Taking Big Data to the Cloud

Enabling cloud computing & storage for big data applications with on-demand, high-speed transport

WHITE PAPER

aspera

an IBM® company

Taking big data to the cloud

aspera
an IBM® company

Enabling cloud computing & storage for big data applications with on-demand, high-speed transport

TABLE OF CONTENTS

HIGHLIGHTS

Challenges

- Transporting large volumes of data securely, at high speed, to, from and across cloud infrastructures
- Moving the data within the cloud, between remote storage and across compute nodes

Solution

- Aspera On Demand for Amazon Web Services
- Aspera On Demand for Microsoft Windows Azure

Benefits

- Removes computing/storage infrastructure as a limiting factor in meeting un-anticipated demand
- Eliminates the need to build IT infrastructure that can handle spikes in activity only to sit idle most of the time
- Pay-as-you-go model reduces upfront investment risks and improves cash flow
- Supports all data, infrastructure and storage types
- Enables efficient, large-scale workflows with applications for ingest, sharing and exchange of big data
- Supports a variety of client options — web, desktop, mobile and embedded
- Seamless deployment within any on-premise, cloud or hybrid infrastructure

INTRODUCTION

Cloud computing has become a viable, mainstream solution for data processing, storage and distribution. Adoption is accelerating — Amazon Web Services (AWS) has gone from 262 billion objects stored in its S3 cloud storage in 2010, to over 1 trillion in 2012. However, companies that work with big data have been unable to realize the full potential of the cloud, due to the inherent bottlenecks of moving big data in, out and across cloud infrastructures.

Aspera pioneered the high-speed enablement of data-intensive workflows throughout the enterprise, and has now brought the same level of innovation to the cloud. Aspera high-speed transport capabilities are now available on demand, enabling efficient, secure, large-scale workflows in the cloud.

Aspera customers and partners who are using the cloud for big data, such as BGI, Netflix, Zencoder, SendToNews and others, are leveraging Aspera On Demand to get their big data onto cloud infrastructure and meet their IT infrastructure needs, so they can focus on what they do best.

Taking big data to the cloud

aspera
an IBM® company

Enabling cloud computing & storage for big data applications with on-demand, high-speed transport

THE CLOUD PROMISE

Cloud computing holds a tremendous promise of unlimited, on-demand, elastic computing and data storage resources, without the large upfront investments required when deploying traditional data centers.

From a business perspective, the cloud offers three key advantages:

- Removing computing/storage infrastructure as a limiting factor in meeting un-anticipated demand.
- Eliminating the need to build IT infrastructures that can handle spikes in activity only to sit idle most of the time.
- Reducing the risk of upfront investment and improving cash flow through pay-as-you-go models, charging only for the resources that are actually used.

THE BIG DATA CHALLENGE

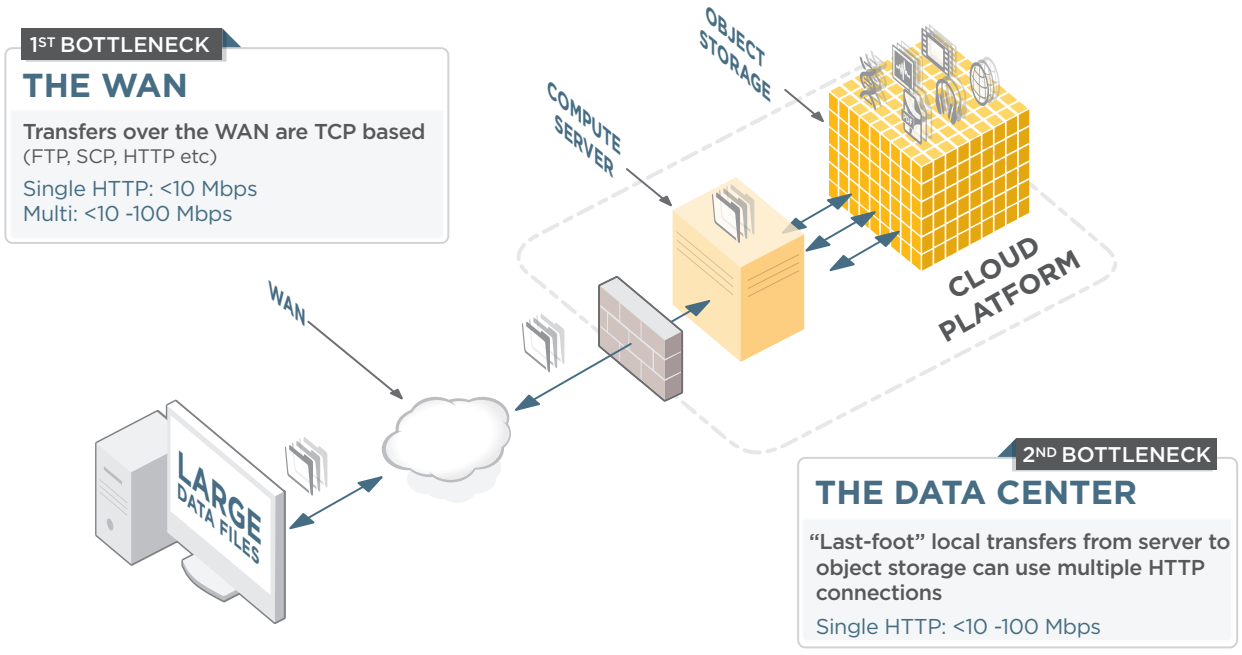
Cloud adoption by businesses has been limited because of the problem of moving their data into and out of the cloud. Often dealing with data sets measuring in tens of terabytes, they have had to rely on traditional means for moving big data:

- Ship hard disk drives to a cloud provider and hope that they don't get delayed, damaged or lost.
- Attempt to transfer the data via the web using TCP-based transfer methods such as FTP or HTTP.
- To become a practical option for big-data management, processing and distribution, cloud services need a high-speed transport mechanism that addresses two main bottlenecks:
- The degradation in WAN transfer speeds that occurs over distance using traditional transfer protocols.
- The "last foot" bottleneck inside the cloud data center caused by the HTTP interfaces to the underlying object-based cloud storage.

ASPERA SOLUTION

Built on top of our patented FASP™ transport technology, Aspera's suite of On Demand Transfer Products solves both technical problems of the WAN and the cloud I/O bottleneck, delivering unrivaled performance for the transfer of large files, or large collections of files, in and out of the cloud.

- Transfers occur at line speed, securely, to and from any location in the world.



Taking big data to the cloud

aspera
an IBM® company

Enabling cloud computing & storage for big data applications with on-demand, high-speed transport

THE SOLUTION

ASPERA ON-DEMAND

- Full client-side r/w of object storage
- Synchronous transfer from Client to object storage (via Aspera On Demand)
- *fasp* transfer speeds end-to-end
- Real-time optimization of HTTP threads, chunk size, interfaced to *fasp*

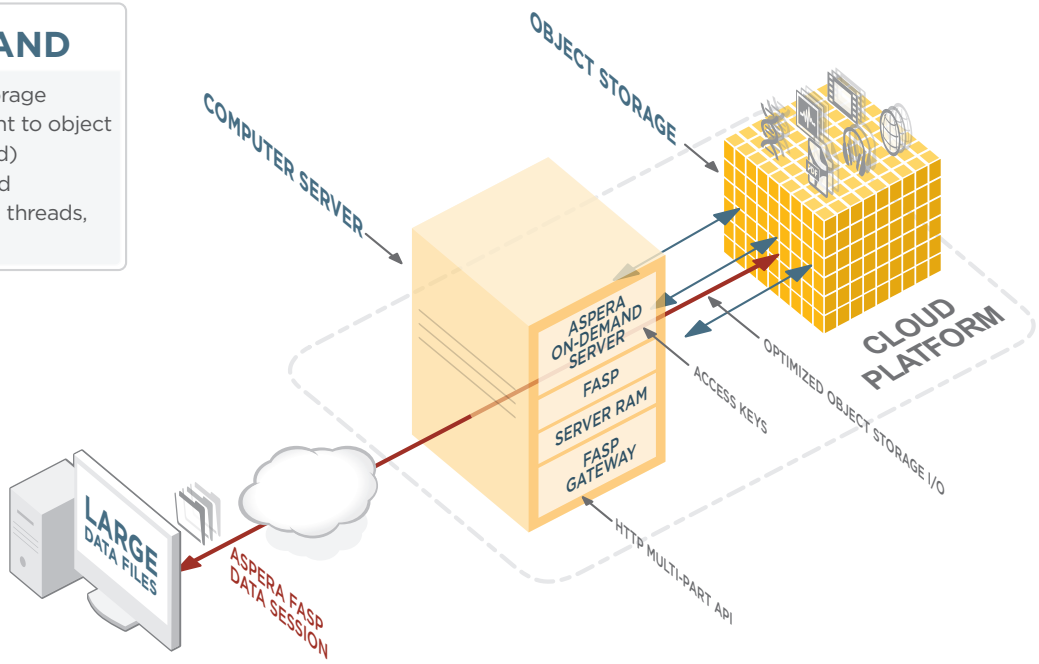
Client Software

- Desktop client
- Web browser plug-in
- Embedded client
- Automatic download widget
- Mobile apps

PERFORMANCE

EFFECTIVE THROUGHPUT

- Up to 1Gbps per instance
- 10 TB per 24 hours



- Files of any size and any format can be transferred at any distance, over any network, under any condition.
- Transfer capacity can easily scale out and back, on demand.
- Aspera On Demand solutions offer full support for all Aspera software and use cases.

