Data Centre and Virtualization

Re-Thinking the Storage Infrastructure for Business Intelligence
Eric Burgener for Infinidat

Eric Burgener
Research Vice President, Infrastructure Systems, Platforms and Technologies, IDC

Biography

Eric Burgener is Research Vice President within IDC's Enterprise Infrastructure Practice. Eric's core research coverage includes Storage Systems, Software and Solutions, quarterly trackers, end-user research as well as advisory services and consulting programs.

Based on his background coverage enterprise storage, Eric's research includes a particular emphasis on flash-optimized arrays, emerging persistent memory technologies, and software-defined storage. He is an active participant in the IT Buyers Research Program at IDC and blogs throughout the year on the topic of Infrastructure and Data Management.

Eric is a regular guest author for Infinidat (https://www.infinidat.com)

Abstract
With digital transformation under way at most enterprises, IT management is pondering how to optimize storage infrastructure to best support the new big data analytics focus. In an article commissioned by Infinidat, Eric Burgener, Research Vice President, Infrastructure Systems, Platforms and Technologies at IDC discusses how successful digital transformation can move organizations towards much more data-centric business models where those that can best drive value for their customers and their own companies out of the data they collect will be the winners.

Introduction
The next generation applications being deployed to leverage data to drive better business decisions are very different from the batch-oriented business analytics of the past – today's analytics depend on artificial intelligence (AI) and machine learning (ML), are mission-critical applications that have to be highly available, often have a real-time imperative, and are more compute- and data-hungry than ever before.

When paired with the accelerated compute so often deployed with AI/ML-driven workloads, traditional storage architectures are severely challenged to keep the processors fed with data and operating efficiently – let alone meeting response time requirements.

Primary research done by IDC in 2020 reveals that today's enterprises, driven by the requirements of digital transformation, are modernizing their IT infrastructure at a rapid rate. Over the next two years, almost 70% of these organizations will be performing a technology refresh on their server, storage, and/or data protection infrastructure to better align their IT and data-centric business strategies.

With respect to modernizing storage infrastructure, enterprises are looking for systems that can deliver both low latencies and high throughput across the different AI/ML-driven big data analytics workloads that require a mix of metadata intensive operations and random and sequential access across both small and large files, coupled with an ability to scale easily into the multi-petabyte range without putting performance at risk – all while providing the type of high availability that mission-critical applications that are driving day-to-day business operations require.

Separate systems have been required in the past to meet the performance and cost-effective massive capacity requirements, leading to data placement strategies that emphasize a smaller, more performant front-end system (tier) and a much

larger multi-petabyte back end system (tier) whose cost structure is optimized for massive capacity, not performance. Data placement strategies fetch active data being used onto the performance tier but strive to keep the less active data on a separate, massively scalable tier that exhibits a much lower $/GB cost – an archive tier.

While this approach may have worked when business analytics were a more batch-oriented operation, newer analytics workloads need rapid access to more data that can be kept in the traditional performance tiers and moving data between tiers lengthens the time to better, more informed decisions.

Newer storage architectures open up the opportunity to re-think traditional approaches to analytics, particularly if the "performance" and "archive" tiers can be cost-effectively combined onto a single platform without giving up either performance, high availability, or multi-petabyte scalability. This scenario is attractive: applications can directly access a much larger amount of data (in fact, in

many cases the entire data lake, depending on its size) without introducing the latencies or the complexities of data movement between two different storage systems. For many of the AI/ML-driven workloads, leveraging more data for analysis drives better insights and decisions, and those applications that have a real-time component benefit from rapid access to the data.

As a system architect, how would you design a storage infrastructure to meet these requirements in a single storage platform? Here are some of the key things you would look for:

- A system that can deliver consistent sub millisecond latencies across consolidated AI/ML-driven business intelligence workloads at multi-petabyte scale;

- An extremely resilient architecture that leverages redundancies and transparent recovery both within a single system and across multi-system configurations to meet the need for disaster recovery;

- A system that supports very high performance host interconnects like FC and NVMe over Fabrics so that the performance capabilities of the system are not frittered away with high storage network latencies.

