

Chapter 1

Data Storage Technologies

GOOGLE CLOUD CERTIFIED PROFESSIONAL CLOUD DATABASE ENGINEER EXAM OBJECTIVES COVERED IN THIS CHAPTER:

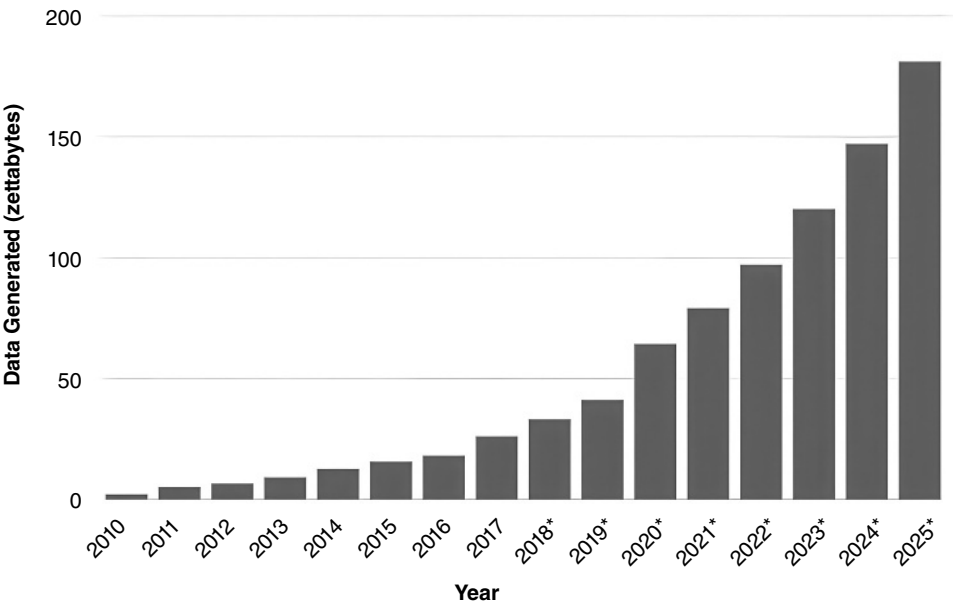
- ✓ 1.1 Analyze relevant variables to perform database capacity and usage planning.
 - Evaluate performance and cost tradeoffs of different database configurations (machine types, HDD versus SSD, etc.).
- ✓ 1.4 Evaluate appropriate database solutions on Google Cloud.
 - Identify solutions to support regulatory and compliance requirements.



The amount of data created globally is exploding. Estimates suggest daily data creation will top 500 exabytes by 2025 (that's 500 million petabytes!), which means that the amount of data being generated is doubling every few years. Figure 1.1 shows the exponential rise in data generated year over year since 2010. As the amount of data generated grows, finding efficient and cost-effective ways to store and access this data is becoming increasingly important.

Traditionally, most data was structured, fitting neatly into tables with defined formats (like customer names and addresses). However, the rise of social media, sensor data, and multimedia content has led to a surge in unstructured data, which doesn't fit easily into predefined formats (like emails, videos, and social media posts).

FIGURE 1.1 Global data generated annually.



Organizations require ways to store this data that are not only secure and reliable, but also adaptable and efficient. Traditional on-premises storage and database solutions, while familiar, can struggle to keep pace with ever-growing data volumes and evolving business needs.

Turning to the cloud presents a compelling solution. Cloud storage and databases options offer inherent scalability, allowing you to effortlessly adjust resources up or down based on demand. This eliminates the need for costly hardware provisioning and management, freeing up IT resources for more strategic tasks. Additionally, cloud providers handle infrastructure maintenance and security, ensuring your databases are constantly protected and up-to-date.

For a Google Cloud Professional Database Engineer, understanding different storage solutions and where to apply them is crucial. In this chapter, we'll explore block, object, and file storage, then delve into structured, unstructured, transactional, and relational databases, highlighting their uses and how a database engineer leverages them as part of an overall architecture.

Data Storage Design Considerations

As a Google Cloud database engineer, designing optimal storage solutions requires careful consideration of various factors. This section explores some key aspects you should define before diving into the design process.

Workload Characteristics

Understanding your workload's characteristics is crucial for effective storage design. Consider factors such as data volume, growth rate, access patterns, durability requirements, retention policies, and any compliance or regulatory obligations. This information will help determine the appropriate storage solution and capacity needs.

For example, imagine that you are an online retail platform handling product catalogs, customer orders, and inventory management. You might identify for the following workload characteristics to inform your planning:

Data Volume: Several terabytes of product information, customer data, and order history

Growth Rate: Rapidly increasing product catalog and customer base, requiring frequent data additions

Access Patterns: High read volume for product catalogs and order lookup; frequent writes for new orders and inventory updates

Durability Requirements: High durability needed to prevent data loss as orders and customer information are critical

Retention Policies: Order data retained for 7 years for tax and compliance purposes; product catalog data updated frequently

Compliance: Adherence to Payment Card Industry Data Security Standards (PCI DSS) for handling payment information

Performance and Scalability

Define the performance requirements for your workload, including necessary IOPS and throughput. Assess the potential for future growth and the ability to scale storage resources to accommodate increasing demands. Establishing performance SLAs will help ensure that your storage solution meets your application's needs.

