



Informatica™

El gran cuaderno del Big Data

Una guía práctica para emprender
su primer proyecto de Big Data

Contenidos

Algo grande	03	Parte 3: Su eficaz cadena de suministro de datos	30
Parte 1: Preparación	04	Su equipo	31
Información imprescindible	05	– Cinco lecciones básicas para formar equipos	32
– Motivos por los que casi todas las empresas implementan proyectos de Big Data	06	– Establecimiento de la gobernanza de datos	35
– Motivos por los que fracasan algunos proyectos de Big Data	08	– Competencias que necesita y competencias que posee	37
– Consejos para conseguir que su proyecto de Big Data funcione	10	– Examen de las herramientas del Big Data	39
Selección del proyecto adecuado	12	Sus procesos	40
– Aspecto del proyecto adecuado	13	– Los ocho pasos del Big Data	41
– Consideración de las repercusiones	15	Su arquitectura	43
– Proyectos tácticos de Big Data: algunos ejemplos	17	– Primeros pasos: zona de pruebas	44
Parte 2: Su estrategia	18	– Arquitectura idónea del Big Data	45
Definición de sus objetivos	19	Plan de su proyecto	46
– Objetivos de negocio	20	– En marcha	49
– Objetivos de TI	22	Siguientes pasos	50
Definición de sus necesidades de datos	24	About Informatica®	51
– Datos necesarios	25		
– Cinco consideraciones clave en torno a los datos	28		

Consejo: haga clic aquí para ir directamente a cualquier sección.



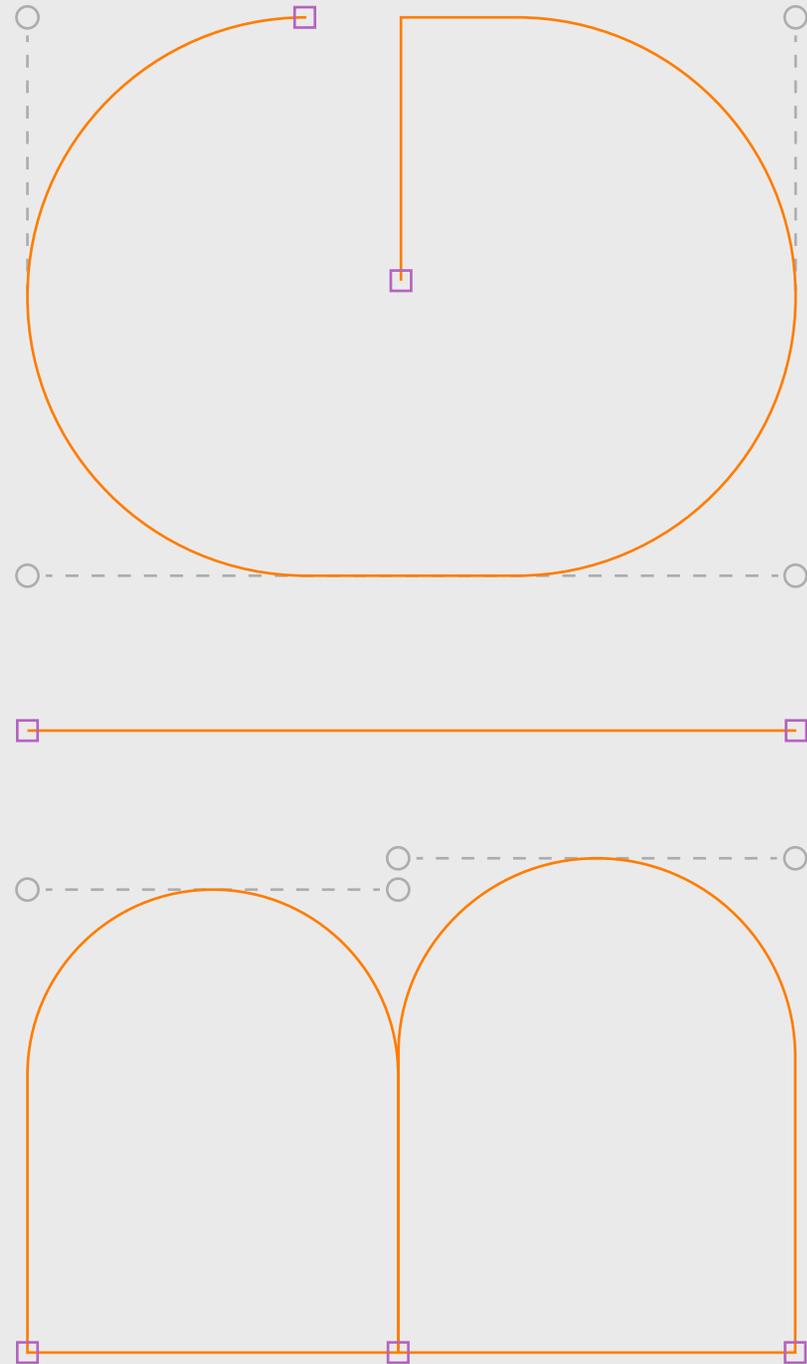
Algo grande

Pocas tendencias tecnológicas han logrado la fama que ha alcanzado el Big Data.

Claro que pocas tendencias tecnológicas han ofrecido a las empresas tanto potencial de transformación. Desde que, a finales de siglo, el software empezó a envolver por completo los procesos de negocio, ha estado claro: los datos cambian nuestra forma de trabajar.

Tanto si empieza una iniciativa táctica localizada como si planea una iniciativa de cimentación que abarque toda la empresa, este cuaderno le servirá como guía práctica para la transición.

Vamos al asunto.



Parte 1

Preparación

Hemos dividido el cuaderno en tres partes. En esta primera parte, vamos a intentar afinar su visión para que escoja el proyecto adecuado.

Información imprescindible

Antes de abordar los aspectos concretos de su propio proyecto, presentamos algunas lecciones que la mayoría de los profesionales del Big Data habría deseado conocer antes de iniciar sus proyectos.



Motivos por los que casi todas las empresas implementan proyectos de Big Data

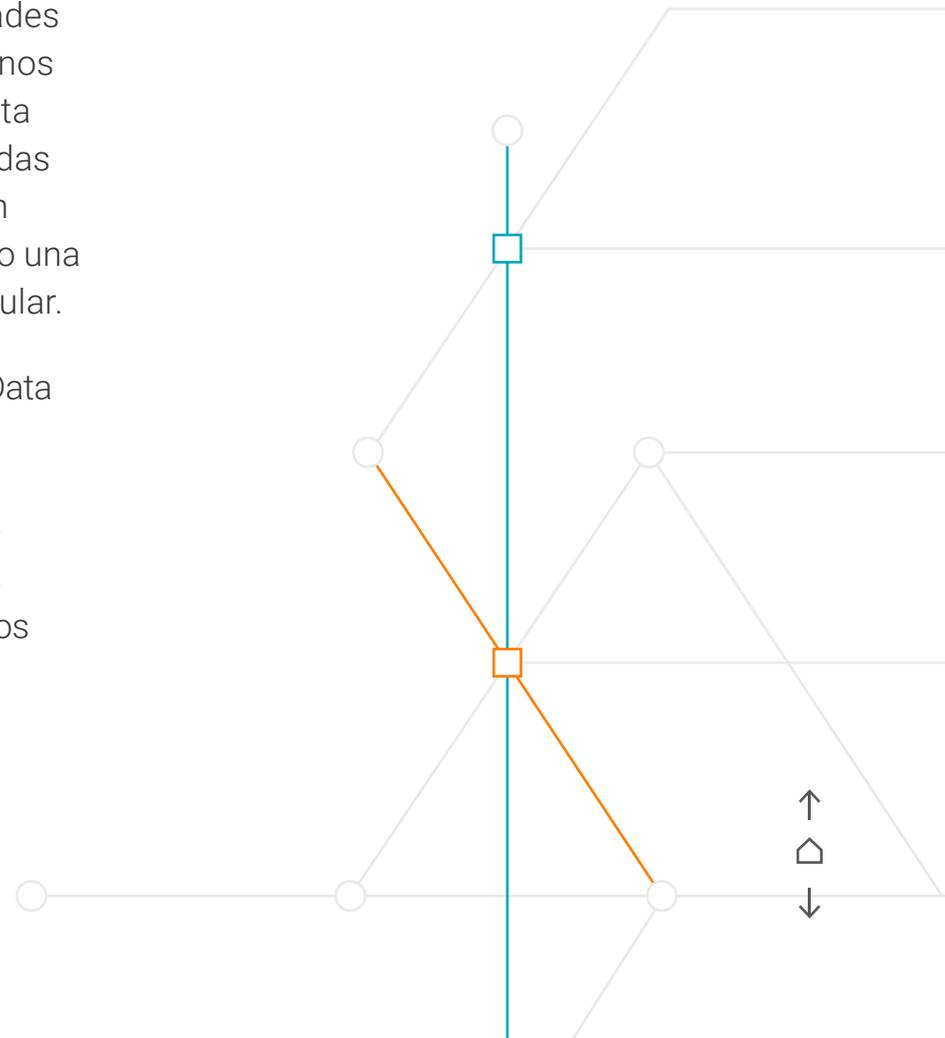
Cuando las empresas se deciden a abordar el Big Data, suele ser por uno de estos motivos.

