

HA Linux containers on Oracle Grid Infrastructure 12c

Technology Preview

*Artem Danielov
CTO, FlashGrid*

*Emil Sildos
VP BD, FlashGrid*

DOAG'16, 2016-11-17

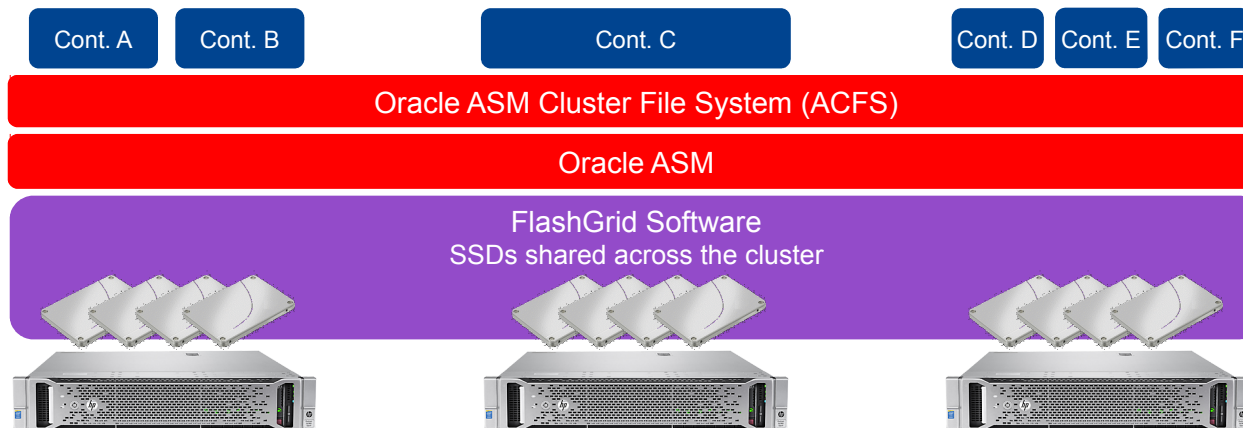
- Consolidation options and why Linux containers
- High availability considerations
 - Fail-over between servers
 - Shared storage on ACFS
 - Live migration of containers
- Hyper-converged platform architecture
 - FlashGrid Read-Local Technology
 - NVMe SSD technology
 - Metro-scale HA with extended distance clusters
 - HA database deployments on Public Cloud

- **Pluggable Databases**
 - Limited to databases only
 - Must use the same database version and patch level
 - Additional Multitenancy feature license needed
- **Virtual Machines**
 - Separate OS and kernel for every database or application
 - Virtualization performance tax (network, storage, fixed memory allocations)
 - Good for small number of databases per server
- **Linux Containers**
 - Advantages of VMs plus bare-metal performance
 - Isolated OS environment for each database, only kernel shared
 - LXC containers on Oracle Linux 7 certified for Oracle Database

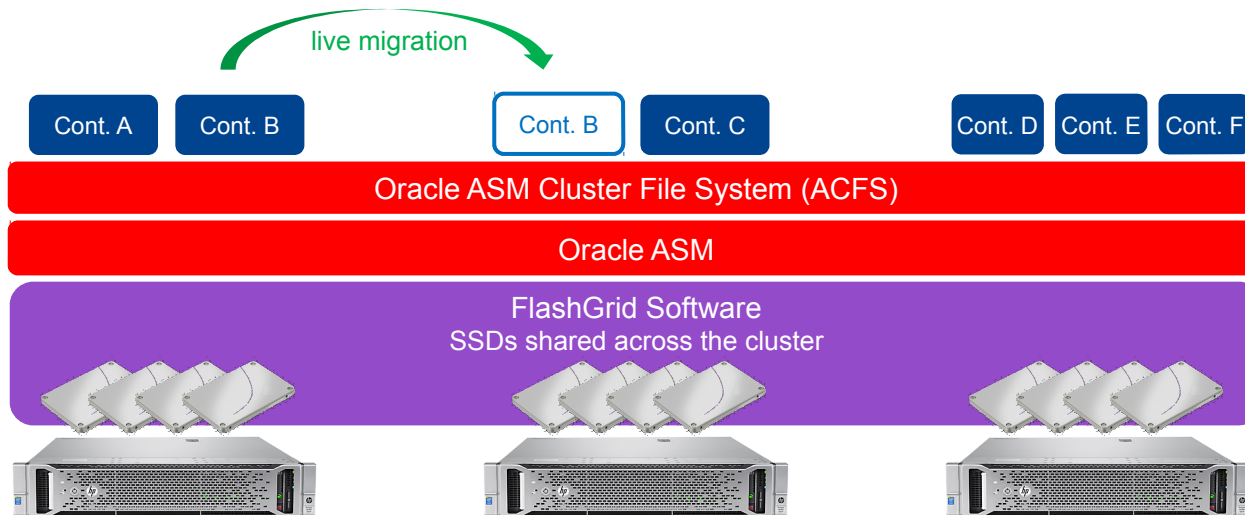
- How to migrate containers and data if a server goes down?
- Need shared storage between the servers
- Need mechanism for automatic fail-over of the containers
- Live migration would be ideal for maintenance



