F5 White Paper

Introducing Data Manager

F5 Data Manager helps enterprises monitor their heterogeneous file storage environments and understand how and why their file data is growing over time. With Data Manager's unique software architecture, process workflow, and reporting tools, organizations can better manage file usage and storage.

by Brad Cain
Architect

Contents

Introduction

F5® Data Manager™ is a storage resource management solution. It provides a detailed understanding of configuration, contents, structure, and usage in file storage environments. Data Manager can be deployed effectively in environments with or without F5 ARX® virtualization, and in environments that contain file servers from multiple storage vendors.

Data Manager presents a *single pane of glass* for file storage environments from which an administrator can view and manage file servers, file systems, and virtual volumes. Data Manager also accelerates ARX deployment so organizations can more efficiently manage file storage and usage.

There are many reasons to deploy Data Manager:

- To visualize the storage environment.
  Data Manager discovers the storage configuration of all file servers in the storage environment. Using this information, the Data Manager GUI displays a wealth of file server detail. For example, for a discovered Windows file server, Data Manager displays CIFS servers, CIFS shares, DFS configuration, Local Groups, file systems, and more.
- To characterize file system usage.
  Data Manager inventories file systems to make detailed statistics on usage, structure, and growth available to administrators. It supports a rich set of reports for viewing file system usage in a variety of ways, for example, bar graphs and charts of file counts and capacity on inventoried shares/exports.
- To accelerate virtualization of the file storage environment.
  Using its discovered configuration information, Data Manager can create an ARX configuration that quickly virtualizes file servers. Thus, it not only simplifies an ARX file virtualization solution, but also accelerates deployment.

How Is Data Manager Deployed in the Network?

Data Manager is a Windows web application that requires no agents and is typically installed on a dedicated Windows server. Its components and licensable modules comprise a centralized, extensible storage management platform.

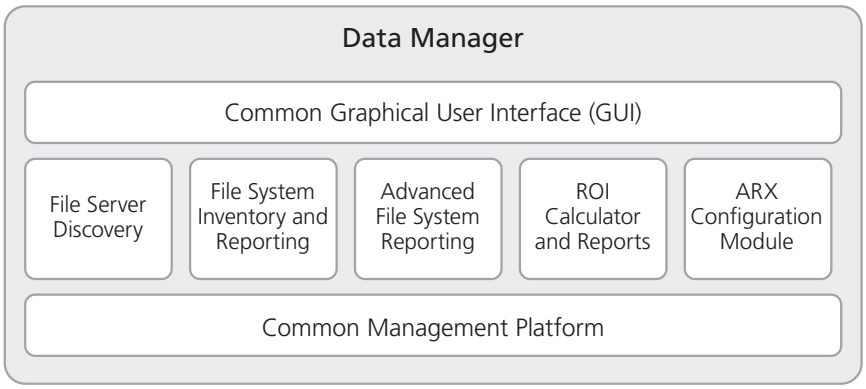

Figure 1: The extensible Data Manager software architecture

Since Data Manager is a web application, both local and remote users can access it through a web browser. The Data Manager Installer configures all components necessary for operation, including its web server and database.

