

BEST PRACTICE

A structured approach to storage optimization and data analysis

Platon Infrastructure white paper



Copyright © 2009 Platon ApS. All rights reserved

Created: October 22, 2008 | Author: Jens Peder Kristoffersen, Senior Principal

Updated: February 4th, 2009

TABLE OF CONTENTS

1	INTRODUCTION	3
2	A STRUCTURED APPROACH TO STORAGE OPTIMIZATION AND DATA ANALYSIS	3
2.1	ESTABLISHING A FACTUAL BASIS FOR DECISION-MAKING	4
2.2	SELECTING REAL OPTIMIZATION OPPORTUNITIES	5
2.3	REALIZING BUSINESS VALUE FROM THE OPTIMIZATION OPPORTUNITIES	6
2.4	MONITORING DEVELOPMENTS AND PROTECTING INVESTMENTS	6
3	CONCLUSION	7

1 Introduction

Since 1999 Platon has been involved in Business Intelligence (BI) advisory, development and implementation of BI-solutions in more than 300 projects world-wide.

Historically, the use of BI has been dominant in the finance function, but over the last several years the demand for factual and valid data has been spreading to all other parts of the modern company. In the companies IT organizations there is a need for factual information on the data volumes and data usage characteristics in order to plan and optimize the storage environment.

Given this need for BI in the IT organization itself, Platon has, based on BI-experience and storage domain expertise, developed a structured model and platform for analysis of structured and unstructured data.

In this white paper, we are describing Platons model for optimizing storage based on factual analysis of the data volumes. A model, which is called Best Practice by Platon, our partners and most importantly our clients.

2 A structured approach to storage optimization and data analysis

Initiating a data analysis project with the objective to optimize the storage environment and operation should contain the following four phases:

1. Establishing a factual basis for decision-making
2. Selecting real optimization opportunities
3. Realizing business value from the optimization opportunities
4. Monitoring developments and protecting investments

This stepwise process ensures that the initial data analysis that identifies the optimization opportunities is followed through – tactically and strategically - to realization of business value, hence a complete optimization process.

The following pages contains a short description of the content of the optimization phases, including a number of examples of how the analyzed storage data can be made available for management and employees in the IT department.

2.1 Establishing a factual basis for decision-making

The first phase in the analysis process is to establish a complete overview with the objective to identify what areas have a high probability for either economic, security or operational optimization potential, which would justify a more detailed data analysis.

Get an overview

Getting and maintaining a factual overview of the infrastructure elements and data relevant to storage professionals is probably the most important activity when operating large amounts of data.

Below is stated the infrastructure, disk capacity and file data detected during the latest Platon ILMAS[®] scanning. Furthermore a set of questions that often are asked by storage professionals as part of their daily work are specified below on this page. I.e. questions when planning capacity requirements or how data needs to be operated now and in the future.

Infrastructure	Disk capacity	Data development
<p>Topic Summary</p> <ul style="list-style-type: none"> Infrastructure summary Infrastructure inventory status <p>Subject Status</p> <ul style="list-style-type: none"> How many domains are monitored? Network domain status How many disk drives do we have? Disk drive status How many servers do we have? Status on number of detected servers 	<p>Topic Summary</p> <ul style="list-style-type: none"> Disk capacity summary Disk capacity inventory status <p>Subject Status</p> <ul style="list-style-type: none"> How much capacity will we need? Forecast on needed disk capacity How is our capacity utilized? Capacity utilization by platform How much capacity do we have? Allocated capacity by platform 	<p>Topic Summary</p> <ul style="list-style-type: none"> Data development summary Development in the data amount <p>Subject Status</p> <ul style="list-style-type: none"> What data is deleted? Status on deleted data How much data will we have in the future? Forecast on the data amount How does our data growth? Historical data growth
Data lifecycle	Data creators	Master data
<p>Topic Summary</p> <ul style="list-style-type: none"> Data lifecycle summary Data lifecycle status <p>Subject Status</p> <ul style="list-style-type: none"> How do we modify our file data? File data by last modification year How do we access our file data? File data by last modification year When did we create our data? File data by creation year 	<p>Topic Summary</p> <ul style="list-style-type: none"> Data creators summary Geography, company and organization <p>Subject Status</p> <ul style="list-style-type: none"> Who creates our data? Data creation by organization Where is our dat created? Data creation by geography What company created the data? Data creation by companies 	<p>Topic Summary</p> <ul style="list-style-type: none"> Master data summary Status on imported master data <p>Subject Status</p> <ul style="list-style-type: none"> What it services is used? Status on used it services What organization data is used? Status on organization master data What geography data is used? Status on used geography master data

To make a first assessment of the optimization opportunities there is no need for scanning individual files. The focus is collection of data on the infrastructure, i.e. factual information on server names, server drives, drives utilization, data growth etc.

By focusing on the basis infrastructure data at first, the cost and implementation time are minimized. At the same time, data actuality and accuracy are ensured through higher scan frequency.

Infra How do the locations utilize capacity?

This report shows the capacity utilization for the different geographical locations. Capacity utilization is shown as a % used capacity out of the allocated capacity.

