TANEJA GROUP
TECHNOLOGY ANALYSTS

P R O D U C T P R O F I L E

An Overview of HP RISS

October, 2005

THE NET NET

In the past 18 months, HP has built one of the most detailed and comprehensive ILM messages in the enterprise IT industry. In a forward looking move, HP positioned its Reference Information Storage System (RISS) as a core technology in their ILM strategy. RISS is a highly adaptable object model disk solution based on HP smart cell technology. As we researched the RISS offering, we came to the conclusion that HP is in the early stages of a far reaching ILM strategy that incorporates both conventional technologies and unique insight: It is possible to leverage an appropriately architected, highly scalable object model storage technology into an open platform for a variety of archival, migration, and backup solutions. With RISS, HP has given itself a very flexible, very scalable platform. We believe that RISS is a product worth understanding as much for what it means to HP ILM as for what it indicates for the future of archiving technologies.

RISS: The Core of HP ILM

Before a detailed discussion of HP's RISS (Reference Information Storage System) platform, it makes sense to frame this technology in its larger ILM context within HP.

While most IT professionals already know that HP possesses one of the largest and widest product families in enterprise IT, many probably do not yet appreciate that HP has worked very diligently over the past 18 months to develop new ILM approaches that federate these distinct technologies in new ways. As a result, HP's take on ILM is unique in the industry for several reasons, going far beyond a mere marketing claim of "ILM" uniqueness.

In the HP worldview, ILM is comprised of 4 distinct functions: Capture, Manage, Retain, and Deliver. In the HP lexicon, the ability to capture a variety of information formats (paper as well as digital) and to find and deliver that information back quickly and easily is key to a complete ILM implementation.

This is different from the conventional ILM functionality promoted by the majority of ILM vendors who usually categorize ILM functions as classify, protect, migrate and archive. In truth, these categories are merely functions of data management and retention. Yet even in this conventional view, HP exhibited keen insight: Every customer will approach these functions in their own way and with changing emphasis over time. The mandate that followed from this insight was the need to build maximum flexibility into the HP ILM paradigm.

In order to achieve this flexibility, HP sought to build technology that customers can leverage across multiple ILM functions and a wide a range of usage cases. If architected correctly, this technology would be radically

scalable, easily integrated with all the necessary 3rd party applications that a customer might deploy in their ILM plans, and also augment other HP storage and data management products already in existence. This core technology would have to support disk-based data protection, archiving, compliance and also integrate directly with top-level applications.

HP's answer to this challenge was RISS. As we will explore, it has a key role in HP's ILM strategy, making it one of the more innovative and flexible technologies in the ILM space today.

The RISS Architecture

HP RISS is an object model disk-based archiving and compliance platform designed for high scale enterprise deployments. Delivered in a 42U rack-mount appliance, RISS leverages industry standard HP Proliant servers with their own internal SATA storage. The minimum available capacity for a RISS deployment is 1.7 TB (mirrored), scaling up to 150 TB capacity, per appliance.

Because RISS is an object model storage solution, it differs in significant ways from traditional disk storage, not least of which is capacity efficiency. One of the inherent advantages of object model storage is the capacity optimizing (i.e. "single instance storage") benefits the approach provides. Because RISS stores information based on attribute values, removing duplicate blocks of data, it reduces the amount of storage consumed. As a result, RISS delivers capacity optimization on the order of 3x to 5x traditional storage approaches.

Even among object model storage solutions, RISS possesses a unique architecture. In most cases, a traditional disk-based archival solution that utilizes a CAS appliance will have centralized indexing capability serving a dedicated application that possesses its own dedicated storage nodes. Effectively, this means that one indexing scheme must be applied to the entire repository for a single application. Because these indexes are typically external to the appliance, they are prone to scale and security issues. Not so with HP RISS. RISS differs from this traditional CAS archival approach by utilizing a distributed back-end for querying and storage of content. This capability is a result of advancements in HP's broad-based smart cell technology initiative.

RISS: Built on HP Smart Cells

The HP smart cell architecture is the polar opposite of the monolithic storage mentality that has dominated the enterprise storage landscape for the past decade. Smart cells take very granular computational and storage resources from industry standard components (e.g. HP Proliant servers, Ethernet connectivity, SATA storage) and federate them with distributed software intelligence. These federated blocks can then be marshaled to achieve very complex, high-scale computational and data storage tasks within the enterprise. Today, RISS is correctly touted as one of HP's flagship efforts in this initiative.

