Rethinking Storage … Again

Analyst: Mike Kahn

Popular wisdom is often transient or cyclical. Was the old wisdom wrong or did something significantly change? Consider the *former* planet *Pluto*. It was a *planet* until 2006, when it was demoted to *dwarf planet* status. Did Pluto change significantly since it was discovered in 1930? (No.) Did we learn a lot more about Pluto in the ensuing years? (Yes.) In the last 80 years, did many (a majority) of astronomers change the way they think about planets? (Most definitely.) Is this the end of this story? (In no way.)

The same has been true for information technology. *Big is bad; small is good. Centralization is bad; distributed is good. Outsourcing is bad; cloud services are good.* **Part of the problem is that we tend to think in absolutes, typically bi-polar ones** – *good and evil*, *healthy and unhealthy*, *share and don't share* – I'm sure that you see what I mean. The reality is that while bi-polar thinking is dramatic (planet versus non-planet) and simpler (no shades of gray allowed), it may not be the best way to consider a range of possibilities. And like the poles of planet Earth, which have reversed many times[1], one's perspective can flip from one conclusion to its polar opposite in a single lifetime or career.

The other part of this problem is that *some things change***.** While the laws of physics (like the speed of light) may not have changed (much), practical implications may be evolving constantly and may have a significant effect on how we might think and analyze, the conclusions that we might draw, and the decisions that we might make. All of this is true about information technology.

The reality is that what was presumed to be false might now be presumed to be true, at least until something else changes. Those of us who have been around the computing industry for forty years can laugh about the physical connectors that first were used to tie a mainframe to its peripherals. (One of the big concerns was making sure that the weight of the connector and cable did not cause it to separate from the computer or cause the whole unit to tip over.) Twenty years ago, all thinking in the data center was limited to the no-more-than 2-meters of connectivity allowed between a server and its wide-cabled SCSI drives. No one could see beyond the physical realities of the present. However, things did change. Conventional wisdom did change. Architectures changed to maximize the new understandings and the revised *rules of the road*. Impossible-to-conceive products appeared and became the accepted norm. (Who doesn't want the ever-expanding functionality of an *iPhone* or specially-designed-for-your-DNA medical treatment?) Could we have conceived of either 10 years ago? For most of us, certainly not.

Having been in the computing industry for more than four decades provides some interesting perspectives. Time and time again, we have said and heard, "I could solve this challenge or problem if only I could operate faster, design smaller, manufacture more accurately, etc." **Time and time again, the physical barriers that were previously insurmountable have been broken. Each time, the rules of the road (i.e., the guidelines of how to operate, based on conventional wisdom) were challenged and replaced with better assumptions about reality and what was possible.** The younger generation of IT professionals may presume that the current rules of the road have been true forever, not recognizing that they were unimagined a decade or two ago and that they too will be supplanted, probably sooner than they expect.

The time has come to challenge some more of the sacred assumptions that we make about storage. Read on to learn which ones and why.

---

[1] For more on the Earth's reversing poles, see http://science.nasa.gov/science-news/science-at-nasa/2003/29dec_magneticfield/.

Tales of Yesteryear

Once upon a time, many our ancestors lived in small rural, communities from which they did not venture very far. They had neither the spare time to do so nor the motivation. Their lives were preoccupied with their day-to-day survival. Their plots of land, plantings, and the animals all needed tending and there always was too much to do. No matter how many sons a farmer had, there was always too much that needed tending. How densely to plant the crops, and what varieties and in what proportions, how to manage the scarce resources like water, and when to rotate the crops to make it all productive – these are the questions that continue to this day.

If I changed a few nouns and verbs, this could be a tale about storage administrators and their farms of data, with many important decisions to make, too much tending to do, and a lot to worry about. **While the changes to the business of storing, sorting, preserving, and reproducing data tend to come more quickly than progress did to agrarian practices, we still seem to have too little spare time to contemplate what might be and, for many, too little motivation to do so, at least until something comes along, rewrites the rules of daily existence, and kicks us reluctantly into the future.**