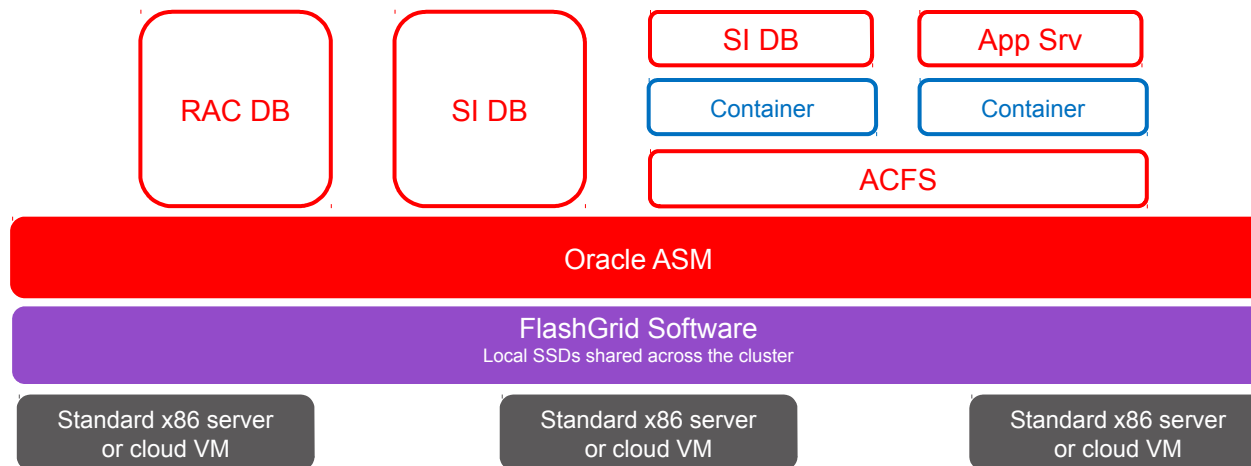
- 2- or 3-node Grid Infrastructure cluster
- Grid Infrastructure provides HA mechanisms for container 'resources'
- Linux container files on ACFS cluster file system
- ACFS provides advanced data services e.g. snapshots
- ASM provides data mirroring across nodes
- FlashGrid software provides shared access to SSDs inside servers

