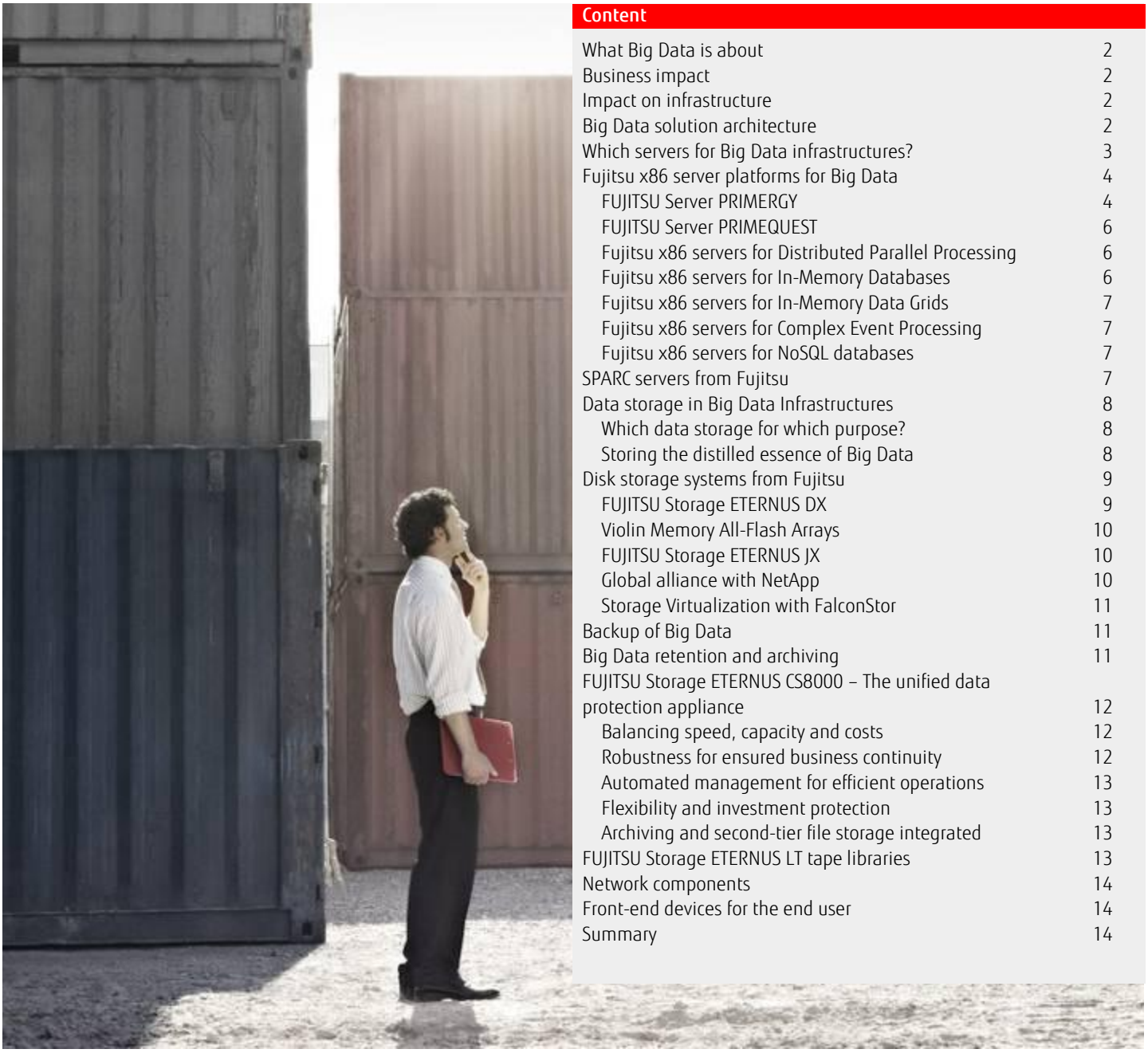
White paper
Infrastructure Products from Fujitsu – Ideal for Big Data

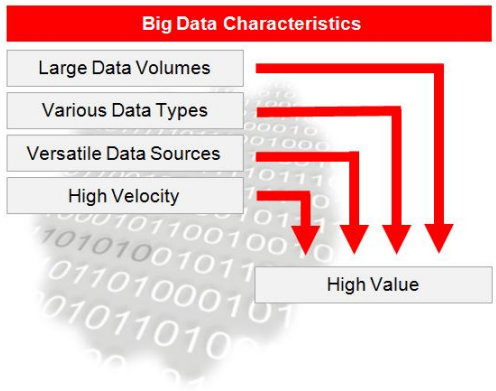
In order to exploit the potential value of Big Data, infrastructure solutions optimally suitable for the customer-specific use case are a mandatory prerequisite. However, infrastructures can only be as good as their ingredients. What should you look out for when building a Big Data infrastructure? What are the requirements with regard to the infrastructure products used? This whitepaper illustrates the most significant aspects.



Content

What Big Data is about

Big Data is about large data volumes in the range of many Terabytes to Petabytes coming from versatile internal and external data sources, such as databases, audio and video files, e-mails, IT logs, web logs, social media and the millions and billions of sensors which are built in everywhere. Especially new sources provide data which rarely is structured, but rather unstructured, semi-structured and poly-structured. Very often this data is continuously generated at a very high velocity, and the same high velocity is also demanded when it comes to analyzing the data in real-time to create value for the organization. Sometimes data expires at the same high speed as it is generated.



Business impact

By discovering facts and insights about customers and suppliers, market conditions, operations and root causes of costs are revealed, that would otherwise remain hidden. This helps you predict things that are going to happen in future, take the right decisions, and initiate the respective actions. Betting on the right horse, skipping useless activities, and optimizing the allocation of resources will increase performance and productivity, accelerate business processes, minimize risks, improve efficiency, effectiveness and profitability and finally give you the competitive edge that you need in today's tough business world.

