facebook

Facebook Storage Tiers

How the fleet is divided:

- Facebook runs three main storage tiers:
  - Type III – MySQL Database (user data)
  - Type IV – Hadoop (site activity analytics and logs, messages)
  - Type V – Haystack (Photo, video, email attachments)
- All storage tiers are currently based on standard 2U12HDD OEM solutions.

(Place image of FS-12TY or DL165Gwhatever in here)

Type III Database

The Foundation

- MySQL running InnoDB engine.
- Main system bottleneck is HDD IOPS.
- FusionIO PCI-E Flash card provides secondary cache to reduce IOPS.
- DB suffers from moving “hotspot”, where a small number of machines get queried intensely compared to overall site activity.

Type III System Components	
CPU	Dual Intel Xeon L5630
RAM	144GB
Storage	8x1TB SATA Enterprise HDD in RAID10 array
Flash	720GB Fusion IO
NIC	1Gb
Chassis	2U12HDD Redundant PSU

Type IV Hadoop

The Warehouse

- Data is stored on HDFS, a distributed file system, servers are used for intense data crunching computation as well as logging large data sets used for the analysis.
- Facebook deployment of Hadoop has 2.2x redundancy. No RAID.
- No explicit system bottlenecks. Balance is maintained between CPU and HDD usage.

Type IV System Components	
CPU	Dual Intel Xeon X5650
RAM	48GB
Storage	12x2TB SATA Enterprise HDD
NIC	1Gb
Chassis	2U12HDD Redundant PSU

Type V Haystack

The Vault

- Custom built file system. Large files made from thousands of appended photos with indices stored in RAM, guaranteeing single disk IOP for any photo retrieval.
- Main concerns are HDD rebuild and full system recovery time. (Rebuilds approaching 1 day, full system recovery 1 week).

Type IV System Components	
CPU	Intel X5630
RAM	16GB
Storage	12x2TB SATA Enterprise HDD in RAID6 array
NIC	1Gb
Chassis	2U12HDD Redundant PSU

What is Knox?

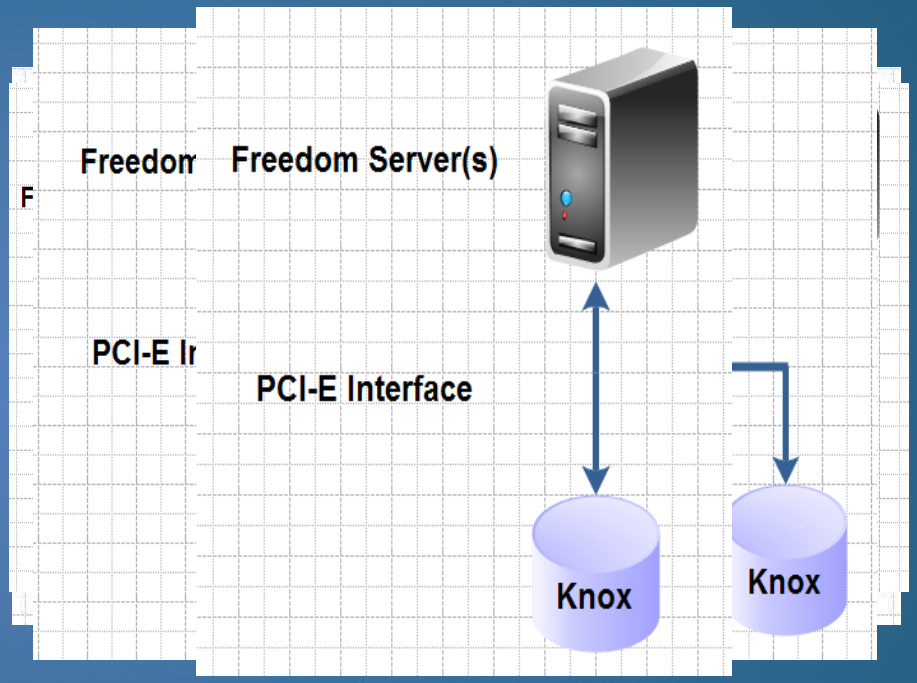
Project Knox goals:

- Design an optimized storage server for Facebook's storage heavy applications :
  - Haystack, Video, Titan and Hadoop
- Leverage already existing Open Compute Server, Power Supply and Rack solutions.
- Create a design that will fit the varying requirements of the different applications.
  - Ensure an assembly, integration and maintenance friendly design.
  - Accommodate potential change in fleet or system architecture.

Introducing Knox

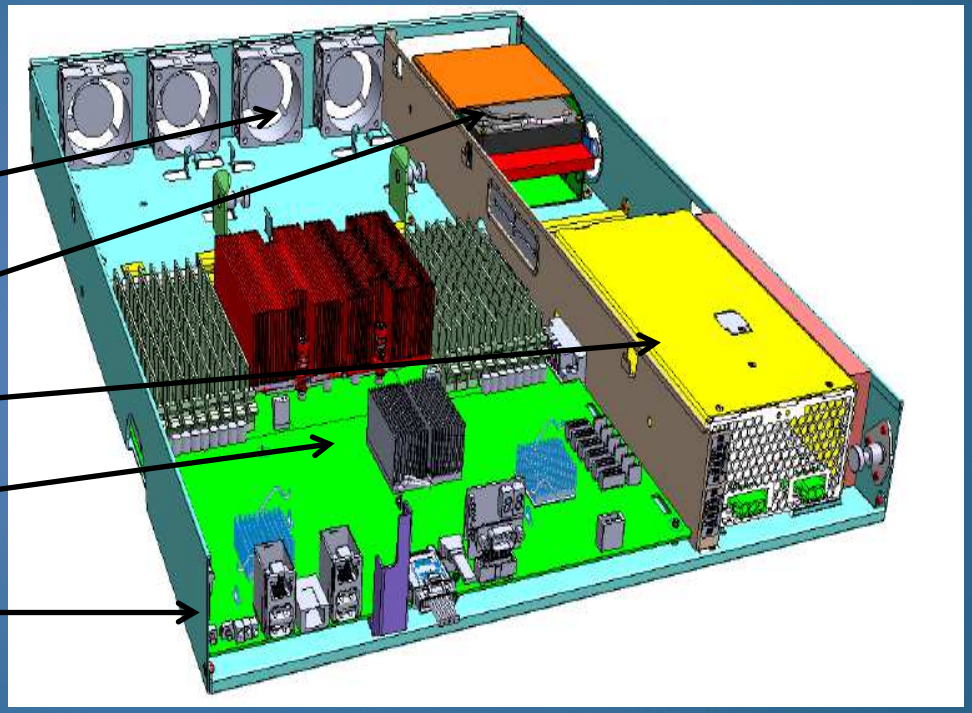
A platform approach to storage

- Knox enables building storage servers with varying compute to storage ratios using the same physical building block.
- PCI-E interface is used to interconnect the compute unit to the storage unit.
  - PCI-E scalability allows adding either more compute or storage nodes independently of each other.