As the basic building block of the RISS architecture, each smart cell should be

conceptualized as an atomic unit, with its own dedicated processor, storage, and content indexing capability. These smart cells can all be grouped together into domains behind any number of RISS portals (protocol handlers) that each directly address a top level application. Further, each RISS domain may be partitioned into smaller repositories that constitute logical partitions across a given domain. Through this structure, RISS can then serve multiple applications or business groups, each with its own assigned access settings and data protection levels.

Two of the major benefits of the smart cell approach are high performance and high scalability. Based in distributed computing principles vetted in the High Performance Computing world, smart cells are designed to balance workload across the entire domain, sharing both query and storage information for a given application across the entire domain to which they are assigned. As a result, for an active disk based archive such as RISS, queries into even extremely large repositories can be conducted nearly instantaneously.

Multiple Applications, One Archive

Unlike traditional approaches to creating an archive, RISS does not utilize a hierarchical file system to interface with each and every top level application it supports. Rather, through its API framework, RISS directly integrates with and manages application semantics across all major protocols (e.g. SMTP, CIFS/NFS, HTTP) through RISS portals. These RISS portals in turn establish their domains at the smart cell level. Recall that because of the smart cell architecture, RISS can simultaneously manage multiple applications. As a result, RISS becomes a dynamic platform to support a range of applications such as email, database and file archiving, compliance, document and content management, and a wide range of 3rd party applications or web services. This ability to leverage a single archival platform for all top level applications constitutes a significant leap forward for data archiving, and translates into direct economic benefit for the enterprise.

A Self-Managing Archive

For any customer creating a large, high growth active archive, minimizing management of that environment becomes critical. Taneja Group has routinely encountered enterprise customers who inadvertently spend more time managing their email or compliance archives than they spend addressing the business problems the archives were built to support.

Because RISS aims to become a consolidated active archive or all data types and their related compliance issues, integrating a wide range of self-management features is critical. RISS has integrated access rights, digital fingerprinting, and signature management into its management framework, making the solution inherently ready for duty as a compliance solution.

Additionally, RISS manages its own health and functions through a range of device security controls, capacity and space management, systems monitoring, load balancing, routing, and fail-over. Because of this automation, when a new smart cell is added to RISS, it automatically joins the assigned domain and contributes to the

overall computational and storage capabilities of the solution.

Protection, Integrity and Security
RISS provides a range of data replication, data integrity and data security functionality to its repositories.

First, each smart cell within RISS mirrors its content to at least one other smart cell. For business continuity purposes and offline archival, RISS deployments can be replicated to another RISS appliance at the file level, based on domain membership. Alternatively, RISS archives can be written off to WORM tape.

To ensure data integrity, especially for compliance usage cases, RISS can leverage the inherent advantages of its object model heritage. Unlike a traditional file system, RISS organizes data based on content values, and so it creates unique content IDs for every data object it stores. Each of the data objects is assigned an ID by a 128 bit hash algorithm which is then joined with a time stamp to comprise a unique digital signature. Because RISS stores two copies of all content, the integrity of each object is thereby guaranteed, and any corruption or attempted tampering is immediately recognizable due to the signature file.

Finally, RISS addresses security issues through physical partitions for each smart cell (i.e. separate physical servers) thereby ensuring no logical intrusion across authorized user groups. At a logical level, RISS provides redundant firewall protection. All node to node communication with RISS passes across a private subnet with no external exposure.

Key RISS Differentiators

After analyzing the RISS solution in depth, Taneja Group believes that there are several key advantages that RISS will bring to the customer's data center, especially when compared to their traditional disk archiving solutions. The following capabilities and features are built into every RISS solution, and add up to a differentiated, powerful object model archiving platform.

Application Integration
HP has taken significant steps to ensure that the RISS API is both open and widely adopted across a range of third party applications. Because of its distributed architecture, a single RISS deployment can support the incremental addition of multiple applications over time. Most importantly, customers can expand application support without any impact to the underlying RISS infrastructure. To date, HP has API support for major email applications (MS Exchange, Lotus), content and compliance management (Ilumin, Orchestria), as well as a growing roster of third party partner applications and an open API for web services.