While some may see this as just the new and improved replacing the old and stale, it is more than the relentless cycle of products enabled by shrinking microprocessors and faster fiber optics, although they both are very important. Most of us in IT measure ourselves and the pace of time by the evolution of technologies. There is nothing wrong with that, but that tends to keep us looking for the short-term solution to the ever-present problems of last month, today, and soon thereafter. **What is needed is an occasional journey to the other side of the looking glass, where there is more time to think about what really might be going on; sort of a journey to examine the accepted precepts of our existence, as might be expressed in terms of** *the new established rules of the road.*

Self-Evident Truths

Among the greatest and most succinct thoughts expressed in the English language comes from the *Declaration of Independence*: *We hold these truths to be self-evident…* Just like the founding fathers of the United States who had specific truths in mind and listed them, we need to consider some self-evident truths and put them to the test of time. While the nature of humanity and morality may be more absolute and enduring, **we know that the IT world has changed and, at least, expanded the meaning of its self-evident truths.**

For example, we all know[2] that data should (or must) be near the applications or users that need to access it. The reasons have been self-evident: For example, our sense of reality confirms this self-evident truth – *distance affects access time*. Data residing in your laptop almost always can be accessed faster than data on an external drive or nearby storage device. If the data is residing across the country in a remote office or somewhere across the Internet, it almost always will take longer to access. This is self-evident. In a self-centered universe, one might demand that all important data be local. Of course, what is local to me may not be local to other users, so there are other factors involved, including

1. The value of the data,
2. The value of the timely retrieval of the data,
3. The importance of the application or user of the data,
4. The cost of storing, managing, protecting, and delivering the data, and
5. The impact of time on all of the above.

1. The Value of the Data

Some data is more valuable than other data, at least some of the time. For example, current medical records are more immediately valuable to the trauma team in the emergency room than they are to medical researchers, who operate on many data point retroactively. While each might deem the same data as important, one has a greater value. Similarly, while the stocking data of every item in a hardware store may be important, those items with a higher monetary value (say more than $50, like a power tool) clearly have a greater impact on the bottom line (than, say, nuts and bolts that cost pennies a-piece).

2. The Value of the Timely Retrieval of the Data

Time is money, but just some of the time. While we all are increasingly impatient, time significantly affects the value of information. If we get access to the data we need in a meaningful timeframe to use it to its full advantage, we are

---

[2] We "know" because (a) it has been taught or told to us, (b) we have observed it on one or more occasions to be true, and/or (c) it seems intuitively obvious to us.

satisfied. If it comes in after the point of decision-making, it value has been reduced significantly. Finding out that the patient has a history of drug allergies is critical before prescribing a new medication. Finding out when the pharmacy is asked to fill the prescription is better than not finding out until there is an allergic reaction, but this later discovery does not get the patient closer to being treated effectively.

3. The Importance of the Application or User of the Data

Let's face it, some applications and users are more important than others. It's not an egalitarian world. Pharmacies have long put calls from doctors' offices ahead of calls from their retail customers. Why? Because the doctors' and nurses' time is considered more valuable.

4. The Cost of Storing, Managing, Protecting, and Delivering the Data

Cost analysis is where most of us involved with storage infrastructure seem to spend a lot of our time. We fret over cutting unit costs, reducing duplicate data, picking the right tier of storage, and oh-so-much more. In recent years, the focus has gone from hardware costs to the costs associated with administration, software, and other operating costs (like energy and floor space).

5. The Impact of Time on All of the Above

Time is involved with each of the four factors just discussed. It's more than just about retrieval time, although that can be critically important. Time influences the value that we associate with data and how we use it. If we are paying for use of data by the number of minutes that we are interrogating a database, then we can see a direct correlation. If the charge is based on time of day or day of week, the direct correlation also is obvious.

Bottom Line on Self-Evident Truths

We tend to think about *what is possible* in terms of *what we know to be true*. **When something** comes along that changes the presumptions that support the self-evident truths, we quietly redefine what is self-evident.

New Possibilities

One of the most sacred assumptions in storage has been that *distance really matters*. While this sometimes is very true, much or most of the time it is not very relevant. There are two reasons for this.

1. We are more concerned about the costs than penalty of data delayed for small periods of time, and
2. We are limited in what we can do about it, i.e., we have learned to live with it in exchange for a lower total cost of storage.

Let's dissect this. First, let's exclude the high-performance requirements, either for transaction processing or large databases.[3] That still leaves the bulk of what is stored in the enterprise (but not on laptops, desktops, etc.). The vast majority of this is data in files (documents, spreadsheets, drawings, photos, etc.) but some block-based data also fits in here, such as archived email and data repositories.

