EnduraData Cross Platform File Replication and Content Distribution (November 2010)

A. A. El Haddi, *Member IEEE, Zack Baani, MSU University*

***Abstract*—In this document, we explain the various configurations commonly used by enterprise customers for remote backup, archive and content aggregation from many sites to a single central site and for data distribution from one site to many remote branch offices.**

***Index Terms*— file replication, data replication, cross platform file replication, content aggregation, data distribution, snapshots**

I. Introduction

In today's global economy, companies and government agencies need to protect and share data on a global scale. Enterprises gain competitive advantages by giving all stakeholder immediate access to data. Furthermore, certain laws and regulations and the threat of litigation make it imperative for corporations to protect their data. As more businesses become global, their exposition to risk increases as terrorist shift from hard to soft targets. Therefore, companies and agencies should take proactive measures to protect their data assets everywhere in the world.

Many companies have data protection strategies and data protection policies in place. However, these strategies are uncoupled, at times, from the global risk management of the entire company. Historical data shows that companies tend to focus on protecting data centers and not the edge of the data center. Mobile workers data and remote branch offices that remain unprotected are a source of data loss. Any data loss will impact the company's revenue and market share. The IT (Information Technology) staff tends to protect critical servers such as email, database and web servers. However, a large amount of critical unstructured data remains unprotected and spread throughout the enterprise.

In addition to improving data protection plans and execution, companies should consider productivity gains across the board, when data protection is combined with data sharing using a unified content data management approach.

To enable companies to protect, share, archive and distribute data, using unified data and storage management techniques, EnduraData released EDWADDS(EnduraData Wide Area Data Distribution Software). EDWADDS is a unified storage management software that combines data protection, data sharing, distribution and archive for the enterprise and for government agencies.

In this document, we present a brief overview of EDWADDS. In the sections that follow, we will summarize the advantages of EDWADDS, highlight a few features and configurations to give the reader some insights about how to combine the various configurations and topologies to protect and share data between all stake holders within an organization and with its authorized business partners.

II. ADVANTAGES OF ENDURADATA EDWADDS

Key features that differentiate EDWADDS from existing products in the market place include:

- Support of large scale data distribution
- Combination of remote disk backup with archive, data distribution and snapshots
- Support of cross platform replication (operating systems and architectures)
- Flexibility of the graphical configuration
- Unified management
- No need for extra file system drivers
- No need for compilers
- Support for embedding in appliances
- Elimination of data loss or data corruption due to faulty file system drivers

---

A. A. El Haddi is a senior software architect and the founder of Enduradata. El Haddi co-founded Constant Data, one of the first real time data replication companies. El Haddi designed data distribution software used by NOAA, NASA, the US Navy, T-Mobile and many large and small companies. Previously El Haddi designed and implemented the replication software now sold by Backbone/Quest. He also designed file replication for Quantum.

- Accelerated data movement and bandwidth control
- Dynamic runtime user interface and menu generation
- Accessibility of data on the remote storage without the use of EnduraData's software agents
- Reduced Total Cost of Ownership(TCO).

III. CROSS PLATFORM ARCHITECTURE

EDWADDS is a cross platform file replication and data distribution software. It allows you to replicate data from one operating system to another operating system. For example you can replicate data from Mac, Windows, Linux or Open Solaris and vice versa. As of this writing EDWADDS supports the following operating systems are supported:

- Linux
- Mac
- Windows XP
- Windows 7
- Windows 2003
- Solaris

Other platforms are supported on demand.

IV. FILE AND DATA REPLICATION TOPOLOGIES

EDWADDS file replication allows system administrators to replicate data using the following topologies:

- Standalone host or server
- One-to-one data replication on a LAN or a WAN
- Many -to-one data replication on a LAN or a WAN
- One-to-many data replication on a LAN or a WAN
- Mesh and grid replication in LANs and WANs.

