las necesidades de la situación persiguiendo la maximización de eficiencias. Surge por tanto la **necesidad** de evolucionar hacia una cuarta opción, un **modelo adaptativo**, ajustable en cada momento al enfoque más óptimo.

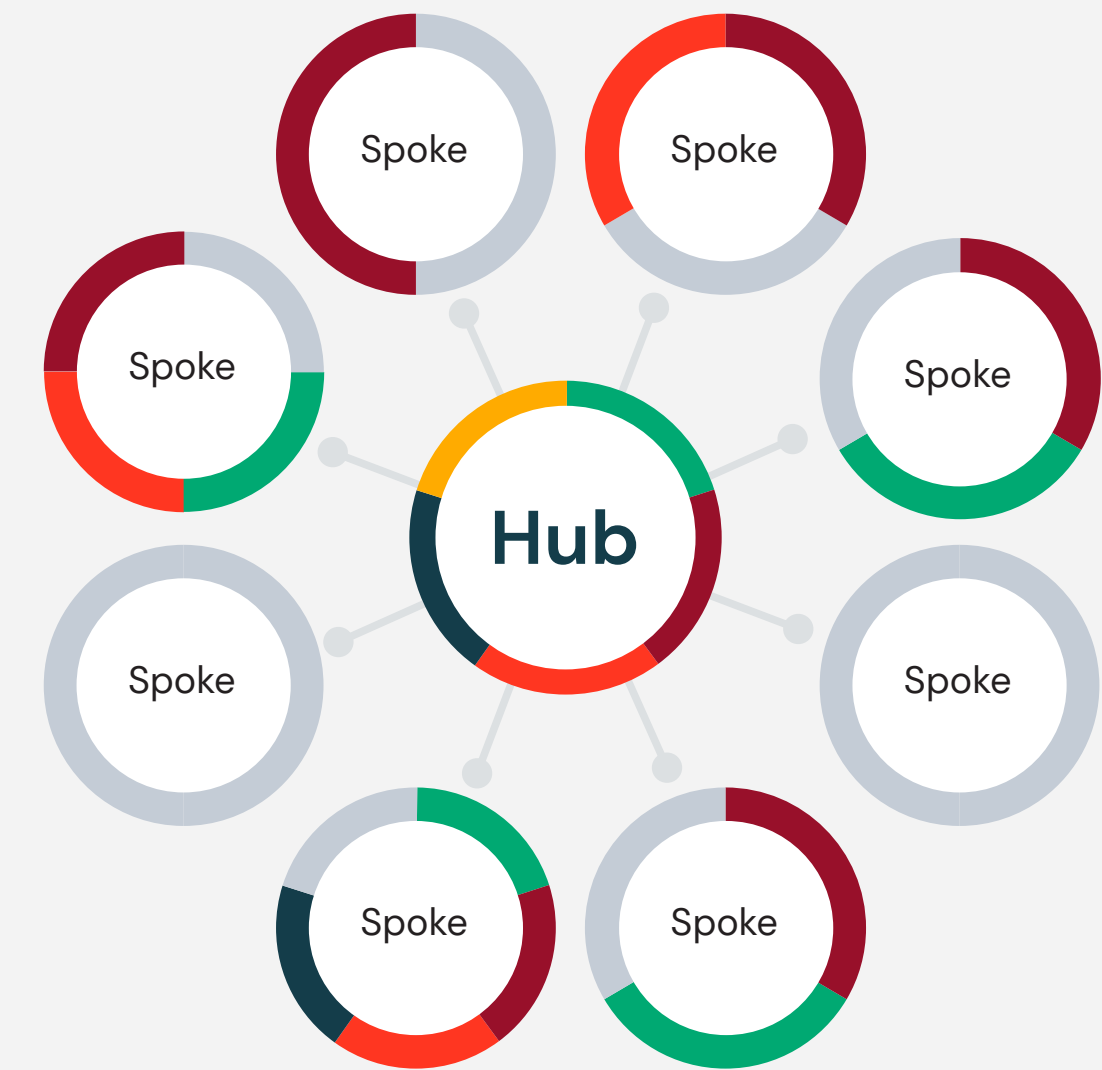


El modelo adaptativo permitirá que el Hub de una organización tome diferente nivel de protagonismo en diferentes proyectos al mismo tiempo y con diferentes spokes. Se trata por tanto de una **organización fluida**, en la que su forma se determina en base a múltiples variables.

- Aplicaciones de visualización
- Análisis ad-hoc
- Analítica avanzada
- Procesos y tecnología
- Gobierno del dato
- Definición requerimientos y puesta en valor de la Información



Combinación de proyectos X



Combinación de proyectos Y

"proyectos X" and "proyectos Y" are a group of projects in different time gap

The distribution of different group projects is different and that implies a Hub constitution volatile, dynamic, and fluid

¿Cómo seleccionamos el enfoque adecuado?



SELECCIÓN DEL ENFOQUE ADECUADO

Para determinar el nivel de descentralización óptimo, es deseable que la decisión se base en criterios objetivos de medición de valor, de beneficios y de costes. Para ayudar en esta tarea, hemos desarrollado una herramienta que permite estimar el mejor enfoque posible para un proyecto informacional en base a la maximización de eficiencias dadas ciertas restricciones iniciales.

Basado en una serie de inputs que definen las características del spoke y del proyecto a considerar, la aplicación devuelve el número de recursos que cada una de las áreas implicadas en el proyecto invierte en el mismo. De esta forma, si repetimos el análisis para cada uno de los enfoques, que actúa como input, deduciremos como óptimo aquel para el que el número de FTEs empleado de forma global sea mínimo.

La aplicación permite su uso, por tanto, como herramienta de gestión de la demanda en modelos Hub&Spoke, donde se puede utilizar junto a otros criterios (valor informacional del dato, distribución de recursos entre áreas, etc) para determinar el nivel de descentralización de los proyectos.

Es igualmente importante destacar que hay ciertos factores que pueden influir en el enfoque de centralización a seguir aun “contradiendo” el resultado de la calculadora. Ejemplos de estos casos son:

- ▶ Proyectos con especial relevancia para la entidad, donde la calidad de la información prima sobre criterios de eficiencia.
- ▶ Proyectos ejecutados en fases iniciales de establecimiento de un modelo Hub&Spoke, donde apenas hay información gobernada y por tanto se requiere una “inversión” para conseguir una funcionamiento estable del modelo.