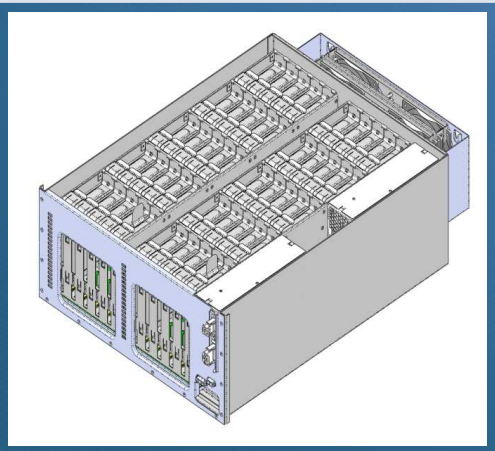
Open Compute System Components

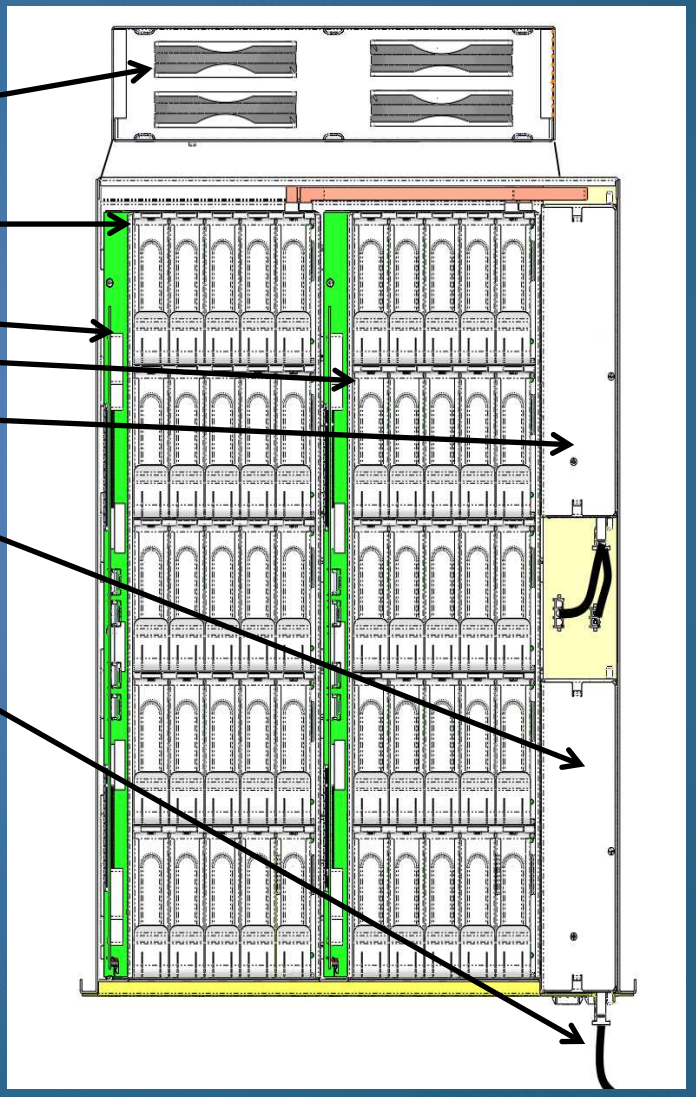
Open Compute System Components
4 system fans + air duct
6HDD slots
OC PSU 277 VAC w/ 48V DC backup
Open Compute Intel motherboard
1.5U open chassis



Knox System Components

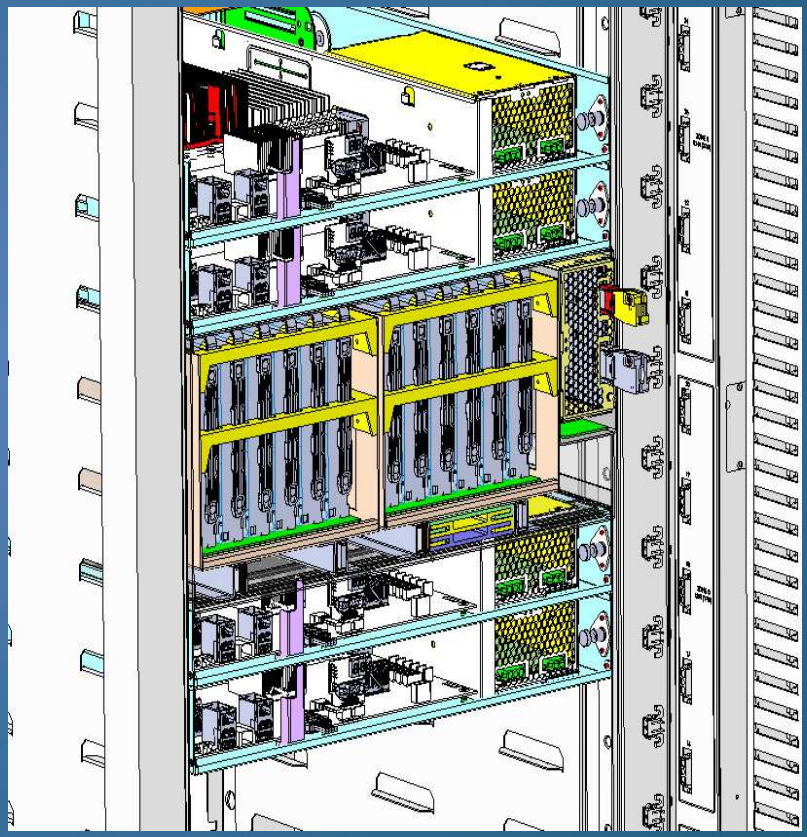
Knox System Components
4X 120mm system fans
Dual 25 hot-plug HDD cabinets
Dual 2xRAID+Expander controller cards
Freedom PSU 277 VAC w/ 48V DC backup (Redundant power supply support)
12V/5V (20A each) power connector & cable
4.5U open chassis





