

Managing Big Data Using Sybase® IQ VLDB Option

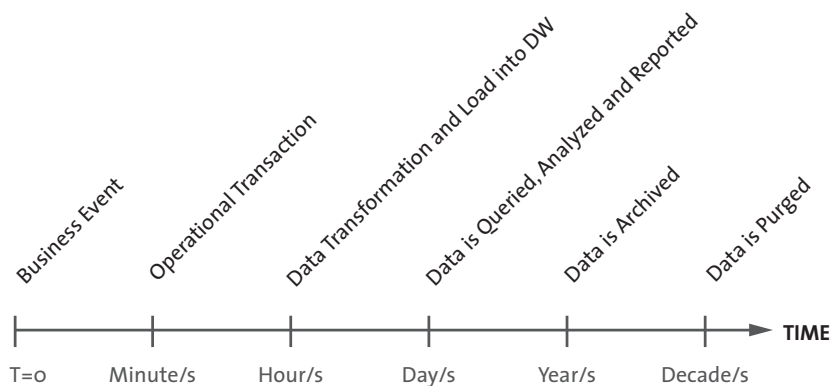
TABLE OF CONTENTS

- 1 Data Explosion
- 2 Sybase IQ 15 VLDB Option
- 2 Storage Layout and Configuration
- 3 Partitioning and Placement
- 4 PowerDesigner® ILM Features
- 5 Summary

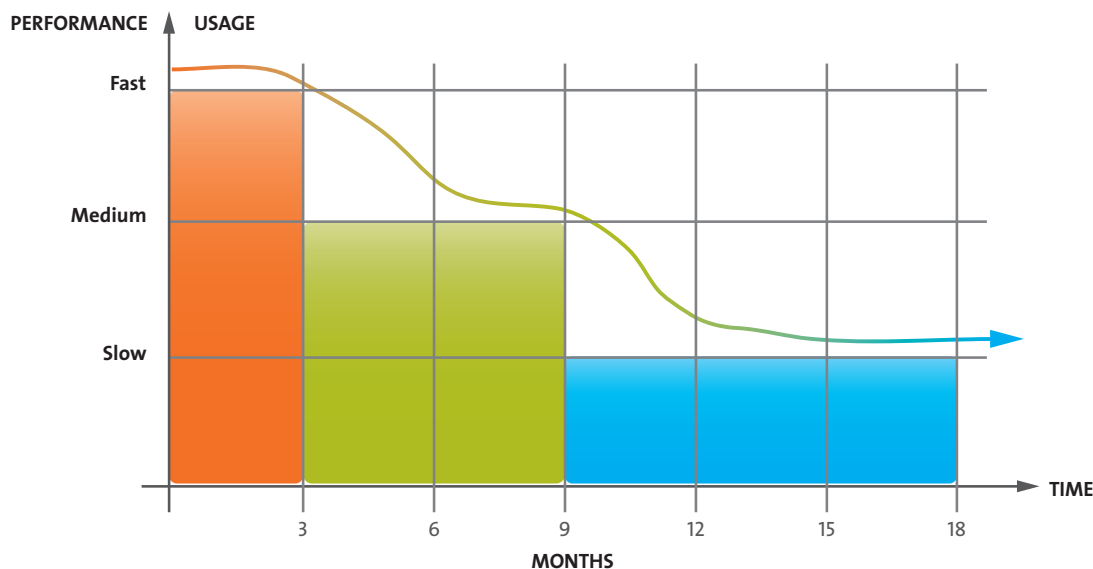
DATA EXPLOSION

As organizations continuously push the boundaries of storage size and storage bandwidth, scalability of an analytics infrastructure is critical. The amount of data that an organization must manage is growing due to business expansion, compliance regulations, and long term data retention and preservation requirements. Information complexity and overload are forcing companies to institute intelligent data management policies and procedures in order to control costs and stay competitive. That means knowing the value of your information, and classifying it accordingly.

