TOSHIBA

Leading Innovation

WHITEPAPER

KEY VALUE DRIVE

1 CURRENT CHALLENGES IN THE USE OF BIG DATA AND STORAGE

The volume of unstructured data has been growing at a tremendous rate, and cloud services have become widespread. Accordingly, both private enterprises and public organizations have started to utilize the accumulated data for various purposes, including:

1) Marketing and product development
2) Improving the production and product utilization efficiency
3) Promoting science and technology, and academic research

Purpose	Description	Examples
Marketing and product development	・Marketing scheme that performs an access analysis to identify keywords on which search engines are likely to hit and directs users to specific websites ・Analyzes consumers' purchase trends using data derived from POS systems and reflects them in the planning of new products	・Search engine optimization (SEO) marketing ・Menus of fast-food and restaurant chains ・Recommender systems of Amazon.com, search portal sites and video sites
Improving the production and product utilization efficiency	・Improves processes using yield data at a semiconductor fab to reduce the defect level ・Monitors the operation of jet engines and turbines and analyzes the accumulated data to help predict a possible failure	・Improvement activities at shop floors using a manufacturing engineering system (MES) ・GE's aircraft engines and power generation gas turbines(*1) ・Monitors Komatsu's large construction machines using GPS systems to know their operating conditions and runs them unattended(*2).
Promoting science and technology, and academic research	・Human genotyping ・Discovery of the Higgs boson ・Research on the mechanism of climate change	・Human genotyping at the National Human Genome Research Institute (NHGRI)(*3) ・Discovery of the Higgs boson using the Atlas Experiment facilities at CERN(*4) ・Activities of the World Meteorological Organization(*5)

*1: http://fisheritcenter.haas.berkeley.edu/Big_Data/Ruh_Fisher%20FINAL.pdf(January ,2012) P5

*2: http://www.komatsuamerica.com/komtrax (April,2014)

*3: https://www.genome.gov/10001634 (May,2014)

*4: https://twiki.cern.ch/twiki/bin/view/AtlasPublic (May,2015)

*5: https://www.wmo.int/pages/summary/progs_struct_en.html (May,2015)

The use of big data has been gaining momentum as shown above because servers with multiple multicore processors and large main memory have shortened the run-times of heavy-duty computation. There is a concern, however, that big-data computing might slow down in the future due to storage and network bottlenecks.

Nonetheless, more and more private enterprises and public organizations will actively utilize big data. In order to solve various issues and to analyze and use the accumulated data effectively, a fresh approach to servers and storage systems is necessary. More specifically, it is necessary to increase storage capacity, ensure scalability, and select the right storage architectures for different types of data.

1.1 Challenges Regarding Storage Capacity and its Scalability

In order to increase the service capability per watt of electricity consumed, hyperscale datacenters operating thousands to tens of thousands of servers are facing the challenge of increasing the server storage density. Servers and its peripherals need to support large-capacity storage from a technical point of view. There are two data storage architectures to increase storage capacity in order to handle data growth: scale-up and scale-out. Generally, it has been considered that a scale-out (distributed) model provides an easier means of adding more resources.

At the same time, it has been pointed out that, as the data volume increases, the conventional scale-out storage architecture will suffer from bottlenecks caused by the processing of metadata and consequently degrade system performance. In order to meet the future needs of not only datacenters but also IT departments of private enterprises, there is a strong likelihood that it will be necessary to combine both high server performance and high storage capacity and yet at the same time achieve high storage density.

1.2 Challenges Regarding Different Types of Data and Storage Characteristics

The presence of diverse data types is also a major consideration. Conventionally, block and file interfaces have been widely used as storage interfaces. This is because most data handled in private enterprises have been structured data, and thus tape and dedicated archive storage systems have been used for long-term data storage and backup. These systems and approaches have been effective in handling structured data, but are not suitable for unstructured data that do not have a predefined data size, come in various types and accumulate at different speeds.
For example, once stored, most unstructured data are seldom accessed. Among such unstructured data are process data collected at manufacturing plants, and data gathered by various sensors. In contrast, the content in Favorites and Hot Video Clips available with video streaming services is accessed more frequently and thus requires high input/output per second (IOPS).