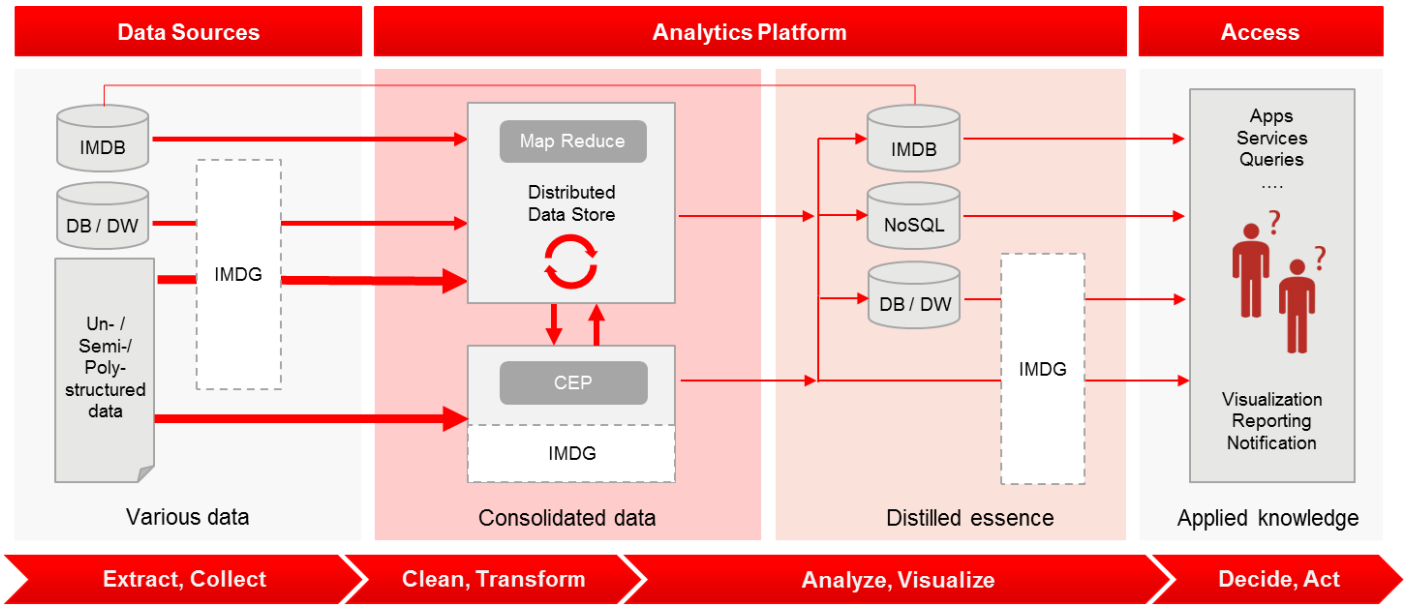
Impact on infrastructure

Keeping processing time constant and short, while data volumes increase, along with meeting real-time demands – all in an affordable way – are serious challenges. These in turn have a strong impact on the infrastructure for Big Data. Traditional approaches with relational databases and server scale up and server scale out will sooner or later reach their limits. Therefore new approaches are required.

By means of Distributed Parallel Processing (based on the so-called MapReduce algorithm) you distribute the data and the I/O load across the local disks of a server cluster and move the computing tasks exactly to the server nodes where the data resides. Everything can then happen in parallel, nothing is shared; and therefore scalability is basically unlimited. This concept is often used to transform data into a shape which is more appropriate for ad-hoc analysis in real-time. The distilled essence of the transformed data can then be exported to a database for rapid access. Using an All-Flash-Array (AFA,) which is designed for flash memory, instead of a hard disk array for the database will accelerate I/O. By using an In-Memory Database (IMDB) you will eliminate I/O totally when analyzing data, because all data is in memory, just as data operations. Thus, ad-hoc queries can be executed in real-time. To accelerate analytics applications with data of any type and any size on disk storage, an In-Memory Data Grid (IMDG) will be the solution to meet real-time demands, because I/O can be drastically reduced. NoSQL databases instead of relational databases enable a faster analysis and even a better compression. For data in motion, e.g. event streams continuously generated at a high frequency, Complex Event Processing (CEP) engines take care of collecting and processing the event streams and triggering the respective action timely.

Big Data solution architecture

The figure below shows the overall Big Data solution architecture. Data at rest of various types is extracted from versatile data sources, collected and loaded into the Distributed Data Store, optionally after having been cleaned. Distributed Parallel Processing based on the MapReduce algorithm is used to transform data into a usable form and for analysis.



The distilled essence of the transformed data can be exported to a SQL or NoSQL database management system.

Real-time analysis is enabled by in-memory data platforms. An IMDB will be the fastest option, especially when new queries are created ad-hoc, again and again. If the size of the distilled essence exceeds the totally available main memory capacity, a disk array with an IMDG in front can be used to accelerate the analytics applications. Another advantage of IMDG is that is applicable to any data objects. Of course, an IMDG can also help reduce I/O, if the analysis is directly applied to the Distributed Data store.

If a traditional disk storage array does not provide the required I/O performance, .a fast storage system, as for instance an AFA can be used to increase throughput and reduce latency, thus accelerating I/O.

Event streams generated at a high velocity are collected in the data store of a CEP (Complex Event Processing) engine, which in turn will look after analyzing the event streams and initiating the respective actions in real-time. As CEP is extremely time-critical, the data store is memory-resident, and in case of large event streams even implemented as an IMDG. If a large number of rules or very complex rules need to be completed, CEP engines will be distributed to several servers in order to enable parallel processing.

Depending on the use case, results of CEP can be forwarded into the Distributed Data Store or the distilled essence for further processing, in the same way as data from the Distributed Data Store is sometimes used for CEP.

An infrastructure can only meet the requirements for Big Data, if the perfect ingredients are used. But what are the characteristics expected from infrastructure products, such as servers, storage systems and networking components?

Which servers for Big Data infrastructures?

A Big Data infrastructure may comprise various infrastructure concepts. Therefore the server requirements will also vary depending on the concept; and even within an individual concept different servers for different purposes can be included.

Let us start with **Distributed Parallel Processing**, where a couple of master nodes distribute the data and the processing tasks across a cluster of many slave nodes. It is evident, that due to their large number, the slave nodes should be space-saving, energy efficient and affordable. Moreover, a modular design with regard to hard disk drives and network connections in order to adapt the solution to new requirements represents an additional advantage. Solid State Disks are at the most an option for buffering temporary results in order to accelerate the Map-Reduce algorithm, but due to its higher costs such use cases might be limited today.

As with Hadoop Open Source Software, which is typically used for Distributed Parallel Processing, data can be replicated to several nodes, high availability is no key criterion for the slave nodes.

This is different with the master nodes which are in charge of managing all the metadata and distributing data and tasks to the slaves. Here, high availability features, such as RAID for their hard disks, redundancy of power supplies or other system components, and an adequate compute performance and main memory capacity to be able to deal with the given number of slave nodes is definitely essential.

For **In-Memory Databases (IMDB)** high compute performance with large main memory capacities, as well as high availability are key requirements. For single-node solutions, fast local disks, e.g. solid state disks (SSD), are required. For multi-node solutions extremely fast connections to the external storage on which the persistence layer resides, are a basic prerequisite. In general, the demands regarding the connection between server and storage are difficult to specify, because this also depends on the respective IMDB. Taking SAP HANA as an example, SAP conducts heavy load tests with appliances of their hardware partners and decide based on the results, whether a certification is justified or not. But be sure, that 10GbE with NAS (Network Attached Storage) respectively 8 or even 16 Gbps with SAN (Storage Area Network) is a mandatory prerequisite. Furthermore, certain IMDB solutions, such as SAP HANA, are only certified for special processors from Intel.

With **In-Memory Data Grid (IMDG)** solutions, main memory capacity and high availability to ensure business continuity without performance losses is of course essential. Furthermore, it is recommended to equip the servers with fast solid-state disks which serve as a persistence layer for the data in the main memory of the server. Due to log files and other information, the required SSD capacity must be considerably larger than the main memory size. In the case of Terracotta BigMemory from Software AG, the ratio of SSD-to-memory capacity should be 3:1. Therefore servers are needed with an optimum balance between a respective configuration and cost. As with the other Big Data infrastructure concepts, data center space and energy efficiency is important also with IMDG.