Intentan llevar a cabo mejores análisis y se dan cuenta de que, para conseguirlo, hace falta una cantidad bastante superior de datos de análisis. Por lo general, es alguna unidad de negocio (como marketing) la que empieza estas iniciativas.

Se dan cuenta de que pueden envolver sus productos con una capa de servicios ofreciendo análisis (a menudo, en tiempo real) que ayuden a los clientes a utilizar sus productos con más eficiencia y eficacia.

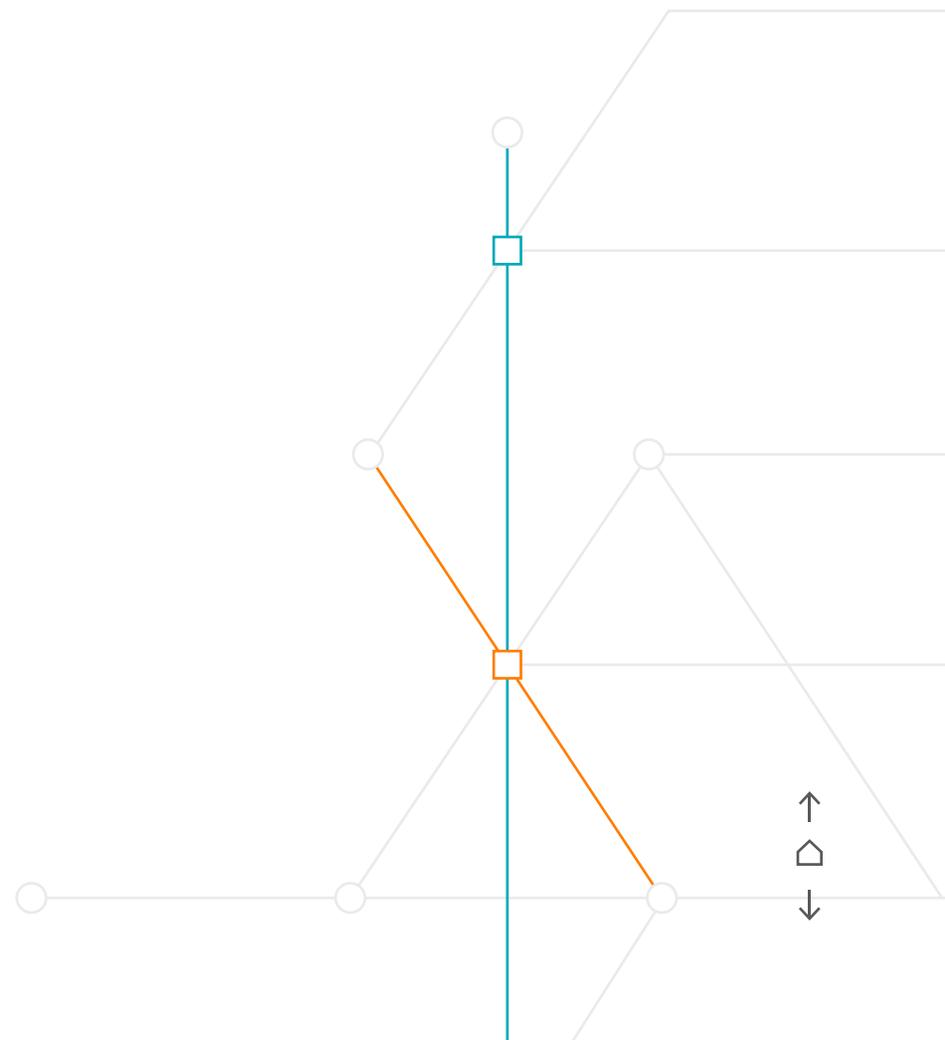
Quieren realizar sus actividades mejor, más rápido y con menos gastos empleando el Big Data con el fin de documentar todas las decisiones que se toman en relación con un proceso o una unidad de negocio en particular.

Se dan cuenta de que el Big Data es fundamental para todas las unidades de negocio de la organización, por lo que intentan establecer las bases para una visión centrada en los datos de todo su mundo.



Saben que deben aprender a manejar el Big Data antes de que sea demasiado tarde, pero todavía no saben cómo hacerlo. El objetivo consiste en aprender y experimentar.

Aunque todos son buenos motivos para interesarse por el Big Data, si desea que sus proyectos resistan el paso del tiempo (y el escrutinio de los distintos departamentos), ha de tener muy claro el motivo que mejor refleja sus intereses.



Motivos por los que fracasan algunos proyectos de Big Data

Según un estudio¹, el 55% de todos los proyectos de Big Data no se termina y muchos otros no cumplen sus objetivos.

Aunque este llamativo porcentaje no es infrecuente en una fase tan temprana de una tendencia tecnológica, sería una insensatez pasar por alto la lección que imparten esos proyectos.

Veamos los cuatro motivos principales por los que fracasan los proyectos de Big Data.

1. Objetivos difusos

El motivo de fracaso que más se cita en el estudio es el "ámbito de aplicación impreciso" del proyecto. Demasiadas empresas desarrollan proyectos ambiciosos (y, al mismo tiempo, demasiado ambiguos) que carecen de objetivos claros, y que luego terminan fracasando cuando llega el momento de tomar decisiones complicadas sobre lo que reviste importancia y lo que no.

2. Expectativas equivocadas

Con toda la fama del Big Data, hay quien llega a algunas suposiciones muy arriesgadas en cuanto a lo que puede ofrecer el proyecto. Aunque resulte tentador hacer promesas valientes para plazos breves, es importante mantener una visión realista de lo que cabe esperar del proyecto, el tiempo que va a necesitar y la cantidad de esfuerzos que requiere. Cuando las expectativas del impacto y de la información son demasiado elevadas, acaba

buscando petróleo en un mar de terabytes de incógnitas. Si las expectativas de la oferta no son realistas, se encontrará intentando cumplir plazos y presupuestos nada razonables.

3. Aumento de costes y retrasos

Si se tiene en cuenta lo novedosa que es esta disciplina para la empresa, no sorprende que casi todos los proyectos de Big Data terminen costando o tardando más de lo previsto. Ello se suele deber a una mezcla de expectativas irreales e incomprensión del procedimiento para crear una arquitectura escalable. En los casos en que se contrata a los escasos y bien remunerados desarrolladores de Java para Hadoop, a quienes se les encomiendan titánicas implementaciones de codificación manual, las empresas no tardan en asumir que resulta imposible salir del entorno de pruebas sin ningún error. La consecuencia es que los proyectos de Big Data terminan como un experimento científico que languidece en el laboratorio y nunca llega a ver la luz.

¹ InformationWeek, "Vague Goals Seed Big Data Failures."



4. Incapacidad de escalar

Con demasiada frecuencia, las empresas se fijan más en la conveniencia a corto plazo que en la sostenibilidad a largo plazo. Aunque no tendría sentido sugerir que se evite siempre esa contrapartida, nunca nos cansaremos de insistir en la importancia de la visión a largo plazo. Para que los datos cuenten con la protección y la gestión apropiadas, es imprescindible supervisar las implicaciones a largo plazo del proyecto.

Las cuatro causas del fracaso del Big Data son preocupantes y demasiado habituales. A continuación, veremos cómo puede evitarlas y crear una implementación duradera.

Consejos para conseguir que su proyecto de Big Data funcione

En vista de que la mayoría de los proyectos de Big Data fracasa por la falta de claridad y la incapacidad para demostrar la utilidad de la iniciativa, debe asumir la tarea de aportar al proyecto el enfoque y las pruebas pertinentes. Siga estos tres útiles consejos para asegurarse de que el proyecto inicia y continúa su andadura.

1. Marque objetivos claros y contenga las expectativas

Si no sabe con certeza qué propósito debe fijar para el proyecto, plantee los objetivos marcados para la infraestructura de datos existente.

Si su organización ya necesita datos para determinados procesos de negocio (como detección de fraudes o análisis de mercado), reflexione sobre la manera en que el Big Data podría mejorar esos procesos o dotarlos de mayor valor. En lugar de afrontar un problema completamente nuevo, solo debería mejorar un proyecto o un proceso existente.

Sin un enfoque claro ni un valor demostrable ante los usuarios de negocio, el proyecto está abocado al fracaso.

