

A Unified Data Mover for Simplifying Large-scale Data Management

SCAsia
Supercomputing 2019

Gathering the **Best of HPC** in Asia

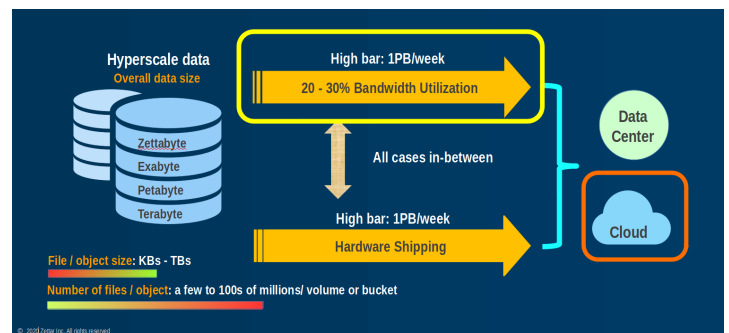
Zettar Inc. (Zettar), the *only* Overall Winner of the Supercomputing Asia Data Mover Challenge, delivers *zx*, a unified data mover. It simplifies large-scale data management, e.g. in a hybrid cloud environment. *zx* transparently supports both file and object storage while *perserving data attributes across storage*. With its unmatched ability for moving data at scale and speed, together with its versatility in tackling diverse data movement needs, *zx* is actually simple to install, configure, and tune. In less than 15 minutes, a *zx* deployment is ready to move PBs of data, fast.

The Problem

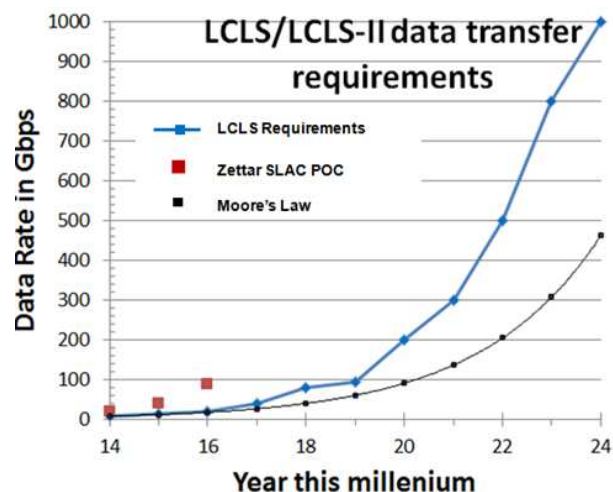
According to Intel Corp., 90% of the world's data has been generated since 2016. *Only about 1% of it is utilized, processed and acted upon.* Imagine the advantage if your business could effectively leverage more data at scale than your competition! Two basic problems of data under-utilization are: 1) the lack of understanding of the role that infrastructure plays in data movement, 2) modern distributed data-intensive organizations have diverse data movement needs, but most dealt with them using many different data movers that are often aged, complex, and inefficient. Some are also very expensive.

The Zettar Solution

Zettar actively contributes to freeware and provides free Webinars, conference talks, papers, and U.S. DOE Technical Reports to help various organizations to gain understanding of their respective infrastructure, especially the critical storage stack. Zettar *zx* is radically simple, scalable, and efficient. It is also the only data mover in the world that is capable of tackling the first three categories listed in the U.S. DOE Technical Report, "**Data Movement Categories**" <https://bit.ly/33RMaHu> The design of *zx* leverages the existing storage, computing, and networking as much as possible with high concurrency and parallelism. It is ready for cost-effective operation and high ROI ownership. It also supports the data movement needs of diverse data management tasks, especially in a hybrid cloud environment.



Problem: Insufficient understanding, complexity, poor efficiency, and high cost



Picture credit: Dr. Les Cottrell, US Department of Energy, SLAC National Accelerator Laboratory, 2016

The Value Proposition

- **Saving tremendous time and money.** With an appropriate infrastructure, zx enables higher utilization of the infrastructure than incumbent competitors (80% or more vs mere 20 – 30%). The value can be illustrated with an example: a tier-1 Life Science business usually has around 200 scientists engaged in research and collaboration involving large amounts of data. Assuming each scientist has a \$300K/year compensation and with zx, they have a 10% productivity improvement, then over a two year period the company saves 2 years x 200 scientists x \$300,000/scientist/year * 10% = \$12M. Once other factors are considered, the saving becomes far more significant.
- **Revenue upside.** Example: assuming the productivity improvement enables a major drug discovery sooner, the revenue acceleration could be in the hundreds of millions of dollars or more.
- **Stay current.** All maintenance releases, updates, upgrades, or features are provided by Zettar directly or indirectly to all licensed users.

The Product Strengths

The aforementioned U.S. DOE Technical Report classifies the commonly seen data movement needs, as observed in data-intensive institutions (*both commercial and non-profit*), into four categories.