Imagine that your client is a high-frequency trading platform. What are some performance and scalability requirements?

Performance Requirements: Extremely low latency, high input/output operations per second (IOPS) for rapid order execution and market data ingestion. Throughput needs to handle thousands of transactions per second.

Scalability: Ability to quickly scale storage resources to handle increased trading volume during peak market hours.

Performance SLAs: Strict latency requirements, such as 99.999 percent of read operations completing within 1 millisecond.

Cost Optimization

To maximize cost efficiency, evaluate different storage options and their pricing models. Implement data life-cycle management strategies to move cold or infrequently accessed data to lower-cost storage tiers. Careful consideration of storage costs will help optimize your overall cloud expenditure.

For example, take a large-scale data warehousing application storing historical sales data. How could you identify cost optimization opportunities?

Storage Tiers: Evaluate different storage options like Nearline, Coldline, and Archive storage tiers offered by Google Cloud Storage.

Life-Cycle Management: Implement life-cycle policies to automatically move older sales data (e.g., data older than 3 years) to lower-cost storage tiers like Coldline or Archive.

Monitor Storage Usage: Analyze storage usage patterns to identify opportunities for optimization and cost savings.

Data Protection and Recovery

Develop comprehensive backup and recovery plans to safeguard your data. Determine the frequency of backups, retention periods, and recovery procedures. Additionally, consider disaster recovery strategies to protect against data loss in case of unforeseen events. Data encryption should also be implemented to protect sensitive information.

For example, some industries will have different tolerances for recovery time objective (RTO), or how quickly you must recover based on the cost of downtime. RTO is about how quickly systems must be restored (speed of recovery), whereas RPO is about how much data loss is acceptable (data currency). We will cover more about RTO and RPO in Chapter 5, “Designing for High Availability and Disaster Recovery.”

What are some ways to design for these data protection and recovery requirements in Google Cloud?

Automated Backups: Leverage native automated backup tools such as Backup for GKE or Cloud SQL Backup.

Persistent Disk Snapshots: Create point-in-time copies of persistent disks for rapid recovery.

Data Encryption: By default, all data stored in Google Cloud is encrypted in transit and at rest.

Security and Compliance

Prioritize data security by implementing robust access controls to prevent unauthorized access. Ensure compliance with data residency regulations and other relevant industry standards. Establish auditing and logging procedures to monitor storage activities and maintain a strong security posture.

Imagine that you represent a large healthcare provider that needs to store sensitive patient data, including medical records, financial information, and personal details. They must comply with stringent regulations like HIPAA, GDPR, and CCPA while ensuring data privacy and accessibility for authorized personnel. What are some ways to address data security and compliance requirements?

Robust Access Controls: Implement IAM policies that limit data access based on user roles and grant users only the necessary permissions.

Data Residency: Store data in specific geographic locations to comply with laws. Replicate data across regions for disaster recovery while adhering to regulations.

While this is not an exhaustive list, documenting some of these requirements will help you start to determine the best storage solutions on Google Cloud.

Storage Technologies

Google Cloud provides a comprehensive suite of storage solutions to cater to diverse data management needs. These solutions are primarily categorized into three fundamental types: block, file, and object storage. Before delving into the specifics of each storage format, it's essential to understand their core characteristics:

Block Storage: Presents data as raw blocks, providing low-level access for applications requiring high performance and flexibility.

File Storage: Organizes data into hierarchical filesystems, offering a familiar structure for applications that need to manage and share files.

Object Storage: Stores data as objects with metadata, optimized for large amounts of unstructured data that requires high durability and availability.

In the subsequent sections, we will explore these formats in detail, highlighting their key features, use cases, and how they are delivered in Google Cloud.

Block Storage

Block storage is a type of digital storage that divides data into small, independently addressable chunks called *blocks*. This structure allows for rapid data access and modification, making it ideal for applications demanding high performance, such as databases and transaction processing systems. Similar to traditional storage systems used in data centers, block storage offers high IOPS and is often used in virtual machine (VM) environments. Google Cloud provides various block storage options as part of its Compute Engine service, catering to different performance and cost requirements. These block storage options are available in three different formats:

Persistent Disk: Dedicated hard-disk drives (HDD) and solid-state drives (SSD) for enterprise and database applications, offering durability and flexibility.

Hyperdisk: High-performance block storage decoupling performance from VM type with volumes that can be dynamically resized, providing scalability and efficiency.

Local SSD: High-performance, low-latency block storage directly attached to the virtual machine, ideal for I/O-intensive workloads.

Let's explore each type of block storage in detail, including key features, configuration options, and use cases.

Persistent Disk

Google Cloud Persistent Disks are block storage volumes that function as virtual hard disk drives (HDDs), solid-state disks (SSDs), or an option that balances price/performance for VMs on Google Compute Engine (GCE) and Kubernetes Engine (GKE).

Key Features

Key features include the following:

- Durability:** Data persists independently of the VM life cycle, ensuring information remains secure even after VM shutdown or termination.
- Flexibility:** Users can dynamically resize persistent disks on the fly to accommodate growing storage needs without impacting VM uptime.
- Scalability:** Multiple disks can be attached to a single VM for increased storage capacity.
- Performance Options:** Users can choose between high-performance SSDs for demanding applications or cost-effective HDDs for large datasets with less frequent access.

