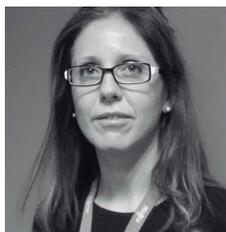


Análisis *real time* y nuevos modelos de negocio

Prueba de concepto de SAP HANA



SAP HANA es una plataforma integrada, compuesta por hardware y software, que combina tecnologías innovadoras de base de datos. Conformar una plataforma in-memory para aplicaciones de alto rendimiento, como herramientas de análisis y simulación en tiempo real. Estas características, trasladadas a un contexto de negocio, ponen de manifiesto las enormes posibilidades que ofrece esta plataforma.



Andrea Toscanini
Manager Business
Analytics

FUJITSU

Los requerimientos de los negocios modernos son cada vez más exigentes. Los mercados cambian a gran velocidad y requieren soluciones que ofrezcan resultados fiables y veraces de forma prácticamente instantánea. SAP HANA plantea un importante paso en esta dirección. Como solución de base de datos, no sólo ofrece almacenamiento tradicional (columnar), sino que también hace uso de tecnologías de optimización, procesamiento en paralelo y diversas técnicas de compresión para ofrecer un nivel de rendimiento diferencial. Todo esto se une a las posibilidades

que ofrece uno de sus componentes, Sybase Replication Server, que se emplea para copiar y sincronizar en tiempo real los datos procedentes del BW, del ERP y del CRM.

In-memory computing

La cantidad y complejidad de los datos que manejan hoy en día las herramientas de BW, BI y similares provocan una carga de procesamiento y tiempos de ejecución tan elevados que con frecuencia pueden llegar a limitar su utilización frecuente. Por ejemplo, si una empresa dispone de bases de datos de 6 o 7 terabytes en sus BW, necesitará varias horas de procesamiento para construir un "cubo" y poder acceder a la información que de él se desprende.

SAP HANA representa un salto cuantitativo en este sentido. Por un lado, el avance que se ha producido en el ámbito de las tecnologías RAM: ahora se dispone de la posibilidad de alojar las bases de datos en memoria de alta capacidad en lugar de hacerlo en discos duros (mucho más lentos). A esto se le llama *in-memory computing*. Esta idea no es nueva, ya existían avances en este sentido como, por ejemplo, el Business Warehouse Accelerator de SAP, que utilizaba esta tecnología hace algún tiempo.

Hoy en día se están barajando ya memorias RAM capaces de alojar terabytes de información. Al eliminar el acceso a discos duros, el

procesamiento *in-memory* consigue tiempos de respuesta hasta 10.000 veces menores y puede procesar datos a una velocidad de 100 GB por segundo.

Hay otro ingrediente fundamental en el *core* de SAP HANA: las innovadoras tecnologías en el tratamiento de la base de datos. Por ejemplo, el almacenamiento de tablas por columnas ofrece posibilidades de compresión de los datos realmente importantes, pudiendo dejarlos en hasta 1/3 de lo que ocupan actualmente. Por otro lado, los algoritmos de procesamiento paralelo (multinúcleo) procesan datos a gran velocidad permitiendo, por ejemplo, consultar 360 millones de registros en apenas un segundo. Pero lo que realmente destaca de SAP HANA es la posibilidad de consultar los datos de analítica del negocio en *real time* y/o *near real time*.

Prueba de concepto

El laboratorio de Fujitsu España ha realizado una prueba de concepto (*Proof of Concept* [POC]) para una empresa de energía renovable que necesita realizar predicciones de producción energética de las plantas eólicas y solares de sus clientes. Para ello combinan modelos de previsión meteorológica con sofisticados algoritmos, obteniendo una predicción de generación energética con alta precisión y fiabilidad para numerosas tecnologías (eólica, solar, hidráulica, mareomotriz, etc.).

El reto era la necesidad de minimizar el tiempo de ejecución de un cálculo complejo. Se trata de un algoritmo que tiene como fin

predecir la cantidad de energía eólica que puede generar una planta específica en un período determinado.

Inicialmente, el tiempo de ejecución era de dos horas por cada planta, para calcular todo un año. Dado que el número de plantas a predecir es de 800, el tiempo de análisis total es muy extenso (superando los dos meses de procesamiento continuo). Tras contactar con el cliente, Fujitsu vio que SAP HANA encajaba perfectamente en la resolución del problema de tiempos de ejecución demasiado largos. SAP HANA demostró que, no sólo es capaz de manejar cantidades ingentes de datos, sino que también permite la rápida integración de diferentes fuentes en tiempo *record* y puede "paralelizar" los diferentes procesos (velocidad, variedad y volumen), además de simplificar un escenario con cierto nivel de complejidad. Permite disponer de los datos a partir de todas las preguntas (valor) que se puedan ocurrir acerca del negocio (visualización, variabilidad y veracidad).

En la tabla adjunta hemos reunido una serie de KPI que resultaron de la prueba de concepto. De ellos se desprende que aún no se han explotado todos los beneficios que ofrece HANA, porque la creatividad que proporciona en su aplicación es prácticamente ilimitada. ●●●

	Indicadores de medición (KPI)	Resultado	Beneficios de negocio asociados a los resultados de la POC
1	Tiempos de implantación de la nueva tecnología basándose en el desarrollo de la POC.	48 días	Agilidad, rapidez y facilidad de implantación de la nueva tecnología, reduciendo costes de implantación en un 400% sobre el BI tradicional (de 2 a 8 meses aprox.).
2	Disminución del tiempo de carga: 5 horas para 300 millones de registros en la predicción actual.	270 segundos	Incremento de la capacidad de procesamiento y flexibilidad en el manejo de históricos para generar nuevos modelos predictivos y mejorar los actuales.
3	Acceso a datos en SAP HANA (antes Oracle)	Mejora x 10 veces	Ajuste de las predicciones meteorológicas de las 800 plantas actuales minimizando el margen de error y optimizando sus resultados operativos y económicos (es el punto crítico actual).
5	Factor de compresión para reducir exponencialmente el almacenamiento en la base de datos.	x 5	Permite un crecimiento a corto plazo de un 250% en el servicio de predicción de plantas de energía renovable para pasar de 800 a 2000 plantas, lo que permite expandir el negocio.
7	Tiempo de cálculo de la ejecución del algoritmo predictivo R (actual en Matlab 2 horas por planta x 800 plantas).	3 horas en total	Generación de una nueva línea de negocio para instalaciones domésticas a nivel mundial, lo que supondría un crecimiento de cuatro o cinco millones de usuarios finales.
8	Tiempo de ejecución de simulaciones para oportunidades de venta (48 horas).	30 segundos	Aumento de la competitividad comercial con simulaciones en <i>near real time</i> para los potenciales clientes.