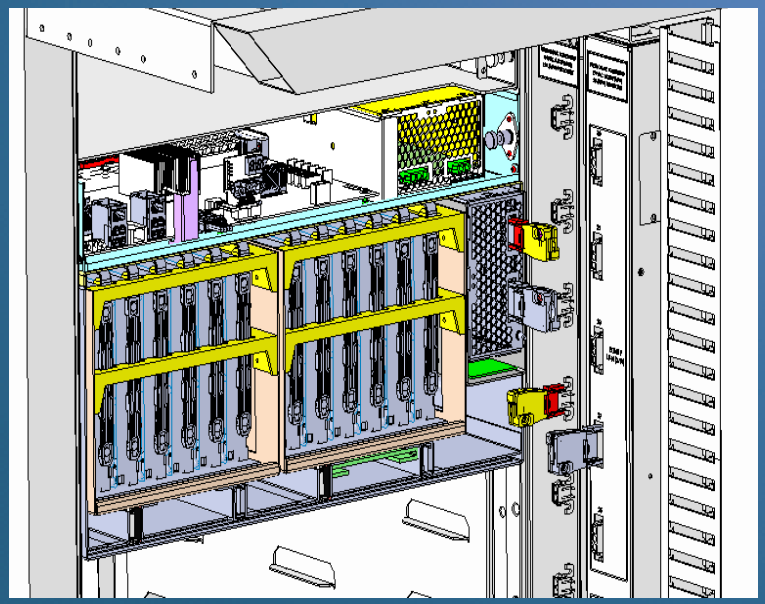
Freedom Knox Type IV Rack Config

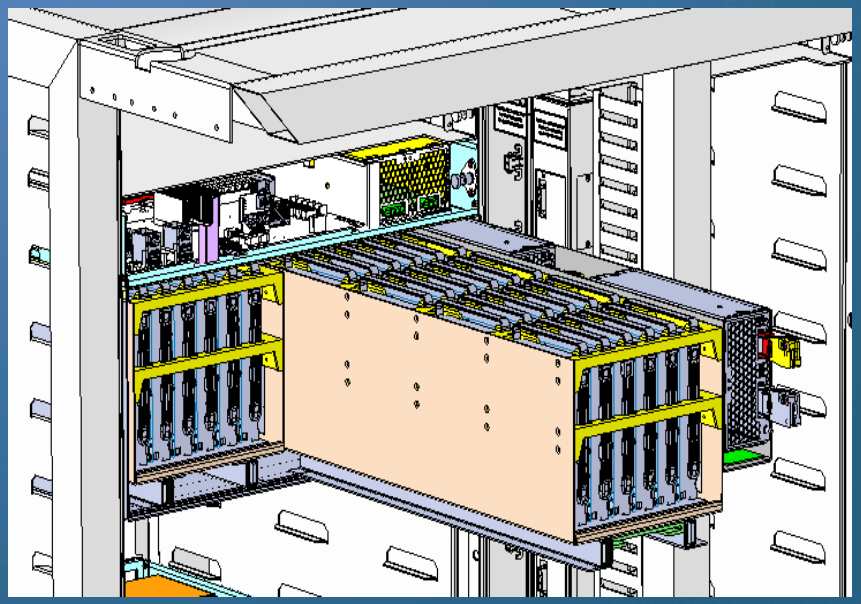
Type IV Rack Configuration	
Unit Height	7.5U (2*1.5U OC + 4.5U Knox)
Units per column:	5
Compute nodes per column:	20
HDDs per column:	250



Freedom Knox Type V Rack Config

Type V Rack Configuration	
Unit Height	6U (1.5U OC + 4.5U Knox)
Units per column:	7
Compute nodes per column:	7
HDDs per column:	350





Knox Architectural Advantages

- Enables designing an optimized storage server configuration with minimal new hardware development.
  - Leverages already existing Open Compute server and its supporting infrastructure.
- Compute to storage ratios aren't fixed - they can be adjusted during integration or after deployment for optimal hardware utilization.
  - The storage building blocks are identical across various tiers, only the number of units per compute node, HDD capacity and the Knox storage controller board stuffing option vary.
  - Enables redeployment of compute (OC server) or storage (Knox RBOD) nodes to other applications after production.

Knox Architectural Advantages

- Enables future design upgrades to be limited in scope to either compute or storage only.
  - A lot of activity surrounding low power processors for the server market is taking place, and may have a transformative effect on the industry. Adoption for storage alone may save millions in server component cost and annual electricity usage.
- First step in the direction of a “Universal Rack”.
  - Capacity planning team desires to develop an infrastructure that will allow adjustment of compute to storage ratios using standard server and storage hardware blocks. Technology to enable this over PCI-E is not mature, but Knox enables a first step in transitioning towards that solution.

Knox Architectural Advantages

- Multiple benefits to the consolidation of more HDDs behind fewer compute nodes.
  - Hotspot effect in the database tier should be mitigated thanks to larger amount of max IOPs available per system.
  - Database and Haystack tiers aren't CPU bound. Reduce CapEx by reducing the number of compute nodes per TB of storage.

