Seagate's Active Drive Technology

Jon Trantham, Principal Technologist, Seagate

Seagate Technology Corporation

March 1st, 2017

Active Drive Technology Background

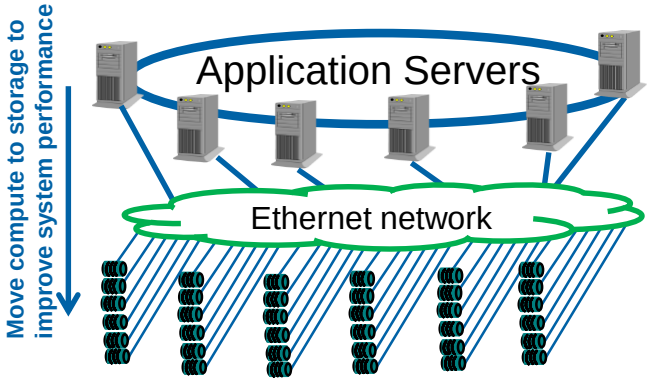
What is it:
Provides the user the ability to run small computer programs (applets) in a disk drive

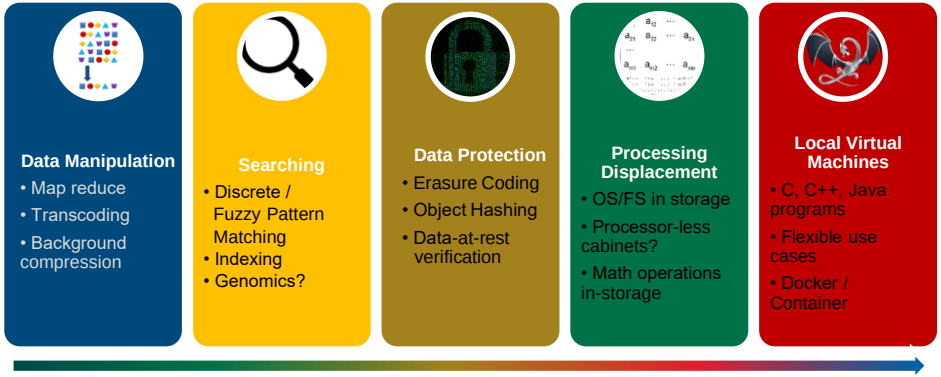
Target Apps:
Customers with large homogeneous datasets (Big Data and High Performance Computing)
Low-cost data-lake systems

Benefits:
Scalable Parallelism - Compute scales with storage
Reduces Network traffic (e.g. 1/1000X)
Eliminates storage controllers
Eliminates metadata storage tier

Technology:
Kinetic drives & chassis
Enhanced Kinetic command set
Compile applets on host machines
Provides execution within drive's ARM μp cores
Send applets to containers and execute locally

