



Butterfly Methodology

An overview of the comprehensive architecture, principles and rules that combine to create the Butterfly effect



by

Ian Smith
Chief Technology Officer
Butterfly Software Limited

April 2011, © all rights reserved





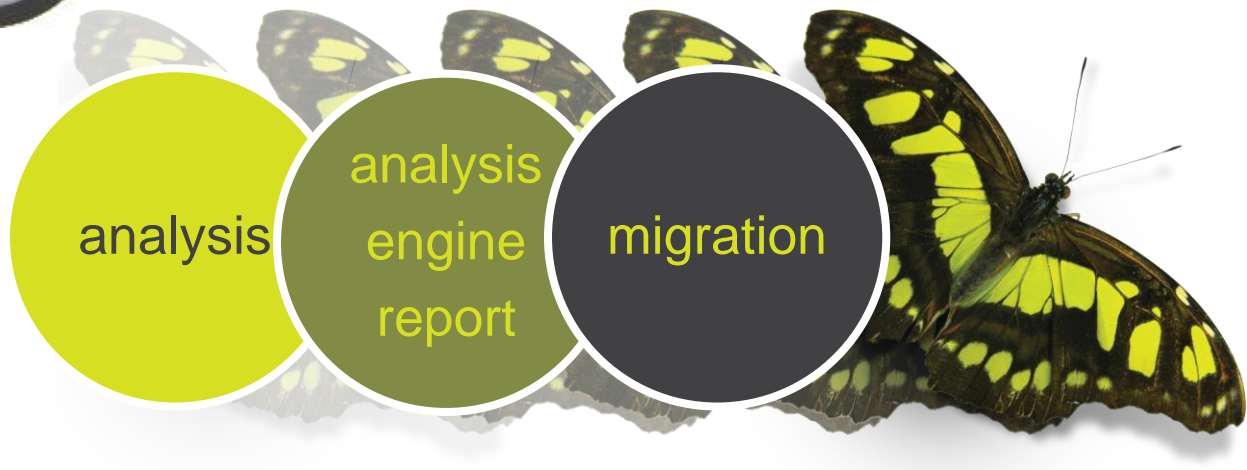
Butterfly Methodology

CONTENTS

BUTTERFLY INCEPTION	2
BUTTERFLY SOFTWARE	2
THE BUTTERFLY CONCEPT	3
BUTTERFLY SOFTWARE SOLUTIONS	3
ANALYSIS REQUIREMENTS	4
BUTTERFLY ANALYSIS ENGINE	5
ANALYSIS GOAL	5
THE UNIQUE BUTTERFLY APPROACH	6
PROPRIETARY COLLECTOR TECHNOLOGY	7
OPTIMISED COLLECTION	7
SECURITY CONTROL	7
THE BUTTERFLY SECRET	7
DEFINING TARGET ENVIRONMENTS	8
THE BUTTERFLY ANALYSIS ENGINE REPORT (AER)	9
THE AER SIMPLIFIES DECISIONS	10
A LEVEL OF ANALYSIS NEVER BEFORE ACHIEVED.....	11
DATA UTILISATION.....	12
MIGRATION PROCESS.....	12
ACTIVE BACKUP MIGRATION.....	14
BUTTERFLY ETHOS.....	14



Butterfly Methodology



BUTTERFLY | INCEPTION

Unlike the myths surrounding many IT entrepreneurial companies, the Butterfly concept was not dreamed up in a kitchen or a garage, or even a back bedroom. It was developed from involvement in corporate business; experience gained from direct exposure to the phenomenal technology advancements in storage capabilities; from the increased business requirements for accessing more information faster and in real-time; and from the legal pressures that required information retention for longer periods and with immediate access.

BUTTERFLY | SOFTWARE

Ian Smith, Chief Technology Officer at Butterfly, is the driving force behind the Butterfly concept. Ian had worked for over a decade with major technology vendors and systems integrators, delivering storage and backup solutions to major corporations. In a changing business world, the data lifecycle was growing beyond the technology lifecycle, meaning that over the lifetime of the data, multiple generations of technology would be in place. The traditional approach to IT was a technology-centric perspective, whereas changing requirements were driving a content-centric approach.