Other configurations are derived from the previous topologies as needed. The configurations can be used for backup consolidation, content aggregation, archive or data distribution.

A. Backup Consolidation, Archive and Content Aggregation

IT staff can configure EDWADDS to replicate data from one host computer to itself (external drive for example) or to another storage server on the same LAN or to another remote location. Figure 1 depicts a configuration used to consolidate backups and aggregate content from many servers and workstations to a single central location. A content aggregation configuration allows users to send data from many branch offices, nomadic and traveling staff or from suppliers to a single point of collection for data sharing purposes, data protection purposes or both.

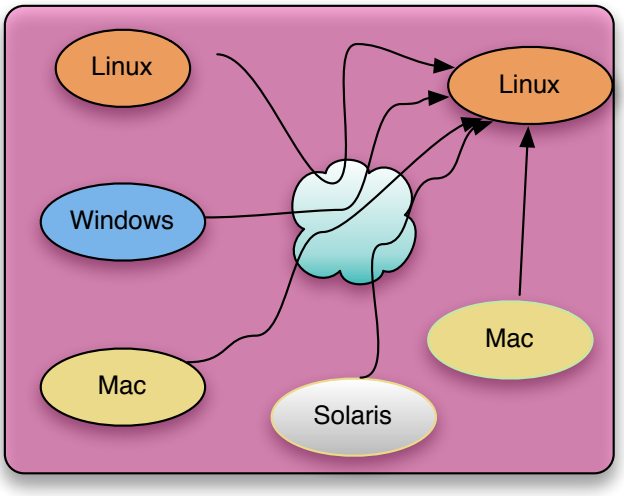

Figure 1: Content aggregation and backup consolidation from many sites to a central site.

EDWADDS has built-in features to enable easy configuration for aggregation and for backup consolidation. One of these features is aliasing for hosts that need to aggregate content in a central location. Aliasing is technique used to create equivalent servers or data sources. The availability of macros to represent various configuration parameters reduces the time required to create a central backup system. EDWADDS allows hardware vendors and value-added resellers to build backup appliances rapidly using their commercial off the shelf hardware and storage. Administrators can also turn an existing server into a backup appliance.

B. Data Distribution

Many companies need to distribute data between sites or within the same site. Data distribution may have several objectives: to make data readily accessible in remote locations for workflow purposes, to automate processes, to balance the load, to protect data, etc. Figure 2 shows a configuration where a Windows server distributes data and content to remote Linux, Mac, Windows and Solaris servers. Figures 3, 4 and 5 show additional configurations of data distribution networks. These figures show data distribution networks of one location to one location, one location to many locations, many locations to many locations and other variations.

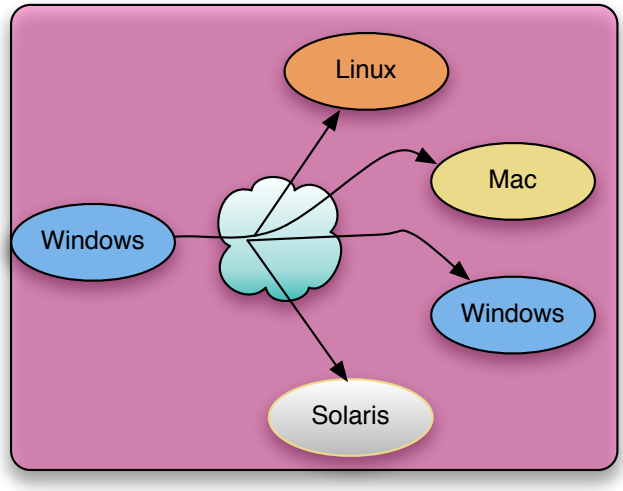

Figure 2: Data distribution from one site to many hosts in remote sites using the Internet.

