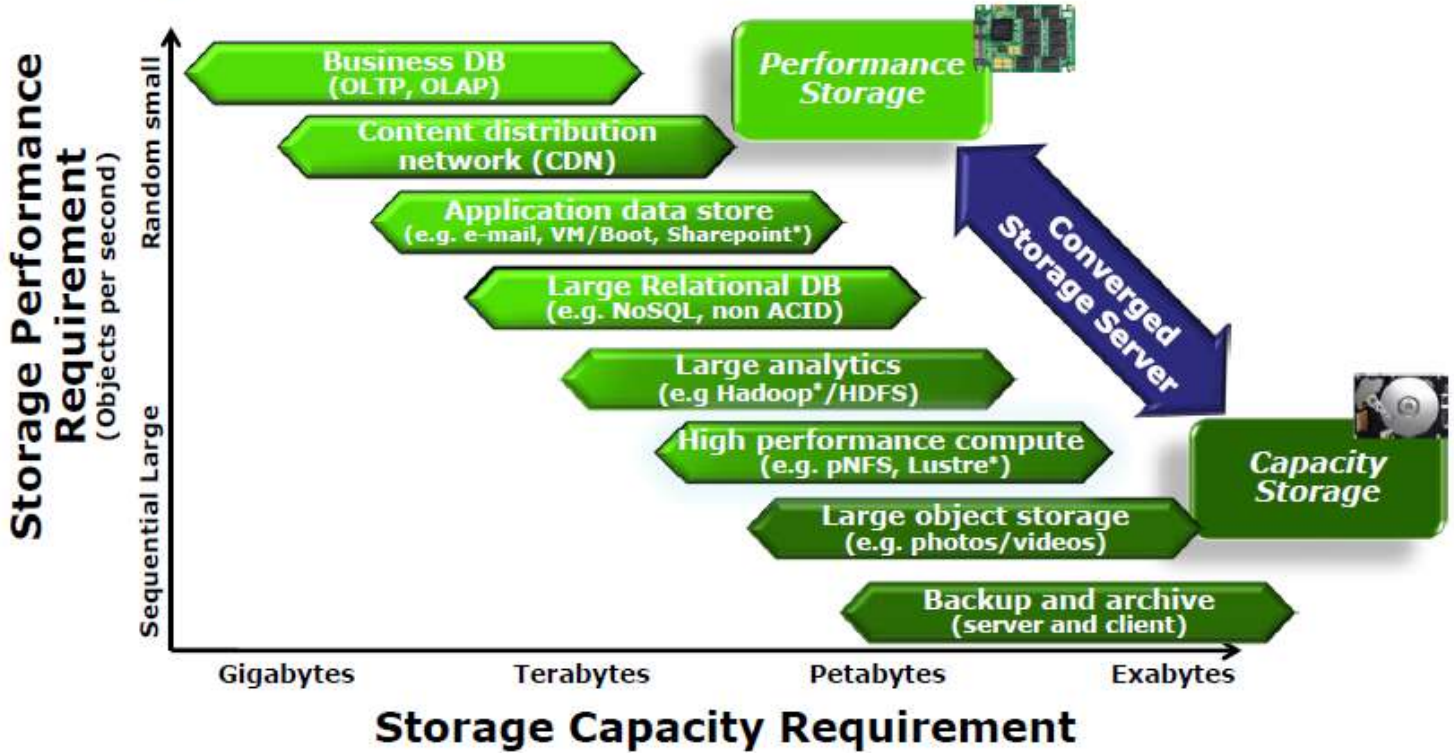
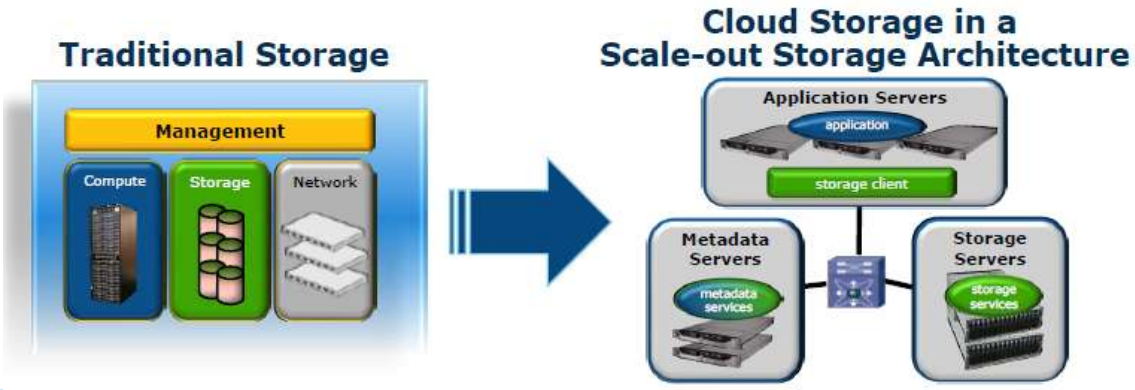
Cloud Storage and EC solution example