In a changing business world, the data lifecycle was growing beyond the technology lifecycle

Delivering solutions in the storage and backup environment revolved around technological advances, with no consideration for actual business requirements. This meant there was no automated, reliable or non-disruptive migration method for the content. These new technology generations and advances never fully realised the capabilities and didn't deliver the business benefits that were promised, because it was deployed with an off-the-shelf approach, without a detailed understanding of the existing environment.

The change in data management (security, retention, business recovery, protection, access) came as the market requirements, spearheaded by the financial community, highlighted the need for stricter controls. New delivery



Butterfly Methodology

channels, such as telephone banking and internet banking generated huge amounts of current and historical data, both structured and unstructured, to be available online. In addition, corporate requirements for data audit trails were further heightened in the early '90s, with events at corporations such as Enron increasing the legal compliance pressures on businesses to retain information for far longer periods.

These requirements increased corporate storage volumes and quickly highlighted a fundamental discrepancy between the lifecycle of the data, which had now exceeded the lifecycle of the technology – the three year cycle. In the early 00's corporate business, instead of reducing storage, power consumption and data centre space, was increasing these elements with multiple, three-year storage silos running side by side, driven by the technology replacement cycle. Each silo, when reaching end of life, was mothballed and new technology configurations were built and used in parallel. These sat side by side, managing data for their own respective active period with their own software solutions for retrieval and access and very little integration or consolidation. Instead of migrating information to the new solution, it became clear that the storage and backup environment as a whole, was out of control and all information was just becoming current information on a range of silos.

Additional backup systems were simply bolted on with a mix of vendor software and hardware backup and retrieval solutions. The spotlight fell on consolidation and migration to reduce online and nearline storage requirements to manageable, cost effective infrastructures. The utopian ideal was to develop an automated, risk free methodology of indexing content. A way to understand easily what needed to move and what could be archived. Instead of simply adding additional capacity for more data, only required information would be migrated to the latest, greatest technology to capitalise on reduced cost, increased performance and improved availability, allowing only one active silo within the customer environment.

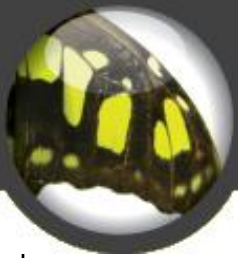
These factors became the driving force behind the Butterfly concept.

THE BUTTERFLY | CONCEPT

To develop a culture where ongoing analysis and migration of backup and storage can be achieved, independent of vendor hardware and software solutions. Allowing the delivery of high availability and performance, while reducing costs and environmental impact across the whole data environment. In addition, analysis and migration needs to be achieved quickly with minimal business and resource impact.

BUTTERFLY SOFTWARE | PRODUCT SOLUTIONS

Butterfly Software provides solutions in the two key backup and storage areas:



Butterfly Methodology

- ANALYSIS
- MIGRATION

Both are inextricably linked to achieve the same goals, but both need to be able to be managed separately to focus on the different requirements in each area.

The key to success is the transparent ability to analyse and understand the source environment, index the stored information, and migrate to a desired key target environment, unencumbered by data type or vendor association.

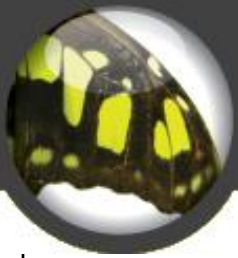
It was migration that became the driving force. From a vendor perspective, the inability to easily understand existing environments and migrate data prevented sales of new technology. Conversely corporations wanted to take advantage of new technology but couldn't achieve or justify the move because of the inability or time associated with understanding and segmenting legacy data.

However, the focus came down to analysing the legacy environment as the key factor. Only when the working corporate environment was understood would migration and the enablement of new technology be achievable.

ANALYSIS | REQUIREMENTS

Traditionally the major vendors in the storage and backup space provided analysis tools. Their perception was that in order to provide 'total analysis' they had to review and analyse the complete system configuration. But there were many drawbacks to their approach in analysing operational environments:

- Commonly they provided analysis only for a single vendor
- Reports were geared towards replacing existing technology with latest technology from a single vendor, rather than a solution that met business objectives
- The breadth and depth of reporting was enormous, running to thousands of different report types which:
 - Often confused business management with too many options and
 - Ignored key business requirements while focusing on technology delivery
- Analysis solutions revolved around introducing software agents placed on all devices across the complete configuration
 - The agents themselves introduced a processing and network overhead
 - With the agents in place, vendors saw the opportunity of gathering additional information and increased the functionality, producing 'bloated' software agents that became bigger overheads
 - Their analysis was, in reality, purely an observation, as opposed to a detailed understanding of the existing environment and recommendations for a new solution.