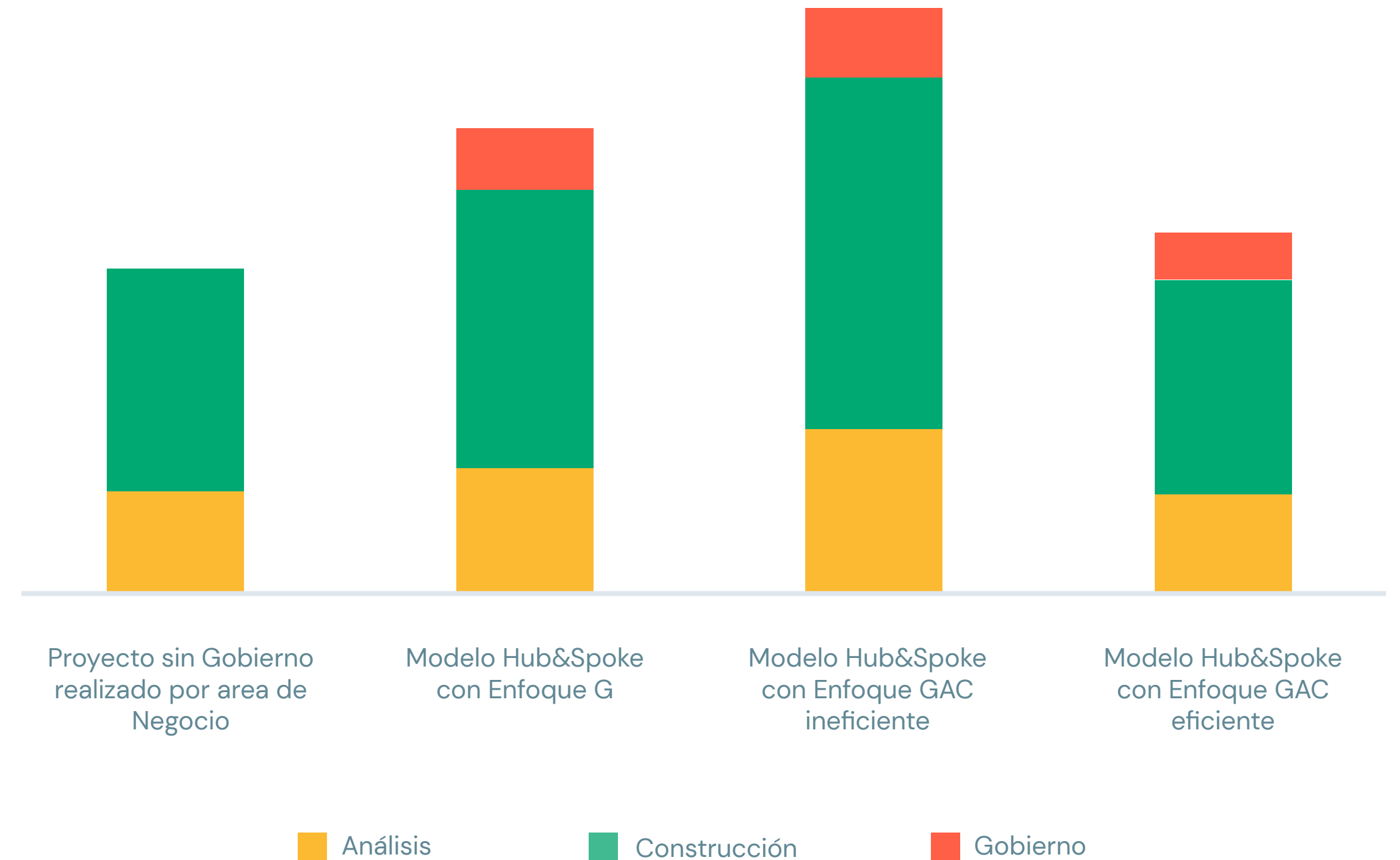
Atendiendo al gráfico siguiente podemos observar la importancia de una correcta toma de decisiones a la hora de determinar el nivel de centralización de un proyecto:

Proyecto sin Gobierno realizado por Área de Negocio: Este enfoque es el más eficiente, puesto que no incorpora una capa de gobierno de la información. Si bien el hecho de no tener este control de la información penaliza algunas fases, la fase de gobierno no compensa el esfuerzo extra en recursos. Sin embargo la capa de gobierno, no tiene como objetivo último esta eficiencia de horas, sino que la calidad y la coherencia de la información generada sea máxima.

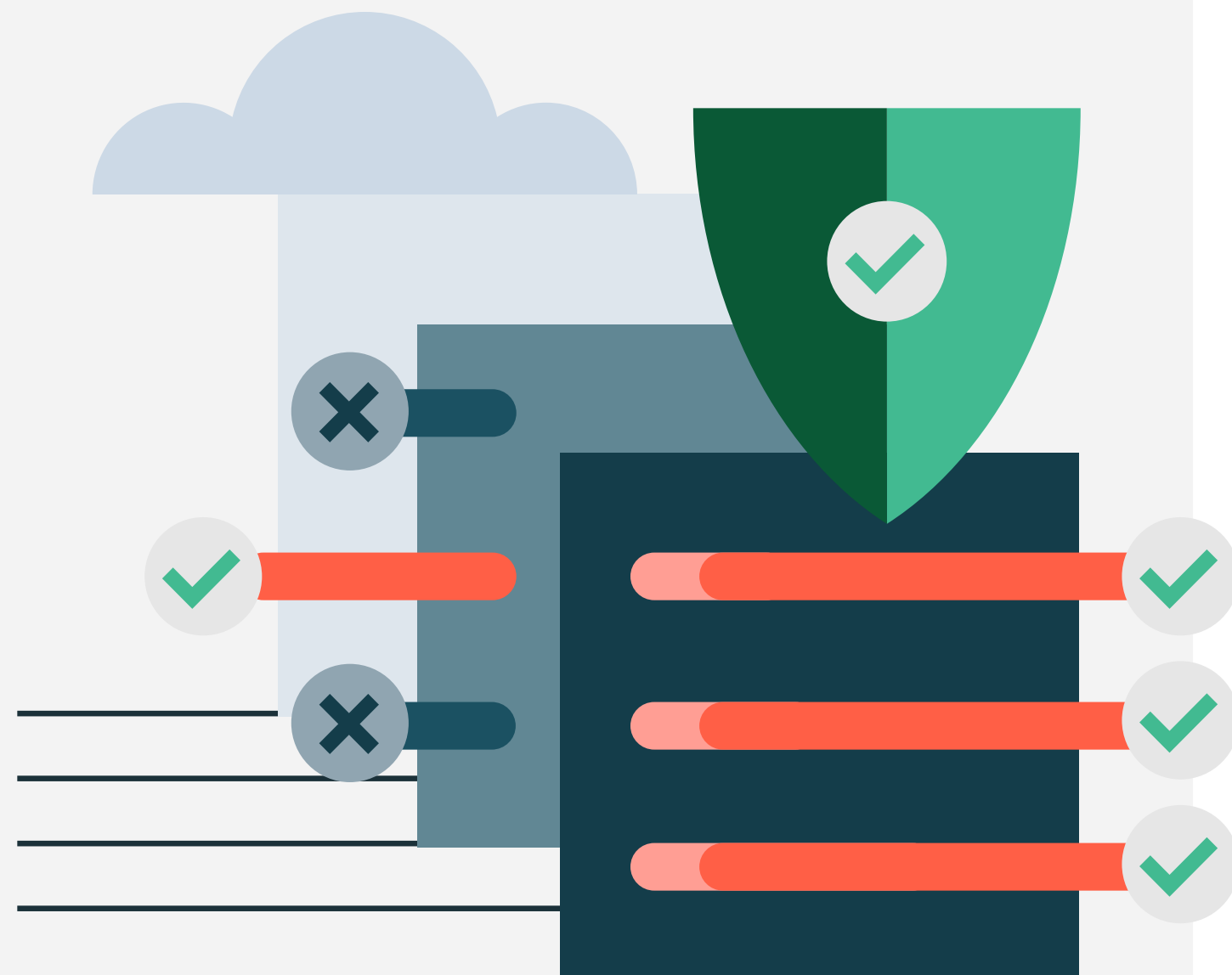
Modelo Hub&Spoke con Enfoque G: En este caso se introduce la fase de Gobierno centralizada, pero el proyecto es definido e implementado por el Spoke. El hecho de incorporar la capa de gobierno supone eficiencias en el resto de fases, pero sobre todo añade un nivel de garantía sobre la calidad de la información generada: coherencia con otras áreas, eliminación de redundancia de procesos, gestión del dato, etc. Dentro de un modelo Hub&Spoke se trataría de la estrategia más descentralizada.