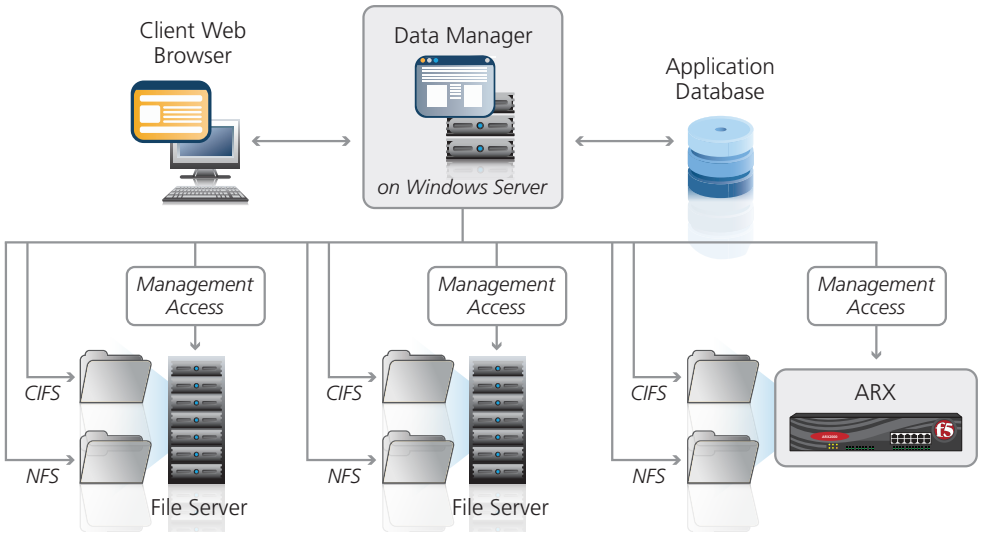

Figure 2: Data Manager deployed in the network

Discovering File Storage Environments

Data centers grow organically over time and are subject to differing financial and application pressures. These pressures often result in the need for heterogeneous storage environments that comprise file servers from different storage vendors. Data Manager helps storage administrators understand heterogeneous environments by supporting discovery of file servers from multiple vendors.

Data Manager begins the discovery process by issuing a series of commands through a file server management interface to find and record file server configuration information. Data Manager then uses the discovered configuration information to present a complete view of the file storage.

The discovery process requires no agents and works as follows:

1. The File Server Discovery module connects to the identified file server through its management interface with administrator-supplied credentials, using SSH, Telnet, RPC, or equivalent.
2. It executes a set of file server–specific, read-only commands that gather configuration details such as the IP addresses, shares/exports, OS version, model, and so on.
3. It normalizes the configuration details and stores them in the application database.

Data Manager supports many of the most popular file storage vendors, including Windows, NetApp, and EMC. For the complete list, see the F5 Data Solutions Compatibility Matrix.

Data Manager also provides a basic discovery process (called Generic) for unsupported file servers.

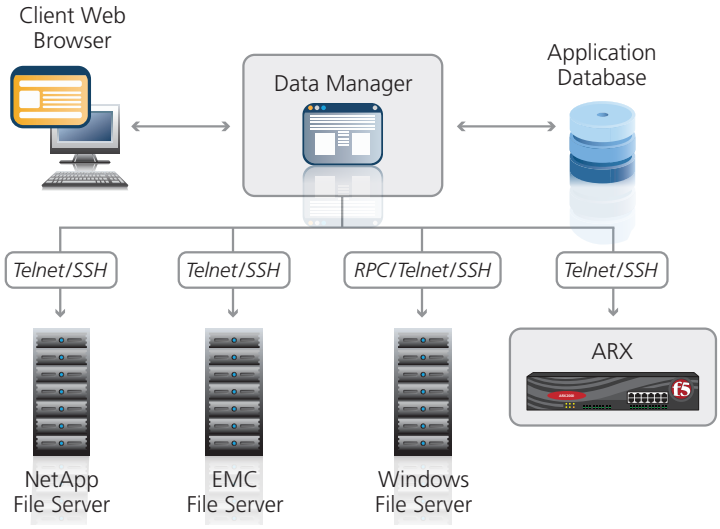

Figure 3: File server discovery in Data Manager

Inventorying and Reporting on File Systems

Due to the rapid growth of file data and the corresponding increase in storage costs, it is more important than ever that storage administrators understand how their storage is used.

Data Manager provides powerful reporting tools that enable administrators to generate reports based on the discovered storage usage information. Using these reports, administrators can better visualize the storage environment and understand the organization's current and future storage needs. With this information, they can more easily develop a plan for the organization's storage usage and set policies for storage tiering and capacity planning.

In addition, administrators can configure Data Manager to gather file system statistics at regular intervals. Thus, changes can be tracked in the environment over time.

Some questions that Data Manager can help answer are:

- What is the file system growth rate?
- What individuals are using the most storage?
- What file types are most prevalent?