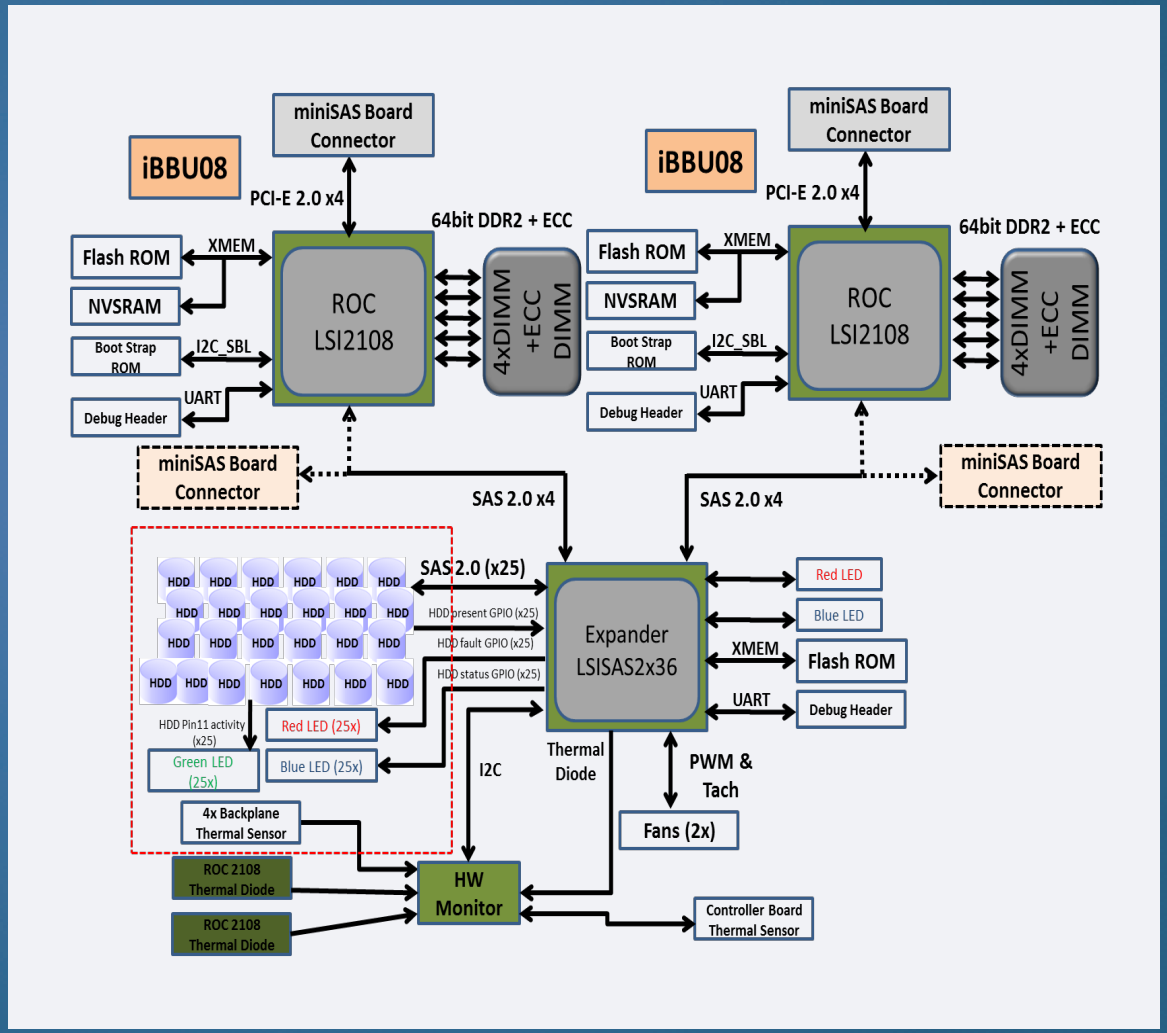
Knox Controller Board Architecture

Knox controller board can be configured with different component stuffing options enabling support for multiple configurations using a single PCB solution.

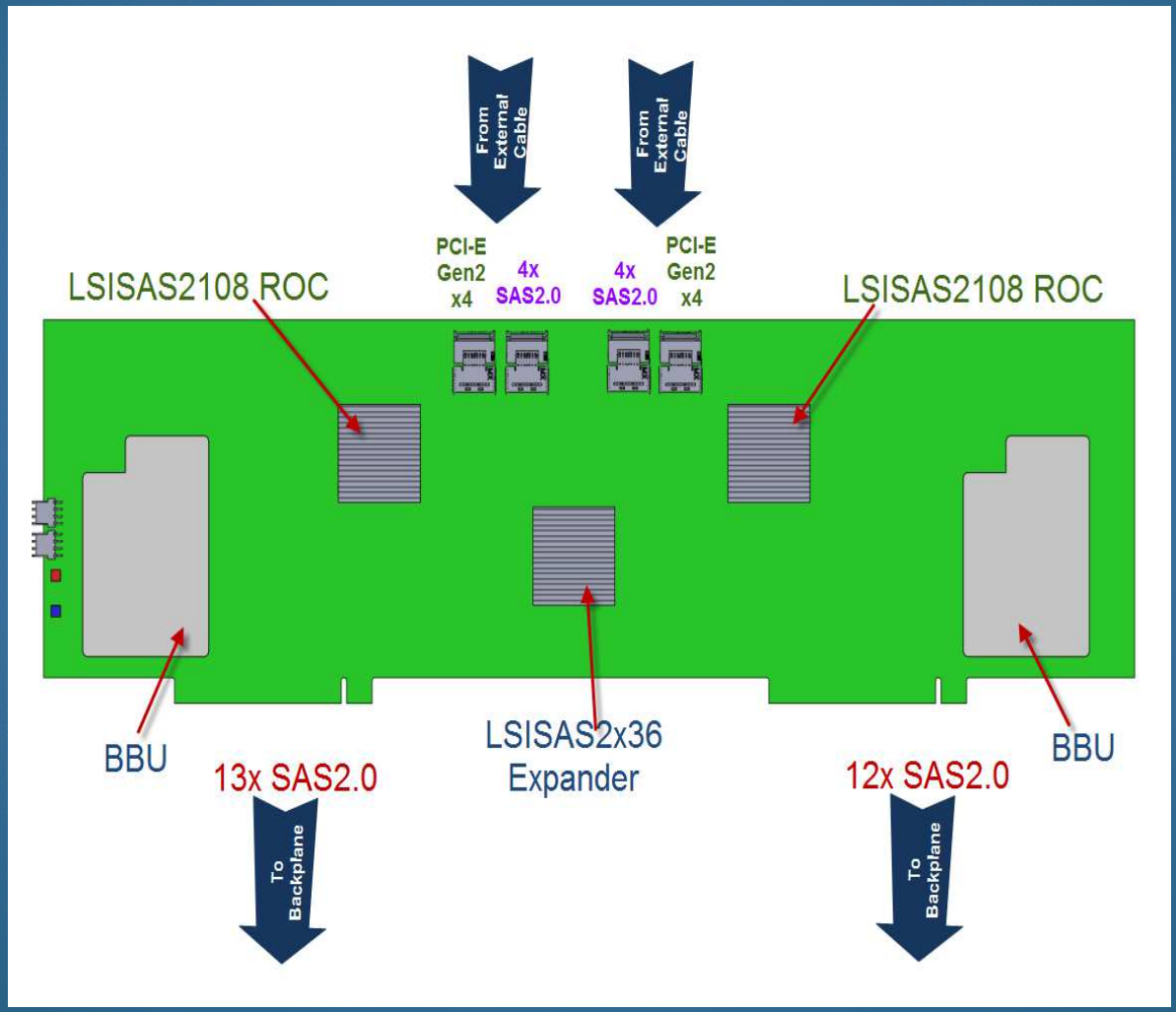
- Single (Haystack) or dual (Hadoop) controllers depending on tier requirements.

