

## La Tecnología CBAT

### CBAT, la pieza clave en el rompecabezas de la explotación de información

Los estudios más recientes acerca de la gestión de la información en las organizaciones, coinciden en que el enorme crecimiento experimentado en los últimos años, es apenas el comienzo de un fenómeno de mayores magnitudes. Esos estudios coinciden en que la información que se generará en los próximos cinco años, superará a toda aquella creada previamente hasta ahora. Y ese incremento no solo será en volumen, sino en complejidad y en diversidad.

Esta rápida expansión de datos en todos los niveles de las organizaciones, ha conducido a la creación de nuevas soluciones de explotación de información, que permitan mejorar la eficiencia de las operaciones, permitiendo realizar análisis complejos sobre un volumen de datos muy grande y de origen heterogéneo.

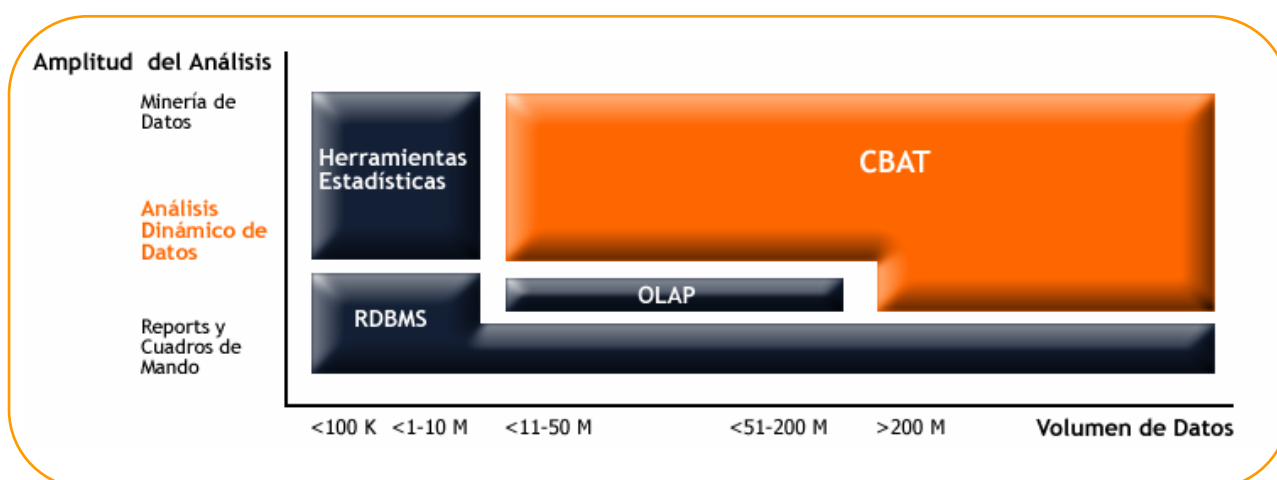
Aunque las organizaciones han estado implantando diferentes tipos de tecnologías de inteligencia de negocio (business intelligence) para administrar grandes volúmenes de información, han comenzado a utilizar una nueva forma de tecnología analítica.

Esta nueva tecnología analítica, llamada CBAT (Column Based Analytical Technology), está permitiendo a las organizaciones realizar análisis de datos a nivel de detalle sobre grandes volúmenes, en forma más eficiente y rentable que las tradicionales. CBAT tiene la capacidad de analizar datos a nivel de detalle, en forma muy rápida, cubriendo así la brecha existente en las tradicionales soluciones de explotación de información.

Las organizaciones, están utilizando actualmente una gran variedad de tecnologías y soluciones, como los almacenes de datos (datawarehouse), cubos dimensionales (OLAP y sus variantes), bases de datos relacionales (RDBMS) y minería de datos (data mining).

Cada una de estas soluciones, tiene limitaciones técnicas para brindar un rendimiento aceptable, por lo que intentan mejorar mediante indexaciones, sumarizaciones, restricciones en el análisis, cortes temporales o trabajando con pequeñas muestras de datos.

Por el contrario, CBAT estructura los datos en columnas y no en filas ni cubos. Su estructura permite consultas muy rápidas sobre grandes volúmenes de datos, permitiendo a los usuarios realizar análisis sobre la marcha, con la máxima granularidad, obteniendo así resultados inalcanzables con cualquier otra tecnología o solución.