Butterfly Methodology

In short, the analysis that businesses required became lost, with the focus purely on delivering the latest technology. The analysis solutions themselves took too long to gather information, became an operational overhead and often caused operating problems and skewed results in the very environment they were analysing.

The analysis became a daunting process involving hardware and software overheads and operational staff. Managing the plethora of software agents distributed across an organisation's network became like the circus trick of managing a whole host of spinning plates on poles – except there became too many plates to manage. Not all agents worked seamlessly. Not all agents could run at the same time. Other network devices and software agents – especially security based – stopped agents running or stopped access to information. With the focus and energy on running '000s of agents, there were always some that were falling 'off the poles' while others were only just starting to spin – they didn't all manage to work at the same time to provide the analysis information required.

This approach was the same across the industry and analysis became a pariah. An unwanted, un-required overhead that ran for weeks and produced inaccurate results that were subsequently challenged and often ignored. When business needs dictate an increase in storage or backup, it generally needs to move quickly, with decisions based on factual current information. This represents a reactive situation, as a symptom of not understanding the current environment and future requirements.

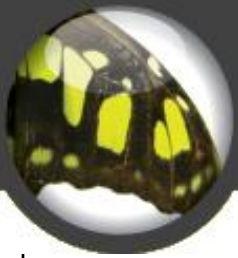
ANALYSIS | GOAL

Butterfly's analysis goal was to develop a methodology and implementation that would minimise operational overheads, negate the requirement for 'human' knowledge and intervention, provide zero disruption on business and deliver the ability to snapshot a real-time data environment in the quickest and lightest touch way.