While large-capacity, low-cost storage suffices to archive seldom accessed data, adequate capacity, high access performance and high availability are necessary to archive frequently accessed data. In terms of investment cost and management, it is not advisable to utilize high-performance and high-availability storage for large, but infrequently accessed data.

Therefore, the IT departments in private enterprises, datacenters and service providers face the challenges of using the right kinds of storage systems according to the types and purposes of structured and unstructured data.

2 Direction of Solutions

A new server storage system is necessary to solve various problems with the conventional storage. As the volume of unstructured data handled in private enterprises is increasing rapidly, they are expected to become predominant. In these circumstances, possible solutions are as follows:

2.1 Increasing Storage Capacity and Ensuring Scalability

In order to continue to increase the per-disk capacity, hard disks that exploit an advanced technology such as shingled magnetic recording (SMR) must be continually released to the market. The deduplication and compression technologies also help increase storage capacity. Incorporating these technologies into a host system has a significant effect. Technologies, such as SMR, are necessary to implement large-capacity storage.
Besides an increase in the per-disk capacity, a new type of storage called Ethernet drives has emerged. An Ethernet drive is a storage node that incorporates a processing unit and a network interface and is directly connected to an IP network. The processing unit can be implemented as a system-on-a-chip (SoC) and run on Linux OS. This way, an Ethernet drive will be able to run application software according to the SoC performance. With the server and storage functions physically combined, Ethernet drives may possibly help increase server rack density.

High density is not the only advantage of Ethernet drives. Since individual Ethernet drives are managed as separate storage nodes, the number of Ethernet drives can be scaled out repeatedly without causing any bottleneck in the processing of metadata, as long as the performance limit of each SoC is not exceeded. Thus, the overall storage capacity can be increased efficiently. Furthermore, in the event of a hardware failure, it is necessary to replace only the failed Ethernet drive. Shutting down a single Ethernet drive for the purpose of maintenance causes less impact than shutting down a server.

Ethernet drives and other new types of storage nodes are required for hyperscale datacenters, and micro datacenters that will be constructed in the future.

2.2 Selecting the Right Types of Storage for Different Types of Data

While solid-state drives (SSDs) consist of high-capacity, high-speed NAND flash memories, hard disk drives (HDDs) are slower but excel in disk cost per gigabyte ($/GB). First of all, it is important to appropriately use them according to the required latency.

Second, the optimal interface should be selected from block and file storage devices, as well as object storage devices that have recently been utilized in datacenters as discussed later.

In order to increase storage capacity to handle data growth, datacenters, service providers and the IT departments of private enterprises have started to use distributed storage systems. In general, distributed storage systems are scaled out rather than scaled up, and support not only block and file interfaces but also an object interface.

A third consideration is software-based storage tiering, which makes it possible to store different types of data into different types of drives. The overall system efficiency and performance can be improved by storing unstructured and structured data into object and block storage devices, respectively.

3. Object Storage

Cloud storage services are rapidly becoming pervasive since they are less costly and provide higher fault tolerance in the event of a natural disaster than on-premises storage systems. In order to meet multi-tenancy requirements and offer high-capacity storage services at low costs, datacenters now utilize object storage devices in addition to the conventional block and file storage devices.

Object storage is configured in a distributed architecture. Since object storage uses Key Value interface commands, it eliminates the need for driver software comprising block devices and filesystems. Furthermore, object storage uses a flat structure and has a wide namespace. For these reasons, object storage devices are better suited for the handling of unstructured data than block and file storage devices.

Regardless of the type of storage media, object storage devices manage data as objects as opposed to block storage devices that use logical block addressing (LBA) and file storage devices that use a filesystem (NFS, CIFS, etc). Moreover, multiple low-cost disks can be configured as a large-capacity storage pool by using a RAID or other disk array controller.