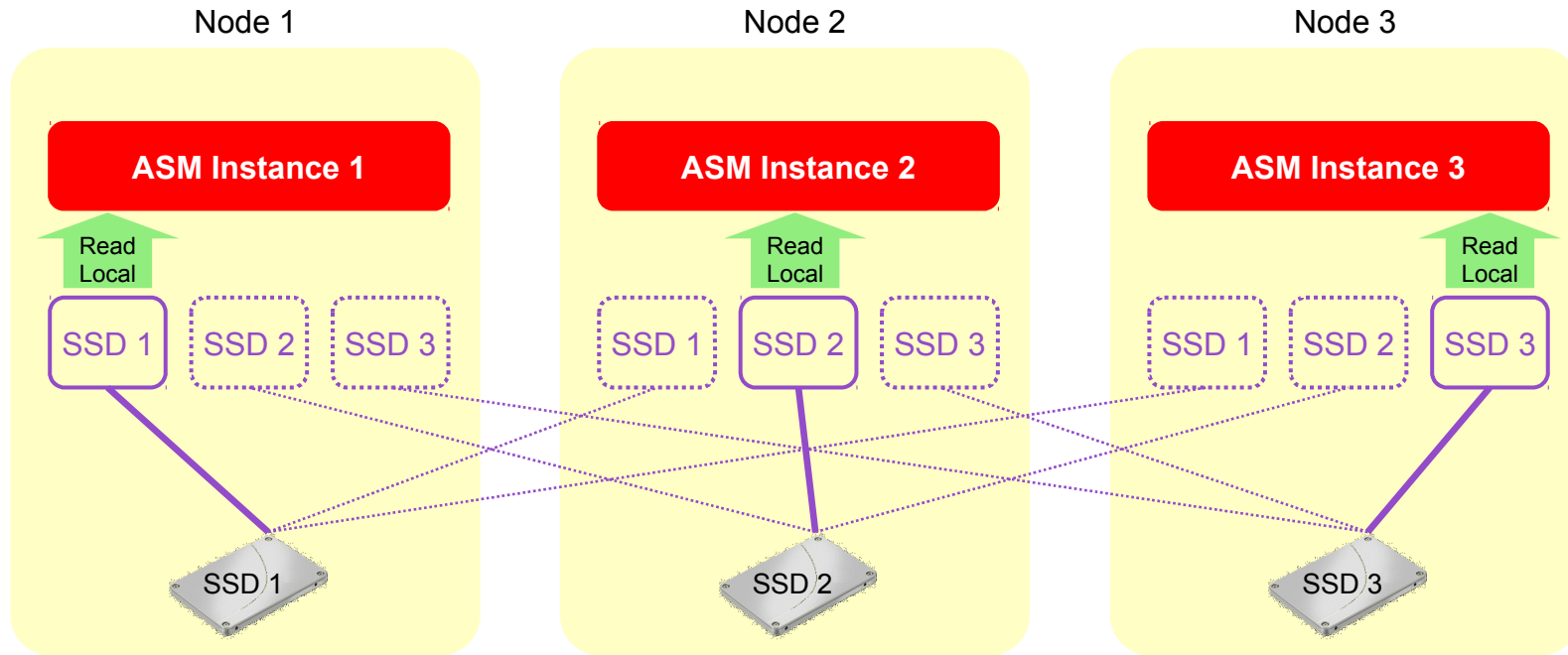


- Tested live migration of LXC container with running Oracle 12c database
- Requires Oracle Linux 7 UEK4 kernel
- Live migration tools for containers with databases developed by FlashGrid (currently in technology preview phase)



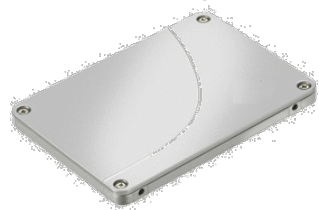
- Universal HA architecture for containers, single-instance DB, RAC DB
- Leverages proven GI and ASM for HA, data mirroring, data management
- FlashGrid Read-Local Technology minimizes storage network overhead
- Choice of underlying infrastructure:
 - On-prem bare-metal servers
 - On-prem virtualized (VMware, Oracle VM, KVM)
 - Extended distance clusters
 - Cloud VMs
 - Bare-metal cloud





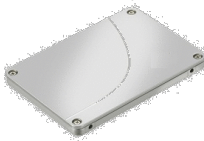
- Minimize network overhead by serving reads from local SSDs at the speed of PCIe
- Accelerate both reads and writes

- Highly efficient replacement for legacy SCSI (FC, SAS, SATA) stack
- High IOPS, low latency, low CPU consumption
- Available from all server and SSD vendors
- 2.5” hot-plug and add-in PCIe card form-factors
- Used in Oracle Exadata*