Modelo Hub&Spoke con Enfoque GAC ineficiente: Este caso se trataría de un ejemplo en el cual se ha seguido una estrategia de centralización errónea. Se trata de un Spoke o un proyecto cuya máxima eficiencia, por diferentes motivos (capacidades técnicas del spoke, nivel de sinergias de los datos, etc) se obtiene cuando se aborda por el área de negocio ya que el Hub no genera sinergias en su implementación.

Modelo Hub&Spoke con Enfoque GAC eficiente: Este enfoque es el ejemplo opuesto al anterior. Se trata de un caso en el cual se ha seguido una estrategia de centralización que permite maximizar los recursos requeridos para el proyecto. Se trata de un Spoke o un proyecto en el que la estrategia de centralización por parte del Hub maximiza la eficiencia de recursos.



¿Qué variables son relevantes en la decisión?

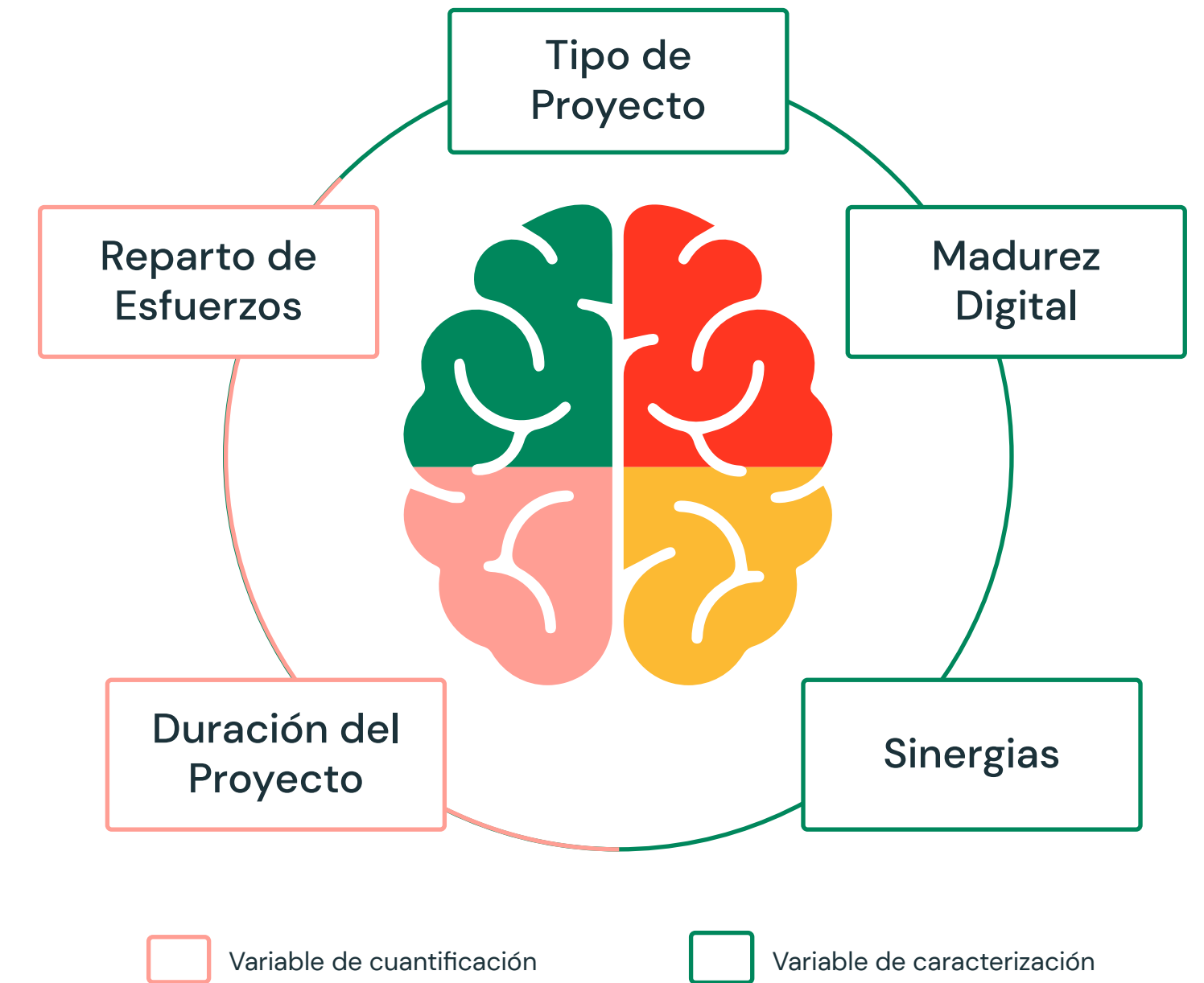


Input

Las variables de entrada que permiten a la herramienta estimar el mejor enfoque a seguir se dividen en dos tipologías:

- ▶ **Variable de caracterización:** Definen tanto el proyecto como las capacidades de las áreas implicadas en el mismo. Son determinadas y fijas para la herramienta.
- ▶ **Variable de cuantificación:** Son entradas modulables en conjunción con las necesidades y disponibilidad de recursos.

Por su parte, el enfoque a seguir en el proyecto, que también es una variable de entrada, y que como veremos luego tiene impacto en otros inputs, no se considera dentro de esta clasificación ya que no se define previamente sino que la aplicación va a realizar una simulación para las 3 opciones posibles y va a identificar aquella que genera una mayor eficiencia en proyecto.



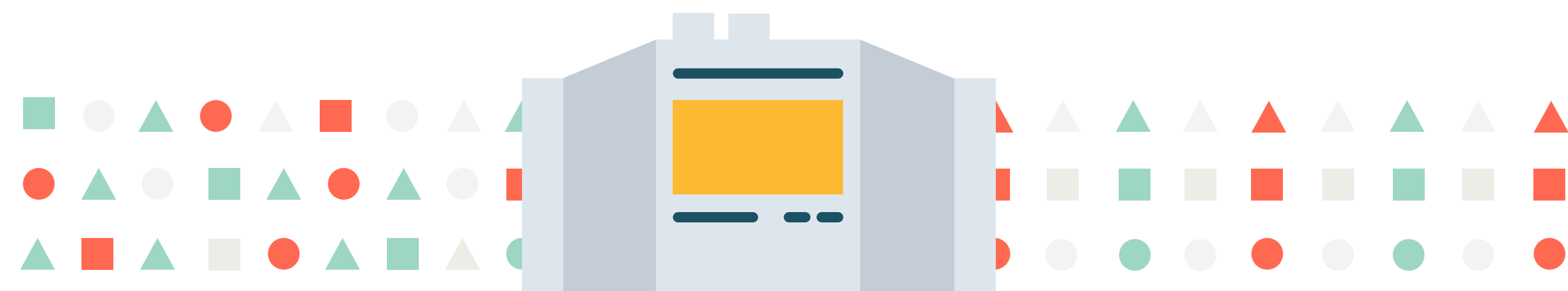
Input

Variables de caracterización

TIPO DE PROYECTO

En función del tipo de salida que genera un proyecto, las fases del mismo varían, y como consecuencia, lo hace la distribución de horas entre los diferentes intervinientes. Adicionalmente, la salida condiciona los perfiles técnicos requeridos para el desarrollo, que, unido a la madurez tecnológica del área de negocio, es un factor relevante a tener en cuenta. En este piloto se han identificado 3 tipos de salida:

- ▶ **Set de datos:** La información se pone a disposición de los usuarios en set de datos o Data Marts. El análisis de información se hace a través de consultas (o queries) o lenguajes de programación (analítica avanzada), por lo que este tipo de salidas estarían pensadas para usuarios con un nivel técnico elevado. Suelen ser las salidas más apropiadas para los modelos analíticos o motores de cálculo.
- ▶ **Plantillas:** El resultado final es el completado de plantillas o templates usualmente utilizados en el ámbito del reporte regulatorio. La plantilla se cumplimenta mediante procesos automáticos transparentes al usuario, que queda liberado de cualquier conocimiento técnico, pero que sin embargo participa en una fase adicional de parametrización de esta.
- ▶ **Cuadros de mando:** El output de la iniciativa es la generación de una aplicación en una herramienta de visualización de datos de forma que el usuario pueda hacer la explotación de los mismos, de forma autónoma y sin conocimientos técnicos avanzados. Requiere, por tanto, de desarrollos técnicos adicionales.



DATA MATURITY

El nivel de madurez del Spoke es uno de los factores más relevantes a la hora de descentralizar un proyecto informacional. Áreas de negocio avanzadas en proyectos de datos pueden asumir tareas técnicas con un alto nivel de eficiencia. De forma general, si tomamos, este factor de forma aislada el nivel de centralización es inversamente proporcional al nivel de madurez “informacional” del área de negocio. Medir este parámetro de forma objetiva siempre es complejo y ha sido objeto de debate en múltiples estudios (cita). Los factores más comunes que suelen citarse para el cálculo serían los siguientes:

- ▶ Porcentaje de datos del área documentados en el Diccionario de términos corporativo
- ▶ Porcentaje de las tareas del área relacionadas con trabajos informacionales
- ▶ Horas de formación de los integrantes del Spoke en BI
- ▶ Número de accesos a herramientas de explotación de datos por departamento
- ▶ Porcentaje de beneficios del área provenientes de herramientas de datos

Para la consideración sobre las competencias técnicas de las áreas de negocio también ha de considerarse la diversidad de herramientas utilizadas durante el ciclo de vida del dato. El traspaso de información entre herramientas no diseñadas a funcionar conjuntamente o los intercambios continuos entre ellas para obtener un resultado es una fuente de entropía que evitará la consecución de una eficiencia óptima. Sin embargo, éste es un factor externo a la calculadora en estudio, que amerita un debate que supera el alcance del artículo.

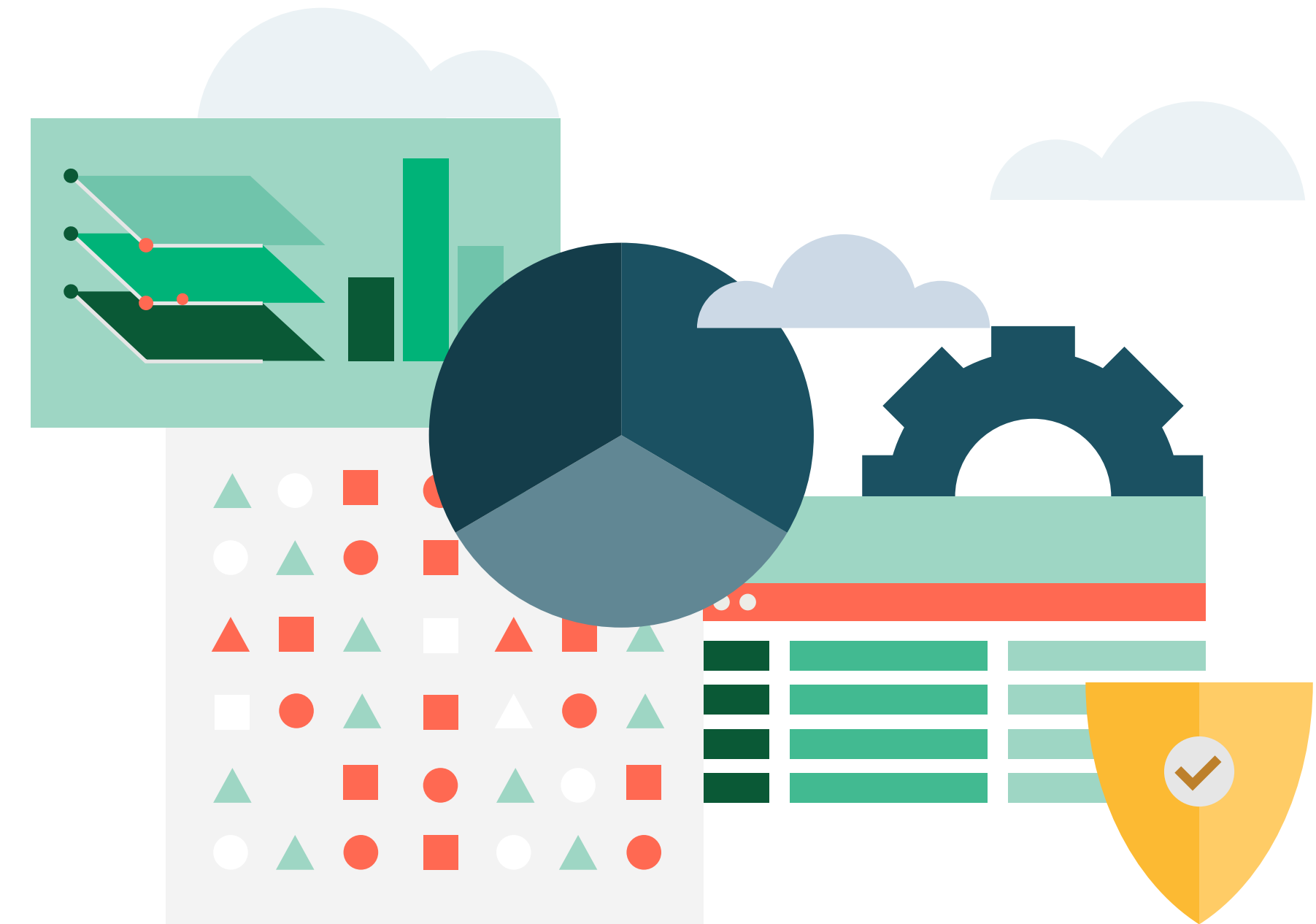
SINERGIAS

Son el grado de eficiencia generado por la centralización de la información en un Hub común. Deriva del conocimiento que el Hub adquiere, como consecuencia de sus tareas de gobierno y de la diversidad de los proyectos informacionales abordados. El principio básico que genera las sinergias es la Reusabilidad de la información. Incluso para las variables no gobernadas, el nivel de **dispersión** de la información condiciona las sinergias. Proyectos que requieren de amplia variedad de fuentes o ámbitos generan una mayor eficiencia desde un enfoque centralizado en un Hub que proyectos con una concentración de la información. Esto es así porque si bien el Hub no tiene gobernada toda la información, si tiene un mayor conocimiento de la distribución de la misma en la entidad.