However, object storage devices have a drawback in that they only determine whether the written data has actually been recorded in a medium and thus provide "eventual consistency" only. In other words, object storage devices check if replica data, which are distributed over several storage nodes for redundancy reasons, are eventually consistent.

Because of the foregoing characteristics, object storage devices are primarily suitable for the management of large but infrequently used data. More specifically, they are well suited for use as archive storage (cold storage) for the cloud, video distribution websites and social networking services (SNS). Object storage is expected to be adopted not only for cloud services but also for IT systems in major enterprise companies.

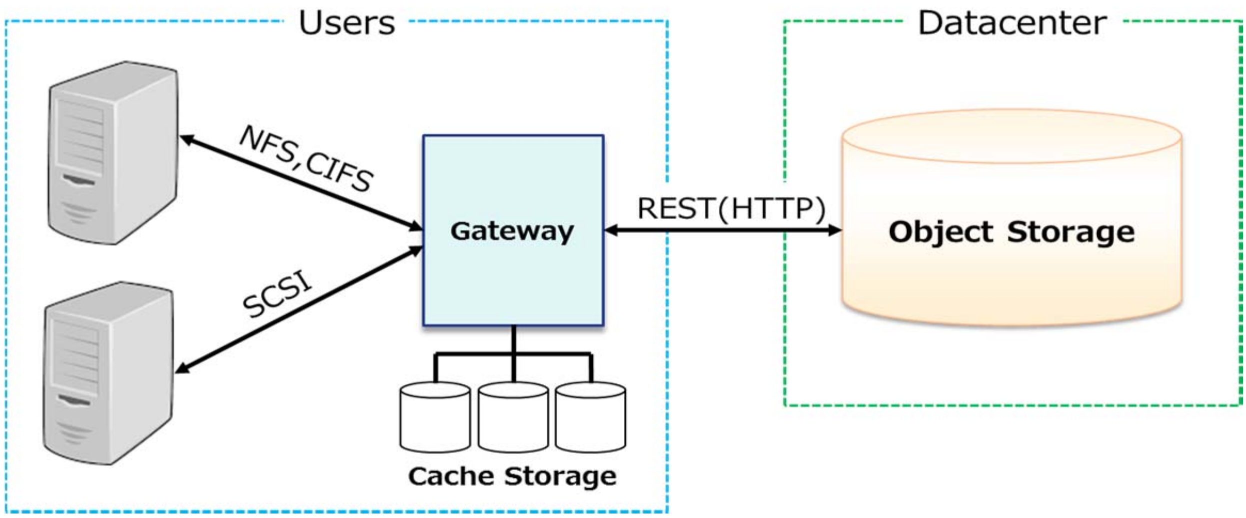
Application software designed for block and file storage devices is not compatible with object storage devices that use Key Value interface commands. Therefore, application software must be modified, or needs to send data to an object storage device via a file or block gateway.

	Block Storage	File Storage	Object Storage
Interfaces	SCSI	NFS, CIFS, etc.	HTTP (REST)
Advantage	High access speed	Ease of access, file sharing	Distributed storage Scalability
Namespace	-	Hierarchical	Flat
Resource Allocation	Not required	Automatically allocated by storage	Automatically allocated by storage
Data consistency	High	High	Low
Security	Low Per volume	High Per file	High Per object
Scalability	Low (up to hundreds of terabytes per volume)	Limited by overhead to maintain namespace consistency	Very large space (hundreds of petabytes)
Applications	Processes involving frequent updates, transaction processing	Shared files, general data	Infrequently updated files Archiving, backup

4 Applications of Object Storage Systems

Major object storage systems include Amazon Simple Storage Service (Amazon S3) of Amazon Web Services, the BLOB Storage service of Microsoft Azure, and Swift of the OpenStack open-source cloud computing software platform.

One of the ways to make it possible for the applications designed for block and file storage architectures to access an object storage device is by adding a gateway to the back-end object storage. The purpose of this is to allow these applications to access a gateway via a file or block interface, and gateway and back-end storage services using REST (Http). In cases where a gateway is provided as a computer appliance, cache storage is provided in the gateway and infrequently accessed data are moved to an object storage device. This way, both investment efficiency and computing performance can be improved.