With this report the storage professional can get a quick overview of where capacity is over- or underallocated, when potentially needing more disk capacity or investigating capacity problems.

Utilization by continent & region	
Europe	54.57%
Eastern Europe	29.85%
Northern Europe	56.40%
Southern Europe	35.17%
Western Europe	41.76%
Americas	40.90%
Caribbean	61.32%
Northern America	41.37%
South America	27.87%
Africa	37.22%
Eastern Africa	37.22%
Asia	35.53%
Eastern Asia	37.98%
South-Eastern Asia	29.15%
Southern Asia	45.50%
Western Asia	9.37%
Oceania	23.66%
Australia and New Zealand	23.66%

Allocated capacity by continent

In order to achieve maximum value from the collected data, it is important to include basic classifications such as server location, application types and data ownership.

The basic classification allows for more detailed and relevant analysis of the collected data, i.e. calculation of utilization rates by location or by IT-services rendered, or by geographical areas.

Often the needed information exists in the clients CMDB database, which of course makes the data collection process easier.

In order to have an optimal and complete basis for decision-making, the collection of metadata about files should be done, in order to analyze data ownership, data usage pattern and data types, which should be included in the evaluation of the optimization opportunities.

2.2 Selecting real optimization opportunities

The second phase of the storage optimization process should contain a factual analysis and evaluation of the potential optimization opportunities, in order not to waste resources in areas with limited economic or risk-lowering potential.

Find Improvements

To secure the right balance between storage cost and business requirements demand an ongoing search of improvements. Below is listed a set of questions that help to find areas that can be improved by technology investments or by operational house keeping.

Data classification	House keeping	Data archiving
<p>Topic Summary</p> <p>🔍 Data classification summary Completeness of basic data classification</p> <p>Subject Status</p> <ul style="list-style-type: none"> 🔍 Is all data classified by ownership? Overview of ownership classification 🔍 Is all data classified by service? Overview of IT service classification 🔍 Is all data classified by location? Overview of location classification 	<p>Topic Summary</p> <p>🔍 House keeping summary Potential in deleting data</p> <p>Subject Status</p> <ul style="list-style-type: none"> 🔍 How much orphaned data do we have? File data from former employees 🔍 How much temporary data do we have? Amount of temporary data on servers 🔍 How much backup data do we have? Amount of local backup data on servers 	<p>Topic Summary</p> <p>🔍 Data archiving summary Potential in archiving unused data</p> <p>Subject Status</p> <ul style="list-style-type: none"> 🔍 Can we archive NSF data? Amount of NSF file data 🔍 Can we archive PST files? Amount of PST data 🔍 Can we archive file data? File data by user activity
Suggested improvements by Platon		
<p>Capacity optimization</p> <p>Topic Summary</p> <p>🔍 Capacity optimization summary Evaluate storage technologies</p> <p>Subject Status</p> <ul style="list-style-type: none"> 🔍 Can we optimize by compressing data? Data reduction by compression 🔍 Can we optimize with de-duplication? Amount of duplicated files 🔍 Can we optimize with thin provisioning? Estimation of stranded capacity 	<p>2008-10-15 01:18:38</p> <ul style="list-style-type: none"> 🔍 Reduce data from PST files by archiving Suggestion regarding capacity decreasing 	

As a starting point, the storage optimization opportunities can be divided into four main groups; data classification, housekeeping, data archiving and capacity optimization.

Optimizing the basic data classification is the basis for a real evaluation of all of the mentioned groups of storage optimization opportunities.

Due to above, this phase should be initiated with a factual evaluation of the existing classification in order to ensure high data quality, reader-friendliness and minimized errors in analysis and reporting.

Based on the analyzed data each of the identified optimization opportunities must be analyzed in order to establish the basis for selection. In addition, the selected optimization opportunities should be compared to benchmarks to evaluate how the company performs compared to other companies, possibly in order to set targets for the optimization (i.e. capacity utilization percent).

As a part of this selection, the corresponding organizational stakeholder(s) should be identified for each optimization opportunity. This is to ensure that the business gets involved in the right way and with the right effort, in order to take business needs and wishes into account.

When the factual basis for decision-making is made according to above, a structured analysis of the storage infrastructure and the file data will accelerate the selection- and decision process for the optimization opportunities, and serve as valid and factual input to i.e. business cases for projects.

2.3 Realizing business value from the optimization opportunities

The third phase of the optimization process should focus on realizing the economic and business values identified for the selected storage optimization opportunities.

Realize benefits

Below is listed all running, planned and closed optimization initiatives that has been approved by the storage operations steering committee.

Running projects	Planned projects	Suggested projects
Project summary ? Running projects summary Active storage projects Project status ? Archive pst files Archiving pst files to new archive solution	Project summary ? Planned projects summary Summary of all planned improvements Project status ? Invest in new storage Sizing of new storage solution	Project summary ? Suggested projects summary Summary of all suggested improvements Project status ? Invest in new storage Sizing of new storage solution

Project portfolio status

Indicator	Status
Invest in new storage	●
Archive pst files	▲

Project status reports

- 2008-10-21 12:02:11**
 ? Archive pst files No. 1
 Status on project progress
- 2008-10-21 11:24:21**
 ? Invest in new storage No. 1
 Status on project progress
- 2008-10-16 23:08:54**
 ? Archiving pst files No. 3
 Status on project progress

When a number of optimization opportunities have been selected for implementation, they should be organized into a portfolio of projects or even an optimization program containing a set of projects.