Adicionalmente, estas sinergias son mayores con el tiempo, al ir creciendo el número de variables gobernadas. De forma inversa al Nivel de Madurez del Spoke, el nivel de sinergias es directamente proporcional al grado de centralización.

Para calcular las sinergias del proyecto hay que tener en cuenta dos factores:

- ▶ **Grado de sinergia global del proyecto:** La forma más fiable de calcular objetivamente esta variable es hallar el porcentaje de datos que se requieren y que ya están siendo gobernados con respecto al total de variables del proyecto. De esta forma, podemos saber el nivel de reusabilidad en que se podría apalancar el Hub.
- ▶ **Reparto de sinergias del proyecto:** La existencia de sinergias en un proyecto informacional no implica un ahorro de esfuerzo homogéneo en toda la vida del proyecto, sino que beneficia especialmente en algunas de sus fases (definición funcional) y no tiene impacto alguno en otras (construcción de un Cdm). Estimar un peso objetivo de las sinergias en cada fase ha de ser una ponderación realizada en base a la experiencia en la realización de proyectos en cada entidad.



Input

Variables de cuantificación

DURACIÓN DEL PROYECTO

La duración del proyecto (sin consideración de ningún tipo de sinergias) depende esencialmente de dos factores principales:

- ▶ **Tipo de proyecto:** En función de si la salida a generar es un set de datos, una parametrización de un informe regulatorio o un cuadro de mando en una herramienta de BI, las tareas a realizar pueden suponer un coste en horas u otro.
- ▶ **Número de variables:** El número de datos a utilizar en el proyecto condiciona el tamaño del proyecto y por tanto el esfuerzo necesario para su culminación.
- ▶ **Existencia de solución táctica:** En caso de que ya existan procesos departamentales o manuales que estén sustituyendo a la versión automatizada corporativa que se está planteando, el esfuerzo dedicado será sustancialmente menor que en el caso de que el proyecto se aborde desde cero.

De forma aproximada, y en base a una parametrización sobre estos factores, se puede clasificar los proyectos en 3 tipologías (duración Alta, Media y Baja) cada uno de ellos con un número de horas total diferente.



REPARTO DE ESFUERZOS

Reparto de esfuerzos por fase: Considerando un esfuerzo total en un proyecto, el reparto de pesos que se estime entre las diferentes fases (definición funcional, construcción, gobierno, validación, etc) y subfases condiciona los beneficios que obtiene cada área involucrada, ya que participa de forma diferente en cada una de ellas. En este sentido, el reparto está condicionado al tipo de salida generada, que como se ha comentado anteriormente, condiciona las mismas fases del proyecto y ha de calcularse en base a la experiencia.

Reparto de esfuerzos por rol: El reparto de esfuerzos tiene dos factores claves.

- ▶ En un primer lugar ha de **definirse una matriz de responsabilidades (RACI)** para cada uno los enfoques posibles. De esta forma se puede medir el nivel de participación de cada interviniente en cada tarea. Una vez definida, en función del enfoque se asignará la distribución correspondiente.
- ▶ El segundo factor es **el reparto de esfuerzos en una tarea en base a la responsabilidad** que se le ha asignado a un interviniente. Es decir, dada una tarea con un esfuerzo asignado X, este ha de repartirse entre todos los intervinientes que participan en ella de acuerdo a su tipo de participación (accountability, responsibility, consulted o informed).

RESUMEN ESFUERZOS

Output

La salida generada por la herramienta determina la distribución de horas necesarias en un proyecto para cada enfoque posible así como la distribución de responsabilidades entre los intervinientes. De esta forma, como ya hemos comentado anteriormente, se identifica como óptimo aquel que suponga un menor requerimiento de recursos totales para su ejecución cumpliendo ciertas restricciones.

Enfoque	% CARGA			HORAS POR PROYECTO Y EQUIPO		
	Hub	Spoke	Ahorro	Hub	Spoke	Total
G	20,00%	75,00%	5,00%	2.520	9.450	12.000
GA	30,00%	63,00%	7,00%	3.780	7.938	11.718
GAC	70,00%	18,00%	12,00%	8.820	2.268	11.088

Enfoque óptimo

* Enfoque óptimo siempre y cuando el Hub pueda asumir la carga de proyecto asignada (8.820 horas)

ENFOQUE: GAC	DEFINICIÓN FUNCIONAL				TAREAS DE GOBIERNO			ENTREGABLE			
Matriz RACI	Análisis	Levantamiento	Asignación de responsable	Definición Spoke	Diccionario	Modelo semántico	Gobierno del dato/uso	Generación DataSet	CdM	Informe	Validaciones
R	Hub	Hub	Hub	Hub	Spoke	Hub	Hub	Hub	Hub	Hub	Hub
A	Spoke	Spoke	Spoke	Hub	Hub	Hub	Hub	Hub	Hub	Hub	Spoke
C											
I						Spoke	Spoke	Spoke		Spoke	

Para la obtención de unas salidas con un alto grado de certidumbre y confianza, cada una de las entradas ha de ser calibrada correctamente en la herramienta. Esta parametrización ha de adaptarse a la entidad de uso y ha de basarse en la experiencia pasada y en la mejora continua.

ENTRADAS

- ▶ Tipo de proyecto
- ▶ Data Maturity del área
- ▶ Sinergias del proyecto
- ▶ Duración del proyecto
- ▶ Reparto de esfuerzo por fase
- ▶ Reparto de esfuerzo por rol

CALIBRADO

- ▶ Asignación de pesos y parámetros internos



SALIDAS

- ▶ Resumen de esfuerzos por área por tipo de enfoque
- ▶ Distribución de responsabilidades en el proyecto para el enfoque seleccionado

Caso de uso

A continuación se muestra el resultado de varias simulaciones para ilustrar como el enfoque adecuado varía de acuerdo las circunstancias de cada caso.

Supongamos dos Spokes diferentes (Área de negocio A y Área de negocio B) y dos proyectos diferentes (Proyecto 1 y proyecto 2) y veamos como las diferentes combinaciones condicionan el enfoque a seguir por el Hub.

En primer lugar, veamos las ponderaciones que se han asignado a cada uno de los inputs comentados en la sección anterior:

PROYECTO	NÚMERO DE VARIABLES	TIPO DE PROYECTO	SINERGIAS GLOBALES DEL PROYECTO (0-1)	OBSERVACIONES
Poyecto 1	850	Cuadro de Mando	0,8	Alto porcentaje de variables requeridas en el proyecto ya gobernadas por el Hub
Poyecto 2	125	Data set	0,4	Bajo porcentaje de variables requeridas en el proyecto ya gobernadas por el Hub

SPOKES	DATA MATURITY (0-1)	OBSERVACIONES
Área de negocio A	0,8	Área muy madura en análisis y visualización de datos
Área de negocio B	0,2	Área poco madura en análisis y visualización de datos

Caso de uso

CASO 1: ÁREA DE NEGOCIO A – PROYECTO 1

En este caso tenemos un Área muy competente técnicamente que tiene que abordar la elaboración de un Cuadro de Mando con altas sinergias con información gobernada por el Hub.