You know that data tends to decrease in value over time:



And you know that newer data is accessed more frequently than older data:



Information Lifecycle Management (ILM) is a set of strategies for managing the growth, storage and retrieval of information over its useful life. Its goal is to insure that frequently accessed data is returned as fast as possible, while controlling storage costs. Sybase IQ includes a set of storage management, data partitioning, and data placement features that enable you to implement ILM.

SYBASE IQ 15 VLDB OPTION

The Sybase IQ VLDB option includes the following capabilities:

- Creation of multiple DBSpaces as logical containers for database objects
- Division of the database into read-write and read-only areas for streamlined administration
- Range partitioning of data on a column value, such as 'date'
- Placement of database objects (tables, indexes, partitions, etc.) into a DBSpace of choice

Sybase's PowerDesigner modeling tool also has some features that support the ILM capabilities in Sybase IQ. Users can build a data model that:

- Defines DBSpaces and their cost attributes
- Configures a table partitioning scheme
- Generates data movement scripts to move data partitions through different DBSpaces over time
- Generates a cost savings report for an ILM scenario

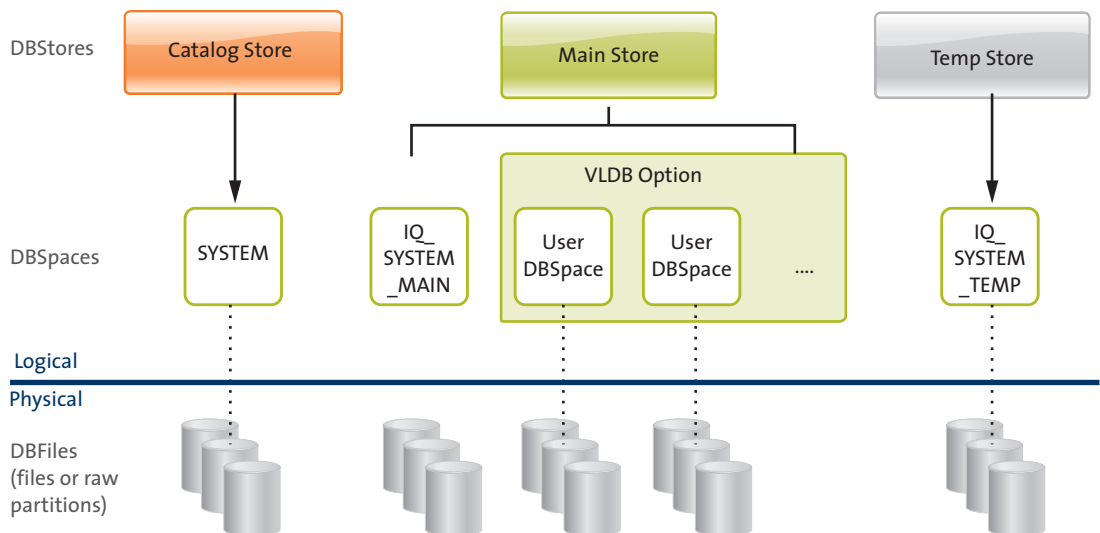
For administrators of very large databases, these features are the tools for classifying, organizing, storing, and migrating data according to its value.

STORAGE LAYOUT AND CONFIGURATION

Sybase IQ 15 lays out a database as a set of logical stores:

- Catalog store: stores database metadata
- Main store: stores system structures and user data
- Temp store: stores temporary data generated during loads and queries (sorting results, etc.) and user defined temporary tables

Stores are composed of logical containers called DBSpaces. A DBSpace is composed of multiple DBFiles which map to individual files or raw storage space. Within the main store, a pre-defined DBSpace, called IQ_SYSTEM_MAIN, holds system structures. IQ_SYSTEM_MAIN is not designed to contain user data. Without the VLDB option, a user may define a single user data DBSpace. With the VLDB option, the user may define an unlimited number of DBSpaces for user data:



A user DBSpace can be *read-only* or *read-write*. A DBFile can also be marked as read-only or read-write. With this separation, the database administrator's job is more efficient. He can:

- Backup read-only data just once
- Backup/restore a read-only DBSpace independently of read-write DBSpaces
- Restore of read-only data can be performed while the system is operational
- Perform data validation on just the read-write portions of the database

A DBSpace can also be *online* or *offline*. The DBA can mark non-openable and non-usable DBSpaces as offline, and IQ will ignore them until they are brought online again.