田口栄治
クラウド・コンピューティング事業本部
インテル株式会社
2012年7月

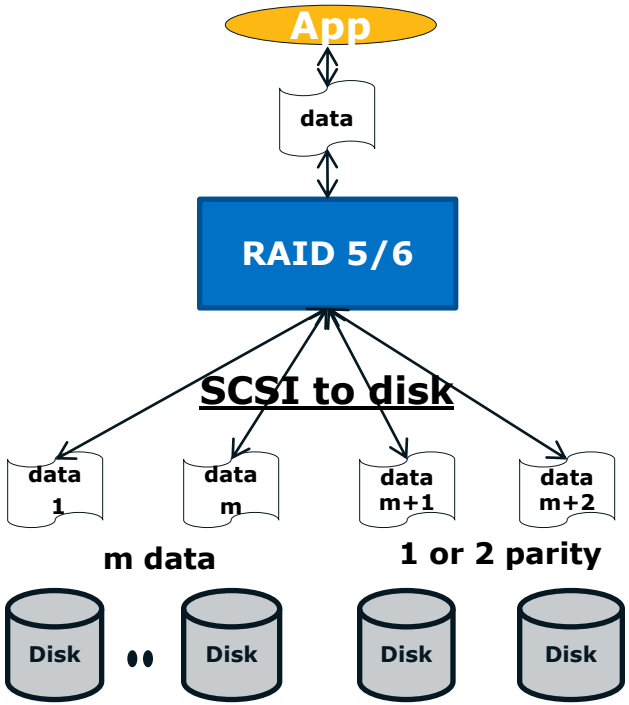
Usage Models Dictate the Solutions



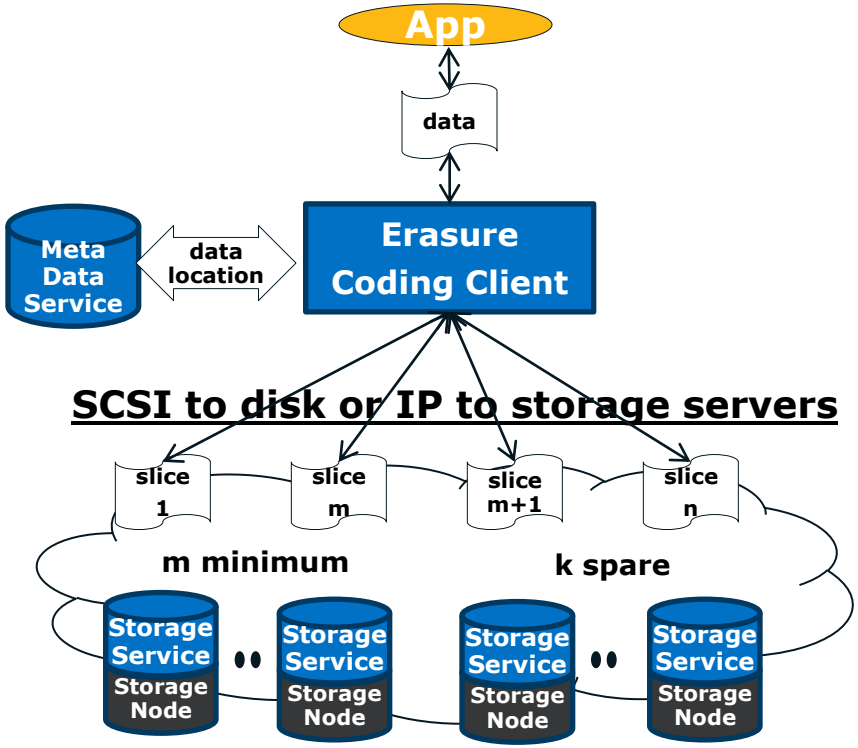


Erasure Coding (EC)