Input to the board can be either SAS or PCI-E.

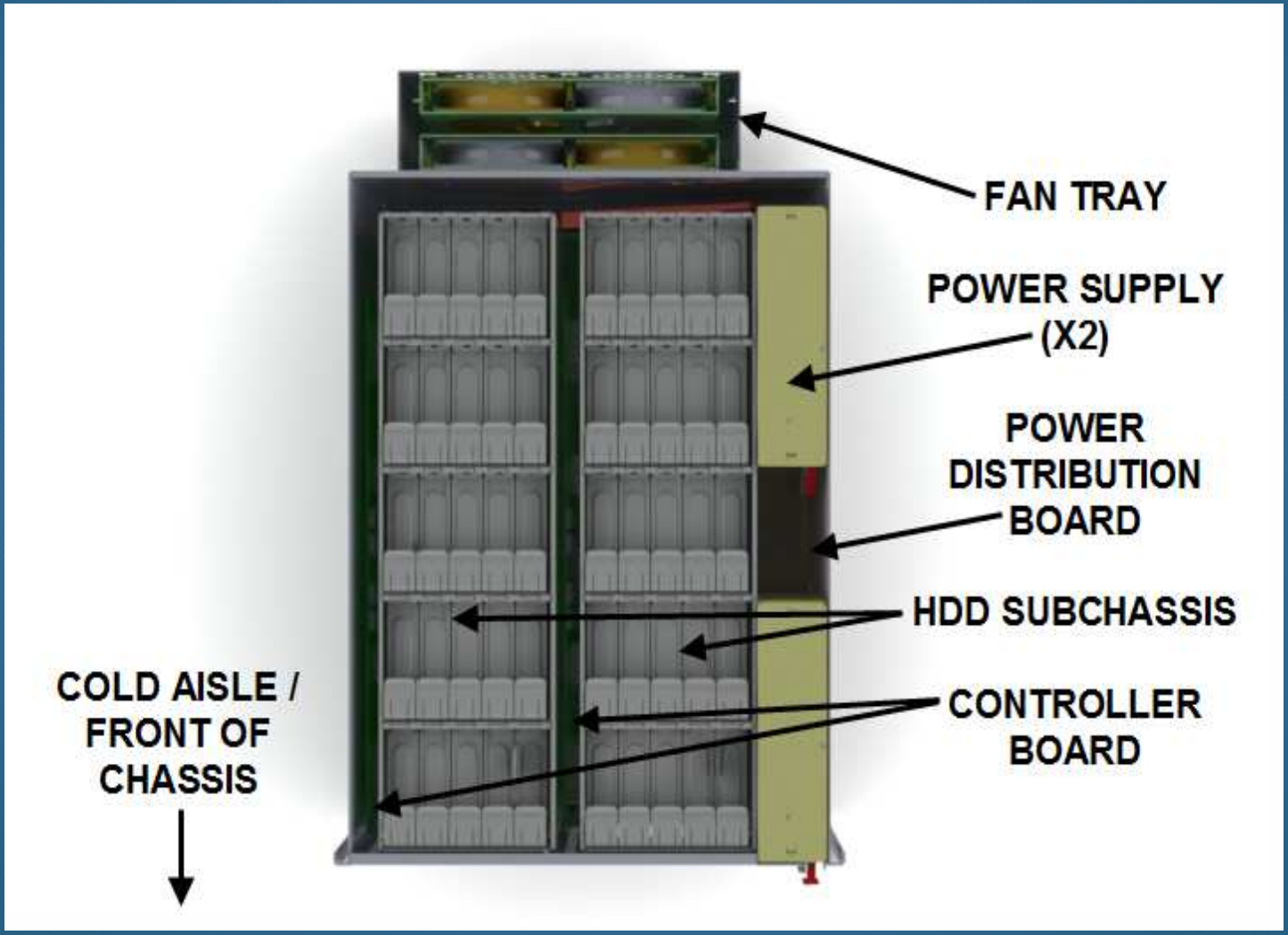
- Use AC caps to create stub free high speed T-routes.



