



INICIO
GRABACIÓN



SANJOSÉ
FUNDACIÓN DE EDUCACIÓN SUPERIOR



Arquitectura De TI

Wilson Cárdenas Cr.
Fundación de educación superior San Jose
Semana 8

A man and a woman are leaning over a desk, looking at a laptop. The man is on the left, wearing a grey t-shirt, and the woman is on the right, wearing a striped shirt. They are in an office environment with other people and desks visible in the background. A large pink circle is overlaid on the right side of the image, and a dark purple circle is overlaid on the left side, containing the text.

CONSULTORÍA SOBRE BASES DE DATOS



Introducción



Las organizaciones se están dando cuenta del volumen considerable de datos que tienen, y necesitan explotarlos para sacarles valor e incorporarlos en la mejora de sus procesos. Para ello, buscan empresas que les guíen sobre qué hacer con esos datos. Aquí es el momento en el que entran en juego las consultorías de Big Data.

El término de consultoría Bases de datos es un término muy amplio que abarca infinidad de procesos. El Big Data es la unión de diferentes tecnologías, pero todas con un objetivo común: extraer información de los datos para generar valor. Lo que nos encontramos habitualmente es que se ha hecho mucha labor de “evangelización” de este mundo del Big Data.

Cuando hablamos de Big Data, no podemos olvidar los parámetros por los que se rigen las características más importantes, las 7 V del Big Data: **Volumen, Velocidad, Variedad, Veracidad, Viabilidad, Visualización**, y la más importante, el **Valor de los datos**. Para todos los parámetros anteriores, el Big Data requiere el uso de un conjunto de herramientas, aplicaciones y marcos para almacenar, procesar y explotar los datos.

Las empresas han oído que los datos son un recurso más y que si los explotas puedes sacar valor y así mejorar tus procesos de negocio.

Llegados a ese punto, las organizaciones se hace la pregunta... ¿por dónde empezar?




¿Qué se hace en una consultoría Big Data?

Normalmente, las consultorías de Big Data comienzan haciendo tareas de entender con el cliente las necesidades que tienen y los datos que pueden aportar. O bien, los datos que se pueden aportar desde la propia consultoría. Por ejemplo, datos libres. Con esto, hay un movimiento en auge del Open Data o datos abiertos.

La expresión Open Data hace referencia a datos que las organizaciones hacen públicos, como pueden ser gobiernos, ayuntamientos, periódicos de investigación, entre otros.

Éstos son un subconjunto de todos los datos que se encuentran disponibles en la red.

Por ejemplo, a raíz de lo que podamos pensar, los datos de Twitter no se consideran Open Data ya que son datos públicos por naturaleza, pero igualmente es una fuente disponible para proyectos de consultoría Big Data.

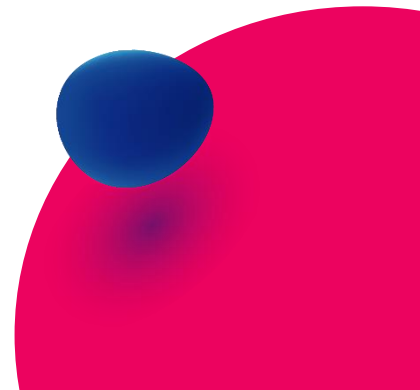




Fases de una consultoría Big Data



SGBD Para poder llevar a cabo un proyecto de Big Data, las consultorías definen una serie de fases donde llevar a cabo todos los puntos estructuradamente, de manera que se llegue a una solución eficaz para la empresa que está requiriendo ese asesoramiento.





Fase 1. Workshops para ver datos y necesidades


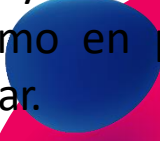


Normalmente, cuando una consultoría recibe un proyecto de Big Data, se realizan una serie de workshops donde se llevan a cabo dos acciones:

Ver los datos disponibles: Se estudian todos los datos que tiene esa empresa y cómo se relacionan entre ellos. También se identifican aquí las fuentes de datos externas que puedan ser de utilidad para la empresa.

Identificar necesidades: Se identifican las necesidades que la empresa pretende resolver con el estudio de los datos. En muchos casos, la experiencia de una consultoría puede establecer cruces que ya se hayan establecido anteriormente en empresas de la misma área o incluso de sectores diferentes.

En ocasiones, el tipo de datos que se obtienen de las empresas son muy variados, desde bases de datos muy estructuradas, documentos de texto como en pdf, imágenes, vídeos, siendo este último el formato más complejo de analizar.

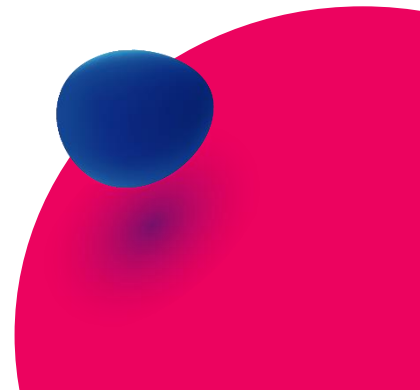




Fase 2. Listado de soluciones



Como resultado de esta primera fase de la consultoría de Big Data y tras analizar los datos, se prepara un listado con las posibles soluciones que pueden implementarse en vista de los datos disponibles y las necesidades de la empresa. Es el momento en el que el cliente toma la decisión de qué solución aplicar.





