

Business Intelligence (BI) Task Force

January 16, 2007

Presentation Outline

- ◆ DSS Data Retention –Overview
- ◆ Cost Trends
- ◆ Performance Trends
- ◆ Initial Concept for data retention

DSS Data Retention - Overview

A data warehouse is designed for storing large amounts of data for long periods of time. But there must be a cost/benefit analysis to determine how long to store that data.

The goal of the BI team is to provide maximum customer value at a reasonable and sustainable cost.

Why is there a need for addressing continued data volume growth in DSS? (Without a retention policy, data volume and associated costs will continue to grow, year over year)

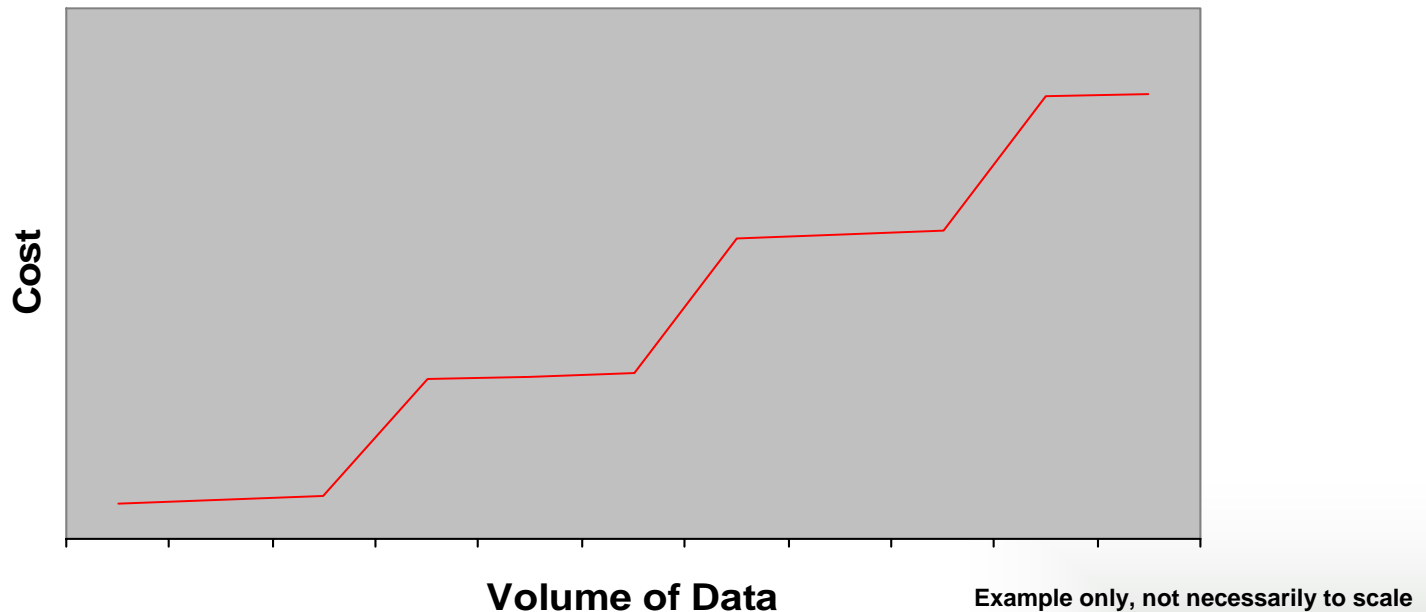
DSS Data Retention - Costs

- ◆ Costs are primarily in two areas
 - *New Hardware costs*
 - Physical disk space
 - Server CPU usage
 - Space in a highly valued, climate controlled area
 - Network/server infrastructure
 - *Maintenance costs*
 - Backups / recoveries require more FTE time
 - Physical maintenance of disk drives/etc.
 - Added DBA costs

- ◆ Cost escalation is not linear.