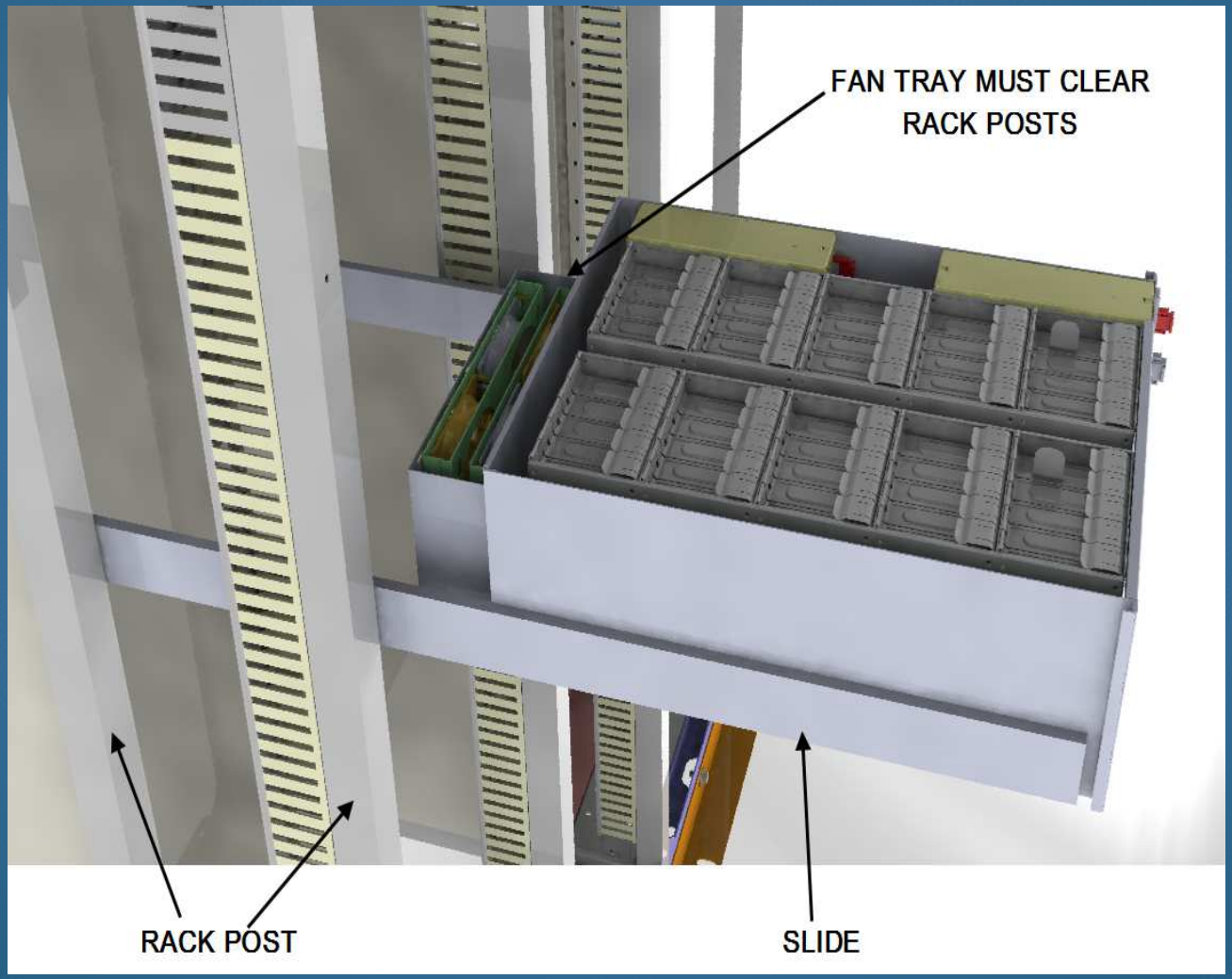
Knox Controller Board Layout



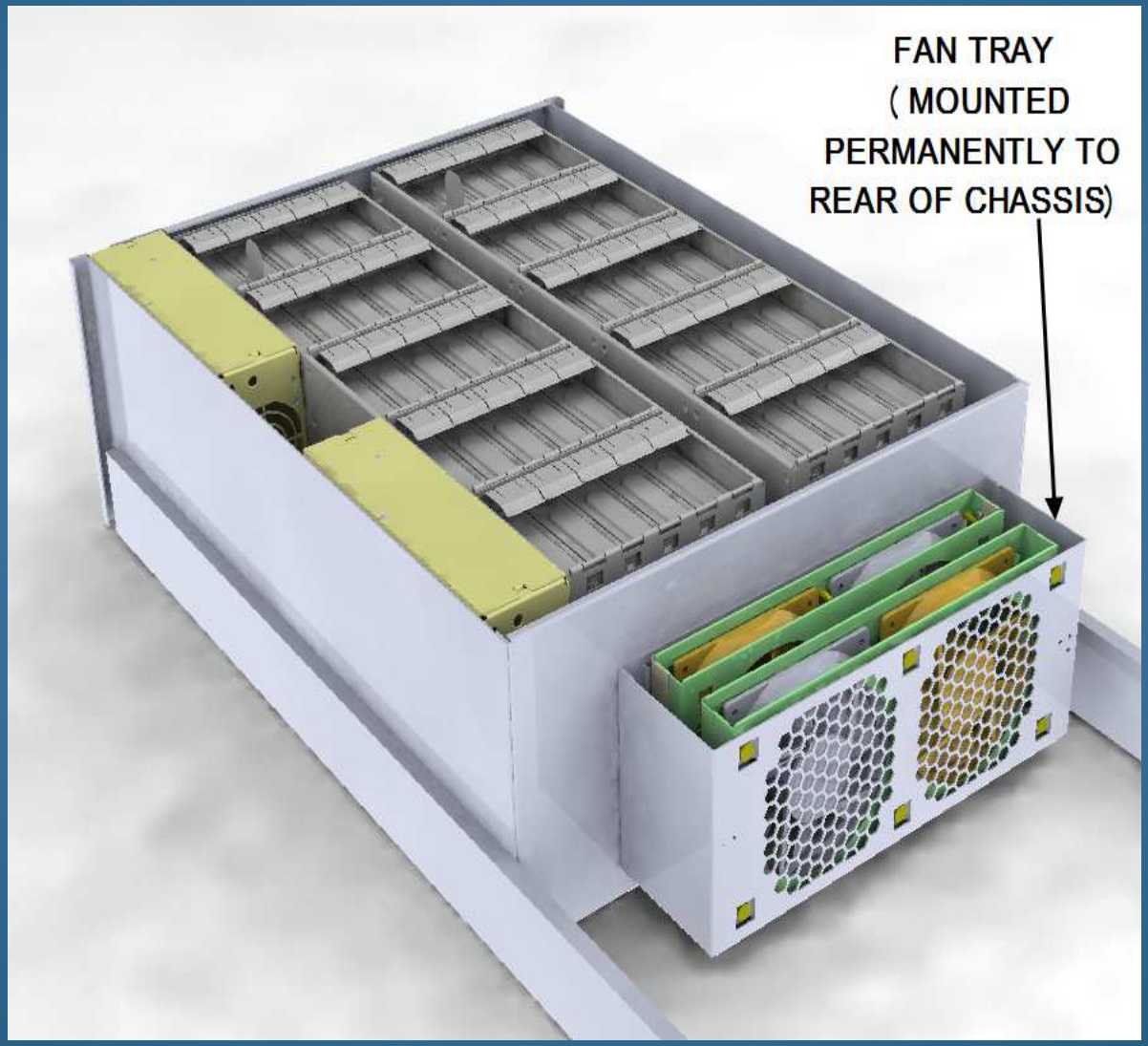