RAID



Erasure Coding



EC extends the data protection architectures of RAID 5/6 to RAID k

k = the number of failures that can be tolerated without data loss:

For RAID 5, k=1; For RAID 6, k=2; For EC, k = n

EMC* Atmos* and Isilon* are example systems using EC

Optimized Object Storage

Object Storage Interface

- http/REST API provides object access
- Objects names are user-specified

Namespaces + Redundancy Policies

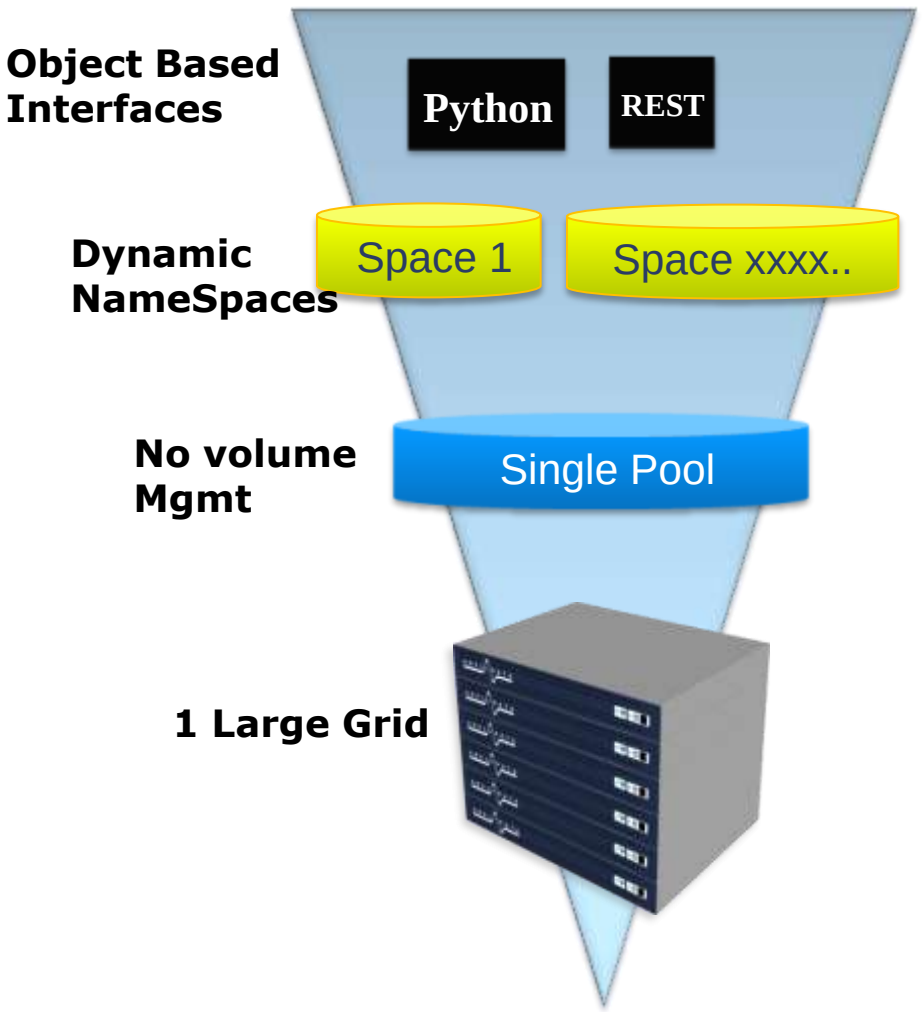
- Virtual containers
- Policies define protection level (e.g. 16/4)

Single System View

- Objects are globally accessible
- No Volume Management

Grid Architecture

- Loose any disk, server, rack or datacenter

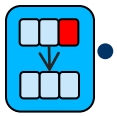


Core Technology Components

AMPLIDATA

BitSpread – Distributed Encoder/Decoder

- RAID replacement technology based on unique variant of Erasure Coding
- "Dial-in" fault tolerance through namespace level policies
  - Namespace1: 16/4 policy protects against any 4 failures in 16 disks/nodes
  - Namespace2: 18/6 policy protects against any 6 failures in 18 disks/nodes
  - Namespace3: 8/2 policy protects against any 2 failures in 8 disks/nodes
  - …
- Provides availability and reliability even during failures
- Policies can be dynamically changed