There is one "pretty-good" assumption that can be made about this data. For almost all of it, the longer it goes without being used the less likely it is to be accessed again. Now, there are exceptions to this generalization, especially data that is used seasonally, like end of month or quarter reporting that compares to the same period a year earlier or when a patient returns to the same doctor for a follow-up visit after, say, six months or a year. Many of these exceptions tend to predictable, allowing time to stage the data "nearby", if that is what is required.

However, the truth of this generalization has a lot to do with economics. Otherwise, we would have kept all of our data on the highest-performing arrays and keep the data replicated at multiple locations. However, we (those of us in IT) are fundamentally cheap, which means that if we can save a lot by deploying tiering, outsourcing, or other architectural frameworks (without too many negatives), we probably will. In IT, like at home for most of us, we don't want to pay for premium levels of service that we don't really need or might not be able to afford. We also have important and evolving requirements – like security and privacy – that move us away from the least costly, bottom-of-the-barrel solutions.

So here we sit, torn between delivering high qualities of service locally and tossing our data over the fence to someone, possibly far away, who will do it for less. Moreover, we now know that the more we subdivide our data into tiers of service, the greater the cost of managing the

---

[3] This restriction is done more to get you to think about the problem without getting bogged down by the details of high-performance and large scale. In many ways, this is an artificial bounding of the problem, but often it is better to first think about the issues and challenges to our assumptions on a smaller scale or in a simpler case.

whole thing.

Maybe what we need is a change in thinking. We need to remove, as much as possible, consideration of the physical aspects of where we store our data. We need to remove the human decision-maker from the process of deciding where data is placed. While storage virtualization did some of this, the storage administrator is still much too intimately involved with the physical storage. S/he just uses "storage virtualization" tools and appliances to reduce the administrative burden, but this is only a stopgap measure. **For all but the most time-sensitive data, we need to give up the good fight for hands-on involvement and turn it over to policy-driven automated software.**

Just as we are beginning to do with applications via server virtualization, we need to cut the cords of dependency on human-handled local placement of blocks and files on physical storage servers. If the application or operating system thinks it needs to know where something is actually stored, give them a virtual location and be done with it. We can't afford to manage the petabytes of stored data in files and blocks the way that we have been doing. **It may sound like a deal with the devil but if you can let someone else (an outsourcer) or something else (automation) take control of 70 to 90% of your less urgent data at a significant savings in the total cost of ownership, make the deal.**

How would this new kind of storage virtualization work? First, it has to be intelligent. It needs to know when something has gone from *nearly dormant* (i.e., rarely used) to *presently hot* (i.e., now in demand), and vice versa. Second, it needs to know about the *capabilities and economics* (including energy consumption) of the storage resources available to it. Third, it must be able to *act transparently* as it transitions data from one storage tier to another and one device to another. Lastly, it must be *policy driven*; new rules of the road must be written and be modifiable. Thus, this does not have to be some static appliance that optimizes to its manufacturer's predetermined priorities and algorithms, although for some this may be all that is required.

In the largest sense, it would be the brains of enterprise storage, distributed as needed to provide protection and performance and always in a state of self-adjustment, based on the requirements for data and the assets underneath. It would replace much of the intelligence that we typically associate with storage arrays. Over

time, the underlying storage could be dumbed down because the specific (and many variations of) local (special-purpose) intelligence would no longer be needed. Provisioning would be simplified, backup and recovery would be done differently, replacement of storage would become totally transparent, etc. It would be a different world, certainly one with different administrative challenges. Can you see this brave new world? I can, and it's going to be very exciting.

But Wait, There is More!

In our old world model, data – when stored as files or blocks or even objects – was truly data of the enterprise sort, such as databases, documents or email. Storage administrators were responsible for the placement, care, and maintenance of this valuable enterprise resource. However, there is more!

We now have a new kind of data, let's call it a *virtual instance* (of an operating system), of which there may be hundreds or thousands at any moment in time. While servers are involved, the fundamental reality is that these virtual instances exist primarily in memory and storage. This is what makes them movable. The storage aspect is what now makes them of interest to storage administrators. You can't move an in-play application (now running as a virtual instance), especially over a long distance, without being able to move, with data integrity and network continuity, the whole virtual container around it. Managing storage cleverly makes virtualized application mobility really possible.