For **Complex Event Processing (CEP)**, the required compute performance strongly depends on the workload to be expected, but a reliable multi-socket server is always recommended. As CEP processes data in the memory, the memory size should be aligned to the amount of data to be processed, and the speed at which data is coming in. With increasing workload, processors and main memory capacity can be added, as long as you have not reached the server's limits (scale-up). If scale-up is not sufficient, many CEP engines allow for scale-out by adding servers.

If the data volume that needs to be kept resident in memory for a certain time window exceeds the maximum memory capacity, an IMDG will help.

In addition to the servers running the CEP engines, it can make sense to run the application that provides input streams to the CEP engines on a dedicated server. Depending on how many input events need to be handled, several of these servers may be required.

How the **database servers** and **application servers** should be equipped, depends strongly on the application and the data volume the application will access. Every server size – 2-way, 4-way, 8-way or even 1-way – is imaginable. It goes without saying that main memory capacity is a critical parameter, and a large memory will always have a positive impact on the system performance. However, also here a balance between configuration and price is desirable.

Finally, your servers of choice should support or even be certified for Linux, because Linux is the preferred and most frequently used operating system in the world of Big Data.

As far as servers for **NoSQL databases** are concerned, it depends on the category of NoSQL databases we are talking about. With document stores, key value stores and column stores we deal with sets of data which are well separated from each other, and can be assigned to the most important transactions, such as user profiles, shopping cart, order, e-book. Most appropriate for this scenario might be dual-socket x86 servers with large local disk capacity.

With graph databases, another category of NoSQL databases, the situation looks totally different. Graph databases are used for strongly cross-linked objects, whose relation to other objects represents the essential information. Examples are web sitemaps, friendships in Facebook, street maps, or net schedules. In order to be able to navigate rapidly along the edges of the graph which represent the relations, a scale-up approach without interaction in a cluster is more promising and easier to achieve than a scale-out approach. Graph databases are typically optimized for representation of nodes and edges without much overhead, compared to a relational representation of graphs; so with billions of related objects fitting into main memory of modern servers scale-up will not hit any limit for most problems. For such NoSQL databases, an in-memory platform based on a multi-processor system with a large memory size would be beneficial.

As the server requirements vary depending on the use case, so much the more important it is to use a server family with a variety of models that can cover all these different requirements and ensures an optimum infrastructure solution according to the customer-specific situation.

Fujitsu x86 server platforms for Big Data

Both x86 server families from Fujitsu, the PRIMERGY and the PRIMEQUEST servers find their place in Big Data infrastructures. Let us first have a closer look at PRIMERGY.

FUJITSU Server PRIMERGY

PRIMERGY servers convince of their rock-solid reliability, leading price performance and energy-efficiency, and are thus also the ideal platform for Big Data. The latest industry standard Intel processor technology is perfectly integrated in PRIMERGY servers, proven by several world record benchmarks.

Choice in any situation

Whatever your operational requirements are, there is a PRIMERGY server that optimally fits your needs. The PRIMERGY product line includes blade systems, cloud servers, rack-mountable servers and stand-alone tower servers. Thus customers can choose the most adequate system for their specific use case. The broad range includes products from one up to four processors with up to 15 cores per CPU.

Business proven quality

Operational continuity is a key element in the durable success of any organization, and to a large extent it is dependent on the reliability of your IT infrastructure. The value delivered by PRIMERGY servers is founded on their business proven quality. This is derived from the close integration of our engineering and production centers and their cooperative quality management. The result is extremely low failure rates. PRIMERGY quality standards are comparable to those of high end Unix servers and ensure high systems availability and long usage cycles, which pay dividends in the form of favorable lifecycle costs.

Performance

If you can operate reliably, then IT performance is the next priority. Of course, CPU performance is a good foundation for an efficient operation. But it is always important that main memory size and I/O throughput are aligned to it. This applies absolutely to Fujitsu PRIMERGY servers. They feature a maximum memory capacity of up to 6 TB. Highest processor performance is only useful in combination with tailored I/O throughput. For this demand, Fujitsu offers a wide range of different connection possibilities to address current and next-generation I/O requirements for standards-based Ethernet, Fibre Channel and InfiniBand networks.

The market-leading overall performance of Fujitsu PRIMERGY servers has been regularly confirmed for many years by top results in the most important system performance benchmarks. PRIMERGY servers also hold leadership positions in terms of overall price / performance, for example in $/tpsE ($ per transaction according to the TPC-E transactional benchmark suite). All these results, achieved across all PRIMERGY server classes, prove that PRIMERGY servers deliver exceptional value for money.

Efficient management

Every server needs to be deployed, integrated into its environment, administered and monitored. Especially for large server clusters, as with Distributed Parallel Processing and Hadoop, this can be very difficult. In order to simplify and accelerate the administration tasks, every PRIMERGY server is shipped with FUJITSU Software ServerView® Suite which enables local and remote management of PRIMERGY based infrastructures, thus improving customer productivity and cost-effectiveness.

ServerView® Suite contains all the elements required for the professional management of server systems throughout their lifecycle. All products of the ServerView® Suite come from one source and are thus optimally integrated. ServerView® Suite even enables a simple integration of PRIMERGY servers with the most relevant enterprise management systems or vendor-specific server management solutions.

Thanks to the integration with Nagios, a popular open source tool for IT infrastructure monitoring, PRIMERGY servers have already got high significance in the open source world, apart from being used in Linux-based Hadoop clusters.

Energy efficiency – more performance per Watt

The more servers are involved in an infrastructure, the more important becomes energy efficiency, because many data centers are restricted as far as the available space, maximum energy consumption and heat dissipation are concerned.

The unique and proven Cool-safe® technology in PRIMERGY servers ensures that up to 25% more air can be "inhaled" than in comparable products. This reduces the risk associated with heat failures and doubles the lifetime of electronic components. Due to Cool-safe®, Fujitsu was able to build the world´s first server using less than 1 Watt per transaction as measured by Watts / tpsE. Cool-safe® Advanced Thermal Design enables the operation in a higher ambient temperature. The temperature in a data center can thus be increased by a few degrees without any problems. Each additional degree means approx. 5-6 % less energy costs for air-conditioning.

Also worth mentioning is the **ServerView Power Management**, which helps maximally increase your energy efficiency by throttling or switching off power consuming components. This is particularly interesting, if not all servers of the Big Data infrastructure are active 24 hours a day.

Low energy consumption means less energy costs and reduced cooling costs at the same time. The energy budget can thus cover much more PRIMERGY server workload without wasting cooling capacities. And not to forget, you will contribute to a smaller carbon footprint.

Supply Chain Flexibility

Flexibility is a key value in the PRIMERGY supply chain. This is supported by Fujitsu's customized production and logistic services approach "made4you". The made4you service can include pre-loading software, the integration of third party components, providing customized consignment stocks and specify rollout arrangements at all stages of the value chain. Our standard in production is Built-to-Order, ensuring that tailor-made PRIMERGY configurations are manufactured, tested and delivered to 100% of the functionality defined in the customer order. This even includes the shipment of fully mounted rack server configurations, so there is no need for your IT team to re-assemble loose servers or components. As a result, cost and time to operation are significantly reduced.

