IBM® Aspera® for Life Sciences Solution Overview

Introduction

Biological data has entered the realm of "big data" with the advent of high-throughput and cost efficient genomic and proteomic technologies. The technologies generating a large portion of this data include next-generation sequencing (NGS) systems for DNA and RNA high-throughput sequencing, high throughput mass spectrometry or cytometry for proteomic data, and digital pathology and other biological imaging pipelines.

Along with an exponential increase in data set size is the geographic dispersion of Life Sciences teams needing access to data repositories for analytic or collaborative purposes. Consequently, Life Sciences organizations often struggle to quickly and reliably move, share, and sync large data sets between on premises or cloud locations. Additionally, for scientific work involving human data for clinical applications, these organizations efforts to comply with strict security and regulatory guidelines are critical.

Legacy data transport methods and technologies are predominant in the Life Sciences community – shipping hard drives and FTP-based transfers are failing to keep up, resulting in increased risk, lower productivity, and the deceleration of time-sensitive discovery or clinical outcomes. As the creator of groundbreaking bulk data transport solutions and technologies, IBM Aspera® uniquely overcome these challenges to securely and reliably move, sync, stream and share data and files at the fastest speeds possible, anywhere in the world, across any network, regardless of file size or network conditions.

How Aspera delivers value:

The following are key examples of how Aspera solutions are used in Life Sciences:

- Facilitates faster and more secure large file sharing between global research teams

- Enables fast, highly controllable, and secure genomic data distribution from nextgeneration sequencing and other service labs

- Accelerates analysis of data with high-speed data uploads to powerful cloud bioinformatics or other analytics platforms

- Enable worldwide follow the sun projects with global file synchronization and sharing for medical device and other product engineering

- Ensure high security transport for all data with clinical relevance

- Provide real-time high definition microscopy and medical image streaming for better evaluation and diagnosis

How Aspera delivers value:

- Files of Any Size – Reliably move hundreds of GB to multiple TB and more per session

- Fast – Fully utilize available bandwidth to move data up to hundreds of times faster than TCP, regardless of file size, network conditions or transfer distance

- Flexible – Support for on-premises, cloud and hybrid environments; clients for desktop, web and mobile users; and SDKs for integration with existing applications

- Reliable – Guaranteed and predictable deliveries with auto-resume from the point of interruption

- Secure – Robust data encryption, integrity checks and authentication

- Control – Real-time, centralized transfer monitoring, control and reporting

- Automation – Orchestration solutions to fully automate transfer workflows

Company Profile

- **Industry:** Healthcare and life sciences

- **Attributes:** Organizations shipping hard drives or using FTP/TCP-based transfer technologies to sync, share, move or stream large data sets across WANs to employees, labs, Business Partners and customers, and/or to migrate data to the cloud. Legacy file transfer tools such as FTP are slow, insecure and unreliable for moving large files or volumes of data over long haul Internet connections, resulting in missed SLAs, low productivity, inability to meet regulatory deadlines, increased CapEx and more.

Entry Points

- **IT:** Difficulty meeting data analytics or processing deadlines because of slow or ineffective file delivery methods. Struggle to overcome traditional transport issues over global distances. Challenged to move large data sets to/from/across cloud platforms. Poor application performance due to slow, unreliable data transfer that impacts end-users experience and performance.

- I**ndividual Scientific Investigator, Researcher:** Need a fast and easy way to transport large data sets to collaborators in a different location. The time to publish is a key driver, so high-speed data transfers can be quite valuable. These buyers might be price sensitive, which is why they commonly use open source or low-end file transfer tools, but these TCPbased technologies (such as FTP) do not offer the same value or benefits of Aspera; for example, security, reliability, speed over distance, ability to move any size files, and more.

- **Application Development (e.g. bioinformatics SaaS platform):** Many of these solutions are being built in the cloud to offer the flexible scale and compute needed to process and manage large biological data sets at a reasonable price. Moving data into these applications can be a challenge. Many commercial software and SaaS providers integrate Aspera into their applications to provide their customers with high-speed transfer that fully uses available bandwidth without negatively impacting the network performance for other business applications.

Success Story

National Cancer Institute streamlines global file sharing

- **Business challenge:** Enterprise Science and Computing, Inc. (ESAC) was selected by the National Cancer Institute (NCI) to launch a collaborative research data center and portal to advance the study of proteomic tumor data. ESAC needed fast, secure and reliable webbased, global file sharing for large, multi-GB mass spectrometry files, replacing costly hard drive shipments and unreliable FTP transfer tools.