Configuration Options

When you configure a persistent disk, you can select one of the following disk types:

- **Standard Persistent Disks (pd-Standard):** HDD-based, cost-effective for workloads with moderate performance needs.
 - Suitable for large data processing workloads that primarily use sequential I/Os
 - Most cost-effective option
 - Backed by standard HDDs
- **Performance (SSD) Persistent Disks (pd-ssd):** SSD-based, provides higher performance and lower latency compared to standard persistent disk.
 - Suitable for enterprise applications and high-performance databases that require lower latency and more IOPS than standard persistent disks provide
 - Designed for single-digit millisecond latencies; the observed latency is application specific
 - Backed by SSDs
- **Balanced Persistent Disks (pd-Balanced):** SSD-based, offers a balance of price and performance.
 - An alternative to performance (pd-ssd) persistent disks
 - Balance of performance and cost. For most VM shapes, except very large ones, these disks have the same maximum IOPS as SSD persistent disks and lower IOPS per GiB. This disk type offers performance levels suitable for most general-purpose applications at a price point between that of standard and performance (pd-ssd) persistent disks
 - Backed by SSDs
- **Extreme Persistent Disks (pd-Extreme):** SSD-based; delivers the highest performance for demanding workloads.
 - Offer consistently high performance for both random access workloads and bulk throughput

- Designed for high-end database workloads
- Allow you to provision the target IOPS
- Backed by SSDs
- Available with a limited number of machine types

Feature Summary

Table 1.1 summarizes the differences between the various Persistent Disk types. The column definitions are as follows:

- Backed By:** Type of physical hard drive: HDD or SSD?
- Performance:** How much IOPS and throughput do you need?
- Scalability:** Do you need to be able to scale your storage up or down easily?
- Cost:** How much are you willing to spend on storage?

Use Cases

Use cases include the following:

- Databases:** High-performance databases like MySQL, PostgreSQL, and Oracle can benefit from the low latency and high throughput of Persistent Disks.
- Virtual Machines:** As the primary storage for Compute Engine instances, Persistent Disks are essential for running virtual machines with various workloads.
- Big Data Analytics:** Persistent Disks can be used for storing intermediate data for big data processing frameworks like Apache Spark and Hadoop.
- Gaming and Media:** Applications demanding high IOPS and low latency, such as online gaming and video editing, can leverage Persistent Disks.

TABLE 1.1 Comparison of persistent disk types

Persistent disk type	Backed by	Performance	Cost
Standard persistent disks	HDD	Moderate	Low
Performance (SSD) persistent disks	SSD	High	Moderate
Balanced persistent disks	SSD	Balanced	Low-moderate
Extreme persistent disks	SSD	Very High	High

Hyperdisk

Hyperdisk is Google Cloud's next-generation block storage solution designed to deliver exceptional performance, scalability, and flexibility for demanding workloads. It represents a significant leap forward in block storage technology compared to traditional Persistent Disks.

Key Features

Key features include the following:

High Performance: Hyperdisk offers substantially higher IOPS and throughput compared to Persistent Disk, making it ideal for performance-critical applications. With Hyperdisk, you get dedicated IOPS and throughput with each volume, as compared to Persistent Disk, where performance is shared between volumes of the same type.

Scalability: Hyperdisk lets you scale performance and capacity dynamically. For some Hyperdisk types, you can independently scale performance (IOPS and throughput) and capacity, providing granular control over resource allocation.

Integration: Seamlessly integrates with Compute Engine and Google Kubernetes Engine (GKE).

Configuration Options

The following Hyperdisk options are available depending on your workload:

- **Hyperdisk Balanced:** (customizable throughput, IOPS, shareable between VMs, bootable)
 - Optimized for most workloads, including databases, web servers, and enterprise applications
 - Offers a balanced combination of IOPS and throughput for general-purpose use cases
 - Provides a cost-effective option for workloads with moderate performance demands
- **Hyperdisk Balanced High Availability:** (customizable throughput, IOPS, shareable between VMs, bootable, synchronous replication for HA)
 - Optimized for most workloads, including databases, web servers, and enterprise applications
 - Offers a balanced combination of IOPS and throughput for general-purpose use cases
 - Protects your applications from a zonal outage by synchronously replicating the disk data across two zones in the same region

- **Hyperdisk Extreme:** (high IOPS/throughput, not shareable, not bootable)
 - Optimized for high-performance workloads, such as high-end databases (Oracle, SQL Server, SAP HANA), in-memory databases, and real-time analytics
 - Delivers the highest IOPS and throughput for demanding applications
 - Comes with a premium price due to its exceptional performance capabilities
- **Hyperdisk ML:** (customizable throughput, not IOPS, shareable between VMs in read-only mode, not bootable)
 - Optimized for machine learning workloads that require the highest throughput
 - Hyperdisk ML has the highest throughput available and the fastest data load times as a result
 - Faster data load times mean shorter accelerator idle times and lower compute costs
 - For large inference and training workloads, you can attach a single Hyperdisk ML volume to multiple VMs in read-only mode
- **Hyperdisk Throughput:** (customizable throughput, no IOPS, not shareable, not bootable)
 - Optimized for scale-out analytics workloads, cold storage, and cost-sensitive applications
 - Prioritizes throughput over IOPS, making it suitable for data-intensive workloads
 - Offers a more cost-efficient option compared to Balanced and Extreme for throughput-oriented applications

Feature Summary

Table 1.2 summarizes the differences between the various Hyperdisk types.