5 Toshiba's Stance on Object Storage

This section briefly describes how Toshiba views object storage and how it is tackling it.

At present, the most commonly used architecture is server-based object storage in which HDDs and/or SSDs are connected to servers. One of the advantages of this architecture is that a system can be configured by combining existing hardware components. Toshiba has sold enterprise and other server HDDs and SSDs, and will continually expand its portfolio of server HDDs/SSDs compatible with object storage architecture.

On the other hand, server-based object storage has a multi-layered hierarchy, including server KVS, filesystem, device driver, and SSD/HDD data management. Because of this, server-based object storage does not always improve efficiency. For users who want to improve the object storage efficiency, Toshiba offers a new kind of product called Key Value drives.

6 Key Value Drives

This section provides an overview of Key Value drives, hardware specifically designed to realize Toshiba's object storage concept.

6.1 What Is a Key Value Drive?

In general data storage, data consisting of a series of fixed-length blocks of 512 bytes each are identified by their logical block address (LBA).

In contrast, in a Key Value drive, data are identified by a key rather than an LBA. Both keys and data (called values in this parlance) can be of variable length (Figure 1).

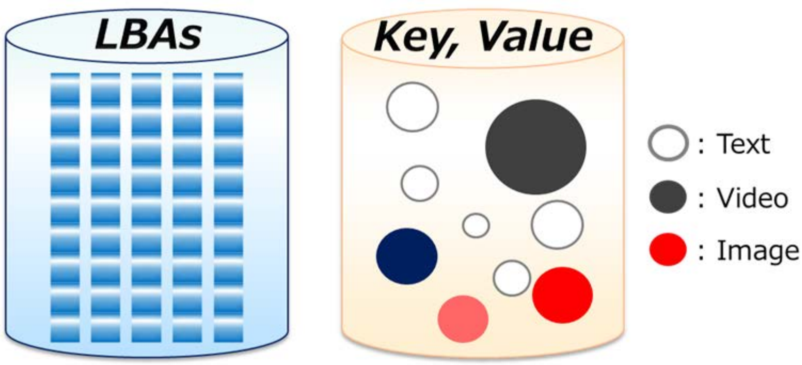


Figure 1 Comparisons between General and Key Value Drives

Whereas general drives are primarily connected to a server via the SATA, SAS, PCIe or other interface, Key Value drives are implemented as a kind of Ethernet drives and thus directly connected to a network via Ethernet. Consequently, each Key Value drive can act as a node and perform part of communications directly with clients without involving a server. In addition to Ethernet, Key Value drives can be configured with PCIe and other interfaces.

6.2 Benefits of Key Value Drives

Compared to server-based object storage, object storage using Key Value drives provides the following benefits:

Replacement for storage servers
Since Key Value drives are directly connected to a network, general data transfers can be directly performed between clients and drives. Servers only need to fulfill the functions that are not available with Key Value drives such as metadata management. Therefore, the number of servers can be reduced.

Small form factor
Key Value drives are available in smaller form factors than conventional servers (e.g., 3.5-inch). Thus, Key Value drives provide higher storage density than server-based object storage.

Fewer storage hierarchy layers
Server-based object storage needs many hierarchy layers as described above. In contrast, object storage using Key Value drives can combine multiple layers into one as necessary in order to improve performance, reduce cost and extend service life.

Early support for new types of media
In order for the conventional storage to efficiently utilize a new generation of flash memory or shingled magnetic recording (SMR) HDDs, host software must be updated or storage interfaces must be standardized. In contrast, only Key Value drives need to be modified to accommodate a new technology; thus it is possible to quickly respond to user needs.