2. Defina las métricas que demuestran el valor del proyecto

Si las métricas están definidas con claridad y se ajustan a sus objetivos, se ahorra muchos problemas. Al fijarse metas realistas y susceptibles de cuantificación, todos a su alrededor percibirán sus progresos.

Aún más importante: sabrán a qué aspira a largo plazo. Plantee el modo de cuantificar la repercusión del proyecto en el contexto de sus objetivos.

Es crucial porque habrá concesiones a corto plazo que tendrá que racionalizar ante los usuarios de negocio, y los objetivos cuantificables sirven para demostrar que les ofrece más valor del que perciben.

3. Adopte una estrategia en cuanto a las herramientas y la codificación manual

Evite la tentación de codificar todo de forma manual y directamente en Hadoop. Recuerde que el objetivo no es crear de la nada y con sus propias manos una implementación que funcione, sino suministrar el valor del Big Data a su organización.

En lugar de realizar la codificación manual de todas las integraciones, limpiar todos los conjuntos de datos y, luego, realizar la codificación manual de todos los análisis, busque herramientas y métodos de automatización con los que acelerar estos procesos.

Más importante si cabe es evitar la trampa de malgastar el escaso y costoso talento en el desarrollo de Java en aspectos que se pueden delegar en otros empleados. Su función consiste en tomar decisiones estratégicas sobre la implantación de recursos limitados de tal manera que se alcancen los objetivos.

Decántese por herramientas que aumenten la productividad del equipo de desarrollo aprovechando las competencias y los conocimientos de sus actuales expertos en ETL, calidad de datos y business intelligence, y reserve para sus superestrellas de Java el trabajo en la lógica específica, para la cual no hay herramientas disponibles.

Además, dado que las tecnologías como Hadoop evolucionan día a día, vale la pena perfilar una capa de abstracción que sirva como protección ante los constantes cambios en las especificaciones de las tecnologías subyacentes.

Ante todo, recuerde que las competencias que necesita escasean, pero siempre hay herramientas disponibles.

Selección del proyecto adecuado

En vista de los retos que deberá afrontar, veamos, a continuación, cómo debe proceder para elegir el proyecto apropiado para su organización.



Aspecto del proyecto adecuado

Si la organización tiene sed de cambios y ya ha aceptado la necesidad de contar con una estructura completa de gobernanza de datos para mejorar los métodos de trabajo, es probable que se pueda saltar esta sección.

Ahora bien, si va a plantear un proyecto táctico localizado que se pueda adaptar con el tiempo a toda la empresa, siga leyendo.

El proyecto adecuado posee las cuatro características siguientes.

1. Valor demostrable

El proyecto adecuado es aquel en que el valor se reparte a partes iguales entre TI y la unidad de negocio a la que pretende ayudar. Eso implica proporcionar un valor claro a un departamento, una unidad de negocio o un grupo, de tal forma que lo perciba.

2. Apoyo

Los ejecutivos que respaldan su visión son esenciales para lograr el éxito del proyecto. En los proyectos de Big Data, hace falta tener apoyos e intercesores en la cúpula que estén dispuestos a defender el trabajo que lleva a cabo.

Por eso, si se ve capaz de crear análisis fantásticos para logística, pero tiene su único respaldo ejecutivo en marketing, cambie el planteamiento. Si cuenta con el apoyo de marketing, dirija su creación a cumplir los requisitos de análisis de marketing. El cambio no se puede forzar. Confórmese con la influencia y sáquele el máximo partido.

3. Efecto dominó

El primer proyecto táctico tiene una importancia estratégica vital.

Aparte de demostrar más allá de toda duda razonable que el Big Data resulta útil para la unidad de negocio en cuestión, debe asegurarse, asimismo, de que su valor se comunica después con facilidad al resto de la empresa.

Por ello, a la hora de escoger el primer proyecto, tenga en cuenta la estrategia.

Una vez que demuestre el valor del Big Data al departamento de marketing, por ejemplo, resultará más sencillo recabar apoyos entre los equipos de logística, los cuales se habrían mostrado reticentes en otro caso.

4. Competencias trasladables

Como señalábamos en el último punto, el valor del primer proyecto le sirve para convencer a otros departamentos de la empresa.

Con ese fin, debe aprender las competencias, las capacidades y las lecciones apropiadas con el primer proyecto. Con más exactitud, tiene que documentarlas de manera que pueda trasladarlas al siguiente proyecto. Recuerde que, si pretende lograr el éxito, debe mirar a los proyectos futuros.

Así pues, prepare condiciones de escalabilidad para poder ocuparse de más proyectos en el futuro. No se trata solo de escalar el clúster. Se trata de escalar las competencias y las operaciones. O bien tiene que descubrir más superestrellas de Java para Hadoop o bien halla la manera de sacar más partido a los recursos de los que ya dispone.

2. Cadencia de beneficios y repercusiones

Al estudiar varios proyectos iniciales, es natural decantarse por los que ofrecen más mejoras y el máximo efecto para el negocio. No obstante, también es relevante abundar en la naturaleza del impacto en el negocio: ¿se apreciará la mayor parte del valor a corto o a largo plazo?

Más importante aún: ¿cuándo notarán los usuarios de negocio dicho impacto? Por ejemplo, si introduce la gestión de datos maestros en el data warehouse, mejorará de forma drástica la eficiencia de su business intelligence, pero los analistas de negocio solo percibirán ese valor cuando se den cuenta de que no tendrán que volver a limpiar datos financieros nunca más.

3. Recursos y restricciones

Teniendo en cuenta el análisis de los dos factores anteriores, piense en los recursos que hay a su disposición. Trataremos este tema a fondo más adelante pero, de momento, tenga en cuenta que, como es natural, su intención con el proyecto es esquilmar cada euro invertido.

La consecución de ese objetivo opera en ambos sentidos. Por un lado, pretende conseguir la máxima repercusión en el negocio pero, por otro, debe trazar una estrategia de inversión del presupuesto. Aunque le seduzca la idea de formar un equipo de especialistas en datos comparable al de Google, ¿acaso se lo puede permitir? Tomar decisiones inteligentes en lo referente a herramientas y personal es esencial para lograr el éxito del proyecto.

Proyectos tácticos de Big Data: algunos ejemplos

El Big Data posee una amplia variedad de aplicaciones. Aunque parezca fascinante, también suena un poco abrumador para quienes no tienen claro con qué proyecto arrancar. Esta es una lista de los proyectos tácticos de Big Data que han emprendido nuestros clientes.

Si aún no sabe por qué proyecto debería empezar su organización, fíjese en los ejemplos siguientes para hacerse una idea más ajustada de lo que el Big Data ofrece a su empresa.

Finanzas

- Análisis de los riesgos y de la oferta
- Recomendaciones de inversión

Distribución

- Interacción activa con los clientes
- Servicios basados en la ubicación

Atención sanitaria

- Predicciones de diagnóstico de pacientes
- Coste total de atención sanitaria
- Desarrollo de fármacos

Fabricación

- Programas conectados a los vehículos
- Mantenimiento predictivo

Sector público

- Cobertura sanitaria
- Mercados de valores
- Optimización fiscal
- Detección de fraudes

Parte 2

Su estrategia

A continuación, seremos prácticos y nos fijaremos en los requisitos específicos de su próximo proyecto de Big Data.

Definición de sus objetivos

Coja lápiz y papel. Como ya hemos señalado, la principal causa del fracaso de los proyectos de Big Data es la falta de objetivos claros. Vamos a asegurarnos de que el proyecto que tiene en mente no se pierda en ambigüedades.



