Summary Paragraph:

Lead by [Jin Li](http://web.archive.org/web/20161107060225/http:/research.microsoft.com/~jinl), Cloud Computing and Storage (CCS) group focuses on Cloud Infrastructure and Algorithm research. They believe that the ultimate milestone of cool system research is a product of significant impact. In addition to pursuing original research and publishing papers in premier venues, they devote time to work with Microsoft product groups to transfer the research into products.

CCS’s recent projects are the [Deep Learning workspace (DL Workspace)](https://microsoft.github.io/DLWorkspace/), and the SoftFlash. DL Workspace is an open sourced toolkit that empower AI scientists to quickly spin up a cloud AI infrastructure (either in public cloud, such as Azure, or in an on-perm cluster ) to manage AI training, interactive exploration, inference, and analytics. DL Workspace supports all major Deep Learning toolkit out of box (e.g., TensorFlow, CNTK, Caffe, MxNet, etc..). It also supports big data analytics works such as Hadoop/Spark as well. It is open sourced, and used in daily production by multiple Microsoft teams (Microsoft Cognitive Services, SwiftKeys, Bing Relevance, etc..). For more information, please take a look at this [video](https://www.youtube.com/watch?v=Xa7exVurUmE&feature=youtu.be).

CCS has assisted Azure to architect and implement the [local reconstruction code (LRC)](https://www.microsoft.com/en-us/research/blog/better-way-store-data/) used in Windows Azure Storage. This is a new family of erasure codes that provide significant reduction in storage overhead and cut down the minimum number of fragments that need to be read to reconstruct a data fragment. It leads to hundreds of millions dollars of savings for Microsoft, a [Best Paper Award](https://www.microsoft.com/en-us/research/publication/erasure-coding-windows-azure-storage/) at USENIX ATC 2012 and a 2013 Microsoft Technical Community Network Storage Technical Achievement Award. CCS has also architected the erasure code used in [Storage Spaces](https://www.microsoft.com/en-us/research/uploads/prod/2016/11/LRC-Erasure-Coding-in-Windows-Storage-Spaces.pdf) in Windows 8.1 and Windows Server 2012 R2. Also, it has architected and implemented the erasure coding used in Lync, Xbox and RemoteFX.

CCS has assisted Windows File Server group to architect and implement the [Primary Data Deduplication](https://www.microsoft.com/en-us/research/blog/eliminating-duplicated-primary-data/) feature in this [Windows Server 2012 Paper](https://www.microsoft.com/en-us/research/publication/primary-data-deduplication-large-scale-study-system-design/) and End-to-End Deduplication for Storage Virtualization in Windows Server 2012 R2. Key contributions include a new data chunking algorithm, a low RAM footprint indexing data structure to detect duplicate data (based on [ChunkStash](https://www.microsoft.com/en-us/research/publication/chunkstash-speeding-up-inline-storage-deduplication-using-flash-memory/)), and a data partitioning and reconciliation technique, the latter two for scaling index resource usage with data size. It leads to major saving to customers (20-82%), and is among top 3 features for Windows File Server introduced at Windows Server 2012. It has received rave reviews ( and there are evidence that some customers upgrading to WIndows Server 2012 for the primary data deduplication feature only).

CCS has exploited the benefits of Solid State Drive (SSD) for storage applications. “[FlashStore](https://www.microsoft.com/en-us/research/publication/flashstore-high-throughput-persistent-key-value-store/)” has implemented a SSD optimized, low RAM footprint key-value store that organizes storage on flash in a log-structured manner. It was tech-transferred to Bing backend. [SkimpyStash](https://www.microsoft.com/en-us/research/publication/skimpystash-ram-space-skimpy-key-value-store-flash-based-storage/) has implemented an ultra-low RAM footprint key-value store. The storage design of SkimpyStash has been incorporated into [BW-Tree](https://www.microsoft.com/en-us/research/publication/the-bw-tree-a-b-tree-for-new-hardware/), a joint project among CCS, [MSR Database group](https://www.microsoft.com/en-us/research/group/database/), and Azure DocumentDB team. BW-tree has been shipped to SQL Server 2014 [(Hekaton)](https://www.microsoft.com/en-us/research/blog/hekaton-breaks-through/) and Azure DocumentDB.

Partner with the Remote Desktop Virtualization (RDV) team, CCS has also assisted to archiect and implement the [RemoteFX for WAN](https://cloudblogs.microsoft.com/enterprisemobility/2012/08/23/remotefx-for-wan-overview-of-intelligent-and-adaptive-transports-in-windows-8-and-windows-server-2012/) feature in Windows 8 and Windows Server 2012, which provides fast and fluid user experience in a remote session running over any WAN and wireless networks [[Paper]](https://www.microsoft.com/en-us/research/publication/optimizing-fec-transmission-strategy-for-minimizing-delay-in-lossless-sequential-streaming-2/).