## ¿Qué es CBAT?

CBAT es una tecnología de análisis a nivel de detalle de alta velocidad. Proporciona una respuesta extremadamente veloz con grandes volúmenes de datos, es muy rápida de desplegar, fácil de administrar y requiere un hardware muy ligero.

Utilizando soluciones CBAT, las organizaciones ahorran dinero, incrementan productividad y ganan en competitividad a través del análisis en profundidad de sus propios datos.

CBAT es la solución ideal para análisis que requieren:

- Análisis y necesidad de información no previsto
- Nivel de detalle sobre un gran volumen de datos
- Gran sensibilidad al tiempo de respuesta

CBAT está diseñado para realizar comparaciones a gran velocidad y manipular grandes volúmenes de datos guardados en columnas, a diferencia de la tradicional tecnología de filas o de cubos; dando respuesta con una granularidad inalcanzable para otras tecnologías.

Para alcanzar ese nivel de detalle, las soluciones CBAT analizan muchos registros en forma individual hasta encontrar aquellos que comparten características comunes, basadas en opciones del usuario.

Las soluciones CBAT, pueden además, identificar subgrupos de registros con características comunes mediante múltiples comparaciones, aspectos interrelacionados de comportamientos y otros datos en columnas al mismo tiempo.

El resultado es un set de datos que se mantiene a nivel de detalle y que puede ser comparado y alineado en muchas columnas de una única forma por cada ítem de la muestra.

Manteniendo ese nivel de detalle, CBAT produce resultados que incluyen reports o que pueden ser utilizados para análisis posteriores, sobre la marcha.

CBAT brinda a las organizaciones una forma diferente de consultar y de procesar esas consultas, de un modo dinámico, interactivo y de alto rendimiento, comparativo y a nivel de máximo detalle.

## ¿Cómo trabaja CBAT?

La unidad básica de guardado de CBAT es la columna, al contrario de las bases de datos transaccionales que utilizan una fila de una tabla como unidad básica. CBAT diseña el guardado en cada columna individualmente y provee una indexación natural para cada atributo.

Esta forma natural de guardado e indexado, es fácil de leer en memoria y de almacenar dentro de índices, indexando solamente lo necesario para mover los datos requeridos.

En el sistema CBAT, los métodos de indexación adicionales se realizan automáticamente, condensando las columnas de datos, eliminando repeticiones, reduciendo necesidad de memoria y mejorando el rendimiento de la lectura de datos.

Muy por el contrario, las bases de datos basadas en filas, escriben filas de datos completas en tablas. A diferencia de la indexación que hace CBAT, los índices de estos sistemas basados en filas están definidos por los administradores de las base de datos. Esos índices se construyen exclusivamente para los campos seleccionados y están guardados en ficheros separados de los datos.

• • •

***“Lo que distingue Alterian de otras aplicaciones basadas en aproximaciones query, es que Alterian utiliza una base de datos basada en columnas que junto con su sistema de indexación hacen único a Alterian, proporcionando niveles de rendimiento que son órdenes de magnitud superiores a los métodos tradicionales.”***

**Philip Howard, Senior Analyst,  
Bloor Research**

• • •

Cuando un índice debe ser utilizado, se deben realizar dos lecturas. Una lectura del índice para encontrar los registros solicitados y una lectura adicional por cada uno identificado en cada una de las filas de datos.

Además, los sistemas basados en filas usan tablas unidas siempre que necesitan resolver consultas, cada fila de datos de cada una de las tablas relacionadas debe ser leída y escaneada en memoria, perdiendo el rendimiento.

CBAT, al seleccionar campos individuales, completa el mismo proceso de tablas relacionadas extremadamente rápido, sin pérdida de rendimiento. Mientras los sistemas basados en filas necesitan un hardware muy potente, y aun así, son muy poco flexibles para temas de análisis que requieran escanear muchas tablas; los sistemas CBAT dan respuesta muy rápida con un hardware muy ligero, por complejo que sea el análisis a realizar.

## Arquitectura CBAT

Las soluciones CBAT, habitualmente están construidas utilizando una arquitectura de cuatro capas.



### Capa de Visualización

Mejora el análisis de datos y provee un sistema de diseño, a través de técnicas de visualización como Mapping, Charting, Plotting, etc.