Use Cases

Use cases include the following:

High-Performance Databases: Databases like Oracle, SQL Server, and SAP HANA can benefit from Hyperdisk's exceptional performance and scalability.

In-Memory Databases: Applications requiring low-latency access to large datasets can leverage Hyperdisk.

Big Data Analytics: Hyperdisk can accelerate data processing and analytics workloads.

Gaming and Media: Applications demanding high IOPS and low latency, such as online gaming and video editing, can benefit from Hyperdisk.

TABLE 1.2 Comparison of Hyperdisk types

Hyperdisk type	Customizable throughput	Customizable IOPS	Shareable between VMs	Boot disk support
Hyperdisk Balanced	Yes	Yes	Yes	Yes
Hyperdisk Balanced High Availability	Yes	Yes	Yes	Yes
Hyperdisk Extreme	No	Yes	No	No
Hyperdisk ML	Yes	No	Yes	No
Hyperdisk Throughput	Yes	No	No	No

Local SSD

Google Cloud Local SSDs are a high-performance, low-latency storage option specifically designed for temporary data within Compute Engine virtual machines. Ideal for workloads demanding rapid data access, such as caching, scratch processing for data analytics or high-performance computing, and temporary data storage like SQL Server's tempdb, Local SSDs offer significantly superior IOPS and latency compared to Persistent Disk and Hyperdisk. This performance boost is attributed to their direct attachment to the host server.

However, it's crucial to remember that Local SSDs are not suitable for long-term data storage. Data on these disks can be lost if the VM stops or terminates. To ensure data durability, consider using Persistent Disk or Cloud Storage for non-temporary data. While Local SSDs provide exceptional performance, their use is limited to scenarios where data persistence is not a primary concern.

Key Features

Key features include the following:

High Performance: Local SSDs offer exceptional IOPS and low latency due to their physical attachment to the host VM.

Ephemeral Storage: Data stored on Local SSDs is not persistent and will be lost upon VM termination or instance reboot.

Cost-Effective: Compared to Persistent Disk, Local SSDs are generally more cost-effective for performance-sensitive workloads.

Scalability: You can attach multiple Local SSD partitions to a single VM for increased storage capacity.

Additional Considerations

Some additional considerations are as follows:

Data Durability: Since Local SSDs are ephemeral, data loss is a risk. Implement backup strategies for critical data.

Workload Characteristics: Carefully analyze your workload to determine if Local SSDs are the right fit. Consider factors like IOPS, latency, and data durability requirements.

Cost Optimization: Balance performance needs with cost considerations. Evaluate the trade-offs between Local SSDs and Persistent Disk for your specific use case.

Use Cases

Use cases include the following:

Caching: Storing frequently accessed data for faster retrieval.

Scratch Space: Temporary storage for intermediate data processing results.

Log Storage: Storing application logs for short-term analysis.

In-Memory Databases: Using Local SSDs as a temporary storage medium for in-memory databases.

High-Performance Computing (HPC): Storing intermediate data for HPC workloads.

Media Processing: Storing input and output files for video editing and rendering.

Comparing Block Storage Options in Google Cloud

Block storage options in Google Cloud are compared in Table 1.3.

Object Storage

Object storage is a type of data storage architecture designed to handle massive amounts of unstructured data.

Unlike traditional filesystems, which organize data in a hierarchical structure of folders and files, object storage treats data as individual units called *objects*. Each object consists of the data itself, metadata describing the data (such as size, creation date, and content type), and a unique identifier. Because these unique keys are in the form of URLs, this means that object storage interacts well with web technologies.

This flat structure makes object storage highly scalable and efficient for storing and managing large volumes of data. It is ideal for handling diverse data types like images, videos, audio files, and log data. Google Cloud Storage is a service that delivers object storage services in Google Cloud, providing features like high durability, availability, and accessibility from anywhere in the world.

TABLE 1.3 Comparing block storage options in Google Cloud

Feature	Persistent disk	Hyperdisk	Local SSD
Performance	High performance, suitable for most workloads	Highest performance, scalable, decoupled from VM	Very high performance, but ephemeral
Latency	Low latency	Lower latency	Extremely low latency
Throughput	High throughput	Highest throughput	Very high throughput
IOPS	High IOPS	Highest IOPS	Very high IOPS
Durability	Highly durable, with replication	Highly durable, with replication	Not durable, data lost on instance termination
Cost	Lower cost	Higher cost, but often cost-effective due to performance	Lowest cost for high IOPS and low latency workloads
Use Cases	General purpose workloads, databases, web servers	High-performance databases, big data analytics, machine learning	Temporary storage, burst workloads
Flexibility	Flexible, can be attached/detached from VMs	Highly flexible, with storage pools and dynamic scaling	Limited flexibility, tied to instance life cycle
Ephemeral	No	No	Yes
Managed vs. Unmanaged	Managed	Managed	Unmanaged

In the last section, you learned about block storage. Let's now take a look at how object storage differs from block storage. Table 1.4 compares some key differences between object and block storage.

