



Technical Overview

JetStream Migrate

Version 3.0, Updated August 2018

Table of Contents

- 1. Introduction 3
- 2. Cloud Migration Challenges 3
- 3. Product Design 5
- 4. Advantages..... 7
- 5. Example..... 8
- 6. Conclusion 10

1. Introduction

Organizations continue to pursue the shift from maintaining their IT infrastructure in traditional on-premises data centers to the cloud. Yet they remain challenged by the difficulty of moving business critical applications from their current infrastructure to cloud service providers' facilities.

The risks of operational disruption can be mitigated by preserving the underlying virtualization infrastructure, moving virtual machines and their virtual disks in their entirety from the on-premises VMware environment to a VMware environment maintained by a cloud service provider. Working with one of the thousands of VMware Service Provider Partners, administrators can be confident that their new cloud environment provides compatibility with their VMs and their datastores, and they can use tools like vRealize to define robust blueprints for network, CPU and memory configurations, load balancing, monitoring and management.

For virtualized workloads, cloud migration can be simplified by replicating a virtual machine's data to the cloud while the VM continues to run on-premises. Migrating a VM's data while it continues to run is called "live migration." Once its virtual disks are fully replicated at the destination, the virtual machine migrates to the destination.

JetStream Migrate is a software application designed to facilitate virtual machine migration to the cloud. It is especially unique in the way that it manages data replication to the cloud, enabling a number of distinctive features, specifically:

- Live migration with minimal downtime for running applications
- Migration estimation to forecast migration duration, network requirements, etc.
- Lightweight deployment, and simple operation
- Fault tolerant data transfer that can resume replication if interrupted
- Data migration supported via data transport device as well as over a network
- Control over the balance between replication duration and VM performance

This overview describes the design of the JetStream Migrate software, and how it provides these features and capabilities.

2. Cloud Migration Challenges

Systems integrators and cloud service providers typically offer professional services for cloud migration. In doing so, they commonly face challenges including:

Uncertainty in Planning: The more sophisticated and bigger the organization, the greater the complexity of the migration. Some applications are closely integrated, and should migrate together. Network addresses have to be preserved, or at least managed consistently. Permissions in LDAP or Active Directory may have to be replicated exactly if applications are expected to be immediately available to their users. And security requirements must be defined

and enforced at the point of origin, the destination, and the data transfer process. The network used for migration should be evaluated in terms of available bandwidth relative to the amount of data to be replicated. The order in which VMs will migrate should be defined and the time required for migration should be understood before the migration begins. Finally, the replication method should be defined – whether fully over a network or primarily by physical shipment of a data transport device.

Operational Disruption: The first on-premises applications to migrate to the cloud were typically test and development systems, which could be stopped for migration. Disruption to these systems entailed some inconvenience, but keeping the systems running continuously was not required. As organizations began to consider migrating business-critical operations to the cloud, “live migration” became a key requirement. In some cases a small maintenance window may be available for migration; in other cases, applications cannot be interrupted for more than the few minutes required for a virtual machine to restart. In addition to the brief interruption when starting the VM in the cloud, live migration can slow down running applications as their data is read from production storage systems. When considering a migration project, the organization must develop a solid understanding of the amount of operational disruption that could result, and prepare appropriately.

Software Overhead: Many cloud migration solutions available today manage data transfer through legacy technologies for disaster recovery (DR). While they may have been updated in some respects for cloud migration, they may retain some limitations. They may be complex to configure or difficult to administer. They may require continuous, uninterrupted connectivity between the source data center and the cloud. They may introduce agents or drivers that are incompatible with (or not fully supported for use with) technologies already in use in the source data center or the cloud destination. They generally cannot support live migration when data is moved on a data transport device.

Storage Incompatibility: Data migration solutions that are based on specific storage infrastructures may not be useful across all workloads. Many enterprises have accumulated a diverse set of storage systems and datastore types, including traditional shared storage arrays on a SAN, file servers, and hyper-converged infrastructure. Not all storage-based data migration solutions can support all datastore types. For example, migration based on file-level synchronization may not easily support block or object datastores. Similarly, distributed data management approaches based on object datastores are fine – for on-premises object datastores. In contrast, a storage-agnostic migration solution will be compatible with the storage at the on-premises source and the cloud destination.

JetStream Migrate was designed to address these challenges. It was purpose-built for cloud migration, and compatibility with VMware vSphere both on-premises and in the cloud.

3. Product Design

JetStream Migrate features four basic components:

- **Migration Estimator** for “pre-flight” prediction of migration duration and migration parameters
- The **Data Replicator** that transfers VMs’ data to the cloud destination and migrates production VMs
- The **Data Transport Device Manager**, that facilitates data transfer via a physical device
- The **Management Server**, used by the administrator to manage and orchestrate migration processes

These four components work together, though the Data Transport Device Manager may not be required if all data can be transferred over a network connection.