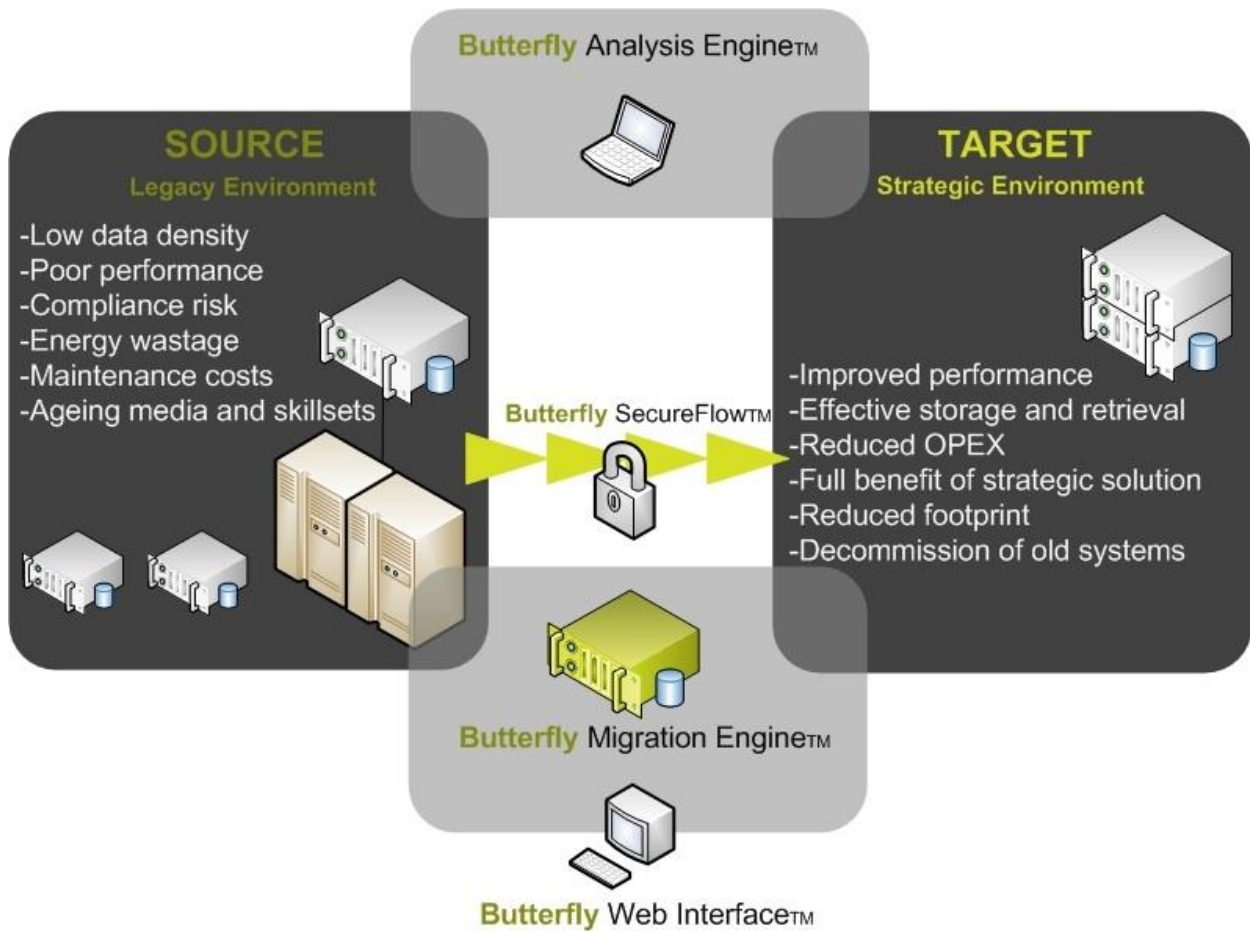
BUTTERFLY ANALYSIS | ENGINE

The Butterfly approach differs significantly from traditional vendor and consultant approaches to analysis:

- Traditionally analysis required information gained from technology and technical teams – people that had personal views based on their experiences and their conclusions.
- Analysis programs ran across weeks and months – while requirements and usage were changing, negating the accuracy of the data.
- Reports ran to hundreds of pages, providing multiple options and upgrade paths.
- Business cases for upgrading revolved around technology and not business objectives and requirements.



Butterfly Methodology



THE UNIQUE | BUTTERFLY APPROACH

- The process turns data collection on its head. Instead of monitoring every device in an operational environment, information is gathered centrally at the storage and backup solutions.
- Information is gathered quickly over a period of hours rather than weeks
- No agents are introduced into the operational environment so there is no processing overhead and zero interruption to business operation
- Database collection resides 'in-the-cloud,' requiring no on-site processing or storage
- All data collected is empirical, without requiring user intervention or drawing on user experience, which ensures pure, factual information for exact analysis
- The agent-free environment enables Butterfly to interrogate all data arrays and storage devices present on a network – no sub-selection of 'mission critical only' devices/data is carried out – all data is included

Analysis is not limited to backup storage and can include primary storage.



Butterfly Methodology

PROPRIETARY COLLECTOR | TECHNOLOGY

Information is gathered utilising a focused 'collector'. This is another unique approach, designed and implemented by Butterfly. Instead of gathering all available information, the collector was designed from the database outwards to fetch the information that the analysis requires. With a full understanding of the environment, Butterfly are able to create a detailed and accurate analysis. This minimises not only data collection but also analysis and sorting time, as all information gathered is utilised for analysis. The overhead on backup and storage data arrays is minimal as information extracted is specific to the analysis and from indexed tables, where possible.

OPTIMISED | COLLECTION

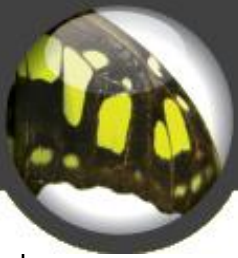
The collector and its operations are optimised to utilise high performance elements of the source database, and use indexed tables for fast retrieval and comparison, enabling Butterfly to gain results and determine the subsequent requirements via the fastest possible route. In addition no technical or management team time is required to supply operational information.

"WITH THE EXACT ANALYSIS INFORMATION GATHERED, BUTTERFLY IS ABLE TO PROVIDE AN ANALYSIS REPORT WITHIN 5-10 DAYS."

