



Scality Data Assessment Service powered by ClarityNow!

Large-scale environments with file-based workflows often experience an incredibly high rate of data growth and turnover as a result of ongoing product development efforts, research, iterative computational analysis, and content generation. These challenges are not unique to any one industry and share commonality across Media, Life Sciences, Energy, Research, and EDA workflows. As the amount of content generated by enterprises continues to proliferate, the strain on enterprise storage infrastructure, business users, IT

administrators and corporate IT budgets continues to grow. While the price of storage continues to fall, the OpEx costs of managing complex data and the CapEx costs of the storage infrastructure are increasing. Scality's data assessment is designed to enable organizations to understand and quantify the growing problems associated with managing their large scale environments that encompass unstructured data across complex heterogeneous storage networks that are required to support these file-based workflows.

Data Assessment @ Scale

The Scality Data Assessment service has a proven track record of scalability with production-based environments consisting of multi-petabyte, multi-billion file systems. Some of the key differentiators of the Scality Data Assessment service are:

Provides immediate actionable information

- Information is provided in business context (as opposed to IT infrastructure context).
Note: the assessment can provide the typical information of file types, age, etc. which is "nice to know" information but difficult to act on.

Eliminates the assessment hurdles typically encountered within large scale environments

- Traditional scanning approaches fall short in large-scale file system environments (either fail to complete, or take so long information is no longer relevant upon completion).
- Immediate knowledge and visibility into large scale environments
- Timely information is critical for environments with high data turnover (Terabytes per day)
- Meaningful information is delivered within 24 hours (as opposed to weeks / months)

Provides ability to incorporate key information and knowledgebase that typically resides within Production / Business Users

- Gathering such knowledge traditionally requires manual processes to correlate data to business objectives. The manual operational processes encounter difficulties within large-scale environments.

Data Management Challenges:

Projects consuming storage for extended periods of time

Data Security

- Piracy concerns are driving organizations to focus on the value of their content
- Traceability of file / data actions and movement is difficult

Reduced “production line” capacity

- Inefficient data management reduces production capacity.

Operational efficiency capacity

- Increased volume coupled with shorter project timeframes increases the operational complexities associated with data management.

Separation of “church and state”

- Many organizations are demanding separation of “production networks” from “IT Networks” due to piracy concerns

Data Assessment Objectives:

Provide a baseline assessment of current environment

- Identify areas for immediate cleanup
- Outline difficulties of ensuring that the right data, is in the right place, at the right time to optimize file-based workflows
- Gain operational understanding of what to archive and when it can be archived
- Understand workflow and operational inefficiencies introduced by data
- Understand where data movement and audit trails are required

Establish empirical cost information associated with data management (or lack thereof).

Provide understanding of how new technologies may optimize OpEx and CapEx costs for organizations with file-based workflows

The Data Assessment Process

The following outlines the requirements for the Scalcity Data Assessment.

1

Web Conference Introduction and Overview [optional]

Prior to actual onsite visit, the Scalcity Data Assessment technical personnel can provide a detailed introduction and overview of the assessment process.

- Customer Time Required: 30 – 60 Minutes via Web Conference

2

Software Installation / Initial Scan

Onsite, the Scalcity Data Assessment technical personnel deploys the assessment software and initiates data collection process. This process includes a discussion and understanding of the customers file-based workflow required to configure basic auto tag rules.

Customer Time Required: 3 hours on site. Ideally the Scalcity Data Assessment personnel would schedule 1-2 days:

- Day 1: Install software, initiate data collection, and basic auto tag configuration (3 hours)
- Day 2: Scalcity Data Assessment personnel present and review preliminary analysis and initial findings (1 – 2 hours)

3

Status Check / Fine Tuning

After the installation and initial review is completed, a follow-up meeting with the customer is held. The follow-up checks the status of the tag and report configuration and makes any adjustments if necessary. Configuration adjustments can be refined and reviewed via an offline copy of the environment using the save debug facility of the assessment software.

- Customer Time Required: 30 – 60 Minutes via email (phone if required)

Data Analysis and Report Generation

Upon the completion of the assessment, the Scalcity Data Assessment personnel delivers an assessment report to the customer with formal findings / report card including empirical cost information (where possible).

- Time Required: 3 – 5 days offsite / remote (Scalcity resources)

“ Get immediate actionable information about your data from even the largest of environments with the Scality Data Assessment service. ”

The Data Assessment Infrastructure Requirements

Collection Server

The following outlines the basic requirements for the data assessment collection server

- Linux based server or virtual machine (all Linux distributions are supported)
- 16GB of RAM (can operate with less if necessary to get started)
- 40GB of disk space
- All file systems of interest mounted read-only on collection server
- By default the collection process runs as root. If root squash is enabled then collection process will be configured to run as user that has broad file system access privileges.

Network Access

The data assessment software uses the following network ports (by default but can be reconfigured) on the collection server:

- TCP port 80 (running HTTP no SSL encryption)
- TCP port 443 (running HTTPS with SSL encryption)
- TCP port 52000 (runs DRMI a proprietary client/server communication protocol)

About Scality

Scality is the industry leader in petabyte-scale, software-defined storage. Founded in 2009, Scality's software-based RING storage delivers billions of files to 200 million users daily with 100% uptime. The RING makes standard x86 servers scale to hundreds of petabytes and billions of objects. The RING's end-to-end parallel architecture provides unsurpassed

performance, while its patented object storage core increases availability, durability and dramatically reduces operational costs. The RING integrates with applications through standard storage protocols such as NFS; S3; OpenStack Swift and Cinder. Customers include the world's largest enterprises, particularly in media, telecommunications, and cloud.



SAN FRANCISCO | Four Embarcadero Center, Suite 2550, San Francisco, CA 94111 USA TEL: +1 (650) 356-8500 FAX: +1 (650) 356-8501
PARIS | 16 Passage Jouffroy, 75009 Paris, France TEL: +33 1 78 09 82 70 FAX: +33 1 78 09 82 71
WASHINGTON, DC | 11911 Freedom Drive, Suite 1050, Reston, VA 20190 USA TEL: +1 (855) SCA-LITY / +1 (855) 722-5489
JAPAN | 1F Place Canada, 7-3-37 Akasaka Minato-ku, Tokyo 107-0052, Japan TEL: +81 3 6894-7427 FAX: +81 3 6894-7701

www.scality.com

@scality