V. UNIFIED MANAGEMENT

Figures 1 and 2 illustrate how EDWADDS supports heterogenous platforms and operating systems. This feature requires a unified management approach that results in significant savings of time and money for IT. EDWADDS allows the IT administrator to manage the entire data distribution and protection network from a single node. To provide this flexibility, EDWADDS architecture is designed to be modular, where each module is responsible for a limited number of tasks. EDWADDS' modules are coordinated by a middleware that ensures that all modules work together in a coordinated manner to manage the data distribution network efficiently. The solution supports many different hardware architectures, operating systems, topologies and storage platforms. EDWADDS uses a low memory footprint and eliminates bottlenecks. The administrator has at his disposition two methods for managing the entire content distribution network from a single node:

- An intuitive graphical interface (GUI)
- An array of command line utilities.

A. The Graphical User Interface

EDWADDS graphical interface allows the system administrator to create complex data distribution and protection networks. This unique feature allows the administrator to see what servers and assets are involved in content distribution. Several flexible controls allow the user to preview what data is sent, where it is sent, accepted for storage, what content is allowed to be sent and when is the content sent, etc. Examples of operations that can be performed using the GUI include:

- Add hosts
- Assign a sender or receiver role to a host
- Create a data link that associates a sender with one or more receivers and that designates the content to be sent and received
- Create a graphical data distribution network with multiple links
- Pause data distribution for a given link or receiver
- Resume data distribution for a given link or receiver
- Create a flexible schedule of a data distribution network
- Manage data distribution jobs: cancel them, set their priorities
- Monitor the progress, statistics, failures and success of the data distribution jobs.

Figures 3, 4 and 5 depict three distinct data distribution networks used for both data distribution and protection networks. Figure 3 shows a cascaded mesh using a combination of a LAN and a global worldwide network. In this diagram, data is distributed from a server called Targua to a website on the same LAN. The web server, in turn, distributes the data to Tokyo and to Madrid, Spain. The server in Spain distributes the data to Casablanca, Dubai and Nairobi. More complex data distribution networks are possible, as shown in figures 4 and 5.

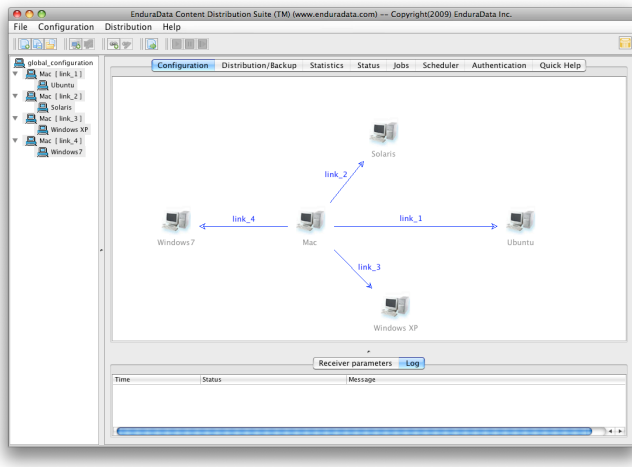

Figure 3: A simple graph of a data distribution network.

The graphical user interface creates simple or complex data distribution networks and monitors them. Once the distribution networks are configured, the user can close the GUI and reduce the resource usage. EDWADDS communication and middleware modules will continue to work in the background independently of the user interface.

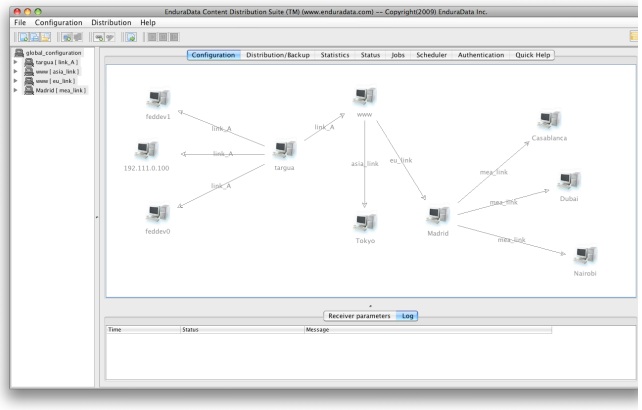

Figure 4: A graph of a cascaded data distribution network.