Fase 3. Ordenación y limpieza de los datos

Una vez que llegamos a este punto, los datos se ordenan y organizan, y se extrae información relevante para la toma de decisiones. Tener los datos ordenados permite la elaboración de conclusiones y la identificación de nuevas vías de desarrollo del negocio.



Fase 4. Prototipo de la solución

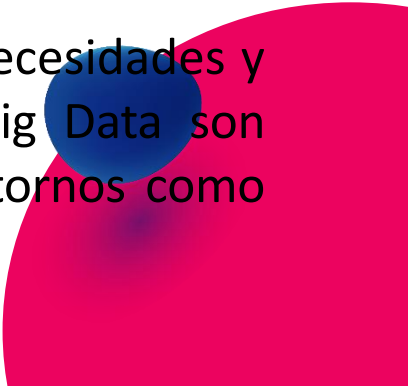


Una vez que el cliente sabe la solución a aplicar ante el problema que tenían, y que los datos se han ordenado, la empresa consultora de proyectos Big Data puede realizar un prototipo sobre la solución que explota esos datos para conseguir mejorar sus procesos empresariales o resolver la necesidad que se ha identificado.

En otras ocasiones, el proceso es más sencillo pues la empresa tiene establecida la solución que quiere mejorar y sabe de qué datos se disponen. En tal caso, en esta fase se compara la nueva solución prototipo con la solución ya existente, midiendo de forma fiable la mejora que se tendría al pasar a la nueva solución.

Terminadas estas cuatro fases estaremos en condiciones ver el beneficio de una consultoría Big Data.

Llevar a cabo este tipo de procesos varía en función de las necesidades y en función del sector. En este caso, las consultorías de Big Data son bastante flexibles, en el IIC damos soluciones en varios entornos como salud, banca, RR.HH., entorno digital, entre otros.





Consultoría sobre MS SQL SERVER



- Consejos para el diseño de tablas
- Consejos para el diseño de consultas
- Revisión de planes de ejecución

<https://www.youtube.com/watch?v=FHdTxyR6rnc&list=PL1vDASG8ZwjBZMurvLLVjowpw2FyWFQNt&index=2>

<https://codigosql.top/mysql/crear-base-de-datos/>





Diseño de tablas

- Diseñe pensando en las consultas, no solo en los datos
- Mueva las columnas nulas o que se consulten poco a otra tabla
- Use los tipos de columna correctos para el tamaño esperado
- Declare primero las columnas de tamaño fijo
- Utilice índices filtrados
- Utilice índices en las vistas
- Cree índices en llaves foráneas cuando estas tiendan a ocupar miles de registros
- Actualice/Recree las estadísticas periódicamente





Diseño de tablas

- Utilice índices filtrados
- Utilice índices en las vistas
- Cree índices en llaves foráneas cuando estas tiendan a ocupar miles de registros
- Actualice/Recree las estadísticas periódicamente

<https://www.youtube.com/watch?v=rHcB0KGoysw>

<https://codigosql.top/sql-server/como-crear-tablas-en-sql-server/>





Diseño de consultas

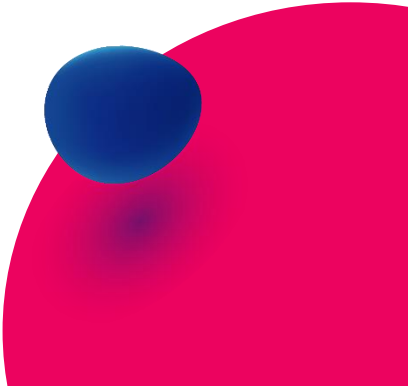
- Usar Exists en lugar de Count()
- No utilizar Not Exists
- Usar joins en lugar de selects anidados
- Utilice CTEs en consultas complejas, pero considere que estos no tienen índices





Diseño de consultas



- Usar siempre consultas parametrizadas, inclusive en las consultas dinámicas ejecutas con “exec”
 - Evitar hacer muchos Joins en una consulta, si es necesario consultar muchas tablas, guardar datos en tablas temporales intermedias o usar CTEs
 - En la clausula Where no use funciones del lado izquierdo del operador
 - No usar Group By sin una función de agregación, en ese caso es mejor usar Distinct
 - No usar Select *, siempre especificar las columnas
 - Evitar usar Order By
- 



Diseño de consultas

- Agregue índices sin nombres en la creación de tablas temporales
- Consulte todos los datos necesarios al inicio del procedimiento, y luego efectúe las operaciones
- Evite los joins o where en columnas sin índices
- Use Hints en casos donde apliquen, ej Recompile







Análisis de planes de ejecución



Siempre analice el plan de ejecución, la mínima instrucción puede generar planes sorpresivamente complejos

Asegurarse que el plan use los índices correctos Revisar en que puntos se hacen escaneos de tabla Comparar Actual Number Of Rows con Estimated Number of Rows Revisar si hay tablas demás involucradas y como podría mejorarse el diseño