SECURITY | CONTROL

Security is a key concern for all businesses and again Butterfly is unique, in that no client data is accessed or collected, only metadata specific to operational devices and software. In addition all metadata is encrypted before transmitting from the collector to the Butterfly secure portal for database import.

THE BUTTERFLY | SECRET



Butterfly Methodology

The key to the analysis is not collecting or storing the data, but in the analysis and comparison of information collected. It is here that Butterfly's proprietary algorithms and approach create a simplified target environment out of a complex multi-vendor, multi-data type, multi-device legacy situation.

All the information gathered is normalised to a central common terminology. Vendor specific terminology is lost; device features and dependencies are converted to a centralised standard, and capacities, availability and performance are correlated to a single standard. Butterfly removes the dependency on vendor specific specialists or experts. No advantages of a specific approach are lost. No elements of a particular technology are removed. No details of unique scripts or applications are removed. All details are extracted, notated and normalised into a central Butterfly 'language'.

This normalisation is the key strength of the Butterfly analysis approach. Essentially all backup and storage environments become equal in terms of assessing their accountability, performance and operation. Analysis becomes a straightforward process of comparing and evaluating information utilising standard metrics. No time is wasted in attempting to compare different methodologies and technical specifications – in gathering and normalising the total information, the comparative standard has been defined and set for a unique analysis experience.

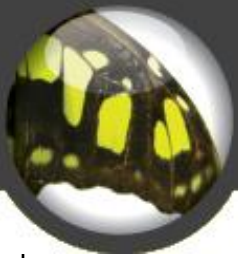
DEFINING TARGET | ENVIRONMENTS

Defining the optimum target environment for a business revolves around a complex reporting analysis process which is simplified by Butterfly to produce an easy to read overview, which highlights all the key deliverables and business advantages.

This reporting can easily be carried out by taking the normalised analysis data and feeding this into Butterfly's proprietary target modules. The target modules are created based on specific vendor technical specifications. Each module represents a vendor's product range with details of capacities, performance and requirements for optimum operation.

The target modules take normalised output from the Butterfly Analysis Engine and translate the information to match the businesses target requirements – this may be a specific product technology or a business objective associated with performance and environmental and operational goals.

Target modules understand the normalised data and translate the input to match the target configuration. This might be specific to a single vendor or may be vendor agnostic and produce a recommendation that incorporates multiple vendor technologies to achieve the correct business goal.



Butterfly Methodology

This process is completed by making logical connections between the source technology and how it currently works and overlaying target technologies over the normalised metrics, to achieve a resolved solution. This solution can also take into account duplicated data and operations to create a consolidated data migration and operational process to a new configuration. Analysis is not limited and encompasses multiple source environments and multiple target environments for a truly flexible approach.

The architecture of the Butterfly software allows the rapid development of source and target modules. These resultant modules can then be added to the Butterfly analysis and migration interoperability matrix to support alternative vendor software and hardware solutions. This rapid development approach ultimately enhances and expands the options available to businesses to fully realise their corporate migration goals.

This is a revolution in migration methods for backup data. Historically, storage was considered to have a manageable route for migration, albeit manual, timely and costly, it could be achieved. Now with the agent-less, light touch approach, primary storage can also be included in the analysis process. Backup consolidation and migration has always been seen to be unattainable because of the plethora of storage media, software solutions, devices required, and duplication of information – in essence too many factors to develop a manageable and timely solution.

Butterfly analysis and migration allows a content-centric approach to data management and migration. This removes any vendor lock-in and gives users freedom of choice over the technology they use to store and backup their data content. This allows a unified approach for both backup and storage.

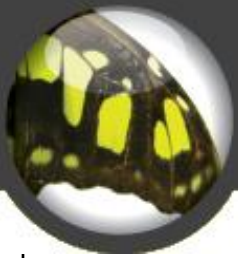
THE BUTTERFLY ANALYSIS ENGINE REPORT | (AER)

When a business looks to analyse its backup and storage data usage it is reviewing a moving target. The very nature of information creation, movement, deletion and updating dictates that business storage requirements change continually. Any analysis of infrastructure utilisation therefore, must be fast, accurate and provide reports that are quick to review and that clearly identify the business benefits of migration.