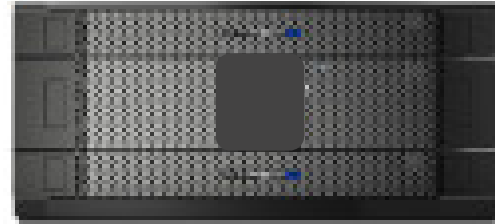


Performance of one NVMe SSD similar to a flash array

3 GB/s
400K IOPS



3 GB/s
250K IOPS



More bandwidth for full table scans and backups

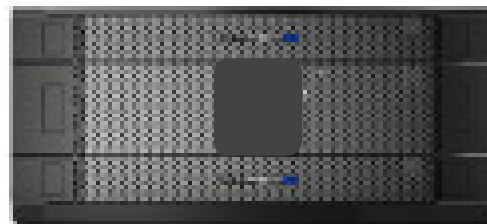
24 GB/s

with 8 NVMe SSDs
inside 2 database servers



+ FlashGrid software

3 GB/s



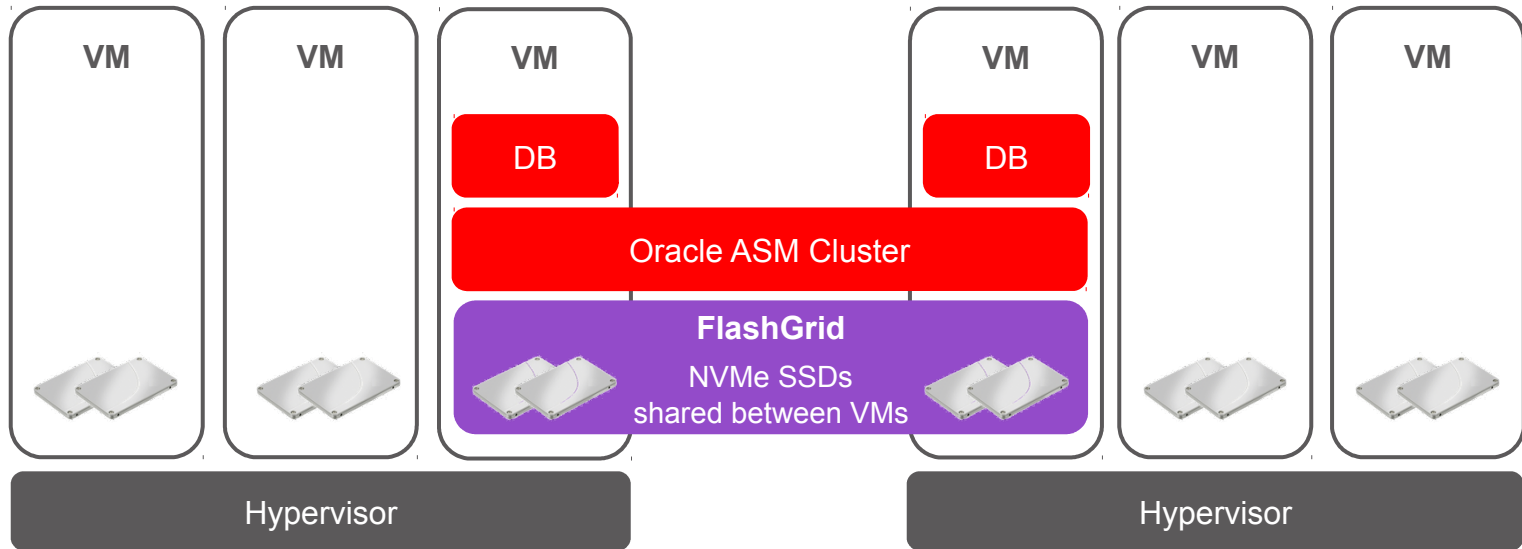
NVMe-Optimized Server Options for On-Prem

Server model	2.5" hot-plug NVMe SSDs			Add-in PCIe card NVMe SSDs		Max total NVMe flash capacity per server
	# slots	Max capacity per SSD	Max capacity per server with 2.5" NVMe SSDs	# PCIe slots available for NVMe SSDs	Max flash capacity per server with 6.4TB add-in card SSDs	
Oracle Server X6-2L	9	3.2 TB	28.8 TB	5	32 TB	60.8 TB
Oracle Server X6-2	4	3.2 TB	12.8 TB	3	19.2 TB	32 TB
Dell PowerEdge R730xd	4	3.2 TB	12.8 TB	5	32 TB	44.8 TB
Dell PowerEdge R930	8	3.2 TB	25.6 TB	9	57.6 TB	83.2 TB
Dell PowerEdge R630	4	3.2 TB	12.8 TB	2	12.8 TB	25.6 TB
HPE ProLiant DL380 Gen9	6	2 TB	12 TB	5	32 TB	44 TB
HPE ProLiant DL560 Gen9	6	2 TB	12 TB	6	38.4 TB	50.4 TB
HPE ProLiant DL580 Gen9	5	2 TB	10 TB	8	51.2 TB	61.2 TB
Supermicro 1028U-TN10RT+	10	3.2 TB	32 TB	2	12.8 TB	44.8 TB
Supermicro 2028U-TN24R4T+	24	3.2 TB	76.8 TB	2	12.8 TB	89.6 TB
Supermicro 2028R-NR48N	48	3.2 TB	153.6 TB	2	12.8 TB	166.4 TB