Google Cloud Storage

Google Cloud Storage (GCS) is a highly scalable object storage service provided by Google Cloud. It allows you to store and retrieve any amount of data from anywhere in the world. Google Cloud Storage is used as an underlying storage solution for a number of services across Google Cloud. Cloud Storage's primary use is whenever binary large-object storage

TABLE 1.4 High-level comparison of object storage vs. block storage

Feature	Object storage	Block storage
Data organization	Objects with metadata	Blocks of data
Access	Through API	Through block device interface
Performance	Lower latency, higher throughput	Lower latency, higher IOPS
Cost	Typically lower	Typically higher
Scalability	Highly scalable	Less scalable
Use cases	Unstructured data, backups, archives	Structured data, databases, VMs

(also known as a “BLOB”) is needed for online content such as videos and photos, for backup and archived data, and for storage of intermediate results in processing workflows.

Google Cloud Storage organizes data into buckets located in specific geographic regions for optimal performance. Google Cloud Storage does not have actual folders. Instead, it uses a hierarchical naming convention for objects within a bucket. This means you can use slashes (/) in object names to create a folder-like structure, but these are simply part of the object’s name and not actual directories.

While there are no physical folders, many tools and interfaces, including the Google Cloud Console, simulate a folder structure for easier organization and navigation. Essentially, it’s a virtual filesystem built on top of a flat object storage system.

Data objects in Google Cloud Storage are immutable, meaning new versions are created upon changes. Bucket versioning can be enabled to track object history, providing data recovery options. When you upload a file to GCS, it becomes an object. If you want to change the object, you create a new object with the updated content. With versioning enabled, GCS keeps track of all object versions, allowing you to revert to previous states if needed.

While objects themselves are immutable, metadata associated with an object (like object name, storage class, ACLs) can be modified without creating a new object. This immutability characteristic makes GCS a reliable choice for storing data that requires high integrity and durability, such as backups, archives, and large datasets.

Key Features

Key features include the following:

- Scalability:** Handle massive amounts of data effortlessly.
- Durability:** Data is replicated across multiple locations for high availability and durability.

Performance: Offers fast data access and transfer speeds.

Cost-Effective: Provides various storage classes to optimize costs based on data access patterns.

Security: Robust security features to protect your data.

Configuration Options

When setting up a Google Cloud Storage bucket, you can define several key parameters to manage cost, access, and data retention.

Storage Classes

A storage class sets costs for storage, retrieval, and operations, with minimal differences in uptime. When you create a GCS bucket, you'll choose if you want objects to be managed automatically or specify a default storage class based on how long you plan to store your data and your workload or use case.

The following Google Cloud storage classes are available depending on your workload:

- Standard
 - Highest performance and availability
 - Ideal for frequently accessed data
 - Best suited for active datasets
- Nearline
 - Optimized for infrequent access
 - Lower cost than Standard
 - Suitable for data accessed a few times per month
 - Incur a small retrieval fee
- Coldline
 - Designed for data accessed a few times per year
 - Lower cost than Nearline
 - Significant retrieval fee applies
 - Suitable for long-term data archiving
- Archive
 - Lowest-cost storage option
 - For data that is accessed very infrequently
 - Highest retrieval fee
 - Ideal for long-term data retention with minimal access needs

Table 1.5 summarizes the differences between the various Google Cloud Storage classes.

TABLE 1.5 Comparison of Google Cloud storage options

Storage class	Name for APIs and CLIs	Minimum storage duration	Retrieval fees	Typical monthly availability	Cost (approximate)
Standard	STANDARD	None	None	99.99%	Highest
Nearline	NEARLINE	30 days	Yes	99.95%	Medium
Coldline	COLDLINE	90 days	Yes	99.95%	Low
Archive	ARCHIVE	365 days	Yes	99.95%	Lowest

Autoclass

When Autoclass is selected, Google Cloud Storage will automatically transition each object to Standard or Nearline class based on object-level activity, to optimize for cost and latency. Autoclass is recommended if usage frequency may be unpredictable and can be changed to a default class at any time.

Location Types

Each storage class is available in one of the following location configurations:

- Region:
 - Lowest latency within a single region
 - Use when data primarily accessed from within a single region
 - Suitable for applications with lower availability requirements
 - Lowest-cost storage option
- Dual-region:
 - High availability and low latency across two regions
 - Data redundancy across two regions
 - Use when compliance requirements mandate data residency in specific regions
 - Ideal for applications requiring high availability and disaster recovery
 - Suitable for applications with moderate latency requirements
 - Typically the most expensive option
- Multiregion:
 - Highest availability across largest area
 - Use when you have global applications with users distributed across multiple regions
 - Ideal for applications that access data with stringent availability requirements
 - Suitable for applications that can tolerate slightly higher latency

- Higher storage price than regions, but lower than dual-regions
- Outbound data transfer charges always apply when reading data

Table 1.6 summarizes the difference between the various Google Cloud Storage location options.

Additional Storage Location Considerations

Additional factors to take into account are as follows:

- Colocate Data and Compute for Optimal Performance and Cost-Efficiency:** Store data in the same region as your applications.
- Choose Storage Location Based on Data Access Patterns and Life Cycle:** Frequent access data should be in the same region as compute resources, whereas long-term archival data can be stored in lower-cost storage classes.
- Consider Data Durability, Availability, and Regulatory Requirements:** Multiregion storage provides higher durability but may incur additional costs. Ensure compliance with data residency laws.
- Implement Life-Cycle Management and Cost Optimization Strategies:** Regularly review storage classes and retention policies to reduce costs.
- Factor in Disaster Recovery and Business Continuity:** Use multiregion storage for redundancy and develop data backup and recovery plans.