Integrated Search & Retrieval
One of the advantages of an object model archive is the ability to conduct nearly instantaneous record searches based on the solution's previously established metadata and content indexes. Analogous to how internet searches are conducted, this manner of "flat space" searching is magnitudes faster than methods utilized by traditional

hierarchical file systems. To that end, HP RISS has integrated a search and retrieval capability that is horizontally applicable across applications deployed on the RISS platform. This means that as the customer adds new applications and capacities to their RISS deployment, they can do so knowing that a unified search interface will be available to locate content, regardless of its top-level application associations. The immediate implications for this integrated search and retrieval functionality will be obvious in both email and compliance customer deployments.

Capacity Optimized Storage

A major consideration for any long-term archive is the cost of disk. In this regard, one of the principle advantages of an object model archive is capacity optimization through the removal of duplicate data, a process sometimes known as "single instance" storing. However, customers should note that different object model solutions conduct single instancing at varying levels of granularity (i.e. file level or block level.) In general, the more granular the duplicate elimination, the more efficient the overall archive store will be. HP RISS utilizes block level capacity optimization across all of its applications. This means that as file content enters RISS, all documents are rapidly checked against pre-existing versions. After indexing, any block level redundancies are immediately eliminated before storing, but the informational coherency of the object is preserved through RISS's indexing scheme. Assuming typical workloads for enterprise file content, Taneja Group observes that capacity optimization can amount to an aggregate 3x to 5x savings over a traditional storage approach for the same data set.

Dynamic and Granular Scaling

Especially in support of ILM initiatives, Taneja Group sees that users are demanding increasingly granular scaling of capacity. As larger numbers of applications are all integrated into a coherent ILM strategy, varying growth rates place increased pressure on archiving solutions to scale according to the needs of individual applications. Taneja Group believes that HP RISS is optimally architected to meet this dynamic scaling challenge. Because the smallest unit of addition within the HP RISS solution is a 1.7 TB unit, with a current maximum capacity per chassis of 150 TB, customers can scale on demand. Additionally, the software within RISS makes it easy to add new capacity.

RISS Usage Models and Impact

Potential RISS customers should realize that RISS is designed from the ground up to become a centralized archive for a wide range of applications. One of the key advantages of object model archives like RISS is the ability to create one unified archive that addresses multiple top level applications, supporting multiple business processes. By selecting and building a disk archive on one platform, significant economic gains are achievable.

RISS and Email Archiving

Taneja Group has seen a significant and rapid increase in enterprise email archiving deployments in the past 18 months. Driven by a combination of compliance concerns,

email/unstructured content growth, the undeniable criticality of email, and the availability of SATA disk as a platform, we have seen active archives become a top priority for many enterprises, large and small. HP has pointed RISS squarely at this customer challenge, making it a cornerstone offering as an active archive platform.

HP provides a complete end-to-end email archiving solution by integrating HP RISS with HP Reference Information Manager for Messaging (RIM), with support for MS Exchange and Lotus Notes. Because the solution integrates directly with the email application it archives, it differs significantly from competing and alternative approaches that require integration with an intermediary email archiving application (e.g. Legato Email Xtender, Veritas Enterprise Vault.) The simplicity of the HP's solution is that its direct API into the email application lets both administrators and end user gain direct control over the email archiving process.

In an unregulated (non-compliance) deployment, RISS will periodically mine the mail server, removing messages and their attachments all selected according to set administrative policies within RISS. Duplicate data is eliminated through capacity optimization, as explained above. From an email application standpoint, these messages are still visible in the inbox, but an icon will indicate their residency within the RISS archive. Users can then retrieve messages directly from their email inbox by clicking on the message, triggering a retrieval action through the RISS portal, back to the appropriate smart cell which serves that data back through the RISS portal. Alternatively, administrators or authorized users can access the RISS web UI to conduct full text or attribute-based searches of the appropriate domains and repositories within the archive.