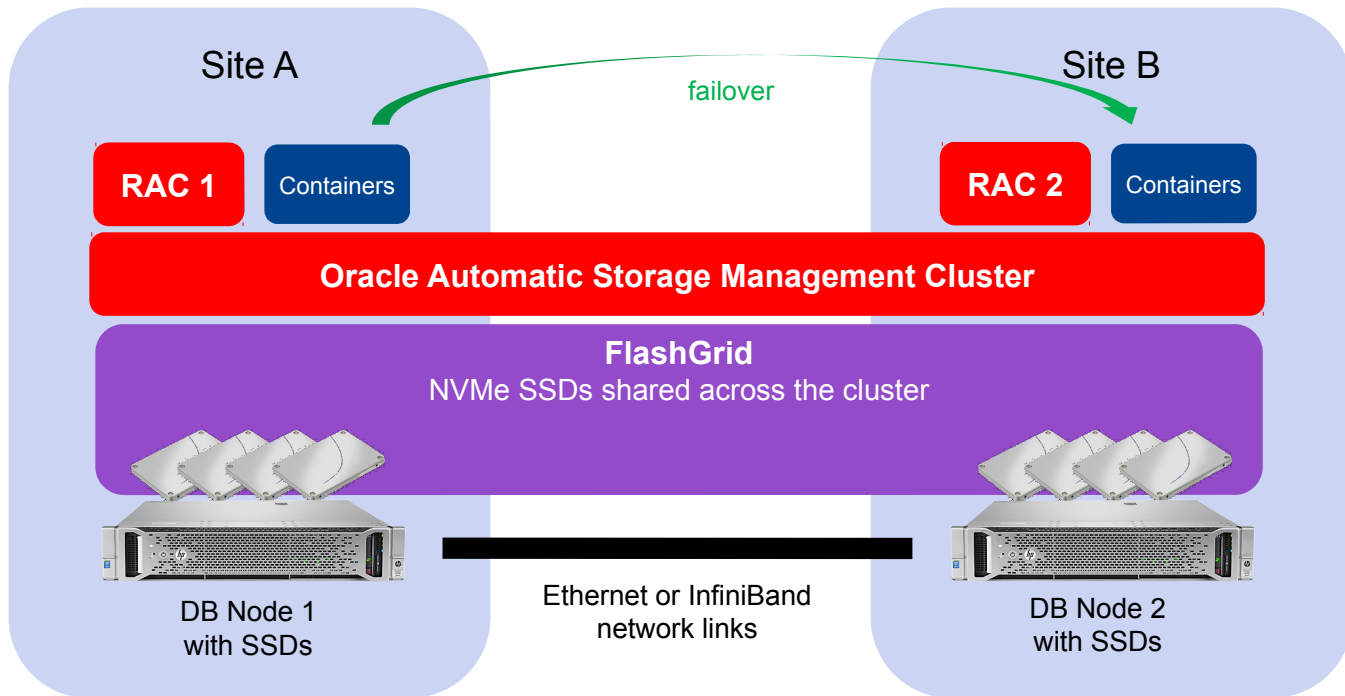
Virtualized Clusters

- Oracle VM, VMware, KVM
- Isolation between clusters: e.g. Production + Pre-production + Test/Dev
- FlashGrid Read-Local Technology minimizes storage network overhead