Aspera's FASP transport protocol eliminates the WAN bottleneck associated with conventional file transfer technologies such as FTP and HTTP. With FASP, transfers of any size into and out of the cloud achieve perfect throughput efficiency, independent of network delays and robust to extreme packet loss.

Users have extraordinary control over individual transfer rates and bandwidth sharing, as well as full visibility into bandwidth utilization. File transfer times can be guaranteed, regardless of the network distance and conditions, including transfers over satellite, wireless, and unreliable long-distance international links. Complete security is built-in, including secure endpoint authentication, on-the-fly data encryption, and integrity verification.

Aspera has developed a high-speed software bridge, Direct-to-Cloud, which transfers data at line speed, from source directly into cloud storage such as AWS S3 or Microsoft Windows Azure BLOB, with no hops or stops in between:

- Enables direct I/O in and out of cloud storage.
- Ensures intra-cloud I/O keeps up with the FASP-based transport over the WAN.
- Transparently handles cloud-specific I/O requirements such as S3 multi-part uploads.

Using parallel HTTP streams between the Aspera On Demand transfer server running on a cloud virtual machine and the cloud storage, the intra-cloud data movement no longer constrains the overall transfer rate. The files are written directly to cloud storage, without a stop-off on the cloud compute server.

DELIVERING ON THE PROMISE

In addition to providing high-speed transport to, from, and between cloud infrastructures, Aspera On Demand Transfer Solutions include a comprehensive suite of applications for efficient, large-scale workflows, available as subscription services with usage-based pricing and optional add-ons.

Taking big data to the cloud

aspera
an IBM® company

Enabling cloud computing & storage for big data applications with on-demand, high-speed transport

Content ingest and sharing for private, public, hybrid clouds

As businesses adopt the cloud for on-demand storage, they are often faced with the choice of where to place their most important digital content. With **Aspera Shares**, they have complete flexibility in where the content is placed. Whether it is stored in an existing data center, a remote office, or public cloud storage such as Amazon S3 or Windows Azure BLOB. With Aspera Shares, companies can ingest and publish large files and directories in multiple locations, or multiple servers in the same location, within their organization or with external customers and partners. A single web interface consolidates browsing across all shared content and a powerful and flexible security model provides a single management point combining authorization, user management, and access control.

Person-to-Person Collaboration

With digital supply chains now spanning the globe, digital media companies need a high-speed transfer platform that can deal with the complexity associated with transferring ever-larger file sizes over longer distances between geographically dispersed teams.

The **Aspera *faspex*™** solution provides an intuitive, efficient way for individuals and groups to leverage the cloud for file-based collaboration. Built for enterprise workflows of any scale, it includes comprehensive user and server administration capabilities and provides enterprise- grade security and optional encryption of the file content over the wire and at rest. Users can choose from a variety of user interfaces and applications to send and receive digital deliveries including an easy-to-use web interface, a Microsoft Outlook client using the Aspera Add-in, an iOS or Android mobile device using the Aspera *faspex* mobile app, or Cargo, a simple automatic download desktop widget.

The Aspera *faspex* Server enables a multitude of business applications such as:

- Distribution of digital assets within and outside the enterprise.
- Digital delivery and collaboration between geographically distributed teams and with external partners.
- Automated distribution of files.
- File-based review, approval and quality assurance workflows.
- Secure file contribution for remote users and partners.

Scale-out process enablement

In addition to unlimited storage capacity, the cloud offers unlimited processing capacity, enabling parallelization of previously serial computing tasks. This is especially useful for scenarios like:

- Single input, multiple outputs process for applications such as video transcoding, where a single large media file is used as an input to generate multiple other files for different formats, resolution, and devices.
- Assembly-line process in applications such as weather modeling or genetic research, where a large number of data files need to be put through identical processing steps.

With high-speed cloud data transfer enablement by Aspera On Demand and parallel processing offered by the cloud, these processes can be optimized to deliver truly breakthrough performance improvements over what can be accomplished on premise with traditional serial computing applications.

About Aspera

Aspera is the creator of next-generation transport technologies that move the world's data at maximum speed regardless of file size, transfer distance and network conditions. Based on its patented FASP™ protocol, Aspera software fully utilizes existing infrastructures to deliver the fastest, most predictable file-transfer experience. Aspera's core technology delivers unrivaled control over bandwidth use, complete security and uncompromising reliability. As organizations turn to the cloud for improved efficiency and unprecedented scalability, Aspera enables data- and processing-intensive workflows with high-speed transfer available on-demand and maximum speed ingest and distribution of big data to and from cloud storage. Organizations across a variety of industries on six continents rely on Aspera software for the business-critical transport of their digital assets.

Learn more at cloud.asperasoft.com