1. **Completely automated data movements**
2. **Partially automated data movements, end user controlled**
3. **Partially automated data movements, controlled by IT professionals**
4. **Strictly interactive data movements and end user oriented**

With its unmatched versatility, scalability, and efficiency, zx is the world's only unified data mover that can effectively tackle the first three demanding categories. For example, it can accelerate the following seemingly unrelated tasks:

1. **Bulk transfer/append stream data over digital connections (Mbps, Gbps, multiple 10/100Gbps; LAN, Metro, WAN)**
2. **Replicating data sets incrementally, even those with hundreds of millions of small files/objects**
3. **Accelerating the data on-loading/offloading + cloud onboarding for data shipping devices (AWS Snowball alike)**

In addition, unlike other commercial offerings that impose various limits, zx comes with the following:

- **All-in-one** – it integrates several products as one for simple deployment, consistent usages, and high efficiency
- **No arbitrary caps** on data and transfer volumes
- **Simple to install, understand, manage, scale, and use. Efficient and fast.**

Since 2015, Zettar has been engaged to support the ambitious data movement requirements of LCLS-II, a premier U.S. DOE Exascale Computing Preparation Project. Zettar software has always been designed and implemented under truly stringent requirements. Comparing other free and commercial data movers to Zettar zx is like comparing the sporting ability of a weekend warrior and an Olympian. For many published product accomplishments, please see the Zettar Website, Company, Publications <https://bit.ly/3ldVqQ0>.

Key Features

The following table selectively lists some user visible zx key features that are unique, powerful, and useful.

Features	Brief Description
Intrinsically scale-out, peer-to-peer architecture	The architecture is intrinsically cluster-capable, i.e. no need to use an "orchestration add-on" or a cluster workload manager. The operation, management, and usage of a multi-node cluster are just as simple as that for a single node
Scale-out encryption	Data traffic encryption power increases with the size of a cluster
Multi-stage parallelism	zx, being intrinsically scale-out and peer-to-peer, provides not only concurrent processing at the node level for storage, computing, and networking but also parallel processing at the cluster (aka swarm) level. Note that as such, zx has the ability to aggregate computing resources at storage, computing, and networking level
Self-load balancing and task distribution	Being peer-to-peer, in a multi-node "swarm", zx instances can automatically distribute tasks and workloads among themselves. There is no need to use an external cluster workload manager like SLURM, TORQUE etc.
Tri-level high-availability (HA)	HA at network interface, node, and connection levels
Intrinsically symmetric	Once two "sites" are connected, both sides can send & receive, even concurrently. Unlike the traditional client & server, which is unidirectional, demanding two applications per end
Data distribution pattern support	Point-to-point, hub-n-spoke, and tree & grid patterns support for modern distributed data-intensive enterprises, facilitating both internal and external collaborations
Multiple parallel transfer tasks	Multiple transfer tasks can be running at the same time (optionally at different priorities). QoS made easy
Built-in Web UI	Each zx comes with a built-in Web UI. Even new users can get up to speed in minutes
Simple yet powerful RESTful API	REST API enables complete control of an entire cluster of zx - integration of existing workflows made easy. API access supports multiple user via a pair of UNIX style username/password, with a basic role control, facilitating task delegations. Easy to meet the requirements from layers 8,9, and 10
Flexible management end-point	Due to the peer-to-peer nature, each cluster is actually a swarm. Thus, the endpoint for management can be any peer - HA for management as well
Check-point of each transfer	If a transfer is interrupted, it can be resumed later using the automatic checkpoint feature - minimize workflow interruptions
Transparent data compression	If a file/data object is sensed as compressible and zx is configured with transparent compression enabled, then the file/data object is compressed before transmitted. Uncompression is done on the receiving end automatically
Unconditional checksumming	For ensuring data integrity, zx employs unconditional checksum calculations for data integrity verification. Multiple hash algorithms supported
Data set version comparison and differential transfers	zx can compare different data set versions and only transfer the differences. No need to use an external content synchronization add-on
Per node or per cluster throttling	Data transfer rate can be throttled at the per node or per cluster level
Latency insensitive data transfers	The performance of zx over LAN, metropolitan, and WAN is about the same. Predictable workflows made easy
Single-site mode	Many data transfer operations can be done by a single "swarm" (aka site), eliminating the need to depend on the traditional client-server model
Object storage support	AWS S3 and true compatible cloud storage. zx is likely one of a very few software data movers that have integrated file and object storage support while preserving data attributes across storage



<http://zettar.com/>
info@zettar.com

Zettar Inc. delivers a break-through unified data mover zx for simplifying large-scale data management tasks involving data movement. With its unmatched simplicity, scalability, and efficiency, it helps distributed data-intensive businesses to defy data gravity and realize data mobility.

Zettar is a registered trademark of Zettar Inc. All other trademarks or service marks are the property of their respective owners. © 2020-2021 Zettar Inc. All rights reserved.