



# Data warehouse

## todavía es el rey

Una pregunta frecuente y que divide opiniones es: ¿son mejores los macrodatos (*big data*) o el almacén de datos (*data warehouse*)? Hay argumentos sólidos para afirmar que el primer concepto es lo de hoy, con un gran volumen de datos no estructurados, su variedad y la velocidad con la que se generan, así como las distintas fuentes externas provenientes de internet y las transacciones de fuentes de datos automatizadas. Sin embargo, ¿cómo negar que el *data warehouse* es tecnología “antigua” que nació alrededor de los años 80? En la edición anterior de *Veritas*, se mencionó en esta columna que incluso el autor Thomas H. Davenport habla del *data warehouse* como la era más antigua de la Analítica o *business analytics*.

A pesar de que las estadísticas dicen que en 2020 la proporción de datos generados en el mundo se compondrá 80% de datos no estructurados, no hay que olvidar el 20% restante, que son los datos estructurados que provienen de sistemas transaccionales centrales de las empresas, por ejemplo, los administradores de facturación, tesorería, inventarios y producción, entre otros.

La fuente principal de conocimiento empresarial proviene de su interior, lo que en inglés se llama *insight*. Las organizaciones buscan generar valor económico y la manera en que lo pueden lograr es al responder estas tres preguntas:

1. ¿Cómo tener una operación más eficiente?
2. ¿Cómo ser más rentables?
3. ¿Cómo generar ventajas competitivas?

El *insight* está en gran medida en los datos y ese 20% no es nada desdeñable. Por el contrario, es información vital para comprender a la organización. La manera tradicional de organizar esos datos se hace con una arquitectura de *data warehouse*, que de acuerdo con Bill Inmon, científico considerado el padre del almacén de datos, es una arquitectura (a diferencia de los macrodatos) que conforma una herramienta tecnológica.

### Ventajas del almacén de datos

1. Es un mecanismo de integración e interoperabilidad.
2. Sirve para la gestión adecuada de datos maestros y de referencia (MDM, por sus siglas en inglés).
3. Implica una correcta gestión de los metadatos, al empezar por el significado en términos de negocio y continuar con las reglas de negocio que le dan sentido.
4. Es una excelente oportunidad para evaluar la calidad de los datos provenientes de las distintas fuentes transaccionales y disparar acciones para mejorarla.
5. Es una práctica madura, con abundante literatura, marcos de referencia y mejores prácticas, a diferencia de los macrodatos.

Desde la perspectiva de la gestión de información, el almacén de datos es la base para la inteligencia empresarial, así como los macrodatos lo son para la ciencia de datos. El primero contesta cierto tipo de preguntas, el segundo, otro. Ambos tipos son importantes para crear valor económico. Lo más relevante es tener la pregunta de negocio para saber a quién cuestionarle. Si no se tiene la pregunta no se comprenderá la respuesta. Desde mi perspectiva, para contestar las cuestiones elementales y por las razones arriba enumeradas, el *data warehouse* todavía es el rey.

Felices datos.📊