Knox Chassis



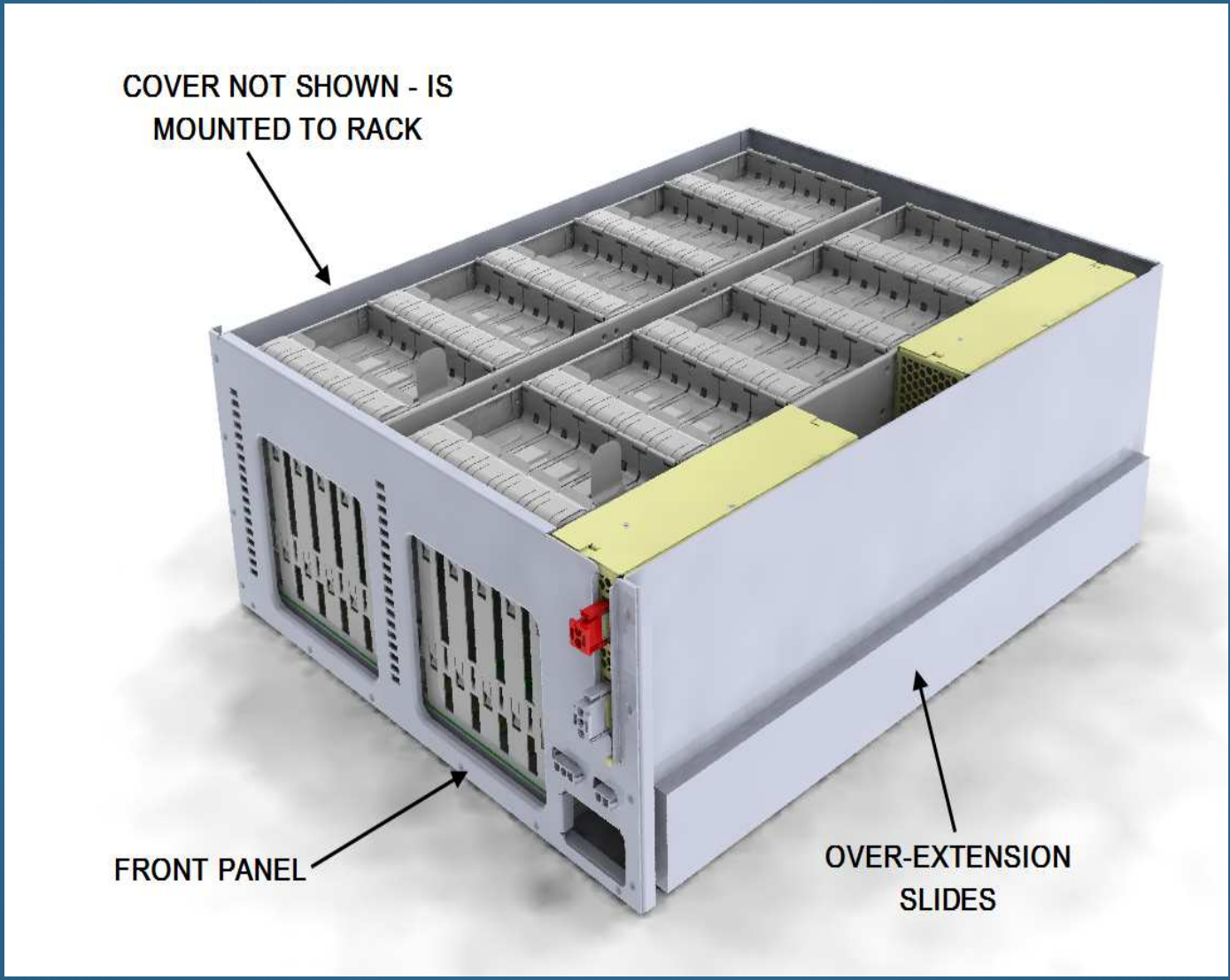
Knox Chassis



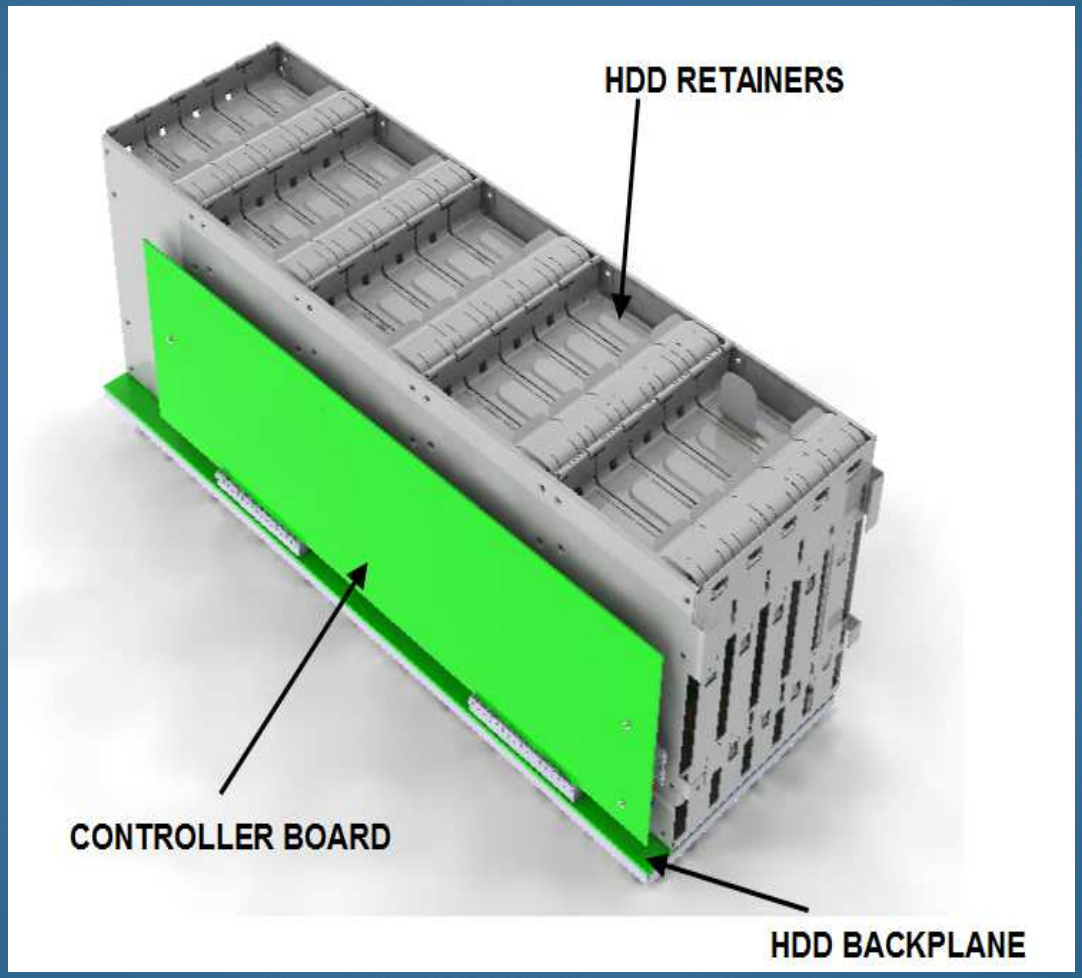