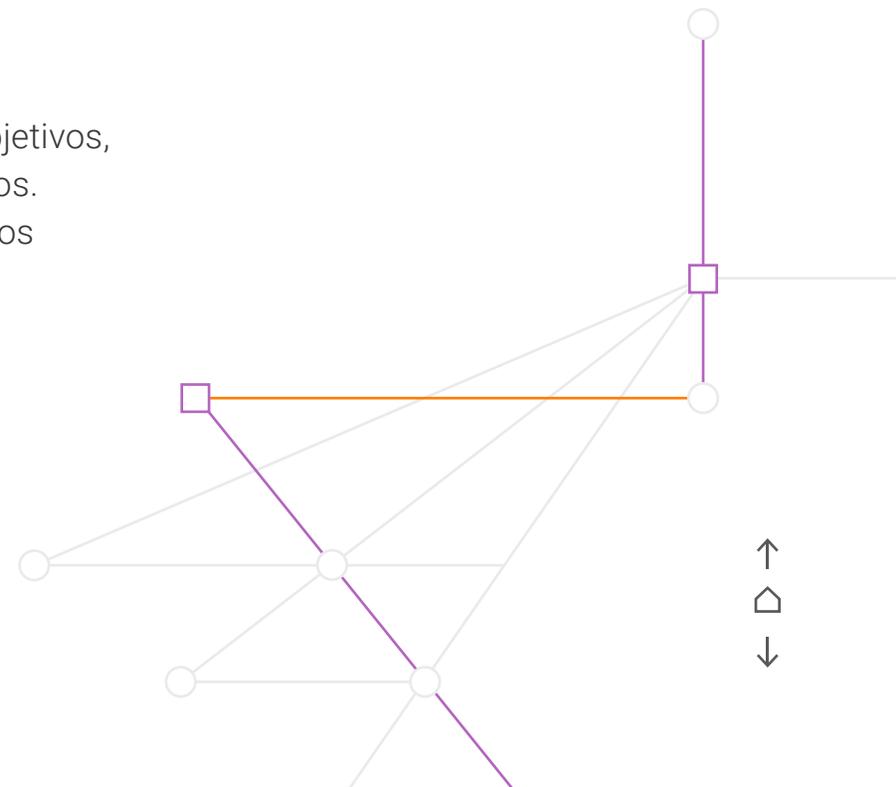
Objetivos de negocio

Empezaremos por el negocio porque, para que el proyecto reciba una buena acogida, sus objetivos han de tener prioridad sobre los de TI.

Los objetivos que se planteen conseguir para el negocio con el proyecto deben ser lo más específicos posibles. Recuerde definir objetivos cuyos efectos se puedan cuantificar.

Por ejemplo, en el proyecto de la interfaz del servicio de atención al cliente que permite predecir la rotación de clientes, no se deben enumerar objetivos tan imprecisos como “mejorar la experiencia del cliente”.

Cuanto más claros sean los objetivos, más cerca estará de alcanzarlos. Cinco objetivos muy restringidos valen más que uno general.



Objetivos de TI

A continuación, fijémonos en los objetivos de TI relacionados con su proyecto.

(Recuerde que, si su proyecto pretende agilizar o mejorar el trabajo de TI, le va a costar vendérselo a los usuarios de negocio. Por eso, conviene comunicar los objetivos de TI junto con los objetivos que ya entusiasman a los usuarios de negocio).

Enumere, por orden de importancia, los objetivos de su proyecto de Big Data que guarden relación con TI. (Indique tantos o tan pocos objetivos como desee).

Ejemplo: Establecer procesos para recopilar, limpiar, controlar y almacenar en tiempo real datos agregados de clientes, datos de uso de tarjetas de crédito, datos de gráficos sociales e indicadores de rotación.

Deténgase, colabore y escuche

Hemos redactado este cuaderno para que le resulte más fácil iniciar su proyecto de Big Data, tanto si trabaja para el negocio como si lo hace para TI. Sea como fuere, no deje los objetivos en el aire. Si precisa orientación concreta sobre las posibles pretensiones, llame a un compañero con experiencia en el campo e inicie ya la colaboración.

Para que el proyecto tenga éxito, es indispensable la colaboración estratégica.

Definición de sus necesidades de datos

Una vez descritos los objetivos concretos de su iniciativa de Big Data, pasemos al meollo del proyecto: los datos en sí. En cualquier proyecto, ha de mantener una mentalidad estratégica acerca de la información necesaria, los conjuntos de datos que cubren esa necesidad, la manera de obtenerlos y la forma de utilizarlos.

Datos necesarios

Antes de nada, fijémonos en la finalidad más básica de su proyecto de Big Data: la información que desea proporcionar a su organización. Responda a las preguntas siguientes con tanta minuciosidad como pueda.

Para alcanzar los objetivos de negocio antes descritos, según los usuarios de negocio, ¿qué deben saber para tomar decisiones fundadas?

Ejemplo: Qué clientes más valorados tienen probabilidades de descartar la empresa y qué comportamientos se relacionan con la rotación.

¿Qué datos sirven para aportar esos conocimientos?

Ejemplo: Historial de compras del cliente, críticas, porcentaje de compras, porcentaje de abandono, porcentaje de rechazo y calidad del servicio de atención al cliente.

¿Qué sistemas de origen contienen esos conjuntos de datos?

Ejemplo: Registros del servicio de atención al cliente, métricas de rendimiento de productos, base de datos de actividad del cliente y gestión de datos maestros de clientes.

Aparte de los datos ya señalados, ¿existe alguna otra información que aporte contexto o más valor a los análisis?

Ejemplo: Encuestas del servicio de atención al cliente, análisis de la competencia, datos meteorológicos y datos de redes sociales.

Cinco consideraciones clave en torno a los datos

Tras describir los datos que va a buscar, ya cuenta con una visión más clara de los retos concretos que le plantea el Big Data.

Hay cinco elementos en particular que debe tener en cuenta antes de seguir adelante, ya que dictan las necesidades tanto de cada conjunto de datos como de conjunto de datos del Big Data.

1. Prepárese para un gran volumen

Prepárese para afrontar la infinidad de datos que va a precisar. En todas las dimensiones, clasifique los datos por su valor (por ejemplo, transacciones de clientes), su uso (frecuencia de acceso), su tamaño (gigabytes, terabytes), su complejidad (datos de máquinas, datos relacionales, vídeos...) y las personas con acceso a ellos (solo los especialistas en datos o cualquier usuario de negocio).

Un inventario concienzudo y organizado de los datos facilita la decisión de cómo gestionarlos. Evalúe la capacidad actual de almacenamiento y procesamiento y busque los métodos más rentables y eficientes para hacerla escalable.

2. Tenga en cuenta la diversidad

El aspecto más complicado del Big Data radica en la multitud de formatos y estructuras que debe conciliar en sus análisis. Tiene que integrar varias fuentes si desea incluir estructuras y tipos de datos nuevos (sociales, de sensores o de vídeos) con las fuentes a las que están acostumbrados (relacionales o mainframes heredados).

La codificación manual de cada integración precisa resulta tan engorrosa que puede consumir todo el tiempo y todos los recursos de que dispone. Aproveche al máximo las herramientas disponibles de integración y calidad de datos para agilizar el proceso y dedicarse a tareas más útiles.

3. Controle la velocidad

Por lo general, al combinar la transmisión de datos en tiempo real con los datos históricos, aumenta el potencial predictivo de los análisis. Por ello, algunos de los datos que le interesan solo tienen valor si fluyen de manera constante hacia sus sistemas.

En efecto, casi todos los análisis en tiempo real se tienen que basar en transmisiones de datos que, a menudo, proceden de fuentes diferentes y tienen distintos formatos. Integre en el proyecto alguna tecnología de análisis de transmisiones y una infraestructura lógica que le permitan gestionar todos los datos.

4. Compruebe la veracidad

Da igual lo relevantes que sean sus análisis: no valen nada si los usuarios no pueden tener una confianza razonable en los datos que incluyen. Cuantos más datos analice, más importante es que mantenga la máxima calidad de datos.

Para que los datos sirvan a un fin determinado, tiene que conocer dicho fin. Si un especialista en datos busca patrones en datos agregados de clientes, la preparación necesaria es mínima. Sin embargo, los datos de los informes financieros y de la cadena de suministro exigen un elevado grado de conservación, limpieza y certificación de precisión y cumplimiento.

Cree categorías basadas en la preparación indispensable que vayan desde datos sin procesar hasta almacenes conservados y controlados de datos limpios, fiables y fidedignos.

5. Tenga en consideración el cumplimiento

Los distintos conjuntos de datos que va a manejar tienen diferentes requisitos y condiciones de seguridad. En cada conjunto de datos, se debe plantear qué hace falta para mantener el anonimato de los datos conforme a las políticas de seguridad.

Montones de datos proliferarán en centenares de almacenes por toda la empresa.

Entérese de dónde residen los datos sensibles, protéjalos en la fuente mediante cifrado y, a continuación, controle quiénes tienen acceso a ellos.

Además de archivar de forma segura e inteligente los datos sensibles, enmascárelos con reglas predefinidas cada vez que los migre o los introduzca en los entornos de desarrollo y prueba.

Aplique estas cinco consideraciones a todos los conjuntos de datos que maneje y no tendrá problemas para superar de forma más realista los retos que plantea el Big Data.



Parte 3

Su eficaz cadena de suministro de datos

Los métodos tradicionales de business intelligence y data warehouse no se escalan para cubrir las necesidades de las iniciativas de Big Data. Por eso, a continuación, veremos la manera de escalar su equipo, sus procesos y su infraestructura.

Su equipo

Su equipo de Big Data representa tanto su mayor reto como su mayor oportunidad. Debe hallar el delicado equilibrio entre quienes comprenden los objetivos de negocio y quienes saben ejecutar los requisitos técnicos.



