Veeva

A VEEVA WHITE PAPER

Modern Infrastructure and Publishing Content in the Cloud

Introduction

Regulatory groups in the life sciences have traditionally maintained that submission publishing applications should be as close as physically possible to the content management application in order to ensure optimal performance. This meant both applications were on-premise and managed within the company's facilities, often times "co-located" with one server atop the other.

This arrangement reflected the very real technical limitations that existed when most publishing systems were initially deployed in the late 1900s and early 2000s. Now, companies are moving to newer, more advanced options such as "cloud-to-on-premise" or "cloud-to-cloud" due to the faster file transfers possible with modern infrastructure.

This paper provides an overview of the recent networking advances that support the move from an on-premise, co-located pairing to a distributed, cloud-based approach, and describes how you can capture file transfer times for your current system.

Change Drivers

Business Perspective

From a business perspective, operating groups simply want applications to work fast and without disruption. In the regulatory world, this translates to being able to view, transfer, and publish content quickly, without jeopardizing submission deadlines. If business teams operated the same way they did ten years ago, the on-premise, co-located model would still be working; but business is changing in two fundamental ways.

First, business teams are producing unprecedented volumes of data known as big data. These are "datasets whose size is beyond the ability of typical database software tools to capture, store, manage, and analyze."[1] Many disciplines within life sciences now generate big data, creating 10s of terabytes of data weekly. Shared for collaboration purposes and stored in regulatory filings to await publication, this big data affects all facets of an organization—including publishing groups.

Beyond big data, there has also been a huge shift in the way organizations work. Operating groups are more distributed and global, connecting businesses with offices and partners all over the world. As an example, it is common for regional offices to publish submissions for the various local markets within their given region.

As a result of these changes, more data is published in more places, creating demand for better data transfer solutions.

[1]James Manyika et al, "Big data: The next frontier for innovation, competition, and productivity," *McKinsey Global Institute*, June 2001.

IT Perspective

Life sciences IT has experienced changes of a similar magnitude.

Until recently, there were significant restrictions on internet bandwidth. In the early 2000s, most connection speeds were only in the "T1" Range (1–10 Mb/s), which is too slow to meet business requirements for data and content transfer over the internet. Since then, speeds have increased on average 50% annually, with speeds in the 100 Mb/s range now routinely provided.[2] As connection speed continues to improve, many IT organizations are improving their infrastructure to provide the "fastest pipe out" possible. Today, large, high-speed transfers have become a reality, as regulatory teams now routinely dispatch submissions, some larger than 100GB, over electronic gateways.

In addition to infrastructure improvements, organizations have looked to reduce burden and costs associated with internally managed IT. Activities such as infrastructure maintenance, application upgrades, and service desk support have moved away from traditional brick-and-mortar IT into the realm of managed services. As a result, IT teams are increasingly pursuing a "cloud-first" strategy, which gives preference to software applications hosted externally, either private cloud or multitenant.

Moving to the Cloud

In order to replace aging systems and adapt to the changing business environment, many life sciences organizations are considering cloud solutions for content management and publishing. For IT groups to evaluate the impact of this change, it is wise to assess the current and proposed mix of hosting options. The following sections provide one approach for testing your system.

Establishing a Benchmark for your Current Environment

When users are asked about performance, the response is something on the order of "fast" or "faster." Although everything should be fast, it's not quantifiable; but "faster" is. To establish a baseline, you'll want to measure how quickly content is transferred between your document management and publishing systems today.

First, establish a small set of sample files (e.g. five files with sizes ranging from one, five, ten, 20 and 50MB). These file sizes should be representative of over 85% of the files in a submission within your organization.

Gathering the Data

Using the sample set, you can start by capturing the upload and download times when manually uploading and downloading through the EDMS interface. The results will vary based on the user's timing and the application's interface. Although not possible with every EDMS, leveraging an API is the ideal method to simulate the process used by the publishing system.

Note: With manual timing, the transfer times of smaller files will typically be rounded to the nearest second, skewing results. It may be best to omit files that will transfer in less than 5 seconds.

[2] Rob Tucker, "Broadband Facts, Fiction, and Urban Myths," *The Institute for Broadband Enabled Society*, 2010.

One and Done?

A single point in time does not provide a good representation of how any system performs. Ideally, tests should be performed at varying times of the day on various days to ensure your benchmark is not skewed by external factors. For example, if the test is performed from an EU office just when the US team comes online, there would be a decrease in system performance.

Capturing transfer times at 9 a.m., 12 p.m., 2 p.m. and 4 p.m. for each day of the week creates a representative sample. Adjust the frequency based on what makes the most sense for your organization.

Time of Day		1 MB	5 MB	10 MB	20 MB	50 MB
9 a.m.	Download					
	Upload					
12 p.m.	Download					
	Upload					
2 p.m.	Download					
	Upload					
4 p.m.	Download					
	Upload					

There is much more that can be done for a true statistical analysis of the data, but we are keeping it simple and only creating a basic average.

Conclusion

Prior to the recent inflection point seen in network delivery speeds, most applications were offered on premise to accommodate the limited upload and download performance. Content management applications were no exception: regionally segregated and on premise. As a result, regional affiliates and those partners working remotely suffered through frustratingly slow transfer speeds. The slow transfers, at times, even blocked normal operations. This forced most organizations into an on-premise, co-located architecture, which was burdensome and costly to maintain.

These challenges, combined with improvements in technology and to the network landscape, have prompted IT and business groups to consider cloud options for content management and publishing applications. Testing your environment with cloud content will show whether triage and troubleshooting are needed for the internal network or if existing speeds are acceptable. Rest assured, with the right configurations, today's current networks deliver impressive transfer speeds for submissions publishing.

About Veeva Systems
Veeva Systems Inc. is the leader in cloud-based software for the global life sciences industry. Committed to innovation, product excellence, and customer success, Veeva serves more than 950 customers, ranging from the world's largest pharmaceutical companies to emerging biotechs. Veeva is headquartered in the San Francisco Bay Area, with offices throughout North America, Europe, Asia, and Latin America. For more information, visit **veeva.com**.

Veeva Systems
4280 Hacienda Drive, Pleasanton, CA 94588
925.452.6500 | info@veeva.com | veeva.com