Knox Chassis



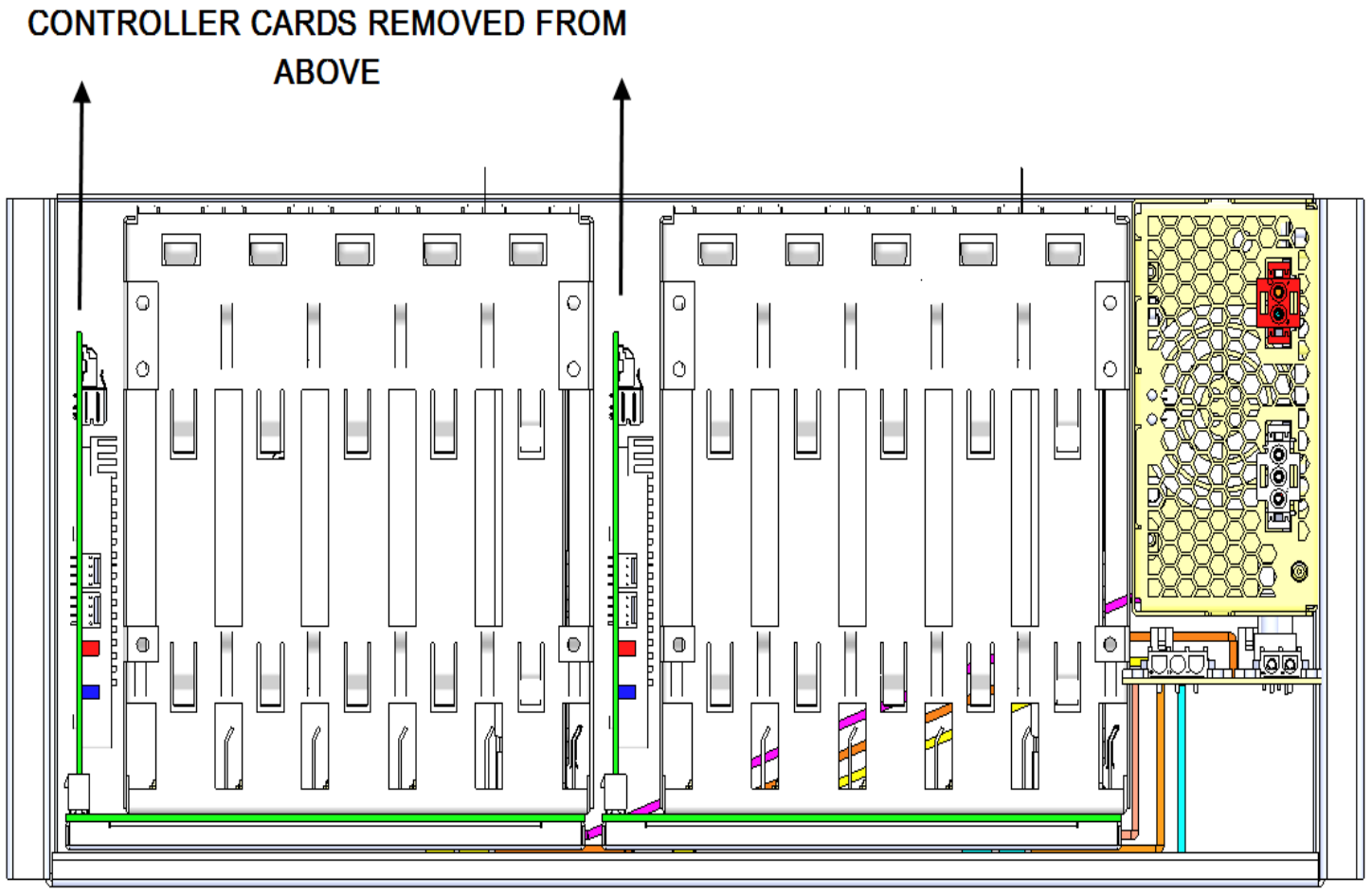
Knox HDD buckets



Knox HDD buckets



Knox HDD buckets



facebook