TIPO DE PROYECTO	SINERGIAS GLOBALES DEL PROYECTO
Cuadro de Mando	0,8
NÚMERO DE VARIABLES	DATA MATURITY
850	0,8

Este es un caso muy paradigmático del modelo Hub&Spoke y un ejemplo perfecto de la necesidad de tener una metodología clara para la decisión de enfoques de colaboración. Contamos con un spoke que es competente en la elaboración tanto de análisis de datos como en el desarrollo de aplicaciones de visualización de datos pero a su vez se trata de un proyecto donde el Hub tiene un gran conocimiento sobre la información a utilizar. En estos casos, la calibración de la herramienta para cada uno de los parámetros de entrada será fundamental y el trabajo de ajuste continuo en base a la experiencia de los proyectos abordados clave.

El reparto de esfuerzo y en el enfoque ideal para este caso serían los siguientes:

RESUMEN ESFUERZOS						
	% CARGA			HORAS POR PROYECTO Y EQUIPO		
Enfoque	Hub	Spoke	Ahorro	Hub	Spoke	Total
G	20,00%	75,00%	5,00%	2.400	9.000	11.400
GA	30,00%	64,50%	5,50%	3.600	7.740	11.340
Enfoque óptimo GAC	90,00%	4,30%	5,70%	10.800	516	11.316

Como vemos, el crecimiento máximo de la ganancia se obtiene al realizar el Gobierno del proyecto, al maximizar las ventajas resultantes de las sinergias con otras líneas de trabajo. Para el análisis y la construcción existe un crecimiento lento pero positivo de forma que es con la realización de la construcción donde los recursos necesarios son mínimos. Sin embargo la dedicación de recursos por parte del HUB es muy grande por lo que tiene que haber disponibilidad para abordar este enfoque.

Caso de uso

CASO 2: ÁREA DE NEGOCIO A – PROYECTO 2

En este caso tenemos un Área muy competente técnicamente que tiene que abordar la elaboración de un Data Set con unas sinergias de información gobernada por el Hub no muy elevadas:

TIPO DE PROYECTO	SINERGIAS GLOBALES DEL PROYECTO
Cuadro de Mando	0,4
NÚMERO DE VARIABLES	DATA MATURITY
850	0,8

El output de la calculadora es el siguiente:

RESUMEN ESFUERZOS						
	% CARGA			HORAS POR PROYECTO Y EQUIPO		
Enfoque	Hub	Spoke	Ahorro	Hub	Spoke	Total
Enfoque óptimo → G	20,00%	75,00%	5,00%	100	375	475
GA	30,00%	65,40%	4,80%	150	327	477
GAC	95,30%	4,34%	4,70%	462	22	483

Caso de uso

Como se ve en el primer cuadro de este caso, en la tarea de Análisis la eficiencia del proyecto decae. Esto es porque al tener el Spoke unas competencias técnicas altas, el grado de sinergia del proyecto (relativamente bajo) hace que este sea más eficiente que el hub en la fase de análisis de información y que esta caída no se compense en la subsiguiente tarea.

Vemos por tanto que el enfoque a seguir para este spoke ha cambiado con respecto al que se ha seguido para el Proyecto 1, siendo más conveniente el utilizar una estrategia más descentralizada en este caso.

CASO 3: ÁREA DE NEGOCIO B – PROYECTO 2

En este caso tenemos un Área muy poco competente técnicamente que tiene que generar un Informe predefinido para un regulador con unas sinergias de información gobernada por el Hub no muy elevadas:

TIPO DE PROYECTO	SINERGIAS GLOBALES DEL PROYECTO
Informe	0,4
COMPLEJIDAD	DATA MATURITY
Media	0,2

Es decir, es un caso análogo al Caso 1 pero a la inversa. Las sinergias bajas hacen que el Hub sea poco eficiente pero a pesar de ello, las pocas competencias técnicas del Área de negocio hacen que cualquier tipo de análisis de la información o las construcciones posteriores sean muy costosas para el Spoke. La decisión de qué criterio pese más se toma en función de la calibración de la herramienta en base a la experiencia de proyectos previos.

Para este caso, la configuración actual proporcionaría los siguientes resultados:

RESUMEN ESFUERZOS						
Enfoque	% CARGA			HORAS POR PROYECTO Y EQUIPO		
	Hub	Spoke	Ahorro	Hub	Spoke	Total
G	20,00%	78,00%	2,00%	100	390	490
GA	30,00%	57,70%	9,30%	165	289	454
GAC	82,00%	2,30%	15,70%	410	12	432

Enfoque óptimo

Como vemos, el enfoque seleccionado en este caso es GAC. El motivo es que pesa más la falta de competencias técnicas de el Área de negocio que las pocas sinergias de variables con respecto al perímetro de información gobernado desde el Hub. Como consecuencia, el porcentaje de ahorro de carga de trabajo crece en mayor medida en la realización las tareas más técnicas.

Consideraciones Tecnológicas

5.1 LA FRICCIÓN TECNOLÓGICA

Como hemos visto anteriormente un factor crítico para el éxito se basa en la capacidad de adaptar de una forma ágil y eficiente el modelo descrito anteriormente en función de diferentes factores como la madurez del área, el tipo de proyecto, sinergias y otros factores.

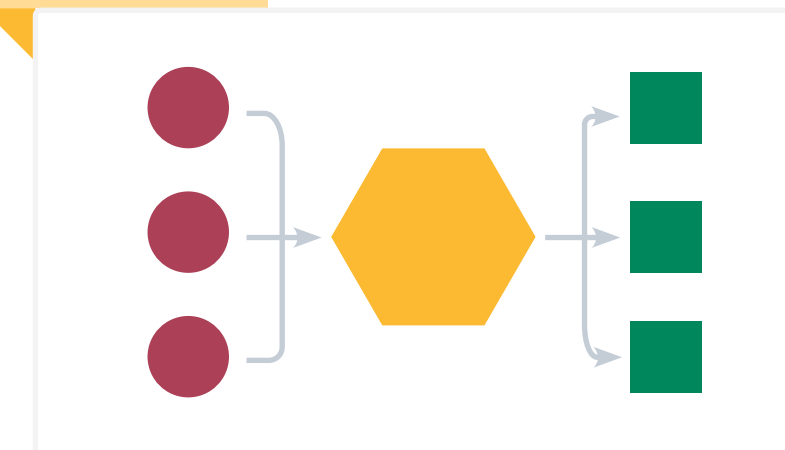
Por tanto se vuelve muy relevante el disponer de una arquitectura tecnológica que eliminen los principales lastres para aumentar la agilidad y productividad. Muchas empresas han identificado como principal capacidad de resistencia a dicha elasticidad de adaptación (entre una aproximación más o menos centralizada) a la propia estructura organizativa y los puntos de sincronización de la información entre sus partes. A parte de la propia estructura organizativa en sí, muchas empresas han definido su modelo organizativo de los equipos de datos en línea con la propia arquitectura tecnológica que influye directamente en la interacción y comunicación.

