

## “Data Warehousing”

Relatório sobre o seminário conduzido pela Eng.<sup>a</sup> Ana Cunha

Traduzido à letra, *Data Warehouse* significa armazém de dados. Logo, falando do ponto de vista informático, é um sistema que trabalha sobre uma base de dados.

No decorrer do seminário pude concluir que existem basicamente dois tipos de sistemas: **sistemas operacionais** e **sistemas analíticos**. O primeiro é um sistema em que se inserem, editam e retiram dados, logo podemos dizer que a sua principal função é a actualização dos dados. Geralmente, as bases de dados destes sistemas estão na terceira forma normal, o que significa que não existe informação redundante, facilitando assim a actualização do sistema, visto só ser necessário fazer alterações num único sítio.

Já os sistemas analíticos trabalham sobre bases de dados não normalizadas, o que permite fazer pesquisas muito rápidas, coisa que o sistema anterior não proporciona. Os **DW** são sistemas analíticos. Estes sistemas são utilizados, por exemplo, para pesquisas de mercado, operações de marketing e outro tipo de operações baseadas em pesquisas por vezes complicadas. Existe informação redundante, uma vez que os sistemas não estão normalizados. Isto implica uma maior ocupação de espaço em relação aos sistemas operacionais. Os sistemas analíticos permitem também o cruzamento de informação.

Uma das maiores vantagens de uma **DW** é a possibilidade de fazer exploração de dados (“data mining”). Esta exploração é efectuada com base em várias ferramentas que podem ser do tipo OLAP, ROLAP, MOLAP, Query Ad-Hoc ou Reporting. Cada uma destas ferramentas é orientada para um grupo de utilizadores e a gama de funcionalidades pretendida.

Uma empresa tem ao seu dispor, normalmente, um sistema analítico e um ou vários sistemas operacionais. Assim, o sistema vai ser actualizado pelo sistema operacional, actualização esta que pode ser muito demorada, logo não pode ser realizada em tempo-real. Esta actualização tem de ser feita de modo que a diferença entre o sistema analítico e o sistema real não seja relevante. O sistema analítico é fácil de pesquisar e difícil de alterar e o sistema operacional é fácil de alterar mas difícil de pesquisar.

A modelação de dados numa **DW** deverá ser feita pelo conceito “Star Schema”. Segundo este conceito a **DW** é assente em tabelas de factos e de dimensão, deste modo é atingido um tempo de resposta óptimo. A coerência dos dados é vital para o desempenho da **DW**, possíveis incoerências são resolvidas em definições claras e partilhadas por todos os intervenientes na introdução de dados e validação semântica dos dados contra as definições.

Conclui-se então que um *Data Warehouse* é um sistema de inquestionável importância para uma empresa de grande dimensão, visto que permite obter rapidamente, e com um nível de fiabilidade elevado, dados vitais para o correcto crescimento de uma empresa.