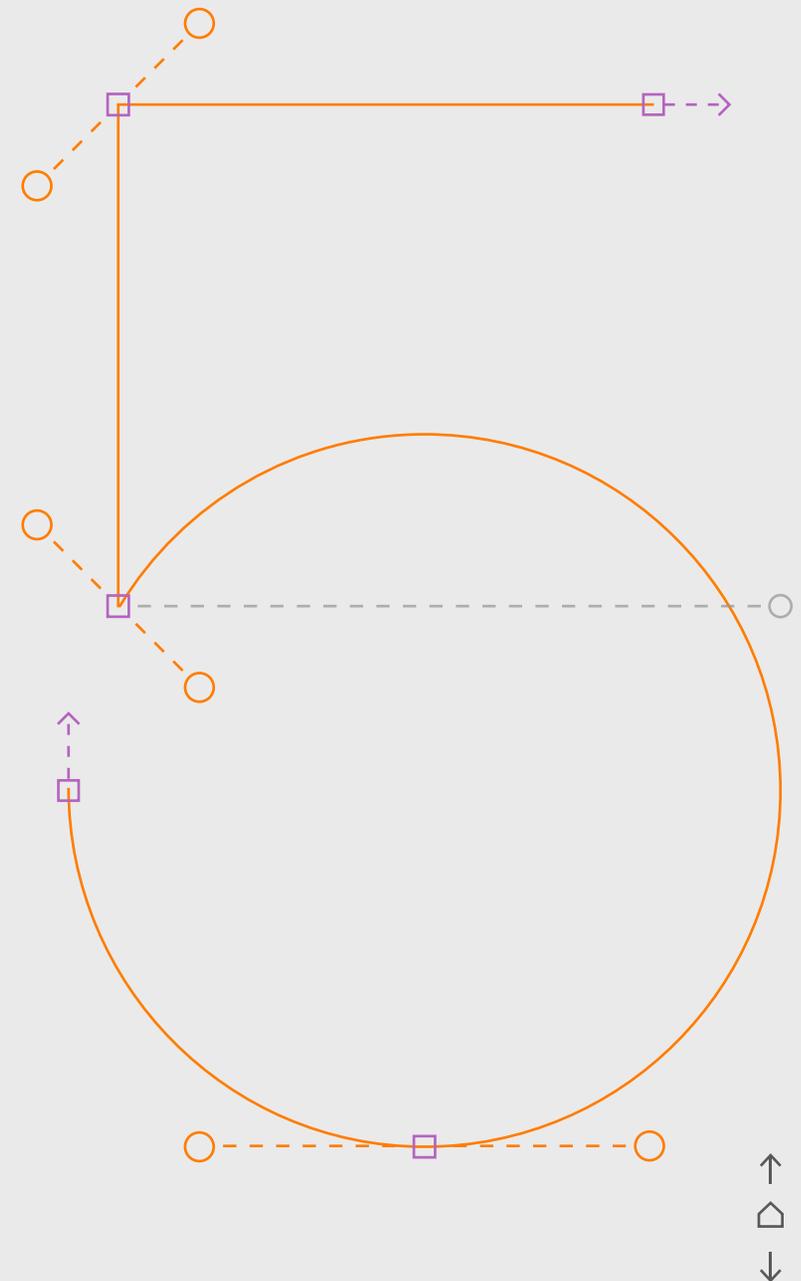
Cinco lecciones básicas para formar equipos

La mayoría de las organizaciones infravalora el grado de competencias que hace falta para aplicar una tecnología nueva como Hadoop.

Las estructuras de datos distribuidas son simplemente difíciles de gestionar. Desde los conocimientos de Java requeridos para el desarrollo con Hadoop hasta las nuevas competencias especializadas en datos que tendrá que contratar, va a tener que reunir un número considerable de competencias nuevas para que su proyecto eche a volar².

Cuando forme el equipo, no olvide incorporar las siguientes lecciones a la estrategia de contratación.

² InfoWorld, "[Hadoop, Python, and NoSQL lead the pack for big data jobs](#)", 5 de mayo de 2014



1. Aproveche las competencias por las que contrató a su personal

Uno de los mayores errores que cometen las empresas cuando contratan especialistas en datos y analistas cuantitativos consiste en obligarlos a hacer el trabajo sucio. Si sus recursos más competentes invierten su tiempo en la codificación manual de las integraciones de datos y en la limpieza de los datos, no solo consigue que se sientan frustrados sino que, además, desaprovecha las competencias que le han resultado tan difíciles de hallar.

Concentre las competencias que escasean en las tareas que de verdad las necesitan. No le interesa que esas personas abandonen el barco, y mucho menos que pierdan el tiempo en trabajos que podría acometer con cualquier herramienta.

2. Mantenga una mentalidad estratégica acerca de la composición del equipo

Si todo sale como es debido, crecerán tanto el ámbito como los recursos del proyecto. Piense ahora en la estrategia para ahorrarse la dura realidad de no poder escalar ciertos procesos con la suficiente rapidez porque solo hay un número limitado de personal con las competencias precisas, incluso en Silicon Valley.

Si se amplía el ámbito de aplicación del proyecto, ¿qué posibilidades reales hay de encontrar a tiempo las competencias que cubran esas necesidades? Por ejemplo, los especialistas en datos son infinitamente más difíciles de encontrar, formar y contratar que los desarrolladores³.

El equilibrio del equipo es crucial. Ha de buscar la mezcla perfecta de la experiencia en gestión de datos adquirida con tesón y el entusiasmo por aprender herramientas nuevas. Además, debe lograr el equilibrio entre el personal con

conocimientos técnicos y el personal con la experiencia en el campo precisa para crear los modelos adecuados.

3. Ajuste los objetivos del proyecto cuanto antes y, luego, comuníquelos

Uno de los errores más comunes que cometen las empresas cuando contratan personal nuevo consiste en olvidar comunicar los auténticos objetivos del proyecto. Desde la primera entrevista y en todo momento hasta el desempeño del trabajo en sí, debe dejar meridianamente claro qué pretende ofrecer a los usuarios de negocio. Aproveche el respaldo de los ejecutivos para explicar la misión y compartir tanto los casos prácticos como los problemas.

Si no entienden a la perfección el valor de negocio del proyecto, corre el riesgo de que los empleados nuevos creen que solo deben tener en cuenta los objetivos de TI.

³ "Big Data's High-Priests of Algorithms," Wall Street Journal, 8 de agosto de 2014



4. Cuando el equipo se amplía, también es mayor la necesidad de gestionarlo

A diferencia de la tecnología nueva, que se puede implantar, implementar e integrar de forma objetiva, el personal nuevo se tiene que acostumbrar al puesto de trabajo, a sus cometidos y al motivo por el que se les encomiendan. Puede ser usted o cualquier otro, pero alguien tiene que afrontar el reto de gestión que plantea un equipo nuevo.

No conviene infravalorar elementos como la cultura o la cohesión. Reflexione largo y tendido sobre el modo de integrar a los empleados nuevos en sus procesos. Quizá no pueda enseñarles competencias pero, sin duda, puede ayudarlos a ser mejores miembros del equipo.

5. Su equipo no puede permitirse quedarse parado

Todos los días surgen tecnologías de Big Data y las que ya existen evolucionan con rapidez. Se trata de un momento fascinante para las empresas que tienen el coraje de adoptar las mejores prácticas enseguida. Ahora bien, también representa el reto definitivo de obtener ventaja frente a la competencia.

Su personal tiene que desarrollar sus competencias tan rápido como cambia el mundo que lo rodea. La buena noticia es que nada motiva más a los buenos empleados que el reto de ir por delante de los acontecimientos. El desafío radica en ofrecer la formación y el debate que precisan para seguir aumentando tanto sus capacidades como las suyas.

Importancia de la estrategia

Una decisión importante que deberá tomar una y otra vez es la de crear las funcionalidades con herramientas automatizadas o mediante integraciones manuales.

La codificación manual ofrece un control completo y preciso sobre el objeto de la creación. A menudo, tiene un valor incalculable y es necesaria si, por ejemplo, desea crear una secuencia compleja para extraer metadatos de una forma que aún no es posible.

Las herramientas, por su parte, ofrecen mayor agilidad y la capacidad de repetir de manera sostenible el mismo proceso. En tareas como la integración y la calidad de datos, son cruciales porque, gracias a ellas, no obliga a los superdotados analistas y especialistas a encargarse del trabajo sucio.

Sea realista con sus recursos. Si no puede formar un equipo tan nutrido ni tan brillante como el de Google, no malgaste sus escasos recursos en el intento.



Establecimiento de la gobernanza de datos

Si (y, esperemos, cuando) emprende una iniciativa de Big Data más básica, debe implantar la estructura de procedimientos para la gobernanza de datos. De hecho, aunque el proyecto de Big Data tenga como objetivo ofrecer valor a un solo departamento, puede ser interesante crear una comisión reducida de gobernanza de datos para aprender a superar los retos únicos que tal organismo presenta.