- What are the fastest-growing projects?
- What are the ages and sizes of the files on the file systems?
- What are the biggest files?
- Are there duplicate files?

Data Manager enhances visibility into file data through a separately licensable File System Inventory module. This module inventories CIFS shares and/or NFS exports on file servers (and ARX devices) to provide information about file system usage and trends. After an inventory is run, administrators can view the resulting statistics and usage data in the GUI or through generated reports.

What types of reports are included?

Data Manager supports multiple report types, including:

- **Detail** reports, which provide extensive statistics about files within a single file system and an overview of how that file system's storage is being utilized.
- **Cumulative** reports, which aggregate statistics across multiple file systems. They provide information about all files on a single file server or in an entire storage environment.
- **Trend** reports, which show statistics for a file system over time. These reports are useful for understanding where, why, and how quickly data is growing. Trend reports use Data Manager's scheduling ability to periodically gather statistics about specified file systems at specified times.
- **Directory Activity** reports, which provide details about the most and least active directories in a file system so administrators can easily recognize storage access patterns.
- **Suspected Duplicates** reports, which provide details about potential duplicates within a file system so administrators can easily identify wasted capacity in a file system.

What types of statistics are included?

Data Manager provides many different file system statistics, including:

- Overall statistics such as total files, total directories, total disk capacity, used capacity, and total file sizes.
- Trends by file count and capacity.

- CIFS file attribute statistics.
- Directory structure information such as fan-out, leaf directories, and so on.
- Size details such as file count by size and capacity by size.
- Age details such as when files were last accessed or modified.
- Top file extensions by capacity and by file count.
- Top file owners by capacity and by file count.
- Top file system directories and directory trees by capacity and by file count.
- Top largest files.
- Most and least active directories by capacity and by file count.
- Suspected duplicate files sorted by wasted space.