Along with flexible management of storage, IQ also offers flexible placement of data within that storage. That feature will be discussed in the next section.

PARTITIONING AND PLACEMENT

In order to perform ILM, you need the ability to *partition* data into meaningful sets, and place those partitions into particular areas of storage. Partitioning divides data into non-overlapping subsets across a dimension, such as "date". For example, you may partition customer order data by the date when the order was fulfilled. Once you have partitioned data, you will want to map partitions to a particular area of storage. For example, the partition "June 2010 Customer Orders" resides in the file "/opt/data/orders/june2010.dat".

The Sybase IQ VLDB option provides partitioning and placement features. A table can be partitioned by specifying a column partition key, and specifying ranges of values of that key. Data within each range of values is relegated to a partition. At this time, range partitioning is the only partitioning scheme that is supported. A data partition includes the base FP (Fast Projection index in IQ) data only, and not the other indexes that exist along with that data.

Partitioning allows data that should be grouped together to be grouped together: e.g., data that shares a common attribute. Partitioning also separates data that should be separated, such as historical data from current data. A partitioned table can be altered to:

- Add/drop partitions
- Split a partition
- Merge adjacent partitions
- Un-partition the table

After partitioning, an administrator can place a database object in a particular DBSpace, or move a database object into a different DBSpace. Database objects are tables, table partitions, columns and indexes. With flexibility of data positioning, frequently accessed data can be assigned to faster storage, and less frequently accessed data can be segregated to cheaper, slower storage. This strategy controls storage costs, while still delivering performance. Here is an example of Sybase IQ SQL commands that partition tables, place the partitions into DBSpaces, and migrate those partitions to other DBSpaces over time:

1) Partition table Orders

```
CREATE TABLE Orders (  
    OrderID INT,  
    OrderDate DATE,  
    Description CHAR(10),  
    PARTITION BY RANGE (OrderDate  
(p2010 VALUES <= '2010-12-31' IN FIBER,  
p2009 VALUES <= '2009-12-31' IN FIBER,  
p2008 VALUES <= '2008-12-31' IN SCSI));
```

Over time, as data is being loaded, start migrating older data to slower, cheaper storage

2) Move p2008 to SATA storage

```
ALTER TABLE Orders MOVE PARTITION p2008 to SATA;
```

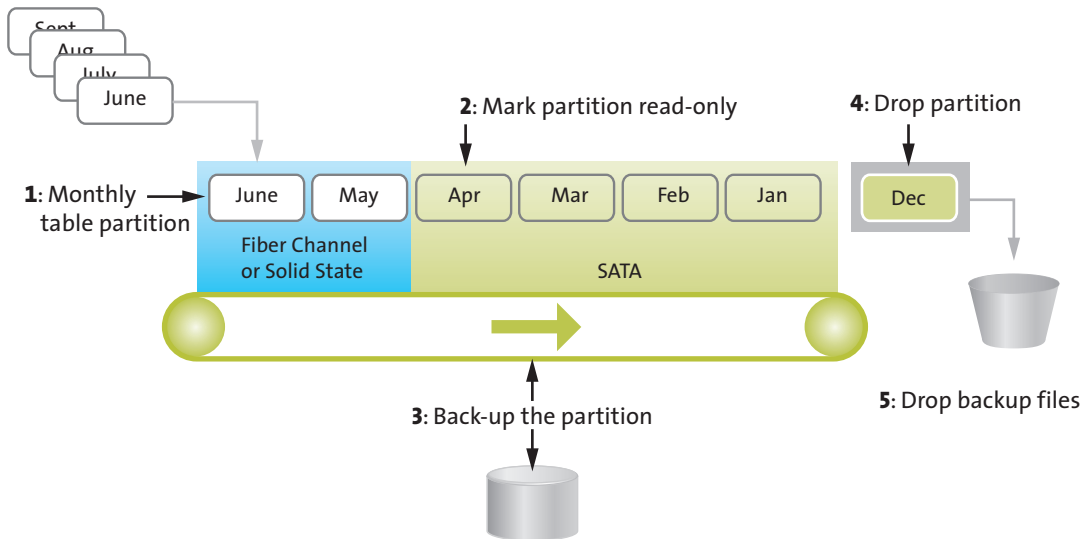
3) Later, drop very old partitions

```
ALTER TABLE Orders DROP PARTITION p2008;
```