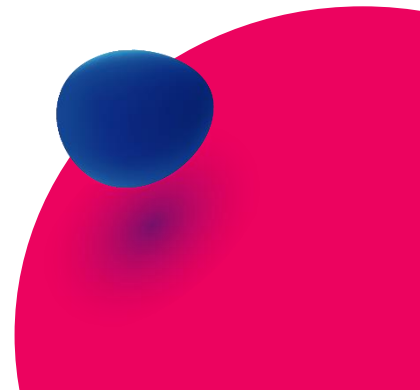
- Asegurarse que el plan use los índices correctos
 - Revisar en que puntos se hacen escaneos de tabla
 - Comparar Actual Number Of Rows con Estimated Number of Rows
 - Revisar si hay tablas demás involucradas y como podría mejorarse el diseño
- 
- 



Análisis de planes de ejecución



- Asegurarse que el plan use los índices correctos
- Revisar en que puntos se hacen escaneos de tabla
- Comparar Actual Number Of Rows con Estimated Number of Rows
- Revisar si hay tablas demás involucradas y como podría mejorarse el diseño





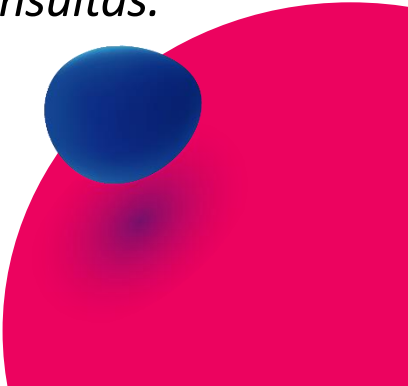

Consultoría sobre ORACLE



¿Que se puede hacer para mejorar el rendimiento de las consultas en Oracle 11?

Tradicionalmente para mejorar el rendimiento de las consultas en SQL contra una BBDD Oracle , una vez optimizada la consulta con técnicas de Tuneling (por ejemplo usando Hints si procede), siempre pasamos a dos técnicas clásicas :

*Reconstruyendo los índices existente en base de datos, para lo cual previamente comprobaremos si esos índices deberían de ser reconstruidos
Actualizando estadísticas sobre las tablas que realizamos las consultas.*



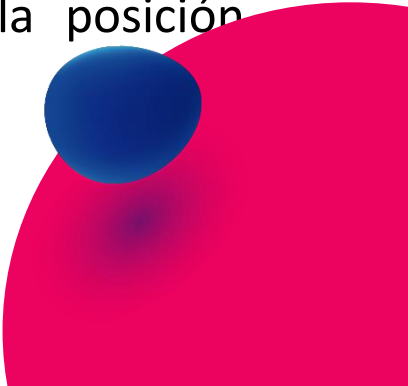



Consultoría sobre ORACLE



Reconstrucción de índices

El índice de una base de datos es una estructura de datos que mejora la velocidad de las operaciones, permitiendo un rápido acceso a los registros de una tabla por lo que se suelen usar sobre aquellos campos sobre los cuales se vayan a realizar búsquedas frecuentes dado que su funcionamiento es similar al índice de un libro: guardando duplas de elemento que se desea indexar junto a su posición en la base de datos, de modo que para buscar un elemento que esté indexado, sólo necesitamos que buscar en el índice de dicho elemento para, una vez encontrado, devolver el registro que se encuentre en la posición marcada por el índice.







Consultoría sobre ORACLE



Los índices son contruidos sobre árboles B, B+, B* o sobre una mezcla de ellos, funciones de cálculo u otros métodos.

El espacio en disco requerido para almacenar el índice es típicamente menor que el espacio de almacenamiento de la tabla (puesto que los índices generalmente contienen solamente los campos clave de acuerdo con los que la tabla será ordenada, y excluyen el resto de los detalles de la tabla), lo que da la posibilidad de almacenar en memoria los índices de tablas que no cabrían en ella. En una base de datos relacional un índice es una copia de parte de una tabla.

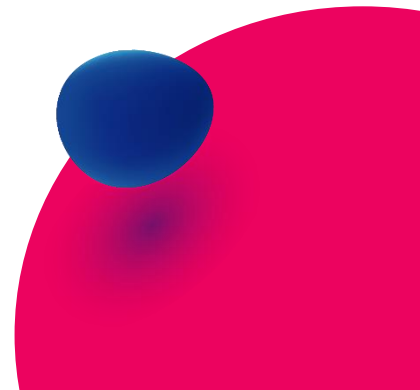




Conclusiones



- Un sistema de base de datos puede reportar muchos beneficios, ya sean estos particulares (el sistema es utilizado por un usuario para satisfacer sus necesidades particulares) o generales (el usuario responde a una compañía u organización que dirige su funcionamiento al trabajo en grupos).
- Con la utilización de sistemas de bases de datos es posible eliminar la redundancia en la información con que se trabaja en una importante empresa u organización, disminuir o bien eliminar totalmente las inconsistencias, aplicar restricciones de seguridad de la información y mantener la integridad e independencia de los datos almacenados.





**FIN DE
GRABACIÓN**