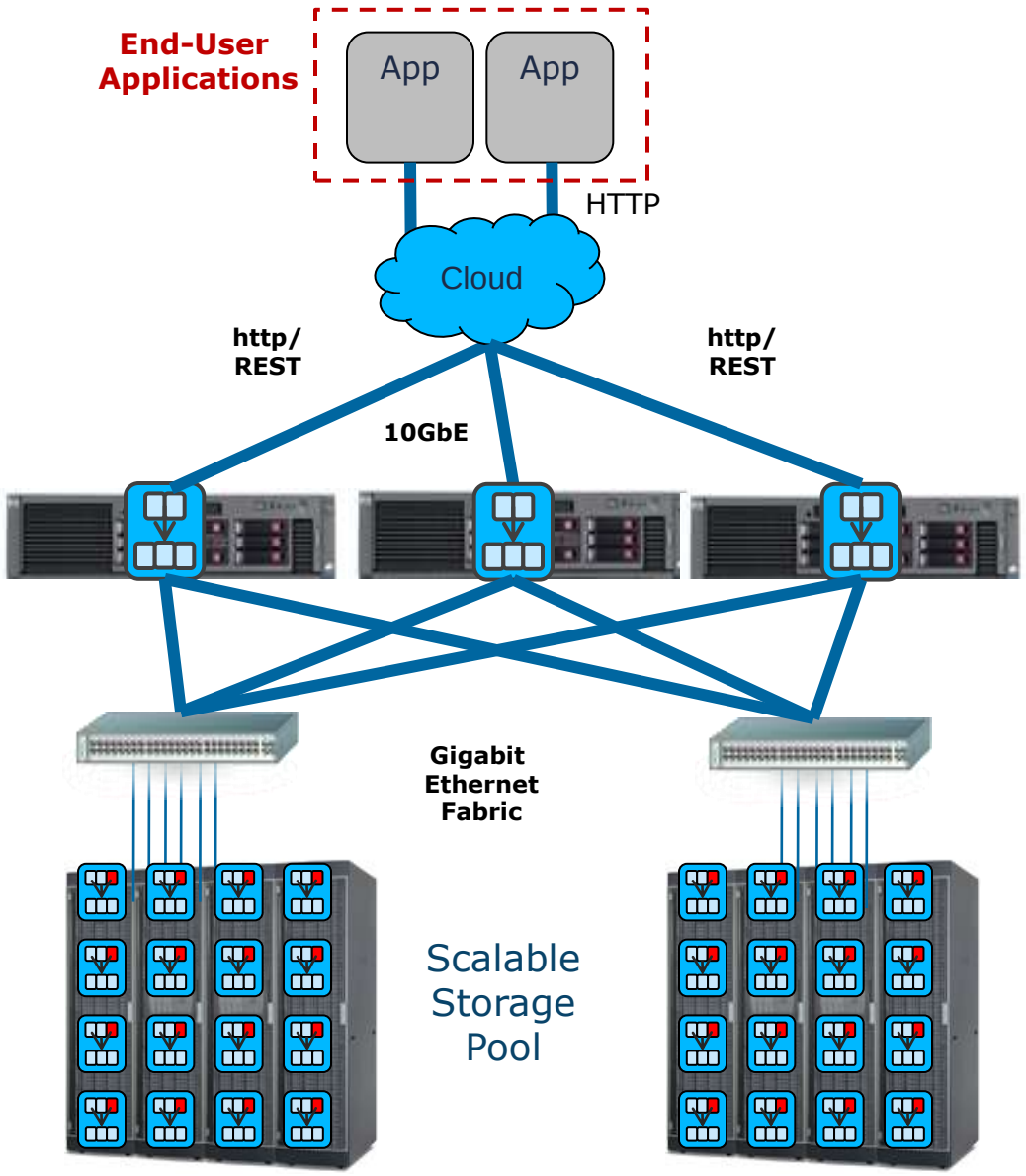


BitDynamics – Maintenance & Self-Healing Agent

- Out of band operations agent for disk monitoring, integrity verification & object self-healing
- Performs automated tasks: scrubs, verifies, self-heals, repairs & optimizes data on disk