DSS Data Retention - Costs

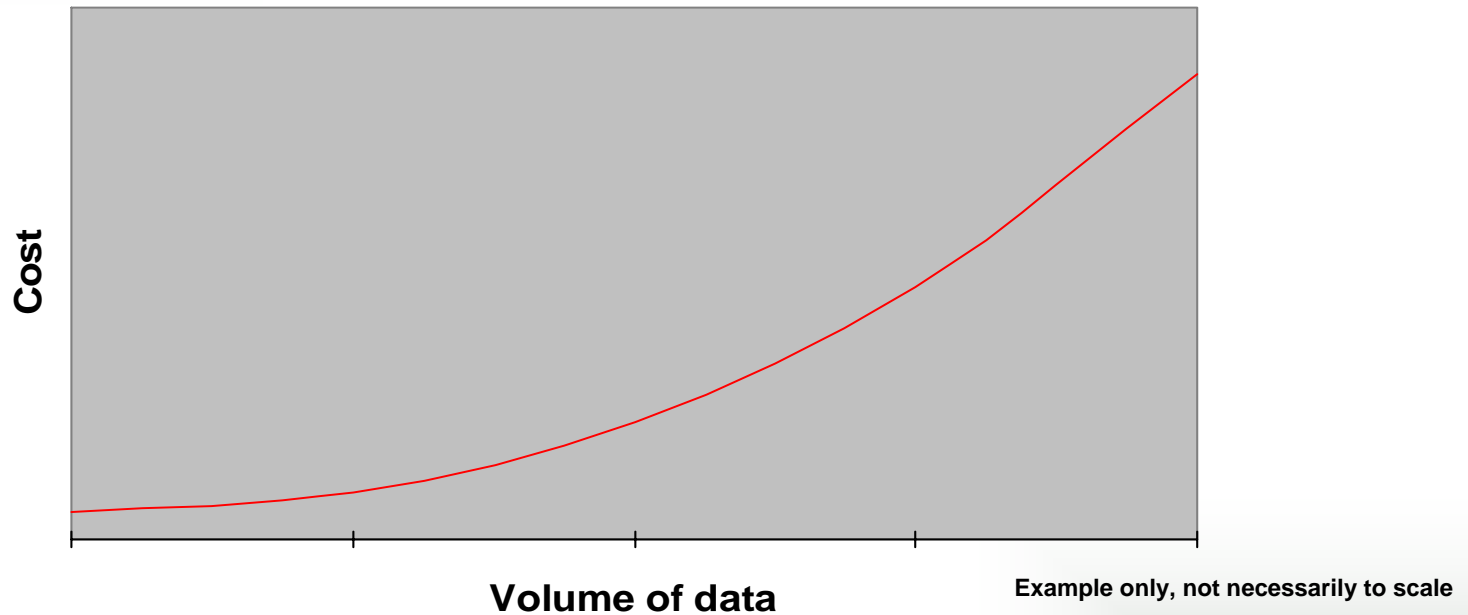
Hardware costs for DSS



Hardware costs for DSS escalate as a step graph, costs grow slowly until a disk array is filled. Then a large capital expenditure is required to increase the available space.