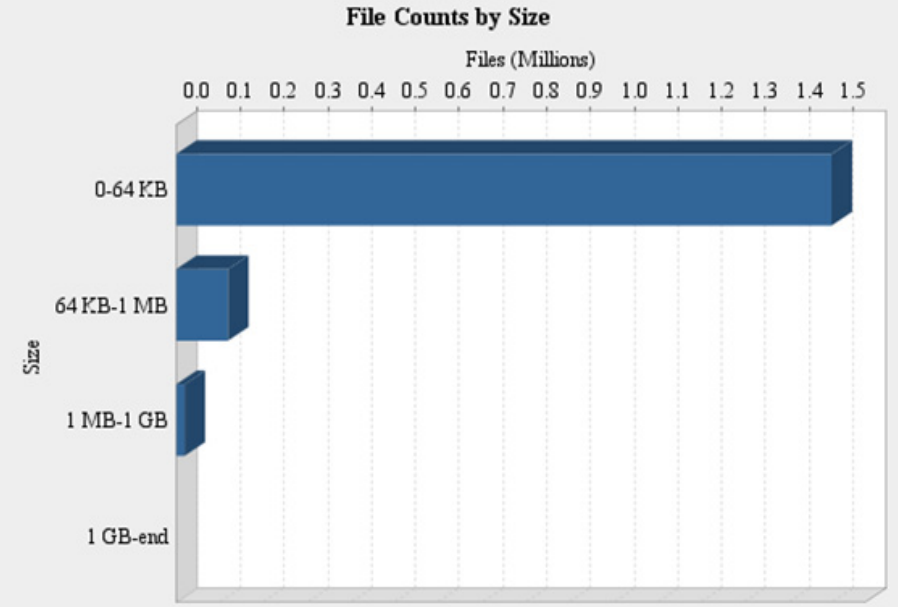

Figure 4: File counts by size in Data Manager

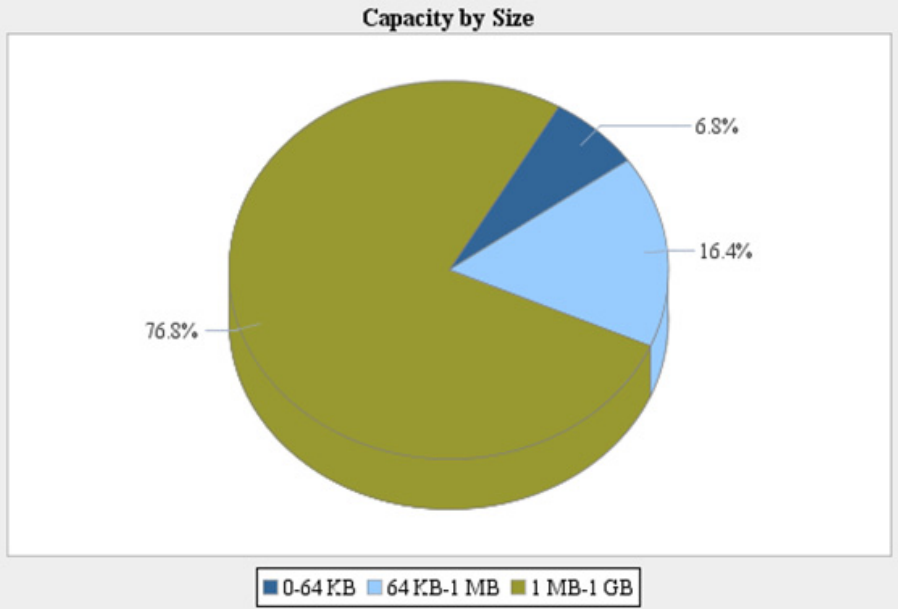

Figure 5: Capacity by size in Data Manager

How does inventory work?

Data Manager inventories file systems using shares/exports on file servers or ARX devices. This process requires no agents and works as follows:

1. Administrators create inventory tasks by entering or selecting a share/export to inventory. Optionally, they can select shares/exports directly from the set found through file server discovery. If inventorying a CIFS share, administrators must also enter the credentials to access the share/export.
2. Administrators select optional parameters for the inventory task such as when to schedule the task, what its scan rate should be, and so on.
3. An inventory task performs a full scan of the share or export. The task reads the file system metadata for all files and directories on the share/export. For CIFS, Data Manager uses the built-in Windows CIFS client. For NFS, Data Manager uses its own built-in NFS client.
4. Data Manager saves all information and statistics in SQL databases, called inventories. Each run of an inventory task produces one of these inventories.
5. Once an inventory task is complete, administrators can generate reports incorporating that inventory.

6. If an inventory task has been configured with an associated schedule, it runs periodically and creates an inventory of statistics each time it runs. Administrators can create detailed Trend reports using these point-in-time inventories.

   Standard reporting enables administrators to create reports based on all files/directories in one or more file systems.

   Advanced reporting enables administrators to create custom reports based on subsets of files in one or more file systems (for example, only Joe's files). Read on to the next section for a more thorough description of advanced reporting.