Access to the right services is just as important as getting the right product. Fujitsu has 80,000 specialists around the world able to provide services in a standardized way on a global basis. These are just some of the many reasons why you can be confident that PRIMERGY servers will deliver an outstanding return on your investment faster than you might have imagined.

FUJITSU Server PRIMEQUEST
PRIMEQUEST are x86 high-end server systems and stand for industry leading x86 quality and superior technology from Fujitsu. The Fujitsu Server PRIMEQUEST 2800B is the successor of the PRIMERGY RX900 S2, offering an unprecedented performance with the help of up to 8x Intel® Xeon® E7 v2 processors and memory capacity combined with high reliability thanks to built-in advanced RAS features for advanced error circumvention. Featuring up to 120 cores and 12 TB of memory, PRIMEQUEST is ideal for big data applications and in-memory solutions like SAP HANA

It is true that Fujitsu PRIMERGY servers feature high availability. But if you are running applications which are absolutely mission-critical for your enterprise and planned or unplanned downtime is not acceptable Fujitsu's PRIMEQUEST servers are the products of choice. They let customers take advantage from the flexibility and the economic benefits of x86 industry standards, without compromising their business continuity demands: FUJITSU Server PRIMEQUEST 2400E and 2800E feature self-healing capabilities and unique features such as Dynamic Reconfiguration which result in outstanding platform reliability, best availability and serviceability. The Dynamic Reconfiguration feature allows the changes of resources while systems are still up and running. Isolated physical partitions enable high availability, where any hardware failures do not have an impact on other partitions. In the unlikely event of a failure, the affected system board can be automatically replaced by a reserved system board and thus instantly recover applications and services.

Fujitsu x86 servers for Distributed Parallel Processing
In standard Hadoop cluster environments, Fujitsu's dual-socket **PRIMERGY RX200 and RX300** rack servers are ideal to be used as slave nodes and master nodes due to their versatile and modular design. Like the whole PRIMERGY family, they excel in quality, great price / performance ratio, cost-effective management and energy efficiency.

PRIMERGY RX200 and RX300 feature a high level of adaptability, thus making cost-optimized system configuration possible in accordance to the respective demands. Hard disk or SSD drives once configured can be expanded anytime. The modular LAN provides a standardized, low-priced extension option to the existing onboard LAN controller; this supports technology changes at any time and allows for an individual configuration of 1Gbit or 10Gbit ports. The modular power supply unit offers standardized power units with an efficiency of up to 94% (80Plus Platinum).

The modular RAID offers a low priced and standardized upgrade option in addition to the onboard RAID solution. Although Hadoop is explicitly designed to cope with slave node failures, a Hadoop slave node will profit in availability from using RAID for the operating system and the local file system. For the master node, RAID is strongly recommended.

Besides PRIMERGY RX200 and RX300, the multi-node server **PRIMERGY CX400** offers maximum compute power per height unit, thus making it ideal for being used as slaves in large Hadoop cluster environments. PRIMERGY CX400 represents a perfect combination of space saving high density in a standard rack cabinet ("4in2U" principle) and highly efficient shared power supply units. Each of the four independent dual socket **PRIMERGY CX250** server nodes within a PRIMERGY CX400 chassis provides high memory capacity and, indispensable for the Hadoop slave server nodes, up to six HDD or SSD drives. High speed inter-communication between the slave nodes (e. g. via 10Gb Ethernet) allows for their optimal collaboration. All in all, utilizing PRIMERGY CX400 as Hadoop slave server nodes leads to an optimum balance between disk capacity and CPU performance, and hence to an optimized operation of Hadoop tied Big Data projects.

For extremely large clusters, our 4-socket server PRIMERGY RX4770 or even a PRIMEQUEST server, which we are going to have a look at in the next section, could be an option as a master node. Running the NameNode of HDFS or the MapReduce JobTracker on PRIMEQUEST will increase the overall reliability and availability.

Fujitsu x86 servers for In-Memory Databases
4-socket and 8-socket x86 servers have always been a preferred platform for in-memory computing and its processing loads, since they combine advanced high availability server features such as memory mirroring support, hot-plug memory boards with socket-overlapping memory mirroring, with effective top performance, rich memory dimension, broad expansion capacities and easy scalability.

The scalable **PRIMERGY RX4770 M1** is a 4-socket rack server which is in every respect a reliable server for in-memory processing purposes. The ideal interaction of integrated redundancy with server management components results in high-level availability and constantly efficient IT production as a character feature of this server platform. The latest generation of Intel Xeon processors E7 v2 equipped with up to 60 cores enable a unique performance boost which however does not bear optimal dividends unless paired with other features: a high expandable main memory capacity (up to 6 TB) with up to 96 DIMMs and a very large number of high performance PCI Express channels provide a balanced high throughput architecture so as best I/O performance to a central (flash) storage is ensured.

As with regard to in-memory databases, Fujitsu offers the single-node and multi-node Power Appliances for SAP HANA based on PRIMERGY and PRIMEQUEST, it is worth mentioning that both server models are equipped with Intel processors as specified by SAP. All Power Appliances for SAP HANA from Fujitsu are of course certified by SAP.

Fujitsu x86 servers for In-Memory Data Grids

For the In-Memory Data Grid software Terracotta BigMemory, where an SSD-to-memory ratio of 3:1 is recommended, the PRIMERGY servers RX200 and RX600 offer the best configuration options.

Fujitsu x86 servers for Complex Event Processing

In general, the PRIMERGY RX300 is sufficient for running the CEP engine, while the PRIMERGY RX200 is the platform of choice to provide input streams to the CEP engine. However, depending on the frequency of event streams being generated, the number of rules to be executed or their complexity, PRIMERGY or PRIMEQUEST models with a larger main memory capacity and higher compute performance can be an option.

Fujitsu x86 servers for NoSQL databases

Most appropriate for columnar stores, key-value stores and document stores are be dual-socket x86 servers with large local disk capacity, for instance Fujitsu's PRIMERGY CX400 chassis with 12 x 4 TB and 2 x PRIMERGY CX270 servers.

For graph databases representing billions of related objects which for performance reasons should fit into main memory, an in-memory platform based on PRIMERGY RX4770 with 6 TB or PRIMEQUEST with even greater memory capacity is recommended.

SPARC servers from Fujitsu

Especially if Oracle databases are involved in a Big Data solution, SPARC servers with the Solaris operating system, the result of Fujitsu's high end server technologies coupled with Oracle's database technologies, will deserve particular attention. The latest members of Fujitsu's SPARC server family, the FUJITSU M10 servers, are based on the 16-core SPARC64 X processor and deliver mainframe class reliability, availability and serviceability (RAS) adopted from Fujitsu's 60 years of mainframe development with maximum scalability for mission-critical workloads. FUJITSU M10 servers scale dynamically from 1 to 64 processors in a modular architecture with core-level CPU activation that allows customers to easily add resources in Building Blocks to meet changing workload requirements without downtime for existing applications.