In addressing these requirements Butterfly turned traditional reporting methodology on its head, by creating a reporting process that provides a business executive report within a single graphical view. This is created from the empirical data, extracted in a matter of hours during the collection. It provides detailed analysis of the current situation, a target solution based on current technology and associated infrastructure business case. In addition, existing operational issues and a high level migration plan are displayed.

THIS IS THE BUTTERFLY ANALYSIS ENGINE REPORT (AER)

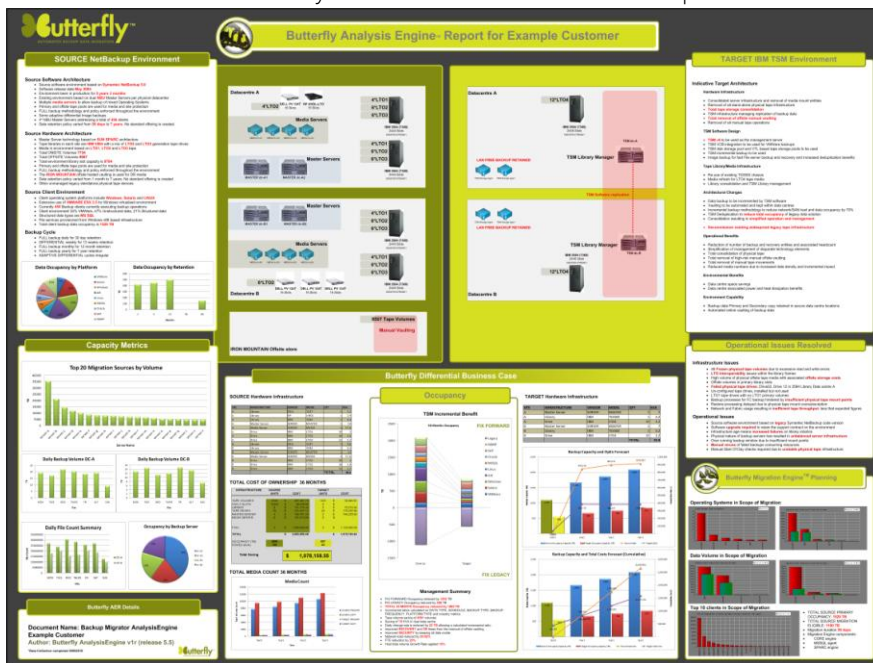
This software provides a completely new way of presenting results with a report based on factual information – no manual intervention is required and no operational overhead is created to produce what is arguably the



Butterfly Methodology

most comprehensive and detailed analysis available. With the target configuration specified, the AER, in essence, provides a transformation plan based on a detailed, automated migration strategy. This is a single, ratified plan and recommendation based on business requirements and not technology. This delivers immediate savings on cost, energy and physical space. In addition it shows how operational efficiencies and performance improvements can be made through the utilisation of contemporary technology solutions.

The AER is divided into key sections as shown on this example:



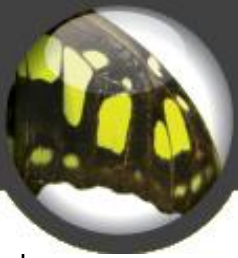
An interactive AER is available on the Butterfly website, which provides more detailed information.

THE AER | SIMPLIFIES DECISIONS

The AER enables a business to make a migration decision based on two routes:

- Target Technology Decision
- Conditional Technology Decision

Target Technology Decision: Enables an organisation to carry out complex 'what-if' analysis quickly - literally at a speed that was previously unimaginable. This flexibility provides a truly free choice of technology by projecting target configurations benefits and drawbacks based on the current installed base and technology options available. The AER compares this against a detailed analysis of the outcome if no changes were made. The occupancy and financial projections cover a three year period to visualise growth



Butterfly Methodology

and expansion effects on the business. Operational issues are examined and architectural considerations defined.

Conditional Technology Decision: In the conditional mode the Analysis Engine automatically defines the target architecture based on solutions to resolve problems that have been reported during the collection, such as hardware faults and failures, application interface errors, operational considerations. In essence it is extending the what-if analysis to evaluate known situations and project the target environment based on repeating historical situations.