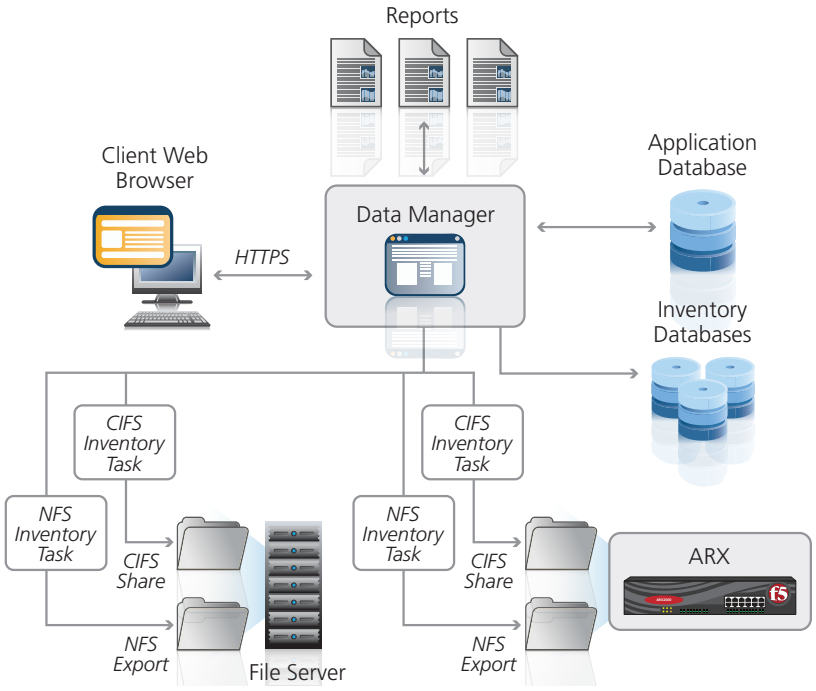

Figure 6: File system inventory and reporting in Data Manager

What is advanced reporting?

Data Manager provides more comprehensive and customizable file data visibility with the separately licensable Advanced Reporting module. Advanced reports are supported for all report types and are useful for:

- Detailed storage usage tracking.
- Storage management policy *what-if* scenarios.
- Additional reporting customization.

With an advanced report, an administrator can report on specific file subsets. This is the essential difference between standard reports and advanced reports: standard reports run against all the files in an inventory, and advanced reports run against a subset.

Administrators define the set or subset that the report will run against by specifying file criteria, such as a specific set of owners, file names, extensions, or sizes, that filter the data being presented to the report engine. For example, suppose an administrator wants to review certain statistics for a marketing collateral project. He might create a file criteria named Active Marketing Projects that contains the following conditions:

- Directory Path condition that filters for the path *public\mydata\marketing-proj-docs*.
- Extension condition that filters for project files with extensions *.docx*, *.txt*, and *.pptx*.
- Relative Modify Time condition that filters for files modified in the past 60 days.

Because the administrator uses the logical operator Match all conditions, the advanced report produces statistics for files in directory marketing-proj-docs that meet all of the conditions listed above.

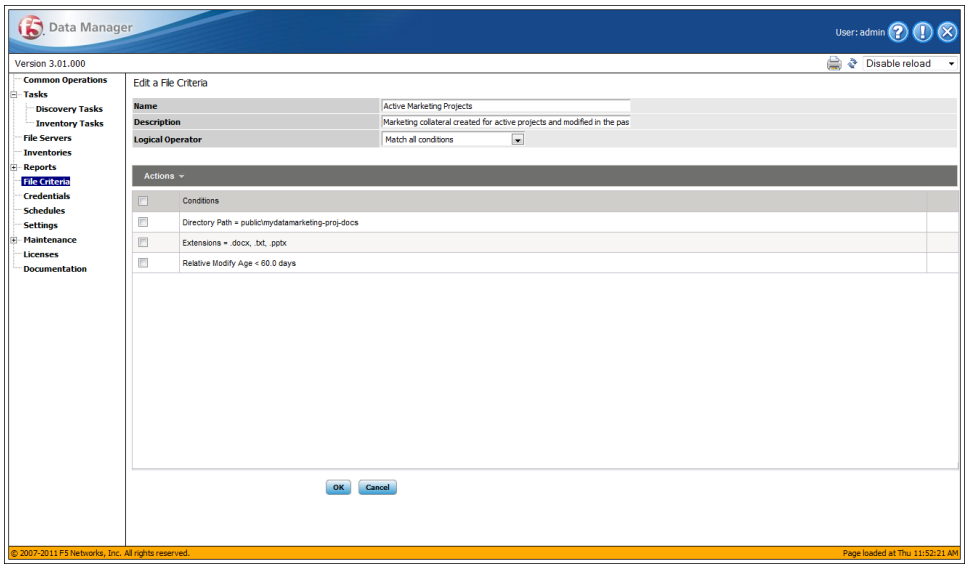

Figure 7: Defining file criteria for advanced reports

In addition, the Advanced Reporting module provides:

- Multi-reporting
  Multi-reporting enables administrators to easily create multiple custom reports (for example, to create several detail reports based on one list of users). Administrators can also generate multiple custom reports from Active Directory groups.
- File groups
  File groups enable administrators to combine statistics in unique ways (for example, a group called *Office Files* that includes all files with .xls, .pdf, .pptx, and .docx extensions). Groups can be used to aggregate statistics in a variety of ways.