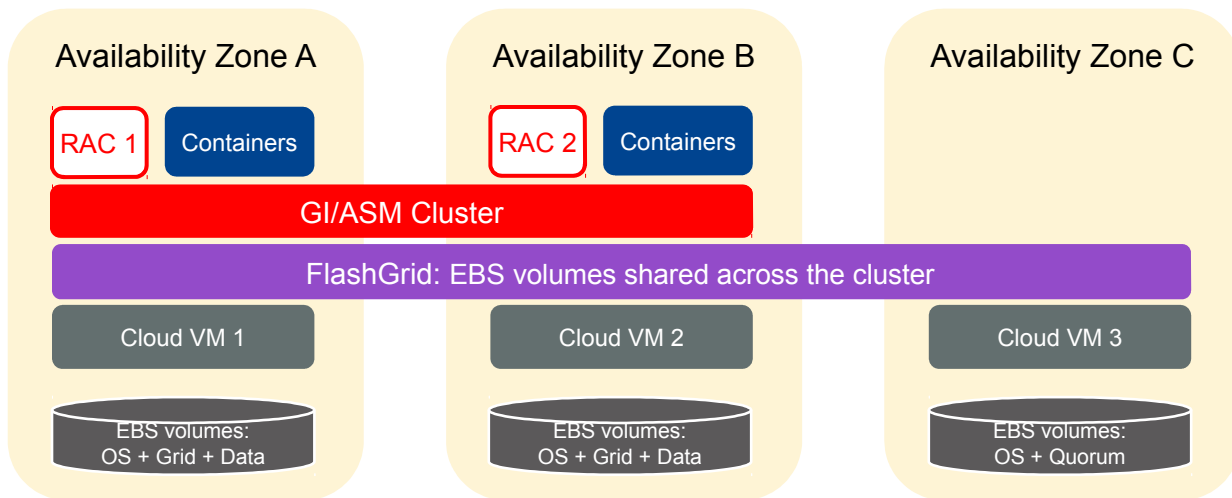
Extended Distance Clusters

- HA/DR solution at metro or campus scale
- Synchronous data mirroring across sites



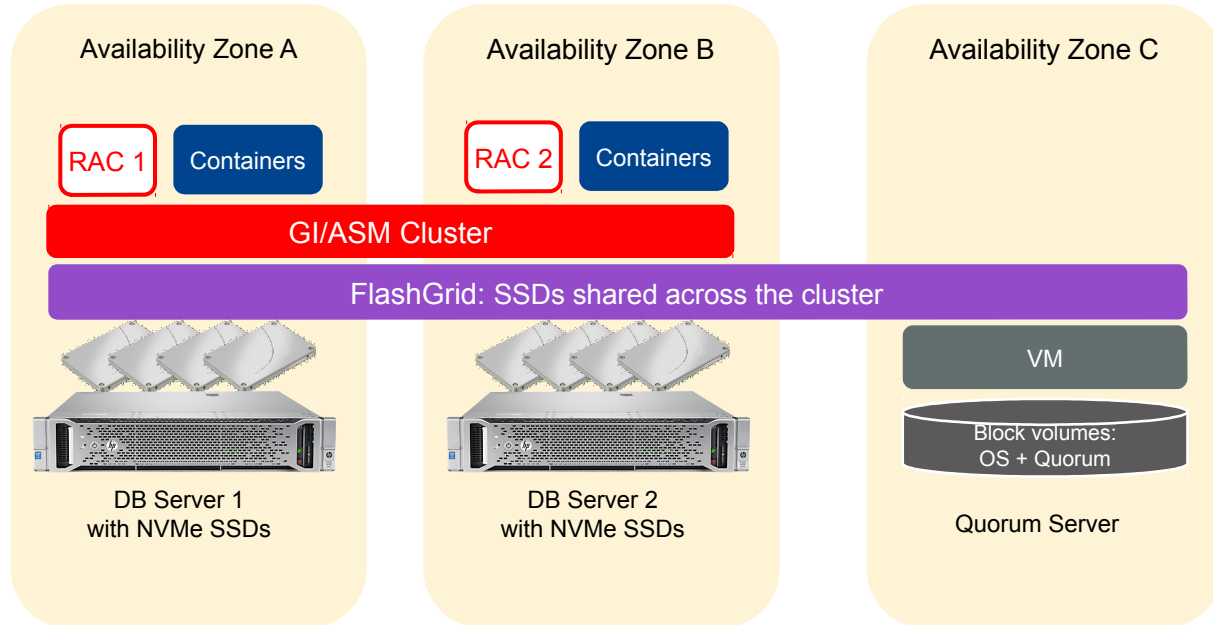
On Public Cloud

- HA databases in the cloud
- Within one availability zone or across two or three availability zones
- FlashGrid Cloud Area Network enables Grid Infrastructure clusters on any public cloud



On Bare Metal Cloud

- HA and high performance databases on bare metal clouds
- Within one availability zone or across two or three availability zones
- Local NVMe SSDs turned into shared HA storage



- It is possible to migrate Linux containers live
- Reference architecture is based on commodity hardware components (bare metal) and standard virtual resources (public cloud)
- Join the discussion and testing

Download FlashGrid software at www.flashgrid.io

Request access to a pre-configured cluster on AWS

Questions?

Emil Sildos

esildos@flashgrid.io