Featuring breakthrough technology including software on chip and liquid loop cooling, FUJITSU M10 servers provide highly flexible system configurations with physical partitioning as well as built-in, no-cost virtualization technologies via Oracle VM Server for SPARC and Oracle Solaris Zones.

FUJITSU M10 servers support both Oracle Solaris 10 and 11 and with the Oracle Solaris Guarantee are binary compatible with all current and past SPARC servers for added investment protection.

Data storage in Big Data Infrastructures

Similar to servers, data storage is also needed everywhere in the Big Data solution architecture. Basically, any flavor of storage system is in the run, be it network storage, such as SAN (Storage Area Network) storage or a NAS (Network-Attached Storage), or even a DAS (Direct-Attached Storage), which includes dedicated JBOD systems (Just a Bunch of Disks) directly attached to a server, as well as local disks of a server. But as with servers, also here the use case matters, when it comes to the question, which data storage to make use of.

Which data storage for which purpose?

Relational databases or data warehouses, no matter if we consider the distilled essence or the ones used as data sources for Big Data, may be stored on SAN storage, on a NAS or a DAS. If data is needed by a single server only, a DAS is the most cost-effective solution. For data access by several servers, a NAS is recommended. Highest requirements in terms of performance and scalability can be met by a SAN.

Purpose	Data storage options
RDBMS / DW	SAN, NAS, DAS
Distributed data store (HDFS)	DAS (SAS, SATA)
Local FS for MapReduce	DAS (SAS, SATA, SSD)
Metadata of Hadoop master nodes	DAS, NAS
NoSQL DB	DAS, NAS, SAN
Persistence layer of IMDB	DAS, NAS, SAN
Persistence layer of IMDG	DAS
History log of CEP engine	DAS, NAS, SAN

A NoSQL database is typically stored on a DAS; however a NAS – especially if the NoSQL database is tuned for NAS - and a SAN can be also an option.

In conjunction with Distributed Parallel Processing, DAS is the only reasonable option. For the distributed data store which either includes a distributed file system (e.g. HDFS) or a distributed database (e.g. Apache HBase), SAS or SATA disks are used. In order to increase the performance of the map-reduce tasks, temporary data and intermediate results can be stored on SSD (Solid State Disks) instead of SAS or SATA disks, however in most of the use cases cost might still be too high for this, at least today. Due to the higher costs, SSD are not a choice for the distributed data store in the near future.

With Distributed Parallel Processing, the master node needs access to metadata which can be stored on a DAS in a single instance scenario. As soon as high availability for the master node comes into the game, the edit log is stored on a highly available NAS.

An In-Memory Databases (IMDB) is fully replicated to a persistence layer in order to guard against failures, such as power outage. This persistence layer can be organized on a DAS, a NAS or even SAN storage. Fujitsu's single-node Power Appliance for SAP HANA uses local disks (Fusion-IO cards, normal hard disks or SSDs) as persistence layer, while the multi-node appliance uses a NAS. The persistence layer of an In-Memory Data Grid (IMDG) is usually a DAS; for performance reasons, SSD are recommended.

CEP engines give you the choice to activate logging of events. The log may reside on a DAS or external network storage (NAS or SAN). You may even move the log into distributed file system for further distributed parallel processing.

In order to forward the log content quickly to other processing instances, it will be an advantage to create the log file on fast local SSD of the server running the CEP engine.

Storing the distilled essence of Big Data

The most critical are the requirements to storage in the distilled essence tier, because meeting these requirements is essential for enabling ad hoc analytics and results in real-time, and thus for creating the expected new business value.

Although the distilled essence has been pre-processed and transformed, it is still about large amounts of data. Therefore the storage systems should be able to cover large capacities and cope with data growth. Existing physical capacities should be optimally utilized. Functions such as data compression usually contribute to this target, however especially with NoSQL databases, such functions are often built into the middleware, and therefore not fully exploited, if implemented in the storage system.

Analytics in real-time requires fast data access, especially when it comes to read operations. But as data is also updated, and sometimes even with a high frequency, the aspect of fast write operations should not be neglected either. That's why the performance of the storage system influenced by processor speed and memory, and the performance of storage connections, such as FibreChannel or Infiniband, due to the reliable bandwidth and latency provided, are extremely important.

Depending on the use case, Big Data analytics can be mission-critical for the business. Downtime can endanger the business or even its existence. This makes high availability of the storage system and fast recovery an absolute must. Often disaster recovery concepts are indispensible, sometimes with long distance support, if for instance compliance demands it.

High quality analytics results can only be achieved with a high quality of the data you analyze. For this purpose, data integrity should be always guaranteed. Moreover, the distilled essence usually represents a high value for the organization, and should therefore be protected from unauthorized access, data theft and data loss. Security features, such as data encryption and access control, should be taken into consideration.

In order to always find the right balance capacity, speed and cost, a flexible choice of media within a storage system is of high significance. Various storage tiers within a storage system and automated storage tiering can also help automate service levels.

And finally, efficiency is always in the center of a customer's consideration; ease of administration and energy efficiency contribute to it.

Disk storage systems from Fujitsu

Various infrastructure concepts, diverse use cases and different demands require a broad range of products. Fujitsu is in the extraordinary situation that for every storage-related problem a solution is in place.

FUJITSU Storage ETERNUS DX

Combining leading performance architecture with automated quality of service management, the Fujitsu Storage ETERNUS DX series aligns storage resources with business priorities, thus enabling higher system utilization and delivering increased system consolidation capabilities and a faster ROI. Unified scalable entry-level and midrange systems, a seamless **family concept** enabling easy system upgrades, and ETERNUS SF, the **unified management** suite across the product line, reduce operational and migration costs. ETERNUS SF provides enterprise-class functionalities in the entry and midrange class and allows flexible disaster recovery concepts for the different model sizes, thus decreasing investment costs. Migration between different models happens in a very efficient way, because disks and their data can easily be transferred, and all the investment of the past is fully protected. An administrator who is trained on one system can operate other models immediately, which in turn keeps training costs low. This is the merit of FUJITSU Software ETERNUS SF, a management suite for the entire ETERNUS DX family.

The broad range of products gives you all the **flexibility** to choose the model that fits best to your requirements. But high flexibility is also provided at a model level. Different disk types, such as SAS (Serial Attached SCSI) and Nearline SAS hard disk drives, as well as Enterprise SSD (Solid State Disks) drives can be combined even in one enclosure, allowing disk capacity, performance and costs to be balanced for various applications in a very granular way. The assignment of disk types to the data can even be automated according to predefined parameters (automated tiering) to serve fast-changing environments that may require self-optimization.

The ETERNUS DX models support all state-of-the-art connection types, both SAN access with free choice of Fiber Channel, iSCSI and FCoE (Fiber Channel over Ethernet) combining the advantages of Fiber Channel and Ethernet, and NAS access with CIFS and NFS. Hence, ETERNUS DX can be flexibly integrated into any existing network environment.