Calculating ROI for an ARX Solution

With an F5 ARX tiering solution, non-critical or infrequently accessed data can be automatically and transparently migrated off expensive storage resources to lower-cost alternatives. For example, data that's modified daily can remain on more expensive, high-performing storage while all other data can migrate to less expensive, slower-performing storage.

Implementing storage tiering with ARX reduces storage costs and, due to the flexibility of virtualization, increases storage utilization. Organizations can achieve further savings by creating optimized backup policies for the primary and secondary storage tiers.

To quantify the savings of an ARX tiering solution, Data Manager provides the ARX ROI Calculator. The ARX ROI Calculator takes multiple inputs, including either customized values or data from inventories, and outputs a report that shows in detail the five-year costs with and without an ARX tiering solution.

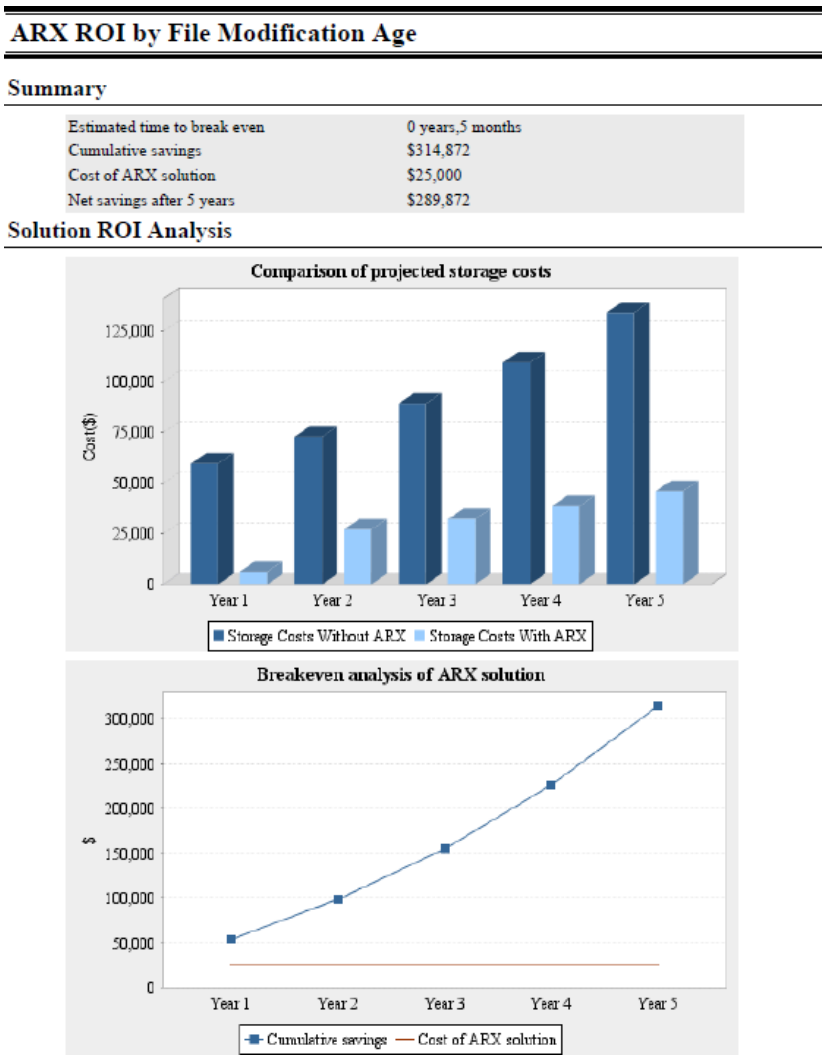

Figure 8: Sample page from an ARX ROI Calculator report

Accelerating ARX Deployment

When an organization decides to implement an ARX tiered storage infrastructure, Data Manager can accelerate deployment and minimize issues with the separately licensable ARX Configuration module.