This constitutes a significant improvement over historical approaches to email archiving. From Taneja Group's perspective, customers should consider the following points when evaluating RISS versus competing email archive solutions:

- Ease of integration
- Total end user self-service
- Dynamic search and query
- Flexible policy-based archiving
- Single-platform scalability

We believe that all of these factors amount to a significant cost savings and a marked gain in management efficiency for the email archiving environment. Without the object model approach that RISS represents, these gains simply would not be possible.

RISS and Compliance

The other major usage model for a RISS active archive is regulatory compliance. Because of the built-in technical benefits of RISS's object model architecture, it inherently satisfies the most fundamental requirements of regulatory compliance. Specifically, several RISS features directly address compliance issues concerning preservation and presentation of data: fingerprinting, time stamping, data retention, and access controls. Additionally, these same features establish RISS as an accepted software-hardware platform for

WORM storage, also a key regulatory requirement.

For archiving in regulated email environments, RISS will archive every message as it enters the email server (as opposed to batch-based mining mode which RISS utilizes in an unregulated email deployment.) As each message enters RISS, it is digitally fingerprinted, time stamped, and signed, and finally encrypted with a 128 bit DES encryption algorithm. This process will then satisfy regulatory requirements for non-tampering and data authenticity.

After the compliance archive is established, a specialized RISS compliance UI can be used by authorized administrators and executives. This compliance tool has been designed to satisfy guidelines established by the major regulatory bodies (e.g. Sarbanes-Oxley, SEC, NASD, HIPPA, DOD, FDA, FCC, FDIC) in order to expedite electronic discovery, quarantining of desired content, and exportation for review. It enables dynamic content and attribute searching and a unified view of contents across the entire archive.

Based on Taneja Group conversations with security officers and IT managers dealing with a wide range of data-centric compliance issues, we know that "smart but simple" has become the benchmark against which all IT vendors are being judged. Given that criteria, we believe the following attributes of the RISS solution make it a differentiated and compelling solution for use in compliance initiatives:

- *Direct API integration with email.* Because RISS integrates directly with a production email environment, this allows enterprises to simplify their compliance archive deployment.

- *Automatic execution of record duplication, separation and time stamping*. Because RISS automatically stores two copies of content in two separate locations on the electronic media, and then verifies their authenticity, it satisfies core compliance criteria without human intervention or management.

- *Integrated Search and Auditing Tools*. Because it has a built in Compliance UI, RISS obviates the need for 3rd party compliance tools for discovery and auditing, greatly simplifying IT's role in compliance exercises.

These simplifying factors indicate that RISS can assist HP customers with quickly and efficiently addressing regulatory compliance challenges to the satisfaction of all major federal regulations.

Taneja Group Opinion

To date, HP has correctly focused the RISS solution on two of the more pressing enterprise challenges in existence: email archiving and regulatory compliance. We believe that RISS demonstrates significant value and flexibility in addressing both of these usage cases. However, it would be wrong to lump RISS in the category of "just another disk archiving platform" and ignore its significantly wider potential for the data center. There are unique factors at play here

that will only become evident in coming quarters.

We find the RISS smart cell architecture both ingenious and uniquely suited to the demands of a unified archival platform. Specifically, as the RISS platform and HP ILM strategies expand to cover an increasingly wide range of applications, the ability to feed all of these applications back to one common infrastructure will prove massively valuable to HP and its customers.

Of course, we believe that the broad-based applicability of the RISS archive extends far beyond the email and compliance areas explored in this profile. In the near future, we would expect to see HP point RISS towards increasingly critical ILM related tasks in areas such as disk to disk backup and recovery, database archiving, and support for advanced file service management. For example, HP indicates that support for continuous data protection (CDP) will be added to HP RISS in 2006, thus extending the technology's unique capabilities in a variety of ways.

Based on what we see to date, HP intends to give RISS a much more ambitious role in its ILM strategy than any other major competitors are giving to their object model technologies. This is a bet on the future, and the question remains, will it pay off?

Within 24 months, we believe that, indeed, HP's bet on the RISS approach will be paying large dividends. By creating a unified object model platform that is open to a wide range of applications, scales freely, and manages itself easily, HP has positioned itself to compete quite effectively in the ILM battles to come. In short, Taneja Group believes that HP's decision to place RISS at the core of its ILM strategy is technologically bold, and potentially a game-changing move if accompanied by strong execution in the field.