Especially in the context of Big Data and data analytics, **performance** matters. For analytics it is the read performance, for the history database of the CEP engine, it is the write performance as well. ETERNUS DX runs with a very lean, real-time type of operating system that has been tuned to use system resources extremely efficiently. The result is stable response times even during peak loads. The midrange systems can optionally be equipped with Extreme Cache modules (PCIe flash cache extensions) - which further boost performance. For years, ETERNUS DX systems have held the top ranking in both, the SPC-1 and SPC-2 benchmarks of the Storage Performance Council, as far as IOPS performance, throughput and price performance ratio ($ / IOPS, $ / MBPS) is concerned.

With its consistent and flexible family design, and its modular architecture ETERNUS DX is prepared for data growth and **scalability**. You may start with an entry level system, and by adding resources you will end up with a midrange or even high-end system. This means investment protection for any customer. ETERNUS DX can be expanded up to 1000s of TB. It is extremely important for us, that I/O performance can keep pace with the large data capacities of. This is taken into account with the number of host connections and by using I/O controllers providing the highest speed in the market.

The ETERNUS DX **Quality of Service** feature with application I/O prioritization enables the consolidation of multiple tiers of applications in a single storage system. It sets performance limits for each connected server according to its priority. By prioritizing data access and dynamically managing any I/O conflict, high performance can be guaranteed for high-priority applications, and at the same time capacity is used more efficiently, thus increasing storage utilization without sacrificing performance.

Using storage virtualization technologies, such as Thin Provisioning, storage **utilization** can be **optimized** even more. You define the logical capacities covering already future requirements, whilst the physical capacity in total only has to match current requirements. The real investment in capacity happens on demand. In order to reduce the size of data copies for backup, **snapshots** can be generated. Data will only be replicated before being updated. Another benefit of snapshots is the fast restore process.

In addition to the snapshot function included in ETERNUS SF, Fujitsu's ETERNUS Snapshot Manager (ESM) which is based on a dedicated OEM version of Simpana snapshot management of Fujitsu's partner CommVault. ESM manages **application integrated snapshots** supporting widespread applications like SAP, Oracle and many more.

As the amount of data is growing exponentially, the **reliability** of the storage system and the **safety** of the Big Data are essential for business continuity. ETERNUS DX has its origins in the mainframe environment, and its development and quality assurance procedures are based on mainframe standards, resulting in leading product robustness.

All critical components are redundant and hot-swappable, data paths are redundant, firmware can be updated without stopping operations, and all relevant RAID levels (0, 1, 1+0, 5, 5+0 and 6) are supported, thereby eliminating single points of failure.

As the volume of stored data increases, data corruption caused by write or read errors becomes more probable. **Data Block Guard** adds a check code to the data before the data is written to the cache or the hard disks, enabling the system constantly to verify that data is unchanged. The check code is removed before data leaves the system when being read.

For performance reasons, online storage systems are equipped with ever-increasing RAID controller caches, whose volatile content often means a risk of data loss. The **Cache Protector** function ensures that data in the caches is written to a non-volatile flash memory after power outage and safely stored there, no matter how long the power outage lasts. The Cache Protector also accelerates the reboot to bring the system up to full power quickly.

In order to protect data against disasters, data can also be replicated to **multiple sites**, even between different ETERNUS DX models. With distances up to 100 km, mirroring can happen synchronously at reasonable cost; asynchronous data replication can be applied to global distances between the sites over a wide area network.

Self-encrypting drives automatically encrypt data before it is written to the hard disk, and thus prevent unauthorized access even after discarding the disks at the end of life, and ensure maximum **security**.

Running ETERNUS DX in **Eco-Mode**, the rotation of a hard disk is stopped, when it is not needed. This reduces power consumption and heat dissipation, expands the lifetime of the disks, and contributes to the outstanding efficiency of ETERNUS DX.

ETERNUS DX systems represent a great choice as a storage platform for the distilled essence of Big Data, but of course also for databases and data warehouses used as data sources in a Big Data solution.

Violin Memory All-Flash Arrays

If an ETERNUS DX although equipped with SSDs and / or Extreme Cache cannot meet extreme velocity requirements when accessing data in the distilled essence at random, Fujitsu utilizes All-Flash-Array (AFA) technology from its partner Violin Memory. In order to ensure long-term data integrity and retention, and prevent any single point of failure from causing data loss, RAID is implemented across flash modules, not just within a module. Another highlight is the distributed garbage collection implemented in hardware which enables highest write performance of up to 40 times higher compared to other solid-state solutions. Violin's flash-based appliances scale to hundreds of terabytes of capacity, millions of IOPS at low latency, high bandwidths at low latency, and are orders of magnitude faster than disk-based arrays when it comes to data analytics.

It goes with saying, that optimization functions such as snapshots, instant cloning, data replication, thin provisioning, compression and deduplication are implemented in Violin AFA.

FUJITSU Storage ETERNUS JX

PRIMERGY servers have ample internal disk capacities. But what happens when internal capacity reaches its limits? As a passive extension of storage capacity, the JBOD (Just a Bunch Of Disks) subsystem FUJITSU Storage ETERNUS JX extends the internal storage capacity of the servers by up to 72 Terabytes using an extremely fast 6 Gbit/s SAS connection to meet the ever growing data volume requirements. In a compact rack-able housing, SAS and SATA hard disks can even mixed. Hot-pluggable and hot-replaceable hard disks and power supplies with redundant fans make the ETERNUS JX a reliable storage subsystem. Attached to the SAS RAID controller of the server, it offers basic data protection through mirroring, striping or adding parity information to the data in all relevant RAID groups. Of course, ETERNUS JX is integrated with PRIMERGY ServerView Suite, thus simplifying management.

In a Big Data solution, ETERNUS JX will be primarily used in a Hadoop cluster or for storing the history database of the CEP engine.

Global alliance with NetApp

ETERNUS DX will be the system of choice when block-level storage is needed. If higher I/O performance at lower latency is demanded to get analytics results in real-time, All-Flash-Arrays from our partner Violin with particular strengths in random I/O can be used, and if a JBOD is sufficient, ETERNUS JX will come into the game.

In case of customers running homogeneous NAS Environments Fujitsu relies on its global alliance partner NetApp and their Filer products, also known as FAS (Fabric-Attached Storage) systems with their ONTAP operating system especially optimized for storage functions. NetApp FAS supports file-based protocols such as NFS, CIFS, FTP, TFT and HTTP.

Storage Virtualization with FalconStor

Looking at medium and large enterprises, it is quite likely that over the years for different use cases different storage systems have been introduced. In the context of Big Data, these storage systems may appear as data sources or even keep the distilled essence. The heterogeneous storage landscape means complexity in terms of data access and management.

The Fujitsu PRIMERGY based FalconStor® Network Storage Server (NSS) Gateway enables storage management in a unified and simplified manner. Storage virtualization abstracts management from the storage systems' physical features. Thus it ensures the availability of critical business applications by leveraging an organization's existing storage investments and provides efficient capacity utilization. Furthermore, it includes features such as thin provisioning, mirroring, performance tiering, data migration, WAN-optimized replication, and application-consistent snapshots.