The Migration Estimator is a utility that helps administrators predict the time and network bandwidth required to move the VMs’ data to the cloud. Running in “observation mode,” it determines the amount of data to be transferred, the actual network bandwidth available for the transfer, and the amount of data that will change during the transfer, as more changes at the origin during migration will require a greater amount of “foreground” data transfer. A key advantage of the Migration Estimator is that it can be used for not only individual VMs, but also “Migration Groups” – sets of interdependent applications and VMs that should migrate at the same time.

The Migration Estimator evaluates the replication network bandwidth, latency and rate of packet loss, and it uses this information to set replication parameters to enable the Data Replicator to fully utilize the network. Using the Migration Estimator, the administrator may determine that it is more efficient to transfer data via physical shipment of a data transport device. JetStream Migrate supports live migration even when data is shipped to the cloud destination on a physical transport device.

The Data Replicator is the mechanism that is fundamentally responsible for getting the data to the destination with minimal disruption to the application as it is running on-premises. The JetStream Data Replicator is integrated with the vSphere hypervisor through an IO Filter, based on the vSphere APIs for IO Filtering. The Data Replicator combines two replication processes to copy data to the destination while the VM and its application continue to run. These processes, call Background Replication and Foreground Replication, are described below.

Data Replication: Key Concepts

Background Replication	The process of reading data from the VDISK at the source and copying it to the destination.
Foreground Replication	The process of capturing data write operations taking place while migration is under way and reconciling the data at the destination.
Write Backpressure	A controlled slowing of the application's write operations, which takes place when network bandwidth is insufficient for the amount of foreground replication that is required due to the application's current write activity.
Replication Backpressure	A controlled slowing of background replication in order to avoid negatively impacting application performance.

Migration Groups are usually identified by the administrator, who manages their data replication process to ensure that all VMs in the Migration Group power on concurrently at the cloud destination. However, for VMs that migrate individually, the administrator can set their migration processes to execute automatically, moving thousands of VMs to the cloud destination in a fully automated process.

The Data Transport Device Manager is used when data will be transported to the destination on a physical device. Sometimes the amount of data to be migrated is very large, and moving data over the network would be impractical. In these situations, transferring the data on a physical device can save a great deal of time. Even when data is transported to the cloud on a physical device, the application and VM can continue to run at the source. JetStream Migrate supports live migration with *both* device-based and over-network data replication.

While the Data Replicator can be used without the Data Transport Device Manager (if all replication will take place over a network connection) the Data Replicator is always needed. Even if almost all data is transported to the destination via physical device, the Data Replicator is needed to execute the foreground replication process for data written while the device is in transit. For maximum efficiency, the Data Replicator begins foreground replication as soon as the device is populated and continues to replicate data while the device is in transit.

The Management Server virtual appliance is a vCenter plugin that is used to manage and orchestrate the migration processes. Regardless of whether data will migrate entirely over a network or primarily via a data transport device, the administrator manages the migration process through the Management Server GUI, which is simply a tab in the vCenter GUI. The Management Server also provides REST endpoints for greater scalability and automation. The migration source and destination vCenter servers can run in Enhanced Linked Mode (ELM), for full administrative control over the destination clusters from the source vCenter. Alternatively, they may be accessed independently if ELM is not possible.

The Management Server orchestrates the migration process. This includes:

- Defining the order of migration
- Automatic creation of the destination VM profile for the final stage of migration

- Identification of the replication network
- Creation of destination VDISKS
- Generation of replication statistics and error reports

4. Advantages

JetStream Migrate was designed to address the following live migration requirements:

- All VMs remain in normal operation while data replication is in progress.
- Data replication does not disrupt or significantly degrade performance for the VMs and applications at the source.
- The rate of data replication may be adjusted if necessary to maintain application performance.
- Standard vSphere functions are supported while data is migrated.¹
- Minimal VM downtime, with configurable downtime settings.
- Migration Group support for collective migration of interdependent VMs.

Due to its unique design, JetStream Migrate offers a number of advantages:

- Live Migration: Data migration to the cloud while the application is running
- IO Filter design: No DR or backup product is required
- No snapshots are used, so replication has less impact on application performance
- Transparent to standard vSphere operations: vMotion, snapshot, cloning, etc.
- Migration Estimator: predicts replication duration and network bandwidth, automatically sets replication parameters and recommends migration strategy
- Support for live migration with virtual disks copied onto a device for off-line transport
- Migration Groups: VMs whose data is migrated together and power on together
- Scalability: automated migration for hundreds or thousands of VMs
- Fault-Tolerance: data transfer can resume if interrupted
- Performance Management: balances replication time and VM performance

¹ There are a few minor exceptions: during data migration the VDISK size cannot be changed, a new VDISK cannot be added to the VM, and snapshot reverts are not allowed.

5. Example

Let's look at a simple example of an organization migrating a substantial number of its virtual machines to a data center operated by a cloud service provider.