Object Storage System Topology

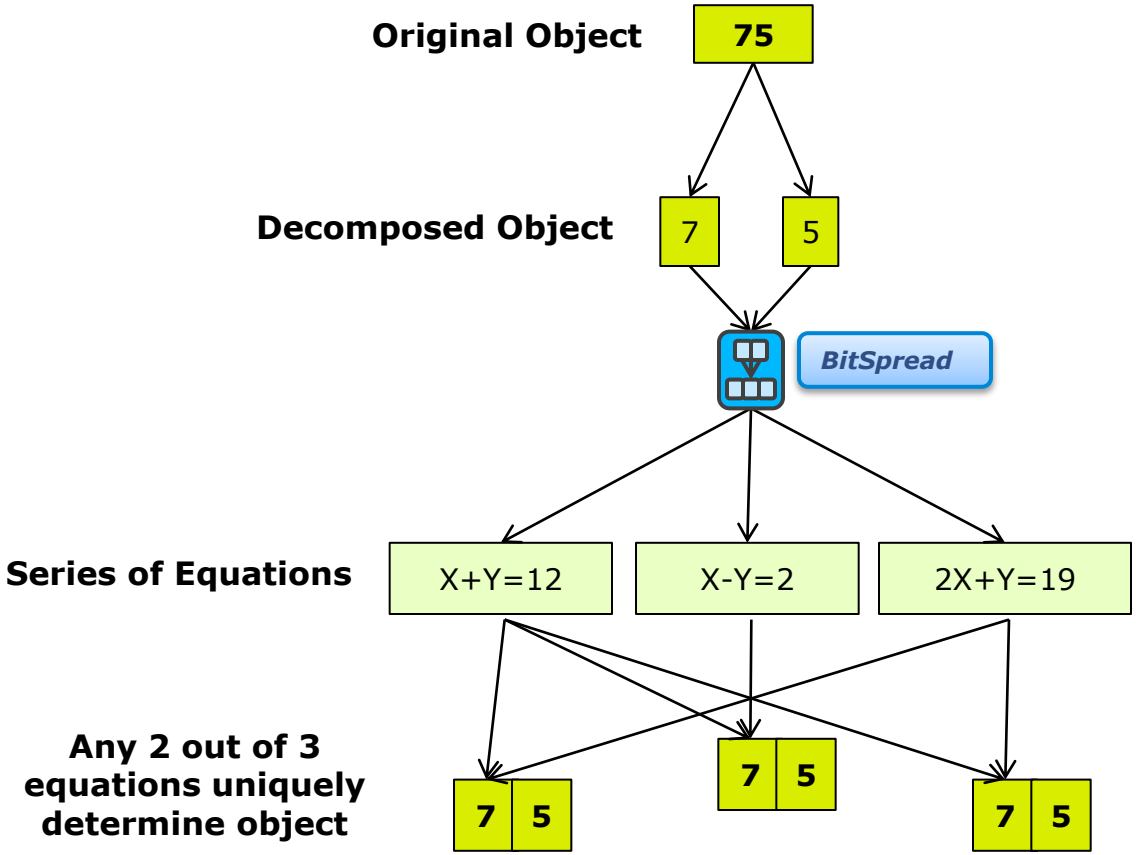
- BitSpread
- BitDynamics



The BitSpread Algorithm – Simple Example

- **BitSpread Encodes data in unrecognizable chunks (actually a series of equations)**
- **Distributes the equations across disks, storage nodes, racks, data centers**
- **Original data can always be uniquely determined from a subset of the equations**
- **BitSpread codec actually uses 4K variables & equations independent of object size**

Simplified mathematics



Large Object Store Rack (Storage Node)