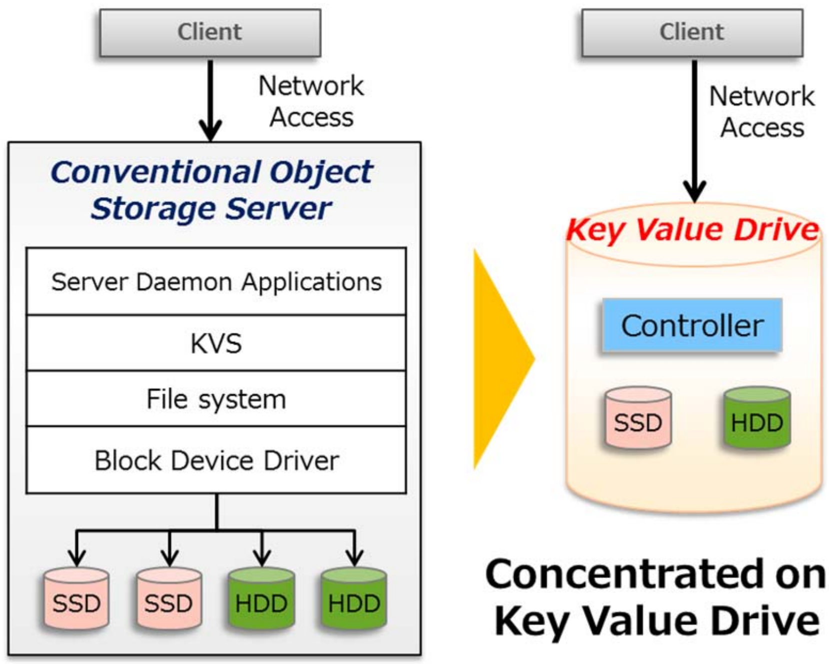


Figure 2 Server-Based Object Storage vs. Key Value Drive

6.3 Toshiba's Key Value Drive Solution

Toshiba is exploring the possibility of offering various Key Value drives to meet different market needs. This section introduces HDD-based drives and hybrid drives.

Key Value Drive C (Capacity)

Key value drive C are a kind of Key Value drives that exploit the high-capacity advantage of HDDs and achieve a lower total cost for HDDs and servers.

Although Key value drive C has the same outline dimensions as a typical 3.5-inch HDD, many Key value drive C can be housed inside an enclosure. Key value drive C is connected to an Ethernet server via a network switch.

Key value drive C supports the OpenKinetic API for server applications.

Key Value Drive P (Performance)

While Key value drive C places higher priority on storage capacity, key value drive P is more suitable for applications that require higher performance.

Key value drive P consists of HDDs and SSDs, which are hierarchically tiered for optimal control. Key value drive P can be read and written at high speed without any concern about the differences between HDDs and SSDs.

Even though key value drive P contains both HDDs and SSDs, it can fit in the 3.5-inch HDD form factor like HDD-based drives. Thus, key value drive P can be housed in the same enclosure as typical HDDs. (In order to improve performance, it is necessary to secure sufficient bandwidth for a network switch on the enclosure.)

Furthermore, object storage software can be run directly on key value drive P instead of on a server. This helps reduce the number of servers required.

7 Software

This section briefly describes the software technology to realize Toshiba's object storage concept.

7.1 Software Organization

In Toshiba's Key Value drives, the OpenKinetic protocol is the major API. Toshiba intends to make active contributions to the OpenKinetic protocol development.

The object storage software that runs on a server uses the OpenKinetic API via a network to perform reads and writes on Key Value drives.

Moreover, key value drive P can also run object storage software. In this case, object storage software performs communications within key value drive P to utilize the OpenKinetic API. Thus, key value drive P can run OpenKinetic-compatible object storage software for servers without any modification.