These requirements may seem simple, but in practice most vendors have been unable to meet them, leaving customers to operate with separate systems to handle the performance and archive requirements. Meeting the performance requirement requires innovations in several areas: highly scalable lock management, an ability to rapidly reference and access any data in the system using trie data structures, and intelligent data placement algorithms that dynamically adapt to workload changes to keep accelerated (or general-purpose) compute fed, coupled with a tiered storage approach within a single system.
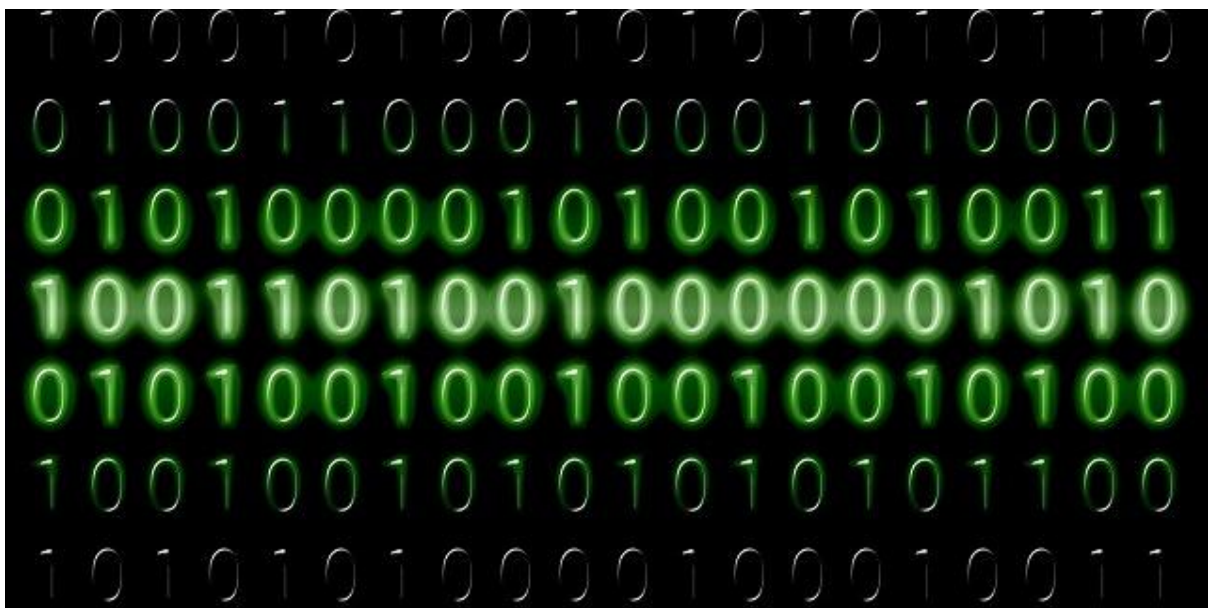
Meeting the massive scalability requirements must include not only the capacity considerations, but a way to be able to directly access all of the storage from any of the controllers in the system (which will have to be redundant to meet the high availability requirements) without having to access storage external to the system. Support for cloud tiering will enhance the value proposition for the platform, but the platform must be able to cost-effectively support petabytes of data in its own right to keep the larger data sets required to drive better business decisions close at hand.

If you don't have access to such a system, you may have to stay with the more traditional archiving model, managing and maintaining separate platforms. But with this type of consolidation for next generation business intelligence workloads, there is not only an ability to drive better business decisions faster – there is an economic benefit from not having to manage a separate archive tier.
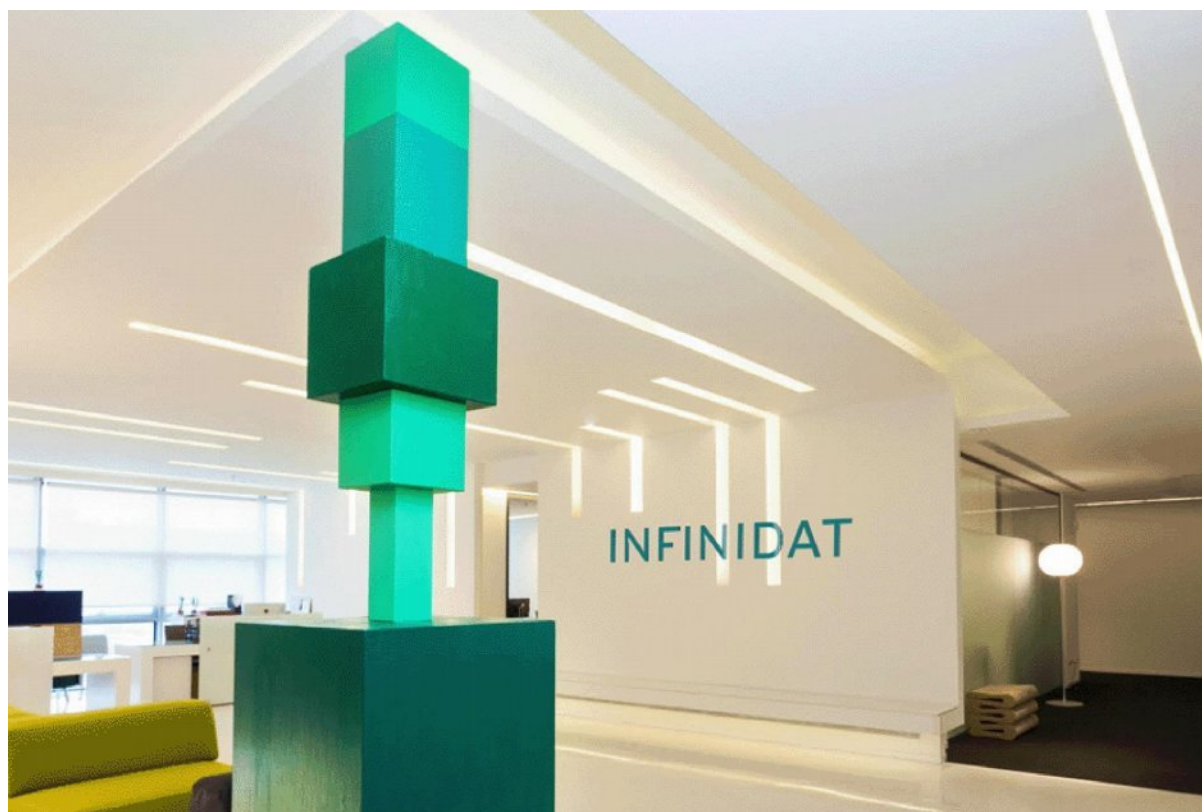
You'll find these capabilities in InfiniBox, a high performance, massively-scalable, enterprise-class storage platform from Infinidat that delivers better than all-flash performance for mixed workloads at $/GB costs dominated by its HDD-based capacity tier, supports over 5 petabytes of effective storage capacity in a single data center rack (assuming its in-line compression), comes with a 100% data availability guarantee which covers both single site and multi-site configurations, and offers both FC and NVMe over Fabrics host connection options.