If you thought that storage administrators were overwhelmed before, then you need to prepare for the screaming that is about to happen as wide-scale enterprise application execution over geographic distances becomes the most important and visible part of the server virtualization quality-of-service equation. If you try to deal with the equation's variables that supposedly are affected by the distance of data from where it is being used by traditional means (especially in real time), be prepared for a turbulent journey. To survive, you probably need to rethink some of the historical assumptions that you have presumed to be true.

The New Bottom Line

If the last ten years have been about squeezing the most our of your storage budget, the next ten years likely will be about trying to keep up with all of the demands that will be placed on storage, especially the ones involving distance. The old economic challenges will not go away,

nor will the storage administration challenges that have kept us up most nights on rollerblades, as we zipped from one most-urgent chore (or crisis) to the next. If you haven't found a way to automate the old problems, you won't have any time to figure out how you are going to deal with and automate the new demands, especially those related to server virtualization and distance.

　　As in most *Captain's Logs*, the focus in this one has been on understanding the situation, what has changed or is changing, and what you need to think about, usually from outside of the box in which you tend to operate comfortably. **There is no one solution to be recommended here, just a recognition that requirements for storage and its management and administration are escalating rapidly and becoming even more critical to the success of the enterprise. The days of managing this by spreadsheet and true grit are over. Policy-driven automation and resource balancing are the only ways to make it through the high seas, strong winds, and changing currents.**

　　<u>*You*</u> *had better get ready!*

About The Clipper Group, Inc.

The Clipper Group, Inc., is an independent consulting firm specializing in acquisition decisions and strategic advice regarding complex, enterprise-class information technologies. Our team of industry professionals averages more than 25 years of real-world experience. A team of staff consultants augments our capabilities, with significant experience across a broad spectrum of applications and environments.

➤ *The Clipper Group can be reached at 781-235-0085 and found on the web at www.clipper.com.*

About the Author

Mike Kahn is Managing Director and a cofounder of The Clipper Group. Mr. Kahn is a veteran of the computer industry, having spent more than four decades working on information technology, spending the last 17 years at Clipper. For the vendor community, Mr. Kahn specializes on strategic marketing issues, especially for new and costly technologies and services, competitive analysis, and sales support. For the end-user community, he focuses on mission-critical information management decisions. Prior positions held by Mr. Kahn include: at International Data Corporation - Director of the Competitive Resource Center, Director of Consulting for the Software Research Group, and Director of the Systems Integration Program; at Power Factor Corporation, a Boston-based electronics startup – President; at Honeywell Bull - Director of International Marketing and Support; at Honeywell Information Systems - Director of Marketing and Director of Strategy, Technology and Research; at Arthur D. Little, Inc. - a consultant specializing in database management systems and information resource management; and, at Intel Corporation, Mr. Kahn served in a variety of field and home office marketing management positions. Earlier, he founded and managed PRISM Associates of Ann Arbor, Michigan, a systems consulting firm specializing in data management products and applications. Mr. Kahn also managed a relational DBMS development group at The University of Michigan, where he earned B.S.E. and M.S.E. degrees in industrial engineering.

➤ *Reach Mike Kahn via e-mail at Mike.Kahn@clipper.com or via phone at (781) 235-0085 Ext. 121. (Please dial "121" when you hear the automated attendant.)*

Regarding Trademarks and Service Marks

The Clipper Group Navigator, **The Clipper Group Explorer**, **The Clipper Group Observer**, **The Clipper Group** *Captain's Log*, **The Clipper Group Voyager**, *Clipper Notes*, and *"clipper.com"* are trademarks of The Clipper Group, Inc., and the clipper ship drawings, *"Navigating Information Technology Horizons"*, and *"teraproductivity"* are service marks of The Clipper Group, Inc. The Clipper Group, Inc., reserves all rights regarding its trademarks and service marks. All other trademarks, etc., belong to their respective owners.

Disclosures

Regarding the Information in this Issue

The Clipper Group believes the information included in this report to be accurate. Data has been received from a variety of sources, which we believe to be reliable, including manufacturers, distributors, or users of the products discussed herein. The Clipper Group, Inc., cannot be held responsible for any consequential damages resulting from the application of information or opinions contained in this report.