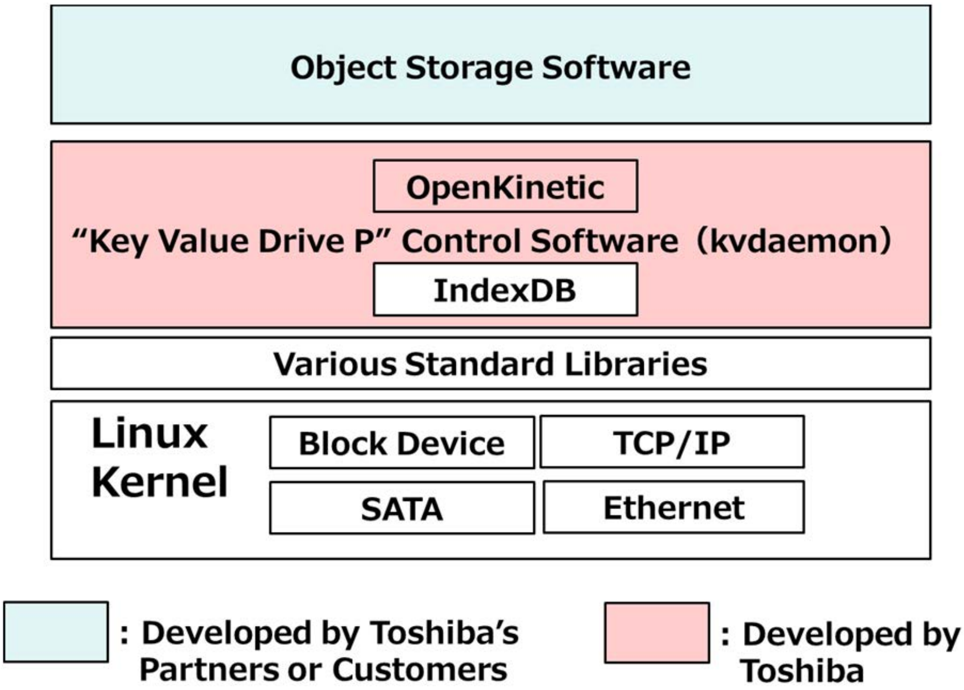


Figure 3 Software Organization in a Key Value Drive P

7.2 Ceph

Toshiba emphasizes the support of Ceph as object storage software to be used on key value drive P. Ceph is an open-source software platform for distributed object storage systems and supports not only an object interface but also block and file interfaces. It is widely used by OpenStack implementors. In contrast to many object storage systems that provide only eventual consistency, Ceph guarantees that all replicas are updated at once. Therefore, Ceph can be used for enterprise applications that demand strong consistency.

7.2 Support of Other Object Storage Software

Toshiba is working closely with several other object storage software vendors. Object storage and other software can be installed and run on key value drive P. Toshiba is cooperating with its partners to make various software available for different applications. For key value drive P, Toshiba provides a development environment to its partners. It is designed to support the development of key value drive P-compatible software from the build of software to its implementation on key value drive P. It aims to make it easier for our partners to develop key value drive P-compatible software.

Note: The platform herein called "Open Kinetic" is formally known as "Kinetic Open Storage Platform.