Key Issues:
Rate of Kinetic adoption
Programming Skills required





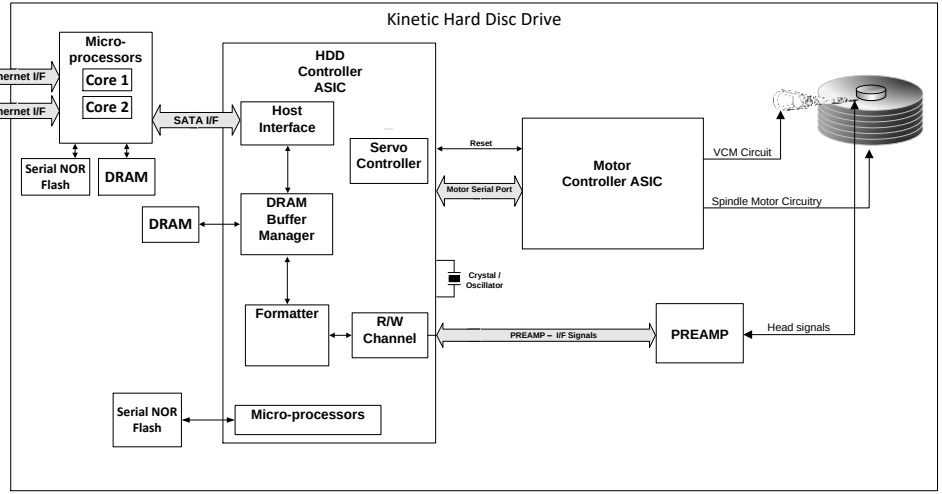
AD Hardware Architecture

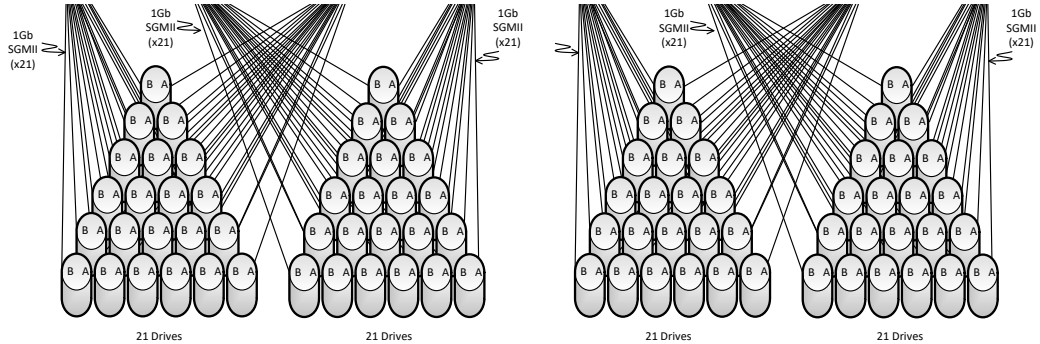
Standard Kinetic Drives in a Kinetic Chassis

- Currently uses separate ARM cores and DRAM
- Hardware is added to traditional HDD PCBA
- Enclosures use low-cost Ethernet switches
- Can be done in a high-availability w/redundancy
- No external RAID controllers or processors!

Questions: Market Applicability

- This architecture works well for traditional M/R
- How sufficient is it for other workloads?
- E.g. will it hold for genomics processing?





Value of Active Drive

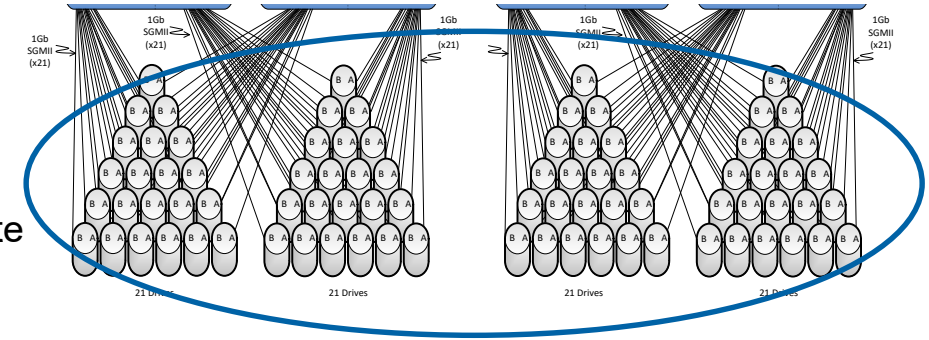
Benefits of Active Drive:

Parallelism:
Applications can run in parallel on many or all drives
Target: performance approaching storage media's data rate
Challenge: distributing data uniformly to storage

Networking:
Significant data traffic reduction (1000X) on network

Processing Requirements:
Computation is performed locally in drive's ARM cores

Cost Reductions:
Kinetic & Active Drive can:
- eliminate metadata tiers
- reduce external processing
- reduce networking bandwidth requirements to the size of the incoming storage data rate



The goal: processing locally in parallel

Goals & Objectives for Seagate-CERN Potential Engagement

SEAGATE

Active Drive Market Applicability

Our Goal: To learn which markeets AD is applicable: e.g. Genomics?

- Is our technology applicable to the genomics market as-is or with augmentation?

- We want to learn more about Genomics storage and processing
  - Data gathering, post-processing, and storage
  - Data formats used (metadata and actual genomic data)
  - Compression
  - Sequentiality of data accesses
  - Percentage Reads vs. Writes
  - Processing used against stored (post-sequenced?) datasets

Key questions:

- Are ARM processors sufficient for any genomics processing tasks?
- Can we embed hardware accelerators in our controller ASIC to assist?
- How uniform are genomics storage and processing across science and industry?

Benefits to CERN

Opportunities for CERN:

- **to be the first to study this topic on an emerging technology**
  - Like the NASA engagement, we will want for you to publish on your findings
- **to shape the future direction of storage hardware**
  - We are looking for any low-cost hardware accelerators that can be added to our ASICs
- **to benefit science**
  - If this technology works, the benefit will be faster, cheaper genomics processing
- **direct access and support from our engineering team**