En esencia, la comisión de gobernanza de datos es el organismo formal de ejecutivos que tiene el cometido de supervisar el enfoque de los datos de la empresa, pero también debe incluir administradores de datos, esto es, personal funcional o de un departamento determinado que se encarga de gestionar los datos procedentes de una unidad de negocio concreta.

(De hecho, algunos de nuestros clientes asignan funciones de administración de datos según el dominio de datos. Eso significa que una persona tiene a su cargo los datos de los productos, otra se ocupa de los datos de los clientes, y así sucesivamente.)

Se debe proponer crear procesos que garanticen que la estructura de gobernanza de datos resulte más positiva que negativa. Trabaje de forma activa para que no se convierta en una carga burocrática asegurándose de que todos los implicados se comprometen a alcanzar los mismos objetivos en los mismos plazos.

La estructura de gobernanza de datos debe poseer las cinco características siguientes.

1. Transversal

Una comisión de gobernanza de datos cuyos miembros tengan funciones parecidas es del todo ineficaz. El objetivo consiste en crear un organismo donde estén representadas las visiones y las necesidades exclusivas de cada una de las unidades de negocio a las que esté dirigido el proyecto de Big Data.

2. Comunicativa

Sin una comunicación fluida entre las funciones, los departamentos y los dominios, es probable que el proyecto quede enterrado en burocracia y malentendidos. Esto sucede con demasiada frecuencia. Asegúrese de que todas las inquietudes se calman o reciben una respuesta apropiada.

3. Eficiente

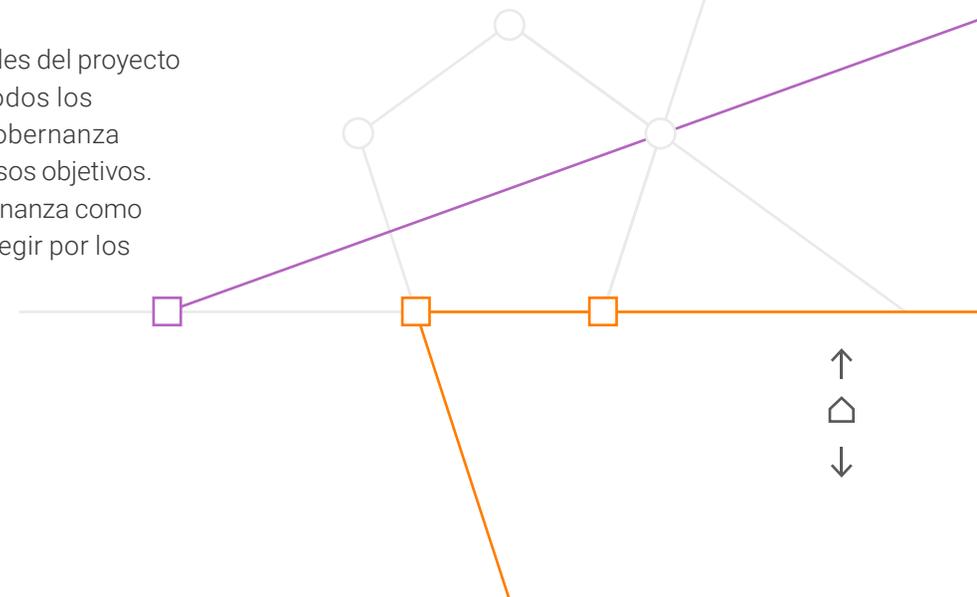
El proceso transversal no debe suponer un obstáculo. Hace falta una gran agilidad para que el proyecto de Big Data llegue a buen puerto. Por eso, integre reglas de comunicación de excepciones y de automatización siempre que sea factible y adopte herramientas de colaboración que mantengan líneas de comunicación oportunas abiertas.

4. Comprometido

Comunique los objetivos principales del proyecto con eficacia y compruebe que todos los implicados en la estructura de gobernanza de datos se dedican a conseguir esos objetivos. Tanto los planteamientos de gobernanza como la toma de decisiones se deben regir por los objetivos comunes.

5. Centralizado

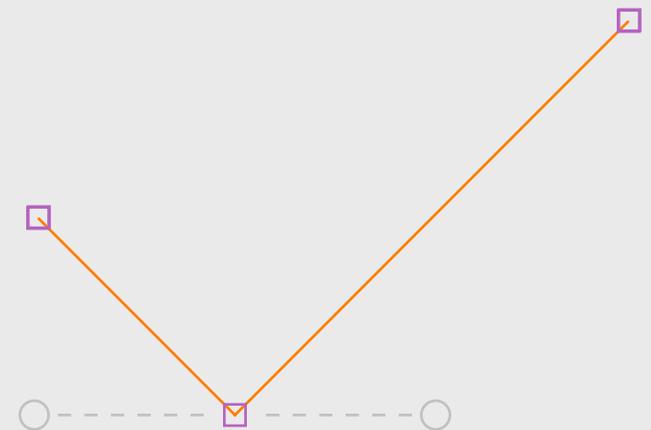
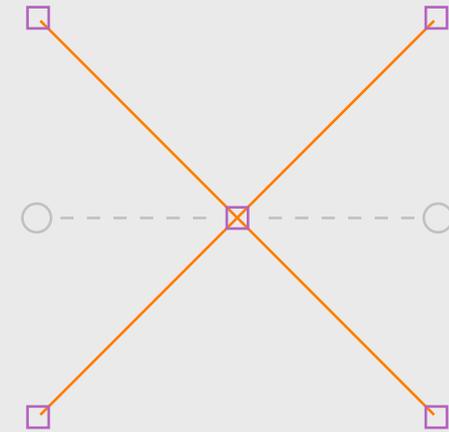
El mayor reto de la estructura de gobernanza de datos se presenta a la hora de dar prioridad a los objetivos de una unidad de negocio frente a los de otra que también cuenta con representación en la comisión. Las decisiones se deben sustentar en los beneficios a largo plazo para toda la comisión aunque ello suponga que los beneficios a corto plazo se aprecien solo en una unidad de negocio.



Competencias que necesita y competencias que posee

Vuelva a coger lápiz y papel. Ahora que ya sabe las diversas oportunidades y los distintos inconvenientes subjetivos que presentará el nuevo equipo, imaginemos su composición.

En la página siguiente, se enumeran las funciones para Big Data basadas en los puestos que han cubierto nuestros clientes. Según el personal de que dispone ahora mismo y del plazo previsto de ejecución del proyecto ([indicado al principio de la sección en la página 25](#)), apunte el número de personas que debe contratar.



Función	¿Tiene ya a alguien capaz de desempeñar esta función?	Es preciso contratar a alguien para esta función	Por el tiempo disponible, es preciso contratar a este número de personas
Especialista en datos			
Experto en dominios			
Analista de negocio			
Analista de datos			
Ingeniero de datos			
Administrador de bases de datos			
Arquitecto empresarial			
Arquitecto de soluciones de negocio			
Arquitecto de datos			
Administrador de datos			
Desarrollador de ETL (integración de datos)			
Desarrollador de aplicaciones			
Desarrollador de cuadros de mando			
Modelador estadístico			
Otro			

Necesidad de una mentalidad integradora

Quando salga en busca de nuevos miembros para el equipo, no se limite a personas con la cualificación correcta. No se equivoque: encontrar a personas con la cualificación correcta ya es un reto en sí mismo pero, aparte, debe buscar gente que esté dispuesta a asumir los objetivos de negocio y que cuente con competencias técnicas.

Nuestros clientes no paran de repetirnos la importancia que reviste que quienes se incorporan a los proyectos de Big Data comprendan la realidad del negocio y estén especializados en datos complejos. Este tipo de mentalidad integradora es importante y difícil de encontrar. Merece la pena fomentarla con formación y los resultados también valen el esfuerzo.

Examen de las herramientas del Big Data

Repase esta lista de herramientas y marque con una equis (X) las más importantes (y pertinentes desde el punto de vista estratégico) para su proyecto.

Análisis

Herramientas y procesos que permiten convertir los datos sin procesar en conocimientos, patrones, predicciones y cálculos sobre el dominio objeto de análisis.

Visualización

¿Puede presentar sus datos y conclusiones de manera que resulten fáciles de comprender y asimilar?

Análisis avanzados

¿Puede aplicar algoritmos analíticos de vanguardia a sus conjuntos de datos a fin de efectuar cálculos complejos?

Aprendizaje automatizado

¿Puede aplicar algoritmos avanzados de aprendizaje automatizado con el fin de identificar patrones y elaborar predicciones imposibles de manejar con su ancho de banda manual?