- **Business solution:** Aspera Connect Server and web-based client plug-in were integrated into ESAC's web portal with custom views for public and private user groups.

- **Business outcome:** Data delivery reduced from days to minutes and hours with complete security and zero data loss, resulting in accelerated research for critical cancer studies and significant savings by eliminating costly global shipments.
Big data bioinformatics enabled in the cloud with high-speed data

Big data bioinformatics enabled in the cloud with high-speed data ingest

- **Business challenge:** Bluebee provides hospitals, research centers and labs with a cloudbased bioinformatics platform for accelerated processing and analysis of large genomic data sets. BlueBee customers produce upwards of 10 terabytes of genomic data a week that need to be uploaded, processed, and analyzed on BlueBee's cloud-based platform, but legacy file transfer tools failed to provide the speed and security necessary to meet strict customer SLAs.

- **Business solution:** After investigating several slow and complicated tools, BlueBee deployed Aspera's Application On Demand. Using Aspera's SDKs, Bluebee natively integrated Aspera into its platform that ran on IBM SoftLayer®, making it possible to go live in a just a few hours.

- **Business outcome:** Bluebee's upload speeds to the cloud increased 35%, even on the slowest networks making it possible for its customers to meet aggressive SLAs and accelerate critical research. The pay-as-you go model and turn-key scalability supports the future growth of BlueBee's business while the robust Aspera security ensures clients are meeting data compliance needs.

Medical Center launches web-based telepathology service to improve patient experience

- **Business challenge:** The University of Pittsburgh Medical Center (UPMC) embarked on a web-based telepathology consultation service for KingMed Diagnostics in China to improve patient diagnoses. Pathologists and doctors in China and the United States needed a method for securely sharing large whole slide image (WSI) sets, from 500 MB to 3 GB in size, in near-real-time to meet the critical, short turnaround time for patient consults.

- **Business solution:** After testing some TCP-based tools that were slow and unreliable, KingMed and UPMC deployed Aspera's Point-to-Point software with hot folders to automatically replicate WSI files across global Internet connections in a fast, predictable fashion.

- **Business outcome:** UPMC achieved transfer speeds of 50–75 Mbps, a 40X improvement over other solutions, with robust encryption and security. Moreover, turnaround time on virtual consultations reduced to under 24 hours, greatly improving the patient experience and outcomes.

Offering Description

- **Solutions:** Aspera offers a portfolio of software and cloud solutions built with its patented FASP® technology – a breakthrough transfer protocol that leverages existing WAN infrastructure and commodity hardware to achieve speeds that are 100s of times faster than TCP. With Aspera, Life Sciences organizations can securely and reliably move, share, sync and stream large files or data sets at maximum speed over global Internet connections, improving productivity, reducing risk and unlocking new ways to do business.

Competition

- **Globus Grid FTP:** Beyond poor performance, Globus has the typical shortcomings of some open source solutions: poor documentation and user interface, complexity, compatibility issues, and installation issues. As an example, to use UDT, users must recompile the Grid FTP binary, a very error-prone process. Another example, installation and set-up is time consuming and complex with little support. It also fails to demonstrate robust security and proposed architectures for Globus Online do not support on-premises deployments.

- **Cloud File Sharing:** Cloud file sharing providers offer online, global file collaboration. These technologies rely on TCP-based transfers, so they are incapable of overcoming the fundamental performance issues that Aspera solves. Large file uploads/downloads using these cloud tools often experience severe reliability and transfer speed issues. These tools also fail to provide infrastructure flexibility, requiring use of their cloud storage. Aspera offers better performance and flexibility for any infrastructure.

- **FTP, Windows® File Sharing, SharePoint:** These are common native or legacy file sharing technologies used by most Life Sciences organizations. All of these tools severely limit data transfer size or struggle, and in many cases fail completely, to move large data reliably over distance or into the cloud. Aspera frequently replaces these inferior legacy technologies.

Additional assets
Case studies

- **GigaScience improves research data accessbility**

  GigaScience is an online open-access life sciences journal, co-published by BGI and BioMed Central, that publishes "big-data" articles covering the full spectrum of biological and biomedical sciences. All of the manuscripts that are published in the journal focus on the use, analysis, or tool-development for large-scale data sets. Content submitters, reviewers, and researchers use the IBM® Aspera® Connect plug-in to upload and download large data sets at maximum speed.

- **EMBL processes large amounts of genomics data for researchers around the world**

  EMBL relies on Aspera's patented FASP® high-speed transfer technology to facilitate collaboration through secure exchange of large research datasets between EMBL research units and globally dispersed partner organizations, including laboratories, hospitals, and academic institutes.