

# From data warehousing to Active Information Integration system

Mukesh Mohania

IBM India Research Lab  
mkmukesh@in.ibm.com

**Abstract.** Enterprises have gathered operational business information from multiple structured data sources and stored it in a central repository, called data warehousing, for decision support functionalities and data analysis. The enterprises are now realizing to integrate their entire information sources, including "unstructured" contents, for deeper and richer information analysis. Several applications, such as processing warranty claims, finding promotional materials in real-time based on user's transaction value, detecting health insurance claim processing frauds in (near) real-time by integrating information from various data sources (some of them may be from the competitors), etc., require integration of both structured and unstructured information based on events and business policies. Thus, it is vital for data warehousing to enable the integration of data and content sources to provide real-time read and write access, to transform data for business analysis and data interchange, and to data placement for performance, currency and availability.

In this talk, we will first review the existing technologies in data warehousing and information integration, and then discuss how the enterprise applications are moving from data warehousing to (Active) Information Integration system. We will also discuss an architecture of a new approach for integrating information based on policies that does not require to defining a global schema (virtualization approach) or any materialization of pre-computed results (warehouse approach). We will finally discuss several applications that require such kind of integration, and show that the current approaches cannot satisfy these applications.