De todas estas herramientas y tecnologías, algunas como la integración de datos, la calidad de datos y la gestión de datos maestros son tan fundamentales para la transición al Big Data que, en realidad, no merece la pena reestructurarlas. No desperdicie las valiosas competencias ni las horas de trabajo del proyecto de Big Data en la cantidad de tiempo y recursos que hace falta para crear esas funcionalidades con sus propias manos.

Recuerde cuáles son los objetivos del proyecto y que estos no incluyen una creación personalizada de absolutamente todo.

Sus procesos

Pasemos a los procesos reales que precisa para enfrentarse al Big Data. Aunque sus procesos sean específicos según sus objetivos y sus requisitos, en esta sección, se ofrece información general sobre lo que cabe esperar y aprender.



Los ocho pasos del Big Data

A partir de la experiencia, estamos en condiciones de afirmar que las metodologías ágiles constituyen un enfoque excelente para los proyectos de Big Data. Garantizan que contenga las expectativas, aprenda de los errores e itere el camino hacia procesos óptimos. Dicho lo cual, el enfoque de su proyecto depende por completo de su situación y de sus preferencias.

En todo caso, los ocho pasos siguientes resultan cruciales para la cadena de suministro del Big Data. Sea cual sea el método elegido, asegúrese de establecer junto con su equipo procesos eficaces para seguir estos pasos.

Repase esta lista de herramientas y marque con una equis (X) las más importantes (y pertinentes desde el punto de vista estratégico) para su proyecto.

1. Acceso a los datos

El primer reto consiste en adquirir todos los datos necesarios. En algunos casos, eso implica capturar transmisiones de datos y, en otros, extraerlos de una base de datos. Configure procesos repetibles y manejables para garantizar que dichos datos se puedan almacenar, a continuación, conforme a los métodos previstos.

2. Integración de los datos

El reto más complejo del Big Data guarda relación con la diversidad de estructuras y formatos de datos. Para lograr una ejecución sostenible de los análisis, debe configurar un proceso para integrar y normalizar todos estos datos. Lo ideal es que requiera el mínimo procesamiento manual posible.

3. Limpieza de los datos

Para que los análisis sean fiables, es imprescindible limpiar los datos a fin de eliminar duplicados, errores y datos imprecisos o incompletos. El proceso ha de garantizar que los analistas y los especialistas más cualificados no pierden su tiempo en “hacer la colada”.

4. Control de los datos

Una forma de mantener una fuente fiable de datos limpios e integrados consiste en establecer un proceso para controlar los datos. La finalidad es crear una nutrida recopilación de datos consolidados, organizados por dominios (productos, clientes, etc.) y enriquecidos con información del Big Data que se pueda suministrar a todos los demás sistemas.

5. Protección de los datos

Ha de establecer dos procesos básicos. El primero consiste en definir las normas y las prácticas de seguridad que exija cada conjunto de datos; el segundo, en detectar los datos sensibles y enmascararlos de forma persistente o dinámica para garantizar la aplicación uniforme de esas normas y mejores prácticas.

6. Análisis de los datos

El proceso de análisis depende de los analistas, las herramientas analíticas y los requisitos relacionados con los objetivos. Resulta esencial mantener una mentalidad que priorice la detección iterativa y la mejora continua ya que conviene que este proceso sea mejor, más rápido, más barato y más escalable con el tiempo y la experiencia.

7. Análisis de las necesidades de negocio

Se trata de un paso que, a pesar de ser fundamental, casi siempre se obvia. Establezca un proceso claro para el análisis de las necesidades de negocio incluso mientras se analizan los datos. Es fundamental porque, si no mantiene el pulso al negocio, se arriesga a dividir los esfuerzos y a minimizar la repercusión en el negocio.

8. Aprovechamiento operativo de la información

Como ya hemos destacado antes en el cuaderno, es indispensable que se perciba el impacto en el negocio del proyecto de Big Data. Cree procesos automatizados para suministrar las respuestas halladas a los usuarios de negocio que más las necesitan. Por ejemplo, es preciso poner los datos sobre los clientes con más probabilidad de rotación a disposición de los agentes del servicio de atención al cliente mediante un cuadro de mando. No olvide incorporar también un bucle de comentarios para saber cómo se recibe la información.

Importancia de la documentación

Si domina estos ocho pasos, su proyecto de Big Data marchará en la dirección correcta. El objetivo consiste en establecer procesos claros, repetibles y escalables en permanente mejora. Con ese fin, la documentación de dichos procesos y las consiguientes mejoras resultan vitales para el equipo.

Las competencias, las capacidades y las lecciones del proyecto de Big Data han de ser trasladables y se tienen que comunicar con frecuencia.

Su arquitectura

Para que la cadena de suministro del Big Data sea eficaz y efectiva, es imprescindible garantizar que la arquitectura es sólida y está concebida de forma estratégica. En esta sección, veremos el aspecto de la arquitectura de Big Data idónea y la manera de implantar la suya por fases.



Primeros pasos: zona de pruebas

A la hora de crear la arquitectura de su proyecto de Big Data, el punto de partida más lógico consiste en configurar un entorno de desarrollo de pruebas en el cual probar datos para asegurarse de que la arquitectura sea viable. Al hacerlo, tenga en cuenta las recomendaciones siguientes.

Empiece con moderación

Si configura una zona de pruebas bien definida en la que ejerza un control absoluto, podrá iterar el método hasta conseguir una implementación correcta. Póngase en marcha cuanto antes y documente las lecciones aprendidas con cada iteración.

El tamaño importa

La diferencia clave entre la zona de pruebas y la implementación real radica en que el entorno de producción será mucho más grande. Por ello, hará falta un procesamiento automatizado para incorporar, integrar, limpiar y distribuir los resultados. Por lo tanto, hacen falta una estructura mucho más sólida así como procesos y componentes de eficacia probada para que el entorno de producción activo sea del todo fiable y flexible.

Enmascare los datos antes de probarlos

Cuando las organizaciones emplean datos de prueba, suelen utilizar una variante de sus datos de producción activos para garantizar que los formatos y las estructuras representen el entorno activo. Lamentablemente, si no se realiza un enmascaramiento adecuado, los datos sensibles pueden quedar expuestos en un entorno nada seguro.

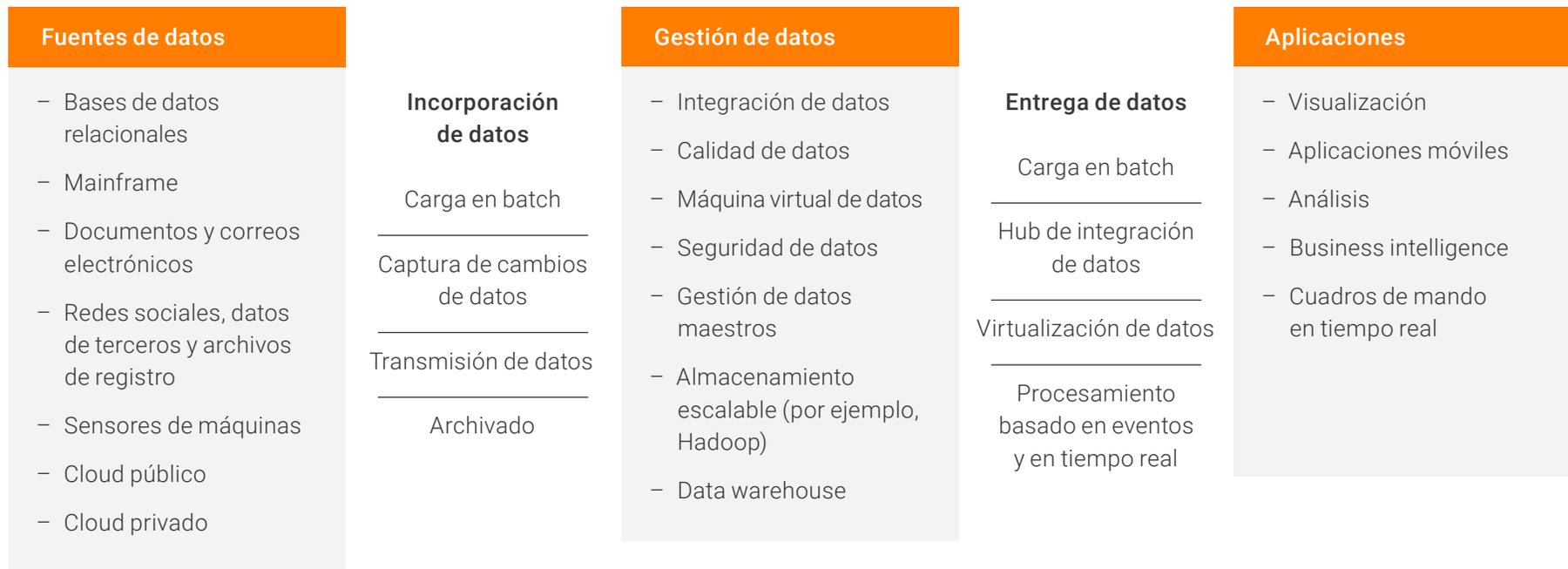
No se extravía en el camino

Una de las causas más comunes del aumento de costes y del retraso de los proyectos de Big Data es que los errores de codificación manual no detectados en la zona de pruebas se convierten en la peor pesadilla del equipo cuando se activa la arquitectura. Por eso, si codifica a mano bastantes partes de la arquitectura, no se sorprenda si tiene que rehacer un montón de código para satisfacer los requisitos de producción y contener las expectativas. También cuenta con la alternativa de emplear de primero herramientas de productividad y automatización que le eviten rehacer el código y subsanar los errores.