The project portfolio or program management should ensure access to funding and needed competencies to implement the individual projects.

In addition, the above means that project owners and targets must be established, as well as a way to track project progress over time, ensuring focus and early-warning of project implementation problems.

Archive pst files to new archive solution

The objective of the "Archive pst files" project is to remove all pst files from the companies files servers to a central e-mail archive solution.

Newest project documents

- ? **Project initiative document**
 Archiving pst files
 Project initiative document
- ? **Project status report**
 Archive pst files No. 1
 Status on project progress

Top five servers

Server name	Data in GB	Files
DKSRV001	430	23,102
DKSRV032	20	1,232
DKSRV043	19	1,948
FISRV012	231	3,420
GBSRV032	323	1,300

Top five users

User name	Data in GB	Files
JDA	21	32
ODA	16	22
NIK	13	13
HKJ	2	7
HKD	1	3

Progress in archiving pst files

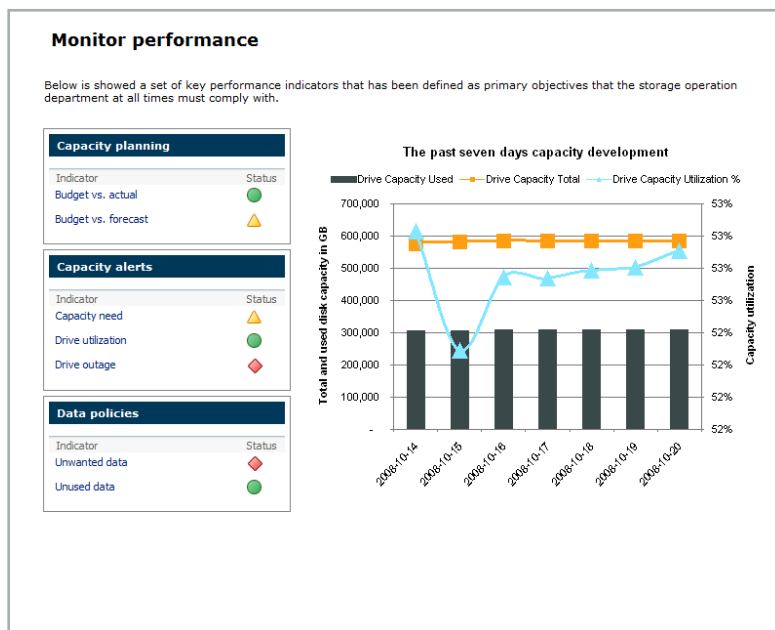
Month	Data in GB (Bar)	Number of files (Line)	Expon. (Number of files) (Line)
August 2008	~2,100	~50,000	~50,000
September 2008	~1,800	~40,000	~40,000
October 2008	~1,500	~30,000	~30,000
November 2008	~1,200	~25,000	~25,000
December 2008	~1,000	~20,000	~20,000
January 2009	~800	~15,000	~15,000
February 2009	~700	~12,000	~12,000
March 2009	~600	~10,000	~10,000

In order to ensure an effective realization of the individual optimization projects, the planning phase should contain an evaluation of what information the project owner and project participants needs on an ongoing basis.

Providing this data will ensure that the project owner and participants get a factual overview throughout the project implementation period, and in effect get a tool to measure the progress of the optimization project and focus on what still needs to be completed in the project.

2.4 Monitoring developments and protecting investments

The fourth phase should contain an ongoing and fact based monitoring and follow-up on the implemented optimization projects, in order to lock-in (and protect) the business value and benefits over time.



The most effective way to ensure ongoing and easy-to-use follow-up is by implementing key performance indicators (KPIs), where the most important figures are measured over time.

Implementing a KPIs concept should include the definition of ownership and responsibilities so that each KPI is linked to a person or a role, ensuring higher probability of a satisfactory result of each of the optimization projects.

An ongoing monitoring of the infrastructure for capacity planning purposes (I.e. using alerts), combined with defined data policies will ensure the IT department that the systems are not “polluted” again, and hence avoid devaluing the realized business benefits of the storage optimization projects.

3 Conclusion

Having completed development and implementation of more than 300 BI-solutions around the world, as well as having completed a number of storage optimization projects, Platon concludes that the above use of factual data analysis and monitoring helps companies in creating insight to their data volumes, selecting optimization opportunities with real business value, and ensuring realization of the economic and business benefits, while protecting investments against devaluation.

For more information contact:

Platon Infrastructure ApS
 Dag Hammarskjolds Alle 13
 DK-2100 København Ø
 Tlf. +45 7028 3000

Author:

Consulting Manager
 Jens Peder Kristoffersen
jpk@platon.net
 +45 5139 3021