TABLE 1.6 Comparison of Google Cloud storage location options

Feature	Region	Dual-region	Multiregion
Availability	Cross-zone redundancy	Cross-region redundancy	Cross-region redundancy
Performance	Best performance within a single region (up to 200 Gbps per region)	Good performance across two regions (up to 200 Gbps per region)	Moderate performance across multiple regions (up to 50 Gbps per region)
Latency	Lowest latency within a single region	Lower latency across two regions	Higher latency across multiple regions
Costs	Lower cost than dual-region and multiregion	Higher storage cost than region. Adds redundancy cost	Higher storage cost than region; cost-effective for high availability needs
Typical Use Cases	Analytics, backup and archive	Analytics, backup and archive, disaster recovery	Content serving

Use Cases

Google Cloud Storage acts as a flexible and scalable data repository that seamlessly integrates with other Google Cloud services, enabling efficient data pipelines, machine learning workflows, analytics, and application development. Here are just some of the use cases that can leverage Google Cloud Storage and when to select each storage class:

- **Standard Storage**
 - **High-Performance Applications:** Serving web and mobile applications with frequent data access patterns.
 - **Real-Time Analytics:** Storing and processing data for immediate insights and decision-making.
 - **Content Delivery Networks (CDNs):** Distributing static content globally for fast delivery.
 - **Data Lakes:** Storing large volumes of raw data for exploratory analysis and machine learning.
 - **Machine Learning:** Store training data and models.
 - **Big Data Processing:** Stage data for processing with tools like Dataflow.
- **Nearline Storage**
 - **Long-Term Data Archiving:** Storing data for infrequent access, such as backup data or compliance requirements.
 - **Data Warehousing:** Storing large datasets for offline analytics and reporting.
 - **Disaster Recovery:** Storing backups of critical data for recovery in case of system failures.
 - **Media Archives:** Storing video, audio, and image files for long-term preservation.
 - **Internet of Things (IoT) Data:** Storing large volumes of sensor data for analysis and insights.
- **Coldline Storage**
 - **Deep Archives:** Storing data for very infrequent access, such as historical data or legal records.
 - **Disaster Recovery Cold Storage:** Storing long-term backups for disaster recovery purposes.
 - **Compliance and Regulatory Data:** Storing data for long-term retention to meet compliance requirements.
 - **Media Libraries:** Storing large media archives for occasional access.
 - **Research Data:** Storing large datasets for long-term preservation and analysis.

- Archive Storage
 - **Regulatory Compliance:** Storing data for long-term retention to meet regulatory requirements.
 - **Disaster Recovery Cold Storage:** Storing very long-term backups for disaster recovery.
 - **Historical Data:** Storing historical data for research or analysis purposes.
 - **Digital Preservation:** Storing digital assets for long-term preservation.

File Storage

File storage uses a hierarchical structure to organize data. This means data is arranged in a tree-like fashion with folders (or directories) acting as branches and files as leaves. File storage stores data in files, organized in hierarchical directories. Access is via file paths. File storage is similar to how your operating system stores and organizes your files. The following is a visual representation of a file storage hierarchy (Figure 1.2):

File storage is ideal for sharing data among multiple users and applications, especially when protocols such as NFS and SMB are preferred. While not as fast as block storage, file storage offers better organization and compatibility.

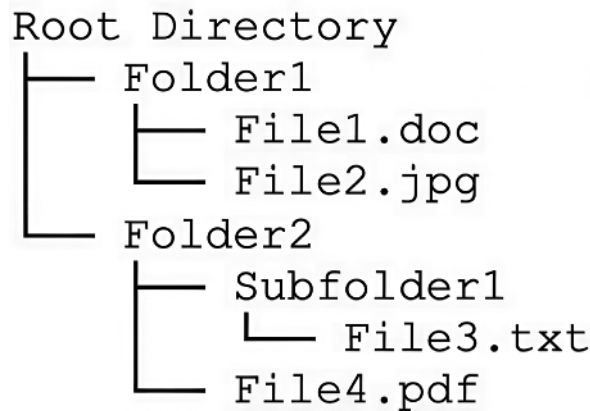
The following are some considerations for when to choose file storage:

Shared Access Is Required: Multiple users or applications need to access and modify the same data simultaneously.

Hierarchical Organization Is Essential: Data is naturally structured in folders and subfolders, making it easier to manage and find specific files.

Random Access Patterns Are Common: Frequent reads and writes to different parts of the filesystem are necessary.

FIGURE 1.2 Visual representation of a file storage hierarchy.



Compatibility Is a Priority: Existing applications and workflows rely on standard filesystem protocols (like NFS or SMB).

Performance Needs Are Moderate: While not as fast as block storage, file storage provides sufficient performance for many applications.

The two main services in Google Cloud that provide file storage services are Google Cloud Filestore and Google Cloud NetApp Volumes.