### Capa de Análisis

Consultas Recursivas con Guardado Intermedio. Permite a los usuarios descomponer complejos procesos de análisis en set de datos que pueden ser utilizados para otros análisis. Utilizando guardados intermedios, es posible reducir pasos y evitar duplicaciones, reutilizando los resultados de pasos previos. Adicionalmente, agregando nuevas columnas, los datos pueden ser rápidamente preparados para modelar para realizar minería de datos o scoring.

Operaciones de Intersecciones. Permite a los usuarios definir sets y subsets de un universo y realizar operaciones con ellos. Esas operaciones incluyen intersecciones, uniones o exclusiones.

Campos Derivados. Permite crear nuevos campos de datos basados en scores, rankings, métricas o grupos de valores derivados de los datos nativos.

### Guardado Basado en Columnas

Los datos son guardados y accedidos mediante columnas, de manera que permiten un alto rendimiento con consultas complejas sobre grandes volúmenes de datos.

Una arquitectura de guardado en columnas evita la necesidad de sumarizar o realizar cortes temporales, permitiendo la máxima granularidad y una rápida respuesta a cualquier consulta.

### Carga de Datos

Tiene la capacidad de integrar índices con una carga muy rápida de datos basados en estructura de columnas para soportar consultas analíticas de alta performance. Esta capacidad hace que CBAT sea extremadamente útil en auditoría y limpieza de datos.

## Propósito de RDBMS (base de datos relacional basada en filas), OLAP (tecnología basada en cubos) y CBAT (Tecnología Analítica Basada en Columnas)

A un nivel simple y con poco volumen de datos, todas las tecnologías de explotación de información se solapan, son similares. Pero con grandes volúmenes de datos es necesaria la tecnología adecuada para dar respuesta a la necesidad de información específica.

CBAT guarda y manipula datos en grandes bases de datos en columnas, no en filas. Esta diferencia es crítica, ya que es lo que permite el rendimiento, la velocidad y la capacidad en comparación con otras tecnologías.

CBAT utiliza los datos a máximo nivel de detalle, no hace agregados o sumalizaciones en cubos, ni cortes temporales. Esto brinda la posibilidad del análisis dinámico con toda la profundidad requerida para transformar datos en conocimiento para las organizaciones.

| Categoría BI | Tecnología         | Propósito   |
|--------------|--------------------|---|
| RDBMS        | Basada en Filas    | Insertar, modificar y borrar transacciones; el objetivo de los sistemas operacionales. Poco flexible para análisis complejos.         |
| OLAP         | Basada en Cubos    | Agregación o sumalización de información, mediante la definición de cubos. No trabaja a nivel de detalle.                             |
| CBAT         | Basada en Columnas | Alto rendimiento, dinámico e iterativo al máximo detalle. Excelente para granularidad, detalle y necesidad imprevista de información. |

### Sobre Ibérica Alterian

Empresa pionera en el análisis dinámico de datos y la explotación de la información, desarrolla proyectos tecnológicos que permiten transformar la información en conocimiento estratégico para empresas y organismos de diferentes sectores privados y públicos.

Alterian ha desarrollado su propia tecnología a la que denomina CBAT (Column Based Tecnology), pilar sobre la que se sustentan sus aplicaciones y herramientas. La tecnología de Alterian permite el desarrollo de múltiples soluciones, tanto a niveles sectoriales como departamentales. Diferentes consultoras independientes como Gartner, Forrester o Bloor suelen mencionar a Alterian como *world class* en soluciones de explotación y análisis de datos.

Entre sus clientes en España se encuentran El Corte Inglés, Telefónica, Vodafone, Travel Club, BBVA, La Caixa, Metro de Madrid, Ministerio de la Presidencia, Orange, Renault, Arbora & Ausonia, Junta de Andalucía, Generalitat de Cataluña, Telepizza, Ifema, Gamesa, etc.

#### Barcelona

Frederic Mompou 5  
Edificio Euro 3  
08960 Sant Just Desvern

#### Madrid

Paseo de la Castellana 210  
8ª Planta, 10  
28046 Madrid

#### Sevilla

Torneo Parque Empresarial  
Avda. Tecnología 26  
Edificio Vilamar 1  
41015 Sevilla

Tel. 902 36 35 03 • [alterian@alterian.es](mailto:alterian@alterian.es) • [www.alterian.es](http://www.alterian.es)