**Title:** Host Managed Storage Solutions for Big Data.

The performance gap between Compute and Storage is fairly considerable. With multi-core computing capabilities, CPUs have scaled with the proliferation of Big Data, but storage still remains the bottleneck. The physical media characteristics are mostly blamed for storage being slow, but this is partially *true*. The full potential of storage devices cannot be harnessed till all layers of I/O hierarchy function efficiently. Despite advanced optimizations applied across various layers along the odyssey of data access, the I/O stack still remains volatile. The problems associated due to the inefficiencies in data management get amplified in multi-tasking Big Data shared resource environments.

***Our effort is to deliver near-ideal performance of storage systems, by identifying issues, designing, and, developing software defined storage capabilities with minimal or no infrastructural change for Data Centers experiencing Big Data.*** Thereby, making changes feasible. Neither do we intend to change application characteristics nor improve storage devices or network infrastructures, but only the way data is managed. Therefore, this research aims to improve the layers along the odyssey of data access environment by understanding the I/O hierarchy and the application needs from storage.

Our contributions have been in three major fields, i.e.,

1) *Operating System optimizations*, dealing with optimizing the OS and extending its competency. We develop solutions, BID-HDD, from the core of the operating system, i.e. a block I/O scheduling scheme to avoid contentions and improve individual storage device (Hard Disk Drives HDDs) capabilities.;

2) *Multi-tier solutions*, which focuses on systems design to incorporate heterogeneous tiers of storage together coupled with value propositions of data being scattered over multiple devices. We manage multiple devices and develop methodologies, BID-Hybrid, to automated tiering using the information obtained at the block interface using SSDs (Solid State Drives) for improving disk performance.;

3) *Workload specific optimizations*, are full-stack data center storage solutions designed and developed to suit workload characteristics. We design and develop methods, LDM- our data management solution, for the complete data center ecosystem using multiple tiers of storage for mitigating the impact of data-dependency in lineage class of applications. LDM amalgamates the information from all the stratas, devices and layers of I/O path.

With theoretical and experimental evaluations, our host managed storage solutions, namely, BID-HDD, BID-Hybrid, and LDM, fulfils our goal of narrowing the gap between what storage is capable of delivering and what it actually delivers in a Big Data environment.

Our research would aid Data Centers to achieve their Service Level Agreements (SLAs) as well as keeping Total-Cost of Ownership (TCO) low. From the Green Computing perspective, our solutions will decrease energy footprint, due to much reduced work to process data across all tiers of computing, i.e. storage, compute (required on storage servers), and network.