What role does replication play in the migration process?

04/04/2019 • 4 minutes to read • Contributors 👤👤

In this article

On-premises datacenters are filled with physical assets like servers, appliances, and network devices. However, each server is only a physical shell. The real value comes from the binary running on the server. The applications and data are the purpose for the datacenter. Those are the primary binaries to migrate. Powering these applications and data stores are other digital assets and binary sources, like operating systems, network routes, files, and security protocols.

Replication is the workhorse of migration efforts. It is the process of copying a point-in-time version of various binaries. The binary snapshots are then copied to a new platform and deployed onto new hardware, in a process referred to as *seeding*. When executed properly, the seeded copy of the binary should behave identically to the original binary on the old hardware. However, that snapshot of the binary is immediately out of date and misaligned with the original source. To keep the new binary and the old binary aligned, a process referred to as *synchronization* continuously updates the copy stored in the new platform. Synchronization continues until the asset is promoted in alignment with the chosen promotion model. At that point, the synchronization is severed.

Required prerequisites to replication

Prior to replication, the *new platform* and hardware must be prepared to receive the binary copies. The article on prerequisites outlines minimum environment requirements to help create a safe, robust, performant platform to receive the binary replicas.

The *source binaries* must also be prepared for replication and synchronization. The articles on assessment, architecture, and remediation each address the actions necessary to ensure that the source binary is ready for replication and synchronization.

A *toolchain* that aligns with the new platform and source binaries must be implemented to execute and manage the replication and synchronization processes. The article on replication options outlines various tools that could contribute to a migration to Azure.

Replication risks - physics of replication

When planning for the replication of any binary source to a new destination, there are a few fundamental laws to seriously consider during planning and execution.

- **Speed of light.** When moving high volumes of data, fiber is still the fastest option. Unfortunately, those cables can only move data at two-thirds the speed of light. This means that there is no method for instantaneous or unlimited replication of data.
- **Speed of WAN pipeline.** More consequential than the speed of data movement is the uplink bandwidth, which defines the volume of data per second that can be carried over a company's existing WAN to the target datacenter.
- **Speed of WAN expansion.** If budgets allow, additional bandwidth can be added to a company's WAN solution. However, it can take weeks or months to procure, provision, and integrate additional fiber connections.

- **Speed of disks.** If data could move faster and there was no limit to the bandwidth between the source binary and the target destination, physics would still be a limiter. Data can only be replicated as quickly as it can be read from source disks. Reading every one or zero from every spinning disk in a datacenter takes time.
- **Speed of human calculations.** Disks and light move faster than human decision processes. When a group of humans is required to collaborate and make decisions together, the results will come even more slowly. Replication can never overcome delays related to human intelligence.

Each of these laws of physics drive the following risks that commonly affect migration plans:

- **Replication time.** Advanced replication tools can't overcome basic physics—replication requires time and bandwidth. Plans should include realistic timelines that reflect the amount of time it takes to replicate binaries. *Total available migration bandwidth* is the amount of up-bound bandwidth, measured in megabits per second (Mbps) or gigabits per second (Gbps), that is not consumed by other higher priority business needs. *Total migration storage* is the total disk space, measured in gigabytes or terabytes, required to store a snapshot of all assets to be migrated. An initial estimate of time can be calculated by dividing the *total migration storage* by *total available migration bandwidth*. Note the conversion from bits to bytes. See the following entry, "Cumulative effect of disk drift," for a more accurate calculation of time.
- **Cumulative effect of disk drift.** From the point of replication to the promotion of an asset to production, the source and destination binaries must remain synchronized. *Drift* in binaries consumes additional bandwidth, as all changes to the binary must be replicated on a recurring basis. During synchronization, all binary drift must be included in the calculation for total migration storage. The longer it takes to promote an asset to production, the more cumulative drift will occur. The more assets being synchronized, the more bandwidth consumed. With each asset being held in a synchronization state, a bit more of the total available migration bandwidth is lost.
- **Time to business change.** As mentioned in the previous entry, "Cumulative effect of disk drift," synchronization time has a cumulative negative effect on migration speed. Prioritization of the migration backlog and advanced preparation for the business change plan are crucial to the speed of migration. The most significant test of business and technical alignment during a migration effort is the pace of promotion. The faster an asset can be promoted to production, the less impact disk drift will have on bandwidth and the more bandwidth/time that can be allocated to replication of the next workload.

Next steps

After replication is complete, staging activities can begin.

Staging activities during a migration