Large object storage

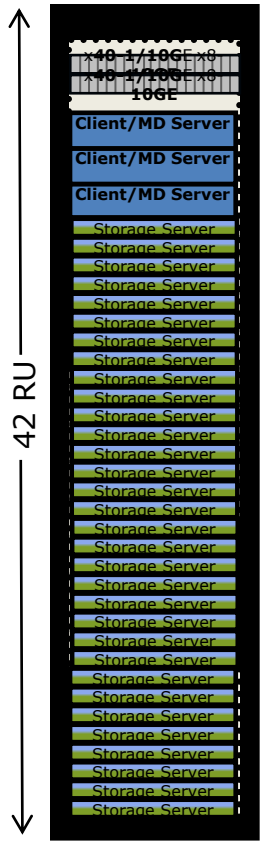
x2 1/10GbE BaseT switches

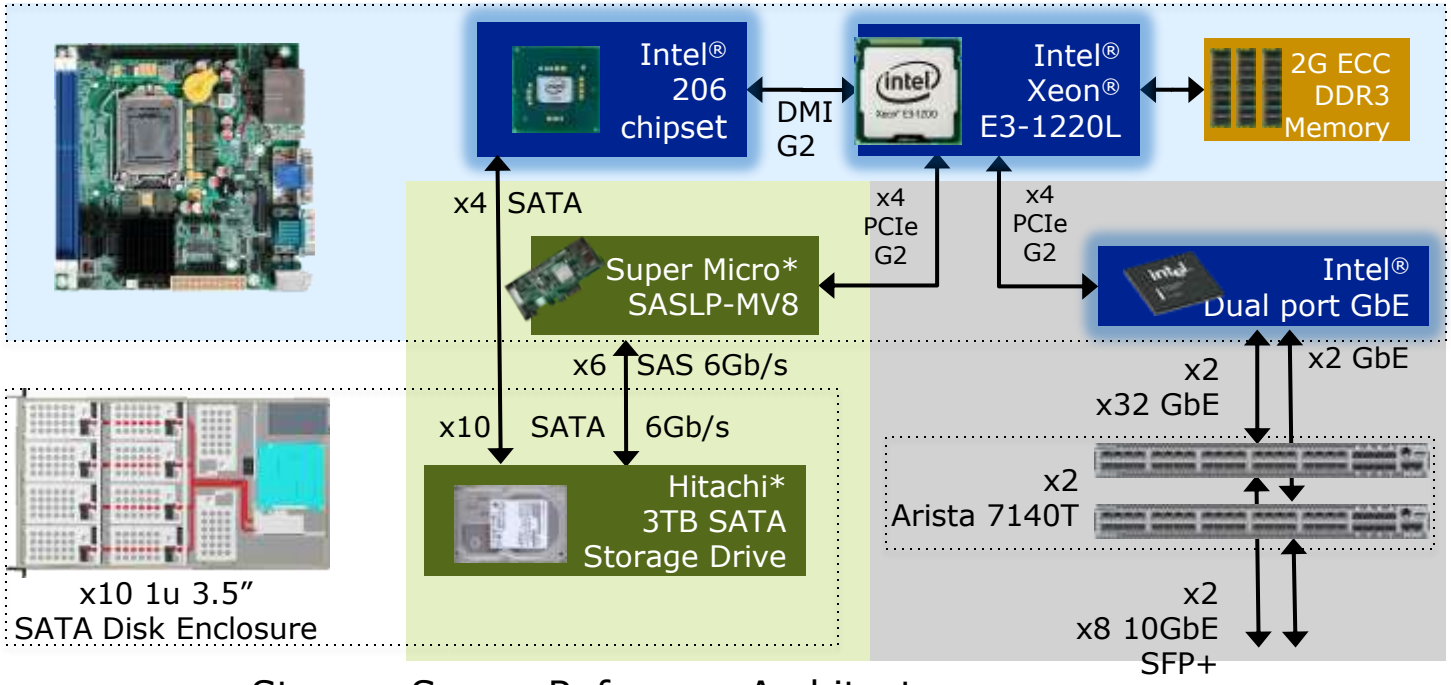
960 TBs storage

x32 storage servers

Amplidata*

X86-64 RedHat* Linux*





Storage Server Reference Architecture

~1PB of raw storage in a 42u rack

High Efficiency, Durability, Scalability with Erasure Coding

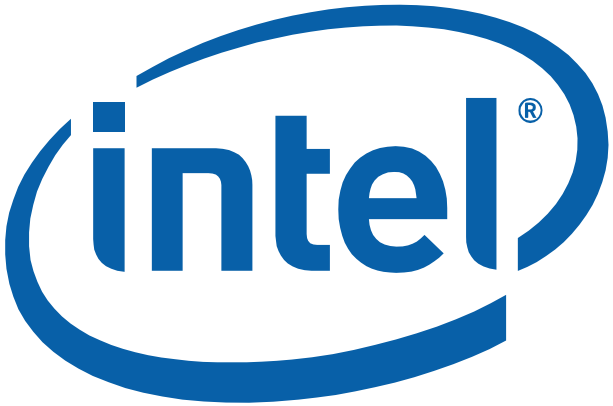
Converged Storage Server with EC Value

(320 Drive, 960TB comparison, no single point of failure[1])