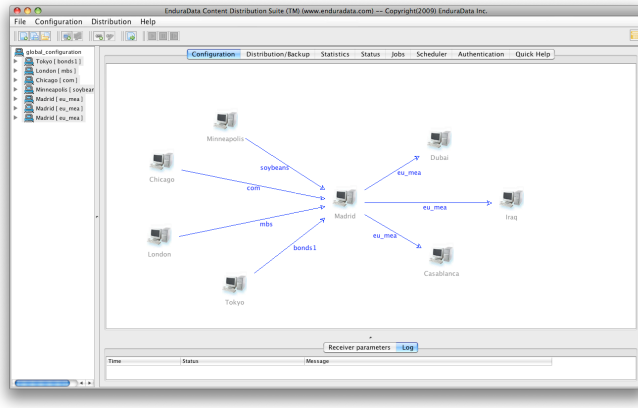

Figure 5: A graph of content aggregation and distribution network.

B. The Command Line Interface

EDWADDS provides the system administrator with a rich set of commands for managing data distribution networks. These include:

- *edq*: Queue content for distribution from one site to one or more sites
- *edjob*: Manage jobs (cancel, change priorities, history, etc)
- *edstat*: Monitor the progress of data distribution and the state of various modules
- *edpause*: Pause data distribution
- *edresume*: Resume data distribution.

The command line utilities allow IT administrators to integrate EDWADDS with other workflow applications and gives them full and flexible control over data distribution strategies.

VI. Authentication and Data Security

Several steps ensure the system's integrity and protect the security of the data and of the distribution networks. These include but are not limited to the following configurable items:

A. The Management Password File

Administrators use a password file to specify a list of hosts allowed to manage the network. Each host list has a password. The host list can be as simple as a host name or IP or a combination of hosts and IP addresses or address string patterns to represent tokens or expressions to be matched with the incoming connection IPs or host names.

B. The link identification

Each link has an identification. A sending link identifier must match the receiver link identifier before communication between the sending side and the receiving side can take place.

C. The Link Password

Each link has a password. The local link and remote link passwords must match before any communication can be established between the sender and the receiver.

D. The Sender Identification

Each receiver has a list of hosts allowed to send data to it. Unless the sender's address or host name is part of the authorized peer list, no communication will be allowed to take place between the two sides.

E. Other Security Precautions.

Other security precautions are built into the software to prevent hostile agents from injecting undesired data or content into the network.

VII. Accelerated Data Movement and Transfer

EDWADDS uses several technologies to ensure accelerated data transfer:

- Segmented non-redundant chunk discovery and de-duplicated transfer
- Data compression
- Multi-threaded and parallel processing
- Parallel input and output exploitation
- Locality of transfers
- Optimized hashing techniques.

VIII. DYNAMIC USER INTERFACE

EDWADDS GUI offers a dynamic menu system. You can tailor the editable menu configuration without the need to upgrade your software. You can add or remove parameters from the list of input items as needed.

IX. PRIVATE USER CONTROLLED CLOUD

With the popularity of cloud computing on the rise, many companies find themselves extremely leveraged to adopt the technology controlled by a few companies, such as Google, Microsoft and Amazon, or to simply ignore the necessity of using the cloud for protection. Several companies provide backup to the cloud. This affords companies the data protection they need unless they need to restore an entire server or even a directory with a few gigabytes of data. A solution for this dilemma is to use corporate existing investments to create private clouds under the full control of corporate IT. EDWADDS enables companies and government agencies to fully control costs while ensuring that the existing infrastructure is used to protect and share data among all stakeholders who need access to data any time of the day and without delay.