Tutorial 1

Data Warehousing and Parallel Data Systems

Instructor

Rick Stellwagen

CTO, Parallel Systems Division NCR Corporation,

Data warehousing has developed into a rapidly growing business. Today, there are hundreds of production-level data-warehouse installations. Parallel database systems and decision-support systems have provided the foundation for most of these production data warehouses.

Currently, thousands of data warehouses and data marts are being planned and installed. A substantial number of new products along with new methods and models have appeared on the market that claim to be data-warehousing "solutions." This has resulted in a wide variety of definitions of the data-warehousing problem.

This tutorial will review the current industry definitions of data warehousing. The various components of data warehouse systems will be outlined, including inflows (transformation), administration, user-information access and database processing. More in-depth discussions will center around data warehouse architectures, the types of user access that are most prevalent in commercial data warehouses, and data models and programming models for achieving performance objectives. Also to be discussed are special techniques that are used to address performance issues based on characteristics of the workload.

This tutorial will define a framework that allows the true potential for re-use and data sharing. This includes optimizations that are maximized with use of a scaleable parallel system as the hub and spokes in a two- and three-tier data warehouse implementation. Several different database and parallel database products will be studied to show how they fit into the various data-warehousing architectures. Finally, the tutorial will consider the role of the internet and multi-media data in this environment.

About the speaker: Rick Stellwagen received his B.S. in Computer Science from S.U.N.Y at Brockport in 1979. He has been active in both research and development of database and information products for the past 17 years.

He produced heterogeneous distributed information systems in early 1980s and was the architect and lead developer for DBSR, NCR’s first relational DBMS in early and mid 1980s. He has built and designed query tools, message-oriented middleware for data and meta-data placement and integration, along with performance, simulation, and modeling tools.

Rick was an architect and director of the Data Navigator and Configurator project (later to be Sybase MPP) that NCR and Sybase pioneered as the first open parallel database system designed for deployment on commodity clusters of servers.

His experience includes being a member of the ANSI X3H2 and RDA committee on databases, and a founding member of the SQL Access committee. Today he is responsible for strategy and architecture of NCR’s Data Warehouse Program.