RESTRICTIONS ON PRODUCT USE

- Toshiba Corporation, and its subsidiaries and affiliates (collectively "TOSHIBA"), reserve the right to make changes to the information in this document, and related hardware, software and systems (collectively "Product") without notice.
- This document and any information herein may not be reproduced without prior written permission from TOSHIBA. Even with TOSHIBA's written permission, reproduction is permissible only if reproduction is without alteration/omission.
- Though TOSHIBA works continually to improve Product's quality and reliability, Product can malfunction or fail. Customers are responsible for complying with safety standards and for providing adequate designs and safeguards for their hardware, software and systems which minimize risk and avoid situations in which a malfunction or failure of Product could cause loss of human life, bodily injury or damage to property, including data loss or corruption. Before customers use the Product, create designs including the Product, or incorporate the Product into their own applications, customers must also refer to and comply with (a) the latest versions of all relevant TOSHIBA information, including without limitation, this document, the specifications, the data sheets and application notes for Product and the precautions and conditions set forth in the "TOSHIBA Semiconductor Reliability Handbook" and (b) the instructions for the application with which the Product will be used with or for. Customers are solely responsible for all aspects of their own product design or applications, including but not limited to (a) determining the appropriateness of the use of this Product in such design or applications; (b) evaluating and determining the applicability of any information contained in this document, or in charts, diagrams, programs, algorithms, sample application circuits, or any other referenced documents; and (c) validating all operating parameters for such designs and applications. **TOSHIBA ASSUMES NO LIABILITY FOR CUSTOMERS' PRODUCT DESIGN OR APPLICATIONS.**
- **PRODUCT IS NEITHER INTENDED NOR WARRANTED FOR USE IN EQUIPMENTS OR SYSTEMS THAT REQUIRE EXTRAORDINARILY HIGH LEVELS OF QUALITY AND/OR RELIABILITY, AND/OR A MALFUNCTION OR FAILURE OF WHICH MAY CAUSE LOSS OF HUMAN LIFE, BODILY INJURY, SERIOUS PROPERTY DAMAGE AND/OR SERIOUS PUBLIC IMPACT ("UNINTENDED USE").** Except for specific applications as expressly stated in this document, Unintended Use includes, without limitation, equipment used in nuclear facilities, equipment used in the aerospace industry, medical equipment, equipment used for automobiles, trains, ships and other transportation, traffic signaling equipment, equipment used to control combustions or explosions, safety devices, elevators and escalators, devices related to electric power, and equipment used in finance-related fields. **IF YOU USE PRODUCT FOR UNINTENDED USE, TOSHIBA ASSUMES NO LIABILITY FOR PRODUCT.** For details, please contact your TOSHIBA sales representative.
- Do not disassemble, analyze, reverse-engineer, alter, modify, translate or copy Product, whether in whole or in part.
- Product shall not be used for or incorporated into any products or systems whose manufacture, use, or sale is prohibited under any applicable laws or regulations.
- The information contained herein is presented only as guidance for Product use. No responsibility is assumed by TOSHIBA for any infringement of patents or any other intellectual property rights of third parties that may result from the use of Product. No license to any intellectual property right is granted by this document, whether express or implied, by estoppel or otherwise.
- **ABSENT A WRITTEN SIGNED AGREEMENT, EXCEPT AS PROVIDED IN THE RELEVANT TERMS AND CONDITIONS OF SALE FOR PRODUCT, AND TO THE MAXIMUM EXTENT ALLOWABLE BY LAW, TOSHIBA (1) ASSUMES NO LIABILITY WHATSOEVER, INCLUDING WITHOUT LIMITATION, INDIRECT, CONSEQUENTIAL, SPECIAL, OR INCIDENTAL DAMAGES OR LOSS, INCLUDING WITHOUT LIMITATION, LOSS OF PROFITS, LOSS OF OPPORTUNITIES, BUSINESS INTERRUPTION AND LOSS OF DATA, AND (2) DISCLAIMS ANY AND ALL EXPRESS OR IMPLIED WARRANTIES AND CONDITIONS RELATED TO SALE, USE OF PRODUCT, OR INFORMATION, INCLUDING WARRANTIES OR CONDITIONS OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE, ACCURACY OF INFORMATION, OR NONINFRINGEMENT.**
- Do not use or otherwise make available Product or related software or technology for any military purposes, including without limitation, for the design, development, use, stockpiling or manufacturing of nuclear, chemical, or biological weapons or missile technology products (mass destruction weapons). Product and related software and technology may be controlled under the applicable export laws and regulations including, without limitation, the Japanese Foreign Exchange and Foreign Trade Law and the U.S. Export Administration Regulations. Export and re-export of Product or related software or technology are strictly prohibited except in compliance with all applicable export laws and regulations.
- Please contact your TOSHIBA sales representative for details as to environmental matters such as the RoHS compatibility of Product. Please use Product in compliance with all applicable laws and regulations that regulate the inclusion or use of controlled substances, including without limitation, the EU RoHS Directive. **TOSHIBA ASSUMES NO LIABILITY FOR DAMAGES OR LOSSES OCCURRING AS A RESULT OF NONCOMPLIANCE WITH APPLICABLE LAWS AND REGULATIONS.**