Google Cloud Filestore

Google Cloud Filestore is a fully managed network file storage service offered by Google Cloud Platform. It provides a scalable, high-performance, and reliable platform for storing and accessing data. Filestore is designed to simplify the management of file-based workloads by handling infrastructure and maintenance, allowing users to focus on their applications. It offers compatibility with standard NFS clients, making it easy to integrate with existing systems. With its ability to scale resources up or down based on demand, Filestore is suitable for a wide range of use cases, including media rendering, data analytics, and content management systems.

Google Cloud Filestore organizes data in a hierarchical structure, similar to traditional filesystems. This means data is stored in files and folders, allowing for easy navigation and management. It's designed to mimic the familiar filesystem structure used on local computers, making it intuitive for users and applications. This hierarchical organization enables efficient data access, sharing, and management within the cloud environment.

Key Features

Key features include the following:

Performance Tiers: Filestore offers different performance tiers to match varying workload requirements, ensuring optimal performance for different applications.

Scalability: You can easily scale Filestore instances up or down to accommodate changes in data volume or performance needs.

High Availability: Filestore is designed for high availability, providing redundancy and fault tolerance to minimize downtime.

Data Protection: Filestore incorporates data protection features such as encryption and snapshots to safeguard your data.

Deeply Integrated with the NFS Protocol: Filestore offers robust support for both NFSv3 and NFSv4.1 versions. This ensures broad compatibility with a wide range of operating systems and applications that rely on the NFS protocol for file sharing.

Integration with Other Google Cloud Services: Filestore seamlessly integrates with other Google Cloud services, such as Compute Engine, Kubernetes, and Cloud Storage, to create comprehensive cloud solutions.

Configuration Options

Google Cloud Filestore offers several service tiers to cater to different performance, capacity, and availability needs. The Basic tiers are cost-effective for development and testing, while the Zonal and Regional tiers provide higher performance and availability for production workloads.

Basic HDD:

- Cost-effective
- Lower performance
- Suitable for file sharing, software development, and use with GKE workloads
- Have a performance increase when the provisioned capacity exceeds 10 TiB

Basic SSD:

- Improved performance over HDD
- Suitable for workloads requiring moderate performance
- Higher cost than Basic HDD
- Performance of basic SSD instances is fixed regardless of the provisioned capacity

Zonal:

- Higher performance than Basic tiers
- Single zone availability
- Ideal for workloads requiring consistent performance within a zone
- Suitable for high-performance computing application requirements such as genome sequencing, and financial-services trading analysis
- When compared with basic SSD instances, zonal may provide a better option for users that require higher availability and lower capacity.
- When compared with regional instances, zonal may provide a more cost-effective option.

Regional:

- Highest availability among the listed tiers
- Replicated across multiple zones providing resiliency to zonal outages
- Suitable for mission-critical applications requiring high uptime
- Optimized for enterprise-grade network attached storage (NAS)

- Suitable for high availability for mission-critical, high-performance computing workloads running within a large namespace
- Supports linear performance that scales with capacity, within the parameters of the specified capacity range

Legacy service tiers

- Filestore's legacy service tiers include the following:
 - Standard
 - Premium
 - High scale SSD
 - Enterprise

When creating new instances, users are strongly encouraged to use these replacements.

- Basic HDD replaces the standard service tier.
- Basic SSD replaces the premium service tier.
- Zonal replaces the high-scale SSD service tier.
- Regional replaces the enterprise service tier.

Use Cases

Here are some use cases:

- Basic HDD
 - **Development and Testing Environments:** Cost-effective storage for early-stage projects.
 - **File Sharing for Small Teams:** Basic collaboration needs within a limited user base.
 - **Web Hosting:** For static content serving with lower-performance requirements.
- Basic SSD
 - **Development and Testing Environments:** Faster performance for iterative development.
 - **Small-Scale Data Analytics:** Handling moderate data processing workloads.
 - **Media Streaming:** For lower-quality or less demanding video streaming applications.
- Zonal
 - **High-Performance Computing (HPC):** For applications requiring high throughput and low latency.
 - **Media Rendering and Transcoding:** Handling large media files and processing intensive tasks.
 - **In-Memory Databases:** As a high-performance storage backend for in-memory databases.

- **Regional**
 - **Mission-Critical Applications:** For workloads demanding high availability and low downtime.
 - **Database Backups:** Storing critical database backups for disaster recovery.
 - **Content Management Systems:** Managing large volumes of digital assets with high availability requirements.

Table 1.7 summarizes the four main Filestore configurations options and compares the key features of each.

Google Cloud NetApp Volumes

Google Cloud NetApp Volumes is a fully managed, high-performance file storage service offered by Google Cloud Platform in partnership with NetApp. It provides a scalable, reliable, and enterprise-grade platform for storing and accessing data. Designed to simplify the management of file-based workloads, NetApp Volumes handles infrastructure and maintenance, allowing

TABLE 1.7 Comparison of Google Cloud Filestore options

Feature	Basic HDD	Basic SSD	Zonal	Regional
Performance Tier	Entry	Standard	High	High
Maximum Throughput	Low	Medium	High	Highest
Workload Suitability	Low-to-medium IOPS, sequential workloads	Medium IOPS, mixed workloads	High IOPS, mixed workloads	Highest IOPS, mixed workloads
Price	Lowest	Low	Medium	Highest
Typical Use Cases	Development/test environments, low-performance file shares	General purpose file sharing, medium performance applications	Production environments requiring high availability and performance	High-performance, mission-critical applications requiring low latency and high throughput
Recommended for	Cost-conscious users with low performance requirements	Users needing balanced performance and cost	Users requiring high availability and performance	Users demanding the highest performance and availability

users to focus on their applications. It offers broad protocol support, including NFS and SMB, ensuring compatibility with a wide range of systems and applications. By providing features like instant capacity adjustments, performance tiers, and advanced data protection capabilities, NetApp Volumes is tailored to meet the demands of complex and performance-intensive enterprise workloads. It offers a hierarchical data organization similar to traditional filesystems, making it intuitive for users and applications to manage and access data efficiently.

