

This is for people taking seminar (and those with an interest in data warehousing).

due to popular demand (to have reading assignment prior to Thursday afternoon), here is the reading for next week's seminar. we'll distribute XML reading soon (probably Friday).

Note that some of the articles are in ps (postscript), not pdf. postscript is a fairly standard text format with unix/academics. since everybody might not have a postscript reader, kate and i are asking our web master to download those ps files, convert them to pdf, and put them on the web site....

For seminar on data warehouse -
read at least three papers (one from each of the following three groups).

If you don't want to take the time to choose, select the first of the group. Note that one of my research contacts has said: "Data Warehousing is considered a 'mature' technology now among researchers". That means that most people believe the interesting, critical research issues are solved, and it's now up to industry to deliver on that, with viable products.

also, as per our pow-wow on Tuesday, we will consolidate the two web xing files, and have fewer initial discussion topics (i.e., not everyone is required to post an idea and this task will be rotating and assigned each week). Matt, Zonke and Dan Rice have agreed to post the reflective pieces to web crossing, which others will read and respond to, prior to seminar. They will post by Friday midnight or sometime Saturday.

Of course, if someone other than those three has a "burning issue" to discuss, he (and there are only he's, except for kate and me!) is free to post that idea.

peruse this site, and learn about Stanford's research effort.
<http://www-db.stanford.edu/warehousing/index.html>
Read one of the following (from the papers at the site):

J. Widom, "Research Problems in Data Warehousing." In Proceedings of the 4th Int'l Conference on Information and Knowledge Management (CIKM), November 1995.

J. L. Wiener. "What is data warehousing and what is Stanford doing about it?" An overview talk given in the Stanford DB Seminar series, Fall, 1997. (Gzipped version.)

H. Garcia-Molina, W. J. Labio, J. L. Wiener, Y. Zhuge. "Distributed and Parallel Computing Issues in Data Warehousing." In Proceedings of ACM Principles of Distributed Computing Conference, 1999. Invited Talk.

Read one of the following two experience papers:

<http://www.cio.com/archive/020101/harrah.html>
Jackpot! February 1, 2001 - CIO

A casino's big payoff came from using IT to manage customer information.

http://www.cio.com/archive/041501/et_revisit.html

Shattering the Glass House BY FRED HAPGOOD

Ten years later, centralized databases still rule—but not everywhere

Read one (or two) of the following industry reports:

<http://www.almaden.ibm.com/cs/quest/papers/whitepaper.html>

Data Mining: Extending the Information Warehouse Framework

IBM

A white paper which explores the technology underlying today's data mining applications.

http://www.datawarehouse.com/iknowledge/whitepapers/knight_138.pdf

Planning for the Next Generation Data Warehouse

Knightsbridge Solutions Corporation white paper

As data-warehousing technology matures, the next iteration of this technology must be prepared to handle enormous amounts of data.

<http://www.kenorrinst.com/datawh.html>

Developing a Data Warehouse Strategy

The Ken Orr Institute, 1998

A white paper that covers data warehousing technology. Includes sections on managing data warehouses, operational vs. information systems and future developments.