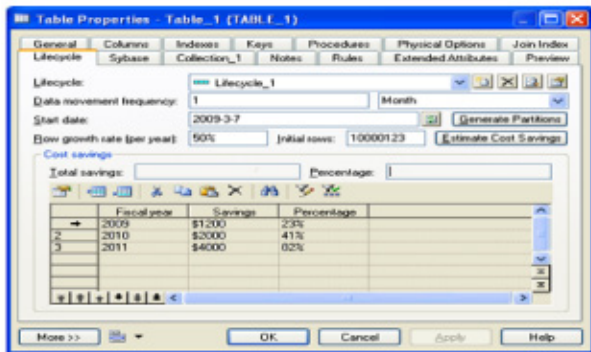
In order to create these kinds of data partitioning and placement commands more easily, the Sybase PowerDesigner modeling tool can automate some of this work. PowerDesigner is a separate product from Sybase IQ, and is not included with the VLDB option, but it has some desirable features that you may want to consider incorporating into an ILM solution. Those features are discussed in the next section.

POWERDESIGNER ILM FEATURES

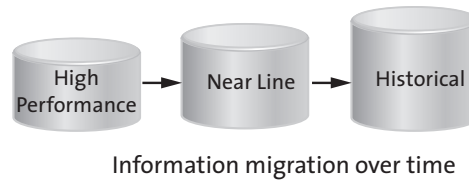
Version 15.1 of PowerDesigner introduces Information Lifecycle Management features in its data model. In PowerDesigner, you can design the metadata for a “roll-on/roll-off” storage management process, where data is cycled through various phases as the data becomes less current:



After you create a data model, PowerDesigner will generate the data partitioning, placement and movement scripts for Sybase IQ 15. PowerDesigner can also generate a cost savings report that delineates your savings over time as you move data to lower cost storage.




PowerDesigner data lifecycle cost simulation



SUMMARY

As databases grow to large sizes, organizations must design and implement ILM processes to manage their data effectively. Sybase IQ supports this requirement with flexible storage management, and data partitioning and placement capabilities. Sybase IQ is a platform that can grow with your business — helping you control storage costs, and continue to maintain performance for your users.



For information on our comprehensive
Consulting and Education Services to
support your Sybase technology initiatives,
visit us at www.sybase.com/consulting.

SYBASE, INC.
WORLDWIDE HEADQUARTERS
ONE SYBASE DRIVE
DUBLIN, CA 94568-7902
U.S.A.
1 800 8 SYBASE

www.sybase.com

Copyright © 2011 Sybase, Inc. All rights reserved. Unpublished rights reserved under U.S. copyright laws. Sybase, the Sybase logo and PowerDesigner are trademarks of Sybase, Inc. or its subsidiaries. ® indicates registration in the United States of America. SAP and the SAP logo are the trademarks or registered trademarks of SAP AG in Germany and in several other countries. All other trademarks are the property of their respective owners. 06/11

SYBASE®
An **SAP** Company