In both instances the AER provides the associated business case in order to support and justify the chosen migration strategy. This ensures that a migration is not based solely on a technology development. Technology forms a central role in the target, but it is not the driving force. It now becomes one of the key factors to ensure operational benefits are achieved in relation to operational and data requirements.

A LEVEL OF ANALYSIS | NEVER BEFORE ACHIEVED

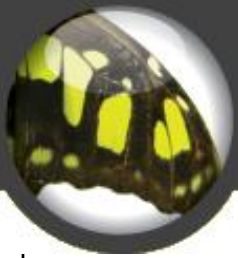
The AER holds a plethora of information that has not previously been accurately analysed and made available. An example of the key information that sets the Butterfly approach apart from alternative solutions is the analysis of Storage Wastage. With the AER, Butterfly produces an Extended Wastage Chart that shows the results of an analysis that delves deeper into storage usage than ever before. It shows the common Raw, Usable and Allocated storage:

- **RAW:** physical storage that is deployed – effectively the total of all available storage
- **USABLE:** available storage after RAID overhead and other array wastage
- **ALLOCATED:** the array reported areas apportioned to systems and applications
- **ALLOCATED VIRTUAL:** storage apportioned to a virtualisation layer, eg thin provisioning

Butterfly's extended wastage chart also includes:

- **DISCOVERED:** the allocated space that is currently discovered by host systems
- **FORMATTED:** discovered storage currently formatted by host systems
- **UTILISED:** the actual storage usage of formatted areas, providing a real formatted/usage ratio

It is the three additional areas of Discovered, Formatted and Utilised, that provide the real picture which can often show that storage is running at a significantly lower efficiency rate than projected. This analysis further defines the actual requirements removing guess work and illustrating exactly what is required in a target environment for continued operation and evolution.



Butterfly Methodology

Butterfly is the only vendor that can create this report, without the need to deploy client agents. The key metrics for these three additional areas can only be achieved by comparing the storage data and backup data, information gained from the two touch points. This comparison is made in the Analysis Engine database, a unique environment with the ability to provide a total data utilisation and wastage picture that covers both physical and virtual. Only with this level of detail, based on analysis that is achieved in hours, can the AER determine the exact level of consolidation and migration of data to a target environment, using storage virtualisation technologies.

DATA | UTILISATION

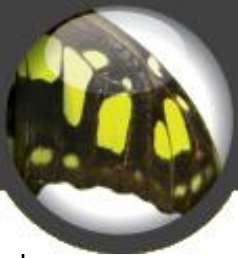
A key factor that is often overlooked when migration is considered is data utilisation. This tends to be ignored because understanding the information that is on storage and in use is complex, and using distributed agent technology it takes time and processing power to extract. With the Butterfly two-touch-point approach of backup and storage analysis, the reports can also consider the data in detail. The analysis provides the AER with details of data type and usage and the relationship between the applications, storage and backup. As the information is in a normalised form, it becomes straightforward for the AER to highlight exact storage requirements. Understanding the content is only achievable through this software.

MIGRATION | REQUIREMENTS

The process of analysing and developing a consolidation plan needs to have the same level of automation and immediacy associated with the selected migration. Utilising the same normalised database, Butterfly produce an automated implementation to migrate, consolidate and transform data to the target environment. An essential requirement of the migration is ensuring that the Migration Engine database is up to date so that when migration takes place, it is based on the current situational analysis. Locating the Migration Engine database in the cloud enables the Butterfly process to automatically schedule data collection at defined intervals. This ensures that all migration processes can be modified to account for any usage changes during migration.

MIGRATION | PROCESS

The analysis process, in normalising the data, also indexes all key details about the source environment, covering key information such as system locations, storage types, server types, applications, management software, operating systems, versions, backup servers and the various manufacturers and model types involved.



Butterfly Methodology

The Butterfly Migration Engine has the ability to migrate all legacy data and retain existing retention and configuration. Migration Engine also has the ability to restructure and reconfigure retention as required by the user. The legacy migration process also takes into account the retrieval requirements and can migrate to multiple devices, can schedule migration to account for the importance of information access and can organise the handling of specific business needs.