Así, una gran parte de las empresas posee hoy día arquitecturas de datos tradicionales implementadas durante las dos últimas décadas (ya sean warehouses, data lakes o la combinación de ambas) caracterizadas por ser monolíticas y centralizadas, por lo que existe una fricción tecnológica que frena no solo la innovación y la capacidad de adaptar el modelo sino que además fomenta la súper especialización de los equipos creando cuellos de botellas en los diferentes puntos de control entre los diferentes equipos de datos y/o unidades de negocio.

Otra empresas usan silos tecnológicos que dificultan una visión del estado del arte, falta de consistencia e incluso una hiper-especialización aún mayor que en el modelo centralizado.

En estos formato monolíticos o de silos los equipos se organizan segmentados en cadena donde los diferentes eslabones de la cadena del dato suele ser perfiles especializados en tareas y herramientas concretas mapeados a silos tecnológicos de cada tarea como herramientas de integración, ingeniería de datos, analistas de datos, científicos de datos, ingenieros de ML. Es común la falta de eficiencia por la duplicación lo que al final se traduce en un incremento de costes y una notable deuda técnica.

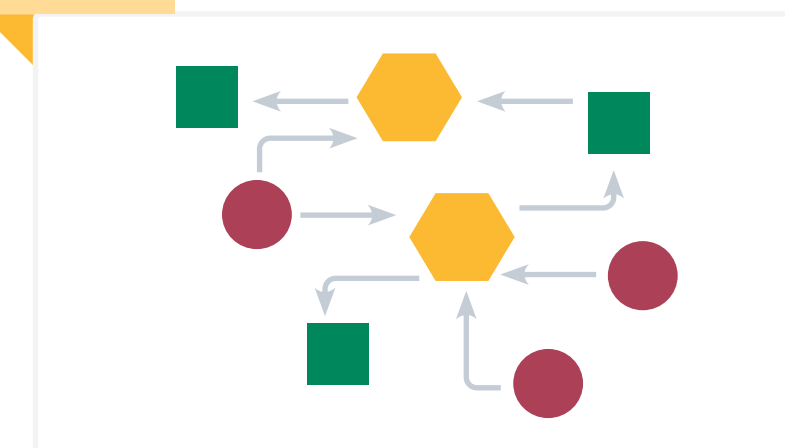
Centralizada



Limitaciones

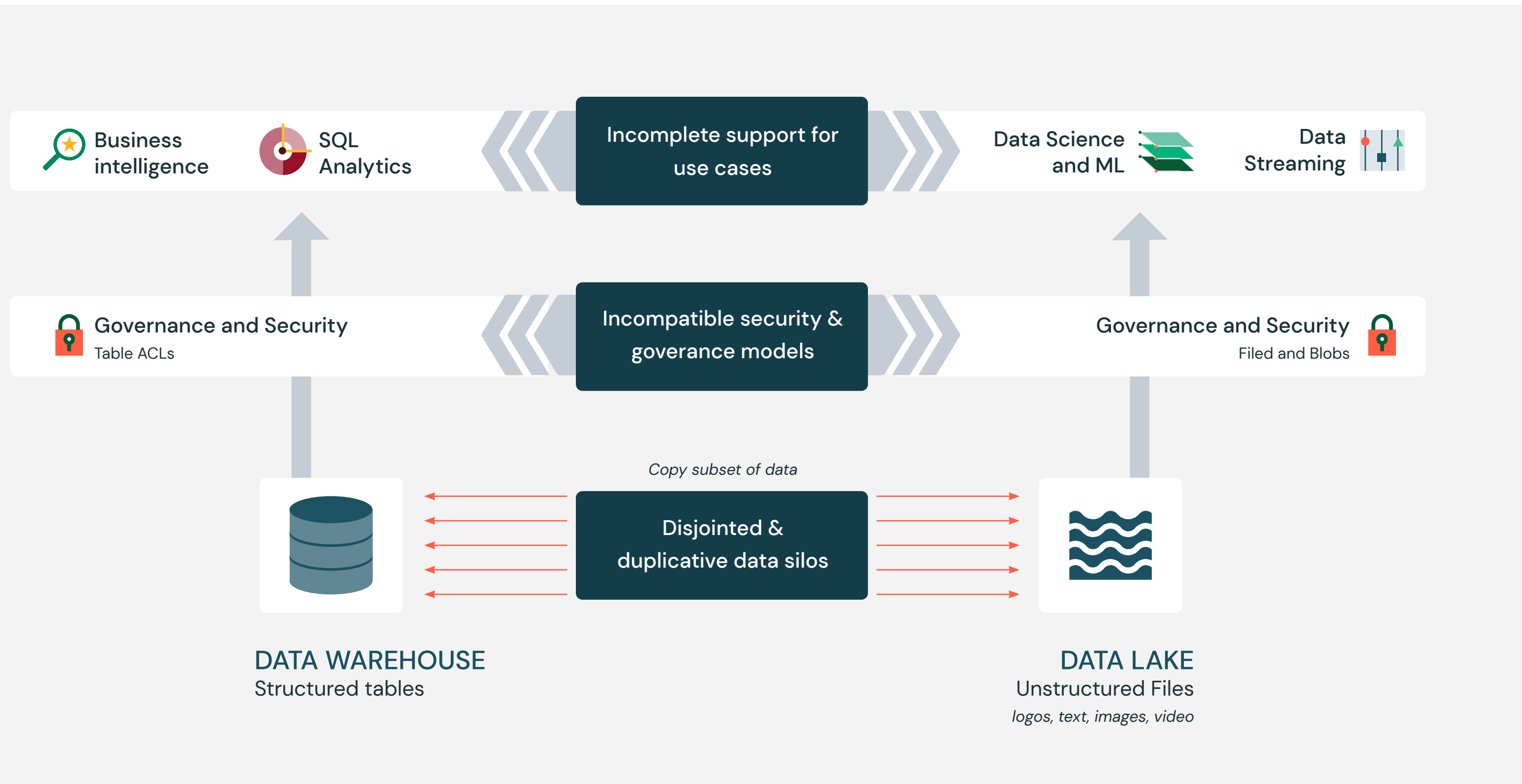
- ▶ Frena La innovación
- ▶ Monolítico
- ▶ Fomenta la especialización
- ▶ Coste por políticas entre unidades de negocio

En Silos



Limitaciones

- ▶ Poca visibilidad global del estado
- ▶ Falta de consistencia, mala calidad
- ▶ Hiper-especialización
- ▶ Duplicación de tareas y roles
- ▶ Incremento del coste
- ▶ Crea una deuda técnica



5.2 LIMITACIONES COMUNES TECNOLÓGICA

Parece claro por tanto que cara a poder adaptar el modelo de una forma elástica y eficiente de cara a obtener de los datos el valor de negocio necesitamos definir un nuevo paisaje tecnológico en consonancia con el modelo operativo y organizativos.