Value	Description	Number nodes=32, 10 drives/node, Cap/Node=30TB				
		EC16 m=10, k=6 16 nodes	RAID0+1 m=10, k=0 2 nodes	RAID5+1 m=9, k=1 2 nodes	RAID6+1 m=8, k=2 2 nodes	RAID 3way m=10, k=0 3 nodes
Efficiency	Raw/Usable Efficiency	63%	50%	40%	34%	33%
	Usable Capacity (TB)	600	480	432	384	320
	Watts/Usable Capacity	53%	67%	74%	83%	1
Durability	lower data loss probability	10^{-8}	2288	1.6	10^{-6}	1

EC is the best efficiency at equivalent durability compared to RAID6+1

[1]Hardware configuration Large Object Reference Architecture


intel

Legal Disclaimer

INFORMATION IN THIS DOCUMENT IS PROVIDED IN CONNECTION WITH INTEL PRODUCTS. NO LICENSE, EXPRESS OR IMPLIED, BY ESTOPPEL OR OTHERWISE, TO ANY INTELLECTUAL PROPERTY RIGHTS IS GRANTED BY THIS DOCUMENT. EXCEPT AS PROVIDED IN INTEL'S TERMS AND CONDITIONS OF SALE FOR SUCH PRODUCTS, INTEL ASSUMES NO LIABILITY WHATSOEVER AND INTEL DISCLAIMS ANY EXPRESS OR IMPLIED WARRANTY, RELATING TO SALE AND/OR USE OF INTEL PRODUCTS INCLUDING LIABILITY OR WARRANTIES RELATING TO FITNESS FOR A PARTICULAR PURPOSE, MERCHANTABILITY, OR INFRINGEMENT OF ANY PATENT, COPYRIGHT OR OTHER INTELLECTUAL PROPERTY RIGHT.

- A "Mission Critical Application" is any application in which failure of the Intel Product could result, directly or indirectly, in personal injury or death. SHOULD YOU PURCHASE OR USE INTEL'S PRODUCTS FOR ANY SUCH MISSION CRITICAL APPLICATION, YOU SHALL INDEMNIFY AND HOLD INTEL AND ITS SUBSIDIARIES, SUBCONTRACTORS AND AFFILIATES, AND THE DIRECTORS, OFFICERS, AND EMPLOYEES OF EACH, HARMLESS AGAINST ALL CLAIMS COSTS, DAMAGES, AND EXPENSES AND REASONABLE ATTORNEYS' FEES ARISING OUT OF, DIRECTLY OR INDIRECTLY, ANY CLAIM OF PRODUCT LIABILITY, PERSONAL INJURY, OR DEATH ARISING IN ANY WAY OUT OF SUCH MISSION CRITICAL APPLICATION, WHETHER OR NOT INTEL OR ITS SUBCONTRACTOR WAS NEGLIGENT IN THE DESIGN, MANUFACTURE, OR WARNING OF THE INTEL PRODUCT OR ANY OF ITS PARTS.
- Intel may make changes to specifications and product descriptions at any time, without notice. Designers must not rely on the absence or characteristics of any features or instructions marked "reserved" or "undefined". Intel reserves these for future definition and shall have no responsibility whatsoever for conflicts or incompatibilities arising from future changes to them. The information here is subject to change without notice. Do not finalize a design with this information.
- The products described in this document may contain design defects or errors known as errata which may cause the product to deviate from published specifications. Current characterized errata are available on request.
- Intel processor numbers are not a measure of performance. Processor numbers differentiate features within each processor family, not across different processor families. Go to: http://www.intel.com/products/processor_number.
- Contact your local Intel sales office or your distributor to obtain the latest specifications and before placing your product order.
- Copies of documents which have an order number and are referenced in this document, or other Intel literature, may be obtained by calling 1-800-548-4725, or go to: http://www.intel.com/design/literature.htm
- Intel, Xeon, Atom, Sponsors of Tomorrow and the Intel logo are trademarks of Intel Corporation in the United States and other countries.

- *Other names and brands may be claimed as the property of others.
- Copyright ©2012 Intel Corporation.

- *Other names and brands may be claimed as the property of others.
- Copyright ©2012 Intel Corporation.