Backup of Big Data

Making Big Data available on the various storage systems and managing it is by far not all we have to consider. High availability features and disaster recovery can help avoid downtime in the event of physical errors, system failures or even disasters, such as natural catastrophes, but they do not protect data from human errors or data corruption. If for instance important data is accidentally deleted or corrupted by software error, replication features will replicate the problem, but not help recover the lost data.

Here a backup is the final insurance for your data. Backup is about creating copies of your data in order to enable a quick recovery if data is lost.

As you are never immune against human failures, backup is basically relevant for data of all storage systems of our Big Data solution architecture, be it a database or data warehouse in the distilled essence tier, be it the persistence layer of an IMDB, be it the pre-processed data in the distributed data store used for Distributed Parallel Processing, or be it the history database which includes the logs of the CEP engine. For the databases and data warehouses used as data sources, a backup solution ought to exist already anyway.

No backup needs to be applied to the persistence layer of an IMDG, because the data stored there is also present in the respective files or databases accessed by the applications which are accelerated by using the IMDG.

The retention time for each backup copy can vary. While it might be sufficient to retain the latest backup version of the distributed data store used for Distributed Parallel Processing, the retention time for other use cases, e.g. the distilled essence, might be typically in the range of several weeks.

Whether a backup is needed for all data or just a part of it strongly depends on the use case, the strategic importance of the data, and sometimes also compliance demands. Most of the restore requirements relate to the recovery of single files. For this use case disk provides the best service levels. Fast and random access patterns are perfectly suited for this media. However, with a retention time of several weeks, the backup copies would just inflate required storage capacities. Therefore the older backup copies, which don't require highest restore performance any more, should be moved to a more price/capacity optimized storage media like tape.

Data inflation can of course also be contained at a certain extent, if you collect only what you need, and if you dismiss irrelevant data already during cleansing. Nevertheless, we will still have to deal with large volumes. Therefore an Information Lifecycle Management (ILM) is highly recommended for your backup data which will help control your storage footprint and costs.

Big Data retention and archiving

If data is not altered any more, or no longer accessed regularly, it might not be justified that it occupies valuable storage space on high-cost online storage systems. However, for the purpose of compliance, but also for other reasons, e.g. post-processing in the future, some of the data need to be stored for a long-term.

Here an archive solution helps. Data is removed from the primary online storage to an archive consisting of more cost-effective media. Archiving frees up a lot of primary storage capacity. Furthermore, the amount of backup data will be reduced dramatically, contributing to a faster backup and a faster recovery process.

Generally there is only one version of the data in the archive, which is separated from the production data and therefore no longer part of the backup process. That's why protection mechanisms for archives can be extremely significant. It may also happen, that data sets in the archive have to be reanalyzed again and again; therefore, data must be easily accessible and timely retrievable.

Which data needs to be archived cannot be answered in general; it will always depend on the customer-specific use case and the requirements and objectives related to it. Usually, data in the essence tier will rather deserve being archived than data of the CEP engine. The archiving relevance of data in HDFS might be somewhere in between.

The retention requirements can range from a few years to decades or even indefinitely.

FUJITSU Storage ETERNUS CS8000 – The unified data protection appliance

There are still organizations in which various islands of backup target systems for various applications are used in parallel, which makes management extremely complex, difficult and expensive. Adding dedicated solutions for Big Data aspects would make the situation even worse.

FUJITSU Storage ETERNUS CS8000 data protection appliance, a worldwide unique solution, introduces a virtualization layer between servers and backup storage systems, to transform the whole data protection infrastructure into one virtualized target system. This allows ETERNUS CS8000 to consolidate complex and heterogeneous data protection environments on a large scale. It is the only solution capable of **consolidating backup and archive** infrastructures into one system and can **connect to all platforms** which might be involved with Big Data, be it industry standard servers, Unix servers or even mainframes.

Balancing speed, capacity and costs

For the backup applications the unified data protection appliance appears traditionally as a virtual tape library (VTL), for the archive applications as a NAS system. The data received from the applications is written to the **internal disks** of ETERNUS CS8000. This provides the full benefit of disk technology, for example, to provide the highest levels of performance, in optimum cases up to 150 TB/h. Many transactions can be handled in parallel. Up to 1,280 virtual tape drives are supported. All data received via the VTL interface is automatically compressed before being saved on disk. In contrary to deduplication technologies, which always come with a performance decrease, the ETERNUS CS8000 compression reduces the amount of storage space required without any loss in performance. Thanks to this, highest performance for backup and restore is ensured to fulfill even the highest service levels in Big Data environments.

As the backup is just a kind of insurance, the entire solution should come at the lowest price possible. **Tape** has many advantages in this regard. To leverage these advantages, the physical tape libraries are connected to the ETERNUS CS8000 at the back-end. The unified data protection appliance takes over the management of tape systems, thus relieving the IT department from the management effort.

As an alternative to the backup-to-disk-to-tape process (B2D2T), ETERNUS CS8000 can be enhanced by the **deduplication** function to act as a final backup-to-disk storage target with deduplication. ETERNUS CS8000 is the only solution on the market which supports **parallel operation of B2D2T and backup to deduplicated disks,** with the same appliance. Data can be easily stored on the appropriate media to fulfill the required service level.

Another option which may play a more important role in the future is storing backup data in the cloud. As a future enhancement, ETERNUS CS8000 will be capable of using the cloud as an additional target via a **cloud gateway**.

To provide flexible SLAs, ETERNUS CS8000 combines the strengths of disks, deduplicated disks and tapes with **rule-based automated management.** Disk capacity can be flexibly utilized as high performance disk storage, as a performance booster for tape storage and as a final disk storage repository with deduplication. This will keep the speed of backup and restore in balance with the costs of media, and support flexible service level agreements. The entire ILM process is automated by the appliance to keep it simple and efficient. Being able to provide the right media for the right kind of data significantly reduces the cost of data protection.

Robustness for ensured business continuity

The **redundant architecture** enables reception of backup data, even if a component should fail. If the write process to the physical tape fails, the data will automatically be directed to another target. Adding and replacing hardware components is automated and non-disruptive.

ETERNUS CS8000 can write data to both local and **remote tape libraries** on multiple sites. This provides an easy way to introduce disaster resiliency with a remote copy on tape. Asynchronous data replication is possible over long-distances in the range of 1000s of kilometers.

It is not just the target systems which can be situated on multiple sites. To ensure highest disaster resilience, one logical appliance can be deployed over two geographically separated sites at a maximum distance of 100 km, denoted as **split-site configuration** with cache mirror. Data can be written to or read from both sites. In case of a failure at one site, workload will be automatically routed to the other site. The backup applications are not aware of this process. The second site even gets the same IP address as the former active site. Backup processes continue as before, in a fully automated manner.

Automated management for efficient operations

If a data copy is corrupt or not available for some reason, the recall request of the backup software will be automatically redirected to the second copy. Of course, this presupposes that the data was at least written in a dual-copy service level. The copy will be provided and the job for the backup software is done. An **automated self-healing** mechanism is initiated by ETERNUS CS8000 that will recover the failed copy. Thus the level of redundancy is restored. Quality checks of duplicate copies can also be conducted automatically. This guarantees defined service levels which require an always available backup copy.