The ARX Configuration module generates the deployment scripts and configuration files necessary to virtualize a file server. Data Manager uses configuration information

collected during discovery to create an ARX configuration and workflow scripts that virtualize the file server with minimal user/client impact.

The ARX configuration process works as follows:

1. Data Manager uses its File Server Discovery module to discover the file server to be virtualized.
2. Data Manager uses the discovered file server configuration information to create an equivalent configuration for ARX. This configuration virtualizes the file server. Administrators can provide input about how the configuration is created.
3. Data Manager generates a final CLI configuration file for the ARX deployment that virtualizes the file server transparently to clients; it also generates a series of workflow scripts that assist in the virtualization process.
4. The administrator applies the configuration file and runs the scripts to virtualize the file server.

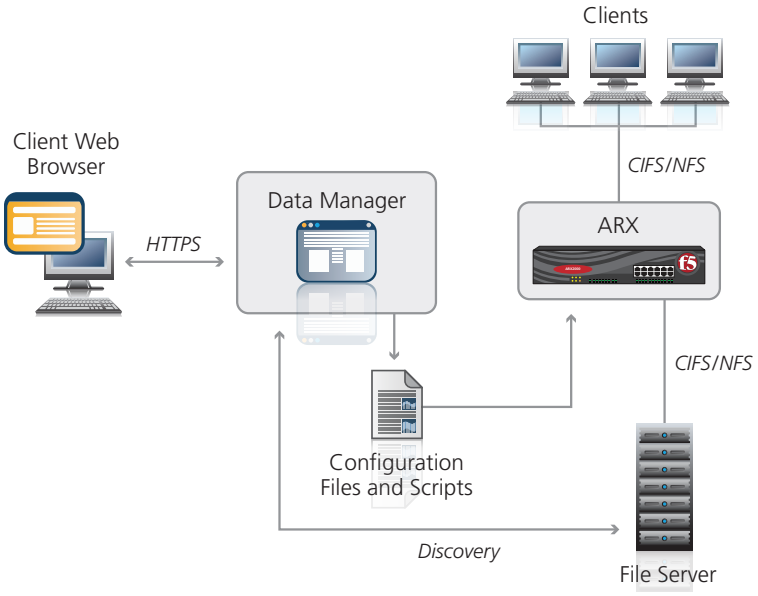

Figure 9: Using Data Manager to configure ARX

Conclusion

Data Manager is a powerful solution for managing heterogeneous file storage environments and for addressing the challenges associated with the rapid growth of unstructured data.

Data Manager gives organizations a detailed understanding of file storage configuration, contents, structure, and usage. Data Manager enables easy access to storage configuration information for all file servers in a storage environment.

With the insight provided by Data Manager file reporting, organizations can improve the accuracy of capacity planning and forecasting, create effective file management policies, and uncover optimization opportunities. Data Manager can also accelerate and simplify the process of virtualizing file storage with ARX. No application can slow the growth of unstructured data. However, with the robust capabilities of F5 Data Manager, organizations are equipped to better handle growth and create efficient and cost-effective strategies to meet the challenges of rapidly growing, heterogeneous file storage environments.

To experience the benefits of Data Manager, download the 90-day trial version. It is distributed as a single Windows executable installation file, which is available from the F5 Downloads web page.

F5 Networks, Inc. 401 Elliott Avenue West, Seattle, WA 98119 888-882-4447 www.f5.com

F5 Networks, Inc.
Corporate Headquarters
info@f5.com

F5 Networks
Asia-Pacific
apacinfo@f5.com

F5 Networks Ltd.
Europe/Middle-East/Africa
emeainfo@f5.com

F5 Networks
Japan K.K.
f5j-info@f5.com

CS01-00059 0911