Arquitectura idónea del Big Data

En el diagrama siguiente, se representa el modo en que recomendamos crear la arquitectura idónea de tecnologías y procesos de Big Data.



Plan de su proyecto

Ya hemos analizado todos los aspectos de su transición al Big Data. A continuación, sírvase de este plan como modelo para gestionar el proyecto de Big Data desde su concepción hasta su implementación.



Plan de su proyecto

Emplee este plan de proyecto como modelo para documentar los detalles y los diversos elementos de su proyecto de Big Data. Luego, sírvase del documento compilado como medio para conseguir el respaldo preciso del resto de la organización. También le resultará de utilidad para convencer a socios externos.

Fase 1: estrategia

Identificar los objetivos de negocio y de TI
Definir las medidas del éxito

Fase 2: datos

Identificar la información necesaria
Identificar los datos y las fuentes para suministrarlos

Fase 3: cadena de suministro

Personal

- Evaluación de las competencias necesarias
- Evaluación de las competencias ya disponibles

Proceso

- Acceso a los datos
- Integración de los datos
- Limpieza de los datos
- Controla de los datos
- Protección de los datos
- Análisis de los datos
- Análisis de las necesidades de negocio

Herramientas

- Informática distribuida (por ejemplo, Hadoop)
- Calidad de datos
- Integración de datos
- Gestión de datos maestros
- Enmascaramiento de datos
- Visualización
- Análisis de transmisiones
- Análisis
- Aprendizaje automatizado

Fase 4: aprovechamiento operativo de la información

Desarrollar cuadros de mando

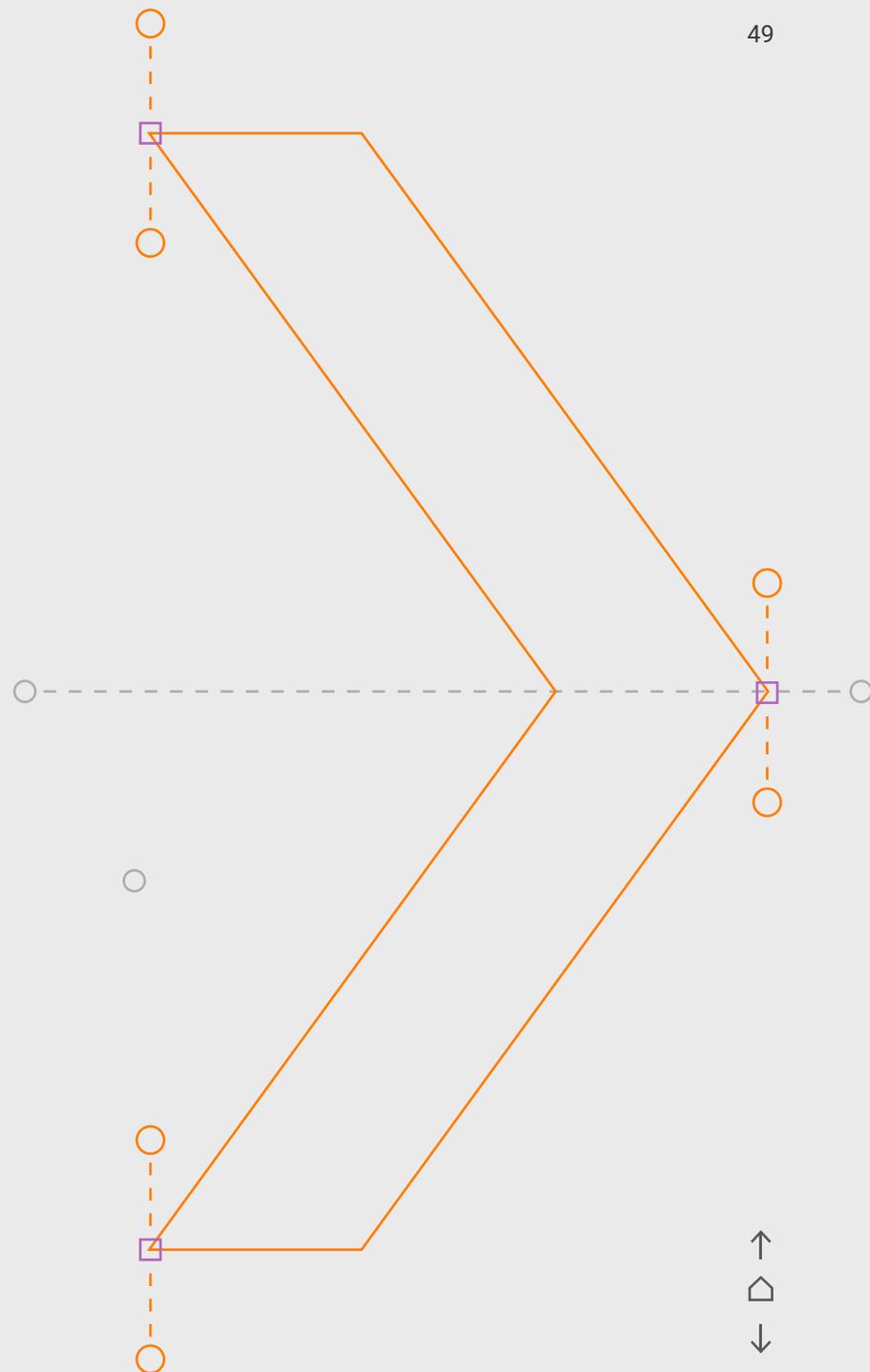
Automatizar procesos para la entrega de datos

Configurar un proceso de comentarios

En marcha

Sírvase de las listas de comprobación, los principios y las directrices descritos en este cuaderno para trasladar el potencial del Big Data a su organización. Sea cual sea (de momento) la envergadura de su proyecto, no nos cabe duda de que cuenta con un mejor equipamiento para salvar los numerosos escollos que surgen en su camino.

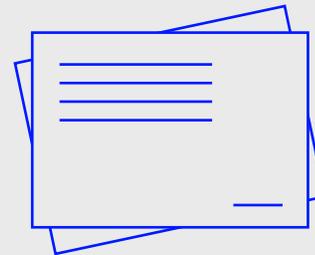
Esto es grande.



Siguientes pasos

¿Está listo para aplicar lo aprendido?

Si es desarrollador de Informática, puede ser desarrollador de Hadoop. Con nuestros servicios, nuestros conectores y nuestras pruebas de software de Big Data, tomará la senda correcta.



[DESCARGAR AHORA](#)

Acerca de Informatica

La transformación digital modifica las expectativas: mejor servicio, entrega más rápida, menores costes. Los negocios deben transformarse para seguir siendo relevantes y los datos tienen la respuesta.

Como líder mundial en gestión de datos de cloud empresariales, le brindamos ayuda para que encabece la marcha de forma inteligente, en cualquier sector, categoría o nicho. Informatica le aporta perspectiva para que aumente su agilidad, concrete nuevas oportunidades de crecimiento o incluso invente cosas nuevas. Al estar completamente centrados en todo lo relacionado con los datos, ofrecemos la versatilidad necesaria para alcanzar el éxito.

Le invitamos a explorar todo lo que puede ofrecerle Informatica y a desatar el poder de los datos para impulsar su próxima disrupción inteligente.

Informatica en España:

José Echegaray 8, edif. 3, PB 3, 28232 Las Rozas, Madrid.

Teléfono: 902 882 062.

Fax: 917 542 950

informatica.com/es

[linkedin.com/company/informatica](https://www.linkedin.com/company/informatica)

twitter.com/Informatica

PÓNGASE EN CONTACTO CON NOSOTROS

IN18-0917-2730

© Copyright Informatica LLC 2015, 2017. Informatica y el logotipo de Informatica son marcas comerciales o marcas comerciales registradas de Informatica LLC en los Estados Unidos de Norteamérica y en otros países.



Informatica™