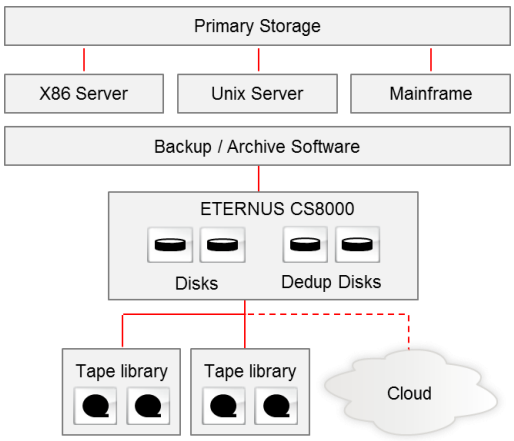
Furthermore, regular **reorganization of tape media**, which have not been accessed for a certain time, refreshes the magnetization of the tapes and prevents age-related loss of the magnetization. This can even be defined as a regularly occurring automatic process.

Flexibility and investment protection

The modular grid architecture delivers extreme **scalability** of capacity and performance. Individual front-end and back-end processor nodes can be added to avoid bottlenecks in the transfer of data between servers and ETERNUS CS8000, and from there to the target system.

ETERNUS CS8000 works seamlessly together with **data protection software** products from the leading market players, such as Fujitsu's alliance partners Symantec and CommVault and many more. Thus it can be easily integrated into existing data protection environments, which means maximum investment protection for the customer.

Due to universal tape interfaces, ETERNUS CS8000 is compatible with all market-leading tape libraries. Even if the libraries or the drives are different, the ETERNUS CS8000 can connect them in parallel. As the virtualization layer decouples target systems from backup software, changes in the infrastructure can be made quite easily. If for instance, a tape library has to be refreshed, no changes are necessary in the backup process, because the application is not aware of the modification. ETERNUS CS8000 also takes care of the **data migration** from the old tape library to the new one.



Archiving and second-tier file storage integrated

However, ETERNUS CS8000 is not just for backup. It provides optional standardized NAS interfaces with NFS and CIFS protocols. The result is a scale-out NAS system, supporting billions of files.

The internal disk storage can act as the final target for the **data to be archived, or as a second-tier file storage** system for less frequently used and aging data. The direct integration with tape enhances the appliance for Hierarchical Storage Management (HSM) to automatically save data on disk and tape, according to defined migration policies and retention periods. To protect the archive against data loss, tape copies, replication scenarios and cache mirror with a split-system are in place.

ETERNUS CS8000 ensures that the files are stored in a secure way, making them **revision-proof** for a defined retention period to meet compliance regulations. KPMG AG confirmed that ETERNUS CS enables storage and retrieval of electronic documents based on the Generally Accepted German Accounting Principles making it the best choice for legally compliant archiving.

Thus, ETERNUS CS8000 is the ideal solution for a unified and optimized data protection infrastructure, capable of consolidating backup and archive infrastructures into one system. By deploying ETERNUS CS8000, enterprises can allocate storage resources for data protection according to business priorities:

- Backup / archive / second-tier file storage
- Right media mix of disks, deduplicated disks, tapes and cloud
- Appropriate data availability levels
- Flexible SLAs
- Automated processes

ETERNUS CS8000 supports SLA fulfillment in terms of speed, capacity and reliability, and at the same time minimizes the investment and total cost of ownership.

FUJITSU Storage ETERNUS LT tape libraries

The affordable FUJITSU Storage ETERNUS LT tape libraries offer impressive scalability and reliability in a compact design. They perfectly meet the requirements of low cost backup and long-term archiving, just as disaster recovery in homogeneous Fujitsu environments or in heterogeneous environments. Moreover, using ETERNUS LT, aging Hadoop data can be easily moved to local tapes in the Hadoop server cluster, sometimes even to prevent local disk overflow.

The systems are equipped with standard LTO-4 / 5 / 6 technologies and can be flexibly connected to the server via the SAS (Serial Attached SCSI) or FC (Fibre Channel) interface. ETERNUS LT systems are enabled for encryption, meeting enhanced security and compliance demands.

Highly automated, simple and remote operation enables usage without any demand for local expert skills. The pay-as-you-grow concept means that customers avoid unnecessary initial investment. And with every new LTO media generation, the costs per Gigabyte decrease.

ETERNUS LT systems are certified for market-leading backup and archiving software.

Network components

In general, a Big Data infrastructure includes many servers and storage systems whose performance is certainly important for data analysis and delivery of results within the desired time interval. However, we should not forget that servers and storage systems are connected to an Ethernet or Fiber Channel network enabling the communication and collaboration of the individual infrastructure components. Especially when dealing with large data volumes, the network must be extremely fast, scalable to cope with the steady data growth, reliable, efficient and easy to manage. Any interruption caused by planned or unplanned downtime or other bottlenecks have to be avoided.

It is Fujitsu's objective to deliver best-in-class solutions that address these challenges adequately. Based on close and well-established partnerships with Brocade and Cisco, Fujitsu offers network components for Ethernet and Fiber Channel networks in Big Data infrastructures. Thus, our customers can benefit from the latest technologies of the leaders in the networking market, designed to meet the critical Big Data demands. And Fujitsu can make sure that its server and storage systems are always connected in an optimum way.

Front-end devices for the end user

It is true, that Big Data solutions require a strong back-end infrastructure including servers and disk storage systems, as well as backup and archive solutions. However, for initiating the right queries and displaying the results, various devices are needed, which meet the requirements of the various end user profiles in an organization.

Fujitsu is in the favorite situation that it can offer the most comprehensive portfolio of end user devices, be it desktop PCs, notebooks, tablet PCs, thin clients and workstations.

Summary

To optimally meet the customer's demands and challenges depending on his specific use case, various solution approaches have to be considered. Each of these solution approaches makes its particular demands regarding the infrastructure products to be involved. This in turn requires a broad diversity of products, if you want to cover the broad scope of use cases by products which perfectly fit.

Fujitsu has got the right answer for any Big Data use case. It provides all infrastructure products that might play a part in a Big Data infrastructure, be it industry standard PRIMERGY servers in all relevant flavors, PRIMEQUEST servers to meet mission-critical demands, ETERNUS DX disk storage systems, All-Flash-Arrays, ETERNUS JX disk storage subsystems, the unified backup and archive solution ETERNUS CS8000 data protection appliances, ETERNUS LT tape libraries, market-leading network components or the various types of access devices. To get an optimum solution, it is not just the right infrastructure approach, but also the perfect ingredients that matter.

Fujitsu provides all these perfect ingredients from a single source, including professional services, managed services and maintenance services, and therefore leaves nothing to be desired.

Contact
FUJITSU Technology Solutions GmbH
Address: Mies-van-der-Rohe-Strasse 8,
80807 Munich, Germany
Website:www.fujitsu.com/fts