Before Migration:

1. The administrator installs JetStream Migrate on the vSphere hosts at the on-premises data center. The installation is through a simple VIB and a vCenter plugin for the management server. Nothing is installed at the destination, because the ISO image will be transferred directly to the destination.
2. The administrator establishes Enhanced Linked Mode (ELM) between the vCenter servers at the on-premises data center and the vCenter servers at the cloud destination. (Note that ELM is optional. Migration can be managed with independent vCenter servers as well.)
3. To ensure secure data replication, the administrator establishes a secure communication channel between the source and the destination. VMware NSX is recommended, and if it is needed, the administrator can install an NSX virtual appliance.
4. At the destination, a virtual network (through virtual switches, etc.) is identified for complete and automatic VM migration.
5. The administrator duplicates essential services including LDAP, Active Directory, DNS, and DHCP at the destination side so that migrated VMs can begin operating immediately.

Migration Estimation

The Migration Estimator is installed together with the Data Replicator and is managed through the Management Server. Once the administrator has identified the VMs to be migrated, the Estimator calculates the amount of data in the relevant virtual disks, and gathers IO statistics to estimate the amount new data that will be generated while migration is under way. If the administrator defines a Migration Group, the VMs' statistics are aggregated for the Migration Group.

To estimate the time required for migration, the Migration Estimator also takes into account the available network bandwidth. The network may be shared or dedicated to migration exclusively. Determining the actual network bandwidth available for migration can be difficult, due to factors related to MPLS networking, general network latency, and rate of packet loss. To determine network bandwidth as accurately as possible, the Migration Estimator runs periodic network probes using various workloads that mimic migration activity.

The statistics collected by the Migration Estimator are saved for off-line analysis in the Management Server database. Some statistical data are used directly by the Data Replicator to set optimal runtime parameters, while others are provided as recommendations (or warnings) for the administrator.

Migration Estimator Statistics

Statistics used directly by the Data Replicator	Optimal number of network connections given the available network bandwidth
	Optimal buffer sizes (both TCP and the Data Replicator's internal buffer)
Statistics used for recommendations and warnings for the administrator	Recommended migration method: <ul style="list-style-type: none">• Full network migration• Data transport Device + network• Standard OVF (with VMs stopped)
	Estimated time required for migration (by VM and by Migration Group)
	Recommendations for write backpressure and/or replication backpressure

Data Replication

The Data Replicator is responsible for actual data replication and the final VM migration. The administrator may choose to manage the final VM migration directly, for example to confirm that all VMs in a Migration Group are operational at the cloud destination. For VMs that can migrate individually, the administrator may set the Data Replicator to migrate them automatically, allowing hundreds of VMs to migrate to the cloud with essentially no administrative overhead.

The administrator uses the Data Replicator in basically the same way whether the data is transferred over a network connection or primarily via physical device. The key difference is whether background replication (copying the VDISKS' data) will be over the network or via device. In either case, the Data Replicator manages foreground replication in the same way.

Below are the steps involved in migration of an individual VM (over a network):

1. The administrator assigns VMs for migration and starts the Migration Estimator.
2. The Migration Estimator collects statistics for the VMs and provides migration recommendations including potential backpressure requirements.
3. The administrator assigns the migration parameters (the target migration duration and maximum VM down time) The administrator then starts the actual migration process.
4. The Data Replicator automatically defines the migration order and starts replicating the VMs' data.
5. Data replication consists of background and foreground replication processes. If necessary, administrator can check migration statistics and apply backpressure. However, if the Migration Estimator did not make any backpressure recommendations, the typical migration process does not require administrative attention.
6. If something does go wrong (network interruption or host failure) or if it becomes clear that the replication duration or VM downtime objectives cannot be met, the administrator is notified (typically via e-mail alert).
7. If the migration objectives can be met, once data has been replicated the VM will be migrated to the destination automatically.
8. When the VM has migrated, (or if the migration fails or is terminated for some reason), the Management Server switches to the next migration task.

Final VM Migration

JetStream Migrate replicates all data before starting the VM at the destination. The final VM migration process stops the VM at the source for very short time and the VM's virtual state is automatically recreated in the cloud. The amount of VM downtime is configurable. At the destination, the network is reconfigured, and the replicated VDISKS are attached to the newly migrated VM. The VM can start automatically, or the administrator can designate VMs for manual start, for example if they belong to a Migration Group.

With migration complete, the Management Server automatically deletes the replication agent at the destination side and releases any cloud resources used for replication. However, the administrator has the responsibility for deleting the original VMs and their VDISKS after confirming that the migrated resources are properly configured and working as expected.

6. Conclusion

JetStream Migrate addresses one of the most important barriers to cloud adoption, the difficulty, uncertainty and disruption of managing data for live migration to the cloud. The unique design of the JetStream Migrate software provides numerous of advantages for organizations that require efficient cloud migration with no interruption to runtime operations.