The migration process revolves around key parameters that take into account the indexed source and target details and using these parameters carries out three specific phases:

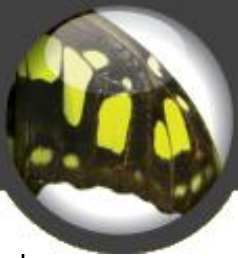
SCOPE – ORGANISE - OPTIMISE

Scoping: This involves looking at the complete source data environment and scoping what needs to be migrated. For example, data that is due to expire will be excluded from migration. Additionally, the process can scope out expired platforms or data types that are not required as part of the target solution. In this way the scoping process compiles a list of included and excluded data. All scoping is user-driven and applied automatically using the empirical data collected from the organisation's source environment.

Organisation: The migration engine automatically builds migration sets of data that needs to be migrated together for efficiency. These migration sets are then grouped together into manageable phases of migration. These phases can be scheduled and managed independently, allowing full control over an enterprise migration program. Phases can be built on a number of configuration points, including individual clients, groups of clients or original backup date. This can be done to match business operation and scheduled to provide minimal overhead and disruption. Phases are then organised to create the total migration plan in line with business requirements. As with scoping, organisation is completed automatically, based on agreed parameters.

Optimisation: During the optimisation phase, collected information is used to ensure the shortest possible migration duration for each phase. Key details and requirements including source technology, media location, infrastructure sharing and data classification are used. These factors enable the time taken to migrate the data, and impact on the source technology, eg tape media, to be significantly reduced. Migration optimisation is tuned for each individual migration phase.

Butterfly utilise a grid approach to migration allocating migration responsibilities across multiple migrators. Migrators are installed on front end client devices and each Migrator is allocated workload from the cloud based Migration Engine. There is no limit to the number of migrators that can be installed, which ensures that the Butterfly elements are not the bottleneck during the migration. The grid based approach ensures that Butterfly always out-performs the speed of the source environment, maintaining data flow.



Butterfly Methodology

ACTIVE BACKUP | MIGRATION

A migration project covers both the migration of legacy data and the active backup operations, ie which environment clients actually backup to. Butterfly Active Backup Migration (ABM) allows the automated reconfiguration of the client environment to address the target solution. This can be done between different vendor software products. ABM structures and schedules this migration in line with the migration of the legacy data. Where possible, ABM avoids the need for a new full backup on the target environment by the automated migration scheduling of legacy data.

BUTTERFLY | ETHOS

Throughout the complete analysis, reporting and migration process, Butterfly has focused on the data and the business needs. This allows the business to harness the benefits of the new technology platform quickly and efficiently. The data is the most important element to the business - customer records, operational log files, ATM machine logs, the list is endless and specific to each business, so the migration to new technology needs to take all these factors into account to maximise performance and capitalise on the technology investment and developments.

Butterfly has introduced a light-touch approach that enables business to take advantage of new technology while continuing operations unhindered. In essence, putting business back in control of how their data is managed for the business, by the business.

Butterfly has enabled businesses to rapidly assess the actual effect and benefit of optimal configuration of vendor technology offerings on their own specific environment. Once accepted, Butterfly offers the next step in realisation of these benefits with automated data migration into the future state architecture.



© 2011. BUTTERFLY SOFTWARE LIMITED. ALL RIGHTS RESERVED.

NO PART OF THIS DOCUMENT MAY BE REPRODUCED IN ANY FORM BY ANY MEANS, NOR MAY IT BE DISTRIBUTED WITHOUT THE PERMISSION OF BUTTERFLY SOFTWARE LIMITED, NOR MAY IT BE RESOLD OR DISTRIBUTED BY ANY ENTITY OTHER THAN BUTTERFLY SOFTWARE LIMITED, WITHOUT

THE PRIOR WRITTEN AUTHORIZATION OF BUTTERFLY SOFTWARE LIMITED.

THIS DOCUMENT IS PROVIDED "AS IS" WITHOUT WARRANTY OF ANY KIND. ALL EXPRESS OR IMPLIED REPRESENTATIONS, CONDITIONS AND WARRANTIES, INCLUDING ANY IMPLIED WARRANTY OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE, ARE DISCLAIMED.

ANY TRADEMARKS REFERENCED IN THIS DOCUMENT ARE RECOGNISED AND REMAIN THE PROPERTY OF THE REGISTERED OWNERS.