DSS Data Retention - Costs

Maintenance Costs of DSS



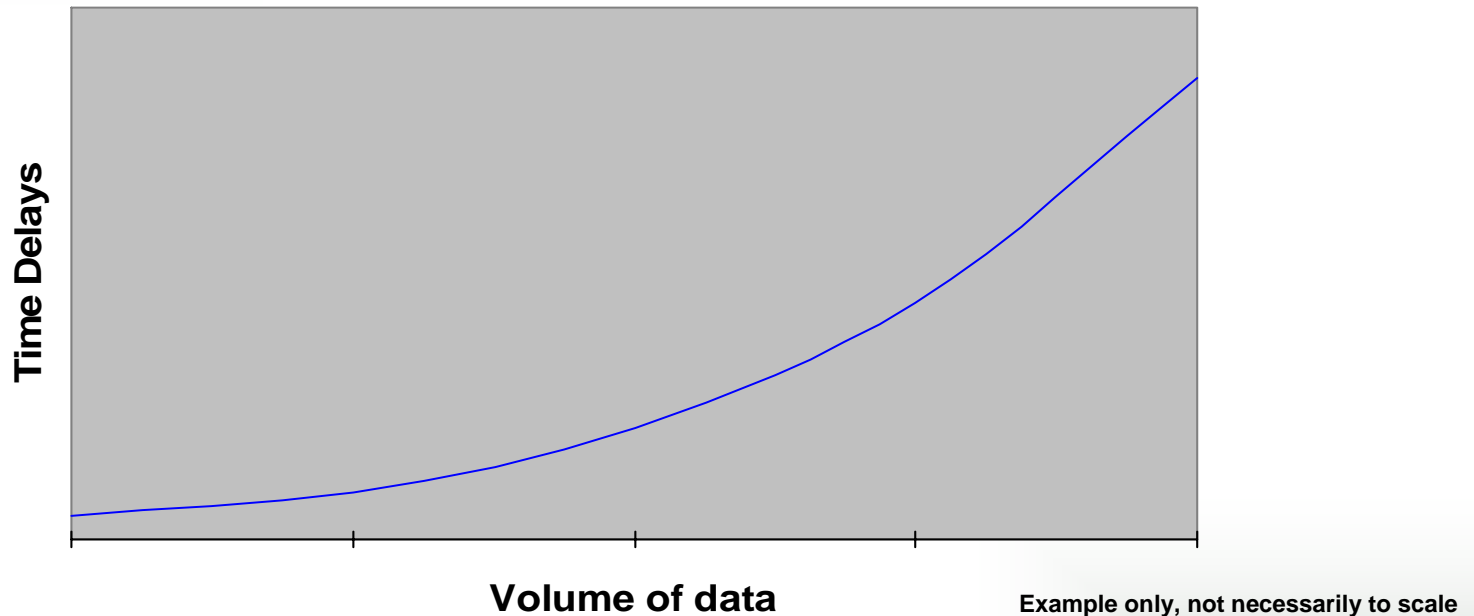
Maintenance costs for DSS escalate on a curve, as data volume increases, costs increase at a greater rate

DSS Data Retention - Performance

- ◆ Performance implications
 - *Overnight load times*
 - It is taking much longer to load the data mart each evening due to volume of data. Tables have billions of rows of information
 - The data mart is sometimes delayed in its availability.
 - *Report/Query times*
 - Applications run slower due to huge table sizes
 - Every report and query takes longer due to increasing data volume.
 - *Performance degradation is not linear.*

DSS Data Retention - Performance

Performance of DSS



Performance degrades according to volume of data. Addressing performance issues involves larger computers, more CPU time and is very costly

DSS Data Retention - Concept

- ◆ There are 2 primary ways to address data volume in a data warehouse
 - *Remove data after a designated time frame*
 - *Reduce the amount of data stored for that designated time frame*

DSS Data Retention - Concept

For discussion only, this is not a policy or a policy recommendation at this time

- ◆ Remove data after a designated time frame, so long as such data removal does not create a compliance problem under the NYISO's tariff.
 - *Retain DSS Customer Settlements data for the current year plus 7 full years.*
 - *DSS currently has 7 years of data.*
 - (This would mean that during 2008 the system could begin removing data from the year 1999.)

*** retain all versions for any invoice under legal challenge*

DSS Data Retention - Concept

For discussion only, this is not a policy or a policy recommendation at this time

- ◆ Reduce the amount of data stored during the designated data retention time frame
 - *Retain all invoice versions and supporting data for up to 6 months after final invoice challenge period, (then remove intermediate versions of invoice while retaining final invoice version)*

*** retain all versions for any invoice under legal challenge*