Key Features

Key features include the following:

- Fully Managed Service:** Eliminates the need for infrastructure management and maintenance, allowing users to focus on applications.
- Rapid Provisioning:** Enables quick creation and deployment of storage volumes to accelerate application development and deployment.
- Performance Tiers:** Delivers consistent performance for demanding workloads, with options for different performance tiers to match specific needs.
- Scalability:** Offers flexible capacity provisioning and the ability to adjust resources on-demand to accommodate changing workloads.
- Multiprotocol Support:** Supports both NFS and SMB protocols and Active Directory (AD) integration, ensuring compatibility with a wide range of systems and applications.
- Advanced Data Protection:** Provides features like snapshots, backups, and replication to protect data from loss and ensure business continuity.
- Hierarchical Data Organization:** Offers familiar filesystem structure for easy data management and access.
- Integration with Google Cloud:** Seamlessly integrates with other Google Cloud services for comprehensive cloud solutions.

Configuration Options

Google Cloud NetApp Volumes offers multiple service tiers to cater to different performance needs:

- **Flex**
 - Highly available, general-purpose storage with up to 16 KiBps per GiB of pool capacity (shared across all volumes in the pool), up to a maximum of 1 GiBps per pool.
 - Ideal for common enterprise workloads like NFS/SMB file shares, SAP shared files, containerized workloads, and Google Cloud VMware Engine.
- **Standard**
 - Highly available, general-purpose storage with up to 16 KiBps per GiB of volume capacity. Maximum of 1 GiBps per volume.

- Suitable for common enterprise workloads like NFS/SMB file shares, SAP shared files, and Google Cloud VMware Engine.
- Premium
 - Highly available, high-performance storage with up to 64 KiBps per GiB dedicated to the volume. Maximum of 4.5 GiBps per volume.
 - Best for performance-critical workloads requiring low latency, like Windows/enterprise NFS, self-managed databases, file shares, VDI, and VMware Engine.
- Extreme
 - Highly available, high-throughput storage with up to 128 KiBps per GiB dedicated to the volume. Maximum of 4.5 GiBps per volume.
 - Optimized for performance-critical workloads needing high throughput and low latency, such as Windows/enterprise NFS, self-managed databases, file shares, VDI, and VMware Engine.

Additional Considerations

Additional factors to consider includes these:

Auto-Tiering: Available for Premium and Extreme tiers, allowing for cost optimization by moving infrequently accessed data to a lower-cost storage tier.

Large Volumes: Premium and Extreme tiers support larger volumes (up to 1 PiB) for specific use cases.

Performance Considerations: Workload characteristics (read/write ratio, block size, I/O concurrency) significantly impact performance.

Use Cases

Here are some use cases:

- Flex
 - **General-Purpose File Shares:** Suitable for shared filesystems used by multiple users or applications with moderate performance requirements.
 - **Development and Test Environments:** Provides a cost-effective option for nonproduction workloads.
 - **Backup and Archive Storage:** Can be used for storing less frequently accessed data.
 - **Home Directories:** Suitable for storing user home directories with moderate file sizes.
 - **Media Streaming:** Can be used for storing and serving media content with moderate performance needs.

- **Standard**
 - **Virtual Desktops:** Provides sufficient performance for basic user workloads in VDI environments.
 - **DevOps Environments:** Supports build and test environments with moderate performance demands.
 - **SAP Shared Files:** Handles SAP workloads with standard performance requirements.
 - **Database Backups:** Suitable for storing database backups with regular access patterns.
 - **Content Management Systems:** Can be used for storing and managing digital assets.
- **Premium**
 - **High-Performance Computing (HPC):** Offers the performance required for data-intensive workloads.
 - **Online Transaction Processing (OLTP) Databases:** Supports databases with high transaction rates and low latency requirements.
 - **Virtual Desktop Infrastructure (VDI):** Provides a better user experience for demanding VDI workloads.
 - **Media and Entertainment Workloads:** Handles high-resolution video and audio editing and rendering.
 - **Enterprise File Shares:** Supports demanding file sharing environments with large file sizes and high concurrency.
- **Extreme**
 - **Real-Time Analytics:** Provides the performance needed for processing large datasets with low latency.
 - **In-Memory Databases:** Supports high-performance in-memory databases with demanding workloads.
 - **Financial Trading Applications:** Handles high-frequency trading with ultra-low latency requirements.
 - **3D Rendering and Animation:** Supports complex 3D workloads with large datasets.
 - **High-Performance Data Analysis:** Ideal for data scientists and analysts requiring maximum performance.