Los datos se han colocado al frente y al centro como un motor del negocio y determinan cada vez más si una empresa tiene éxito o fracasa. Los sistemas para administrar y procesar datos no han evolucionado con el crecimiento explosivo de los datos y la llegada de nuevos modelos como el aprendizaje automático y la IA. Los datawarehouses funcionan bien para analítica tradicional pero son una tecnología surgida ya hace 40 años que no fue diseñada para un mundo donde la escala masiva, nuevos tipos de formatos y los nuevos casos de uso como ML son la norma. Los lagos de datos ofrecen mayor flexibilidad y eficiencias económicas, pero no satisfacen las necesidades de calidad y rendimiento exigidas y tienden a convertirse en poco más que volcados de datos. ¿El resultado? Un mundo donde las empresas improvisan soluciones híbridas que combina ambos mundos generando complejidad e ineficiencias.



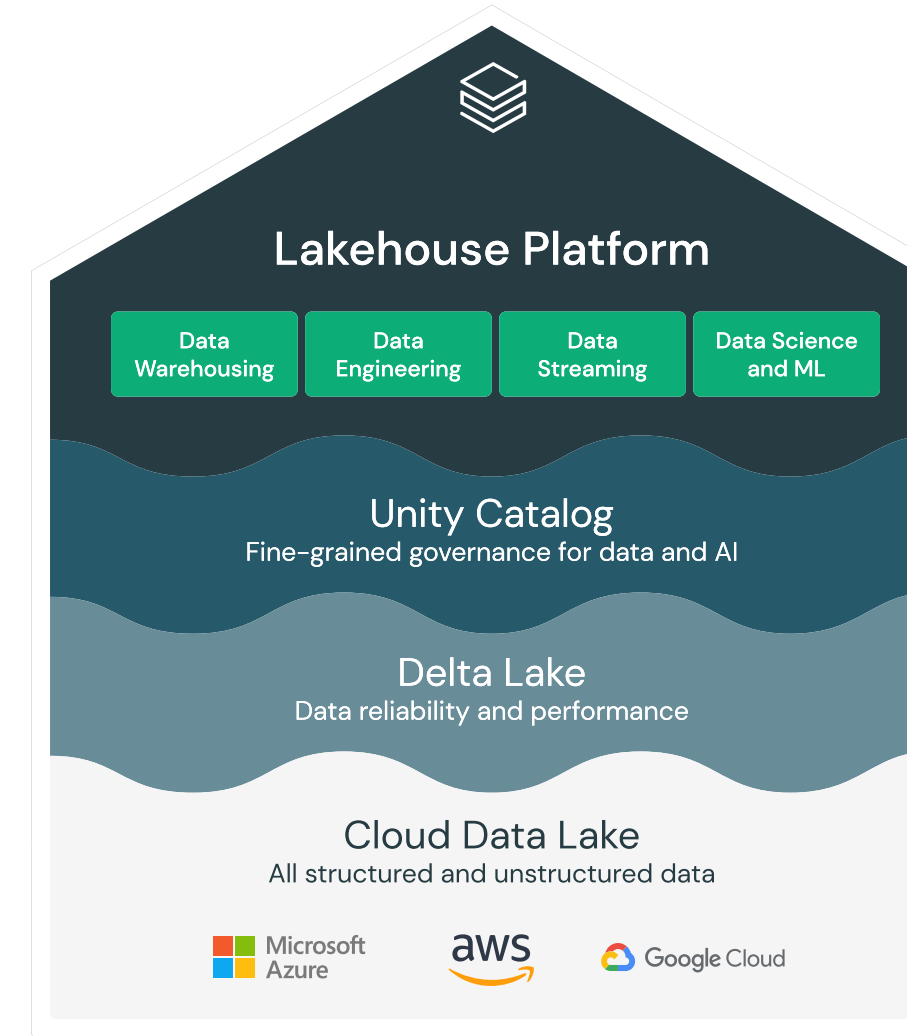
5.3 UN NUEVO PARADIGMA: LAKEHOUSE

Lakehouse es un paradigma emergente en el ecosistema de las plataforma de datos nacido en la nube y que marca la próxima generación, donde se combina lo mejor de los data warehouses (como fiabilidad y el rendimiento) con la flexibilidad y economía de los lagos de datos en la nube. Así, una aproximación Lakehouse simplifica la plataforma de datos y unifica los mundos de la analítica tradicional con la analítica avanzada.

Estas arquitecturas dotan a los equipos de datos de flexibilidad y elasticidad para cubrir múltiples casos de uso en una sola plataforma. A destacar la capacidad de gestionar y explotar cualquier formato de dato con fiabilidad y calidad y dentro de un

modelo de gobierno unificado entre todos los equipos de datos para BI, ML e IA lo que acelera no sólo la obtención del valor. Esto se traduce en una aproximación más simple y colaborativa obteniendo mayor agilidad y reducción de costes frente a las plataformas de datos tradicionales.

Por último destacar la flexibilidad de usar una aproximación lakehouse permite centralizar todas aquellas capacidades genéricas y/o compartidas por los dominios de datos entre áreas (acceso autoservicio a cómputo, seguridad y gobierno para garantizar el cumplimiento regulatorio) y al mismo tiempo la flexibilidad para con una misma arquitectura cubrir las diferentes etapas de la vida y gobierno tanto de los datos como del código y modelos producidos ya sean para analítica tradicional o avanzada.



5.3 EJEMPLOS

En Databricks vimos que el cambio a una arquitectura lakehouse se avecinaba durante mucho tiempo, dado que las necesidades de nuestros clientes en cuanto a análisis e IA convergían y su arquitectura actual era demasiado complicada para mantenerse al día.

Como creadores de esta nueva categoría Lakehouse y siendo la empresa de datos e inteligencia artificial que colabora con más de 5,000 organizaciones a nivel mundial — como Comcast, Condé Nast, H&M, y más del 40% de las Fortune 500 — y como creadores de esta nueva categoría Lakehouse vemos a diario como no solo nuestros clientes han confiado en un Lakehouse para unificar sus datos, la analítica

e IA sino como el propio ecosistema de proveedores de tecnología ha comenzado a adoptar el Lakehouse como la nueva generación de plataforma de datos como su reconocimiento entre analistas como Gartner.

Conclusión

Las compañías sitúan cada día al dato en una posición más central dentro de su modelo de negocio. Como resultado, se dedican cada vez más recursos al ecosistema que rodea al ámbito informacional (IA, modelos analíticos, EDW, modelos predictivos, Cuadros de Mando, etc), ya sea en inversión en tecnologías o en capital humano (formación, contrataciones de perfiles específicos, etc).

Es por tanto fundamental que dicha apuesta se maximice mediante un uso eficiente de estos recursos. Como contribución a este esfuerzo, hemos tratado de diseñar una metodología para la selección del enfoque más adecuado dentro de un modelo Hub&Spoke.

Esta metodología, materializada en una herramienta de ayuda a la toma de decisiones, permite, según se ha visto, adecuar el grado de centralización ideal para un área de negocio o Spoke determinado dado un proyecto específico. Para apoyar lo aquí expuesto, se ha ilustrado su funcionamiento con una serie de casos de negocio en los que se ha podido comprobar como las características de los proyectos y de las áreas sponsorizadas por el Hub condicionan el enfoque a seguir, sin ser ésta una relación rígida que podamos conocer de antemano.

De esa forma podemos establecer un criterio objetivo adicional, a considerar en unión con otros (criticidad del proyecto, capacidad de recursos,...) para encontrar la solución idónea para cada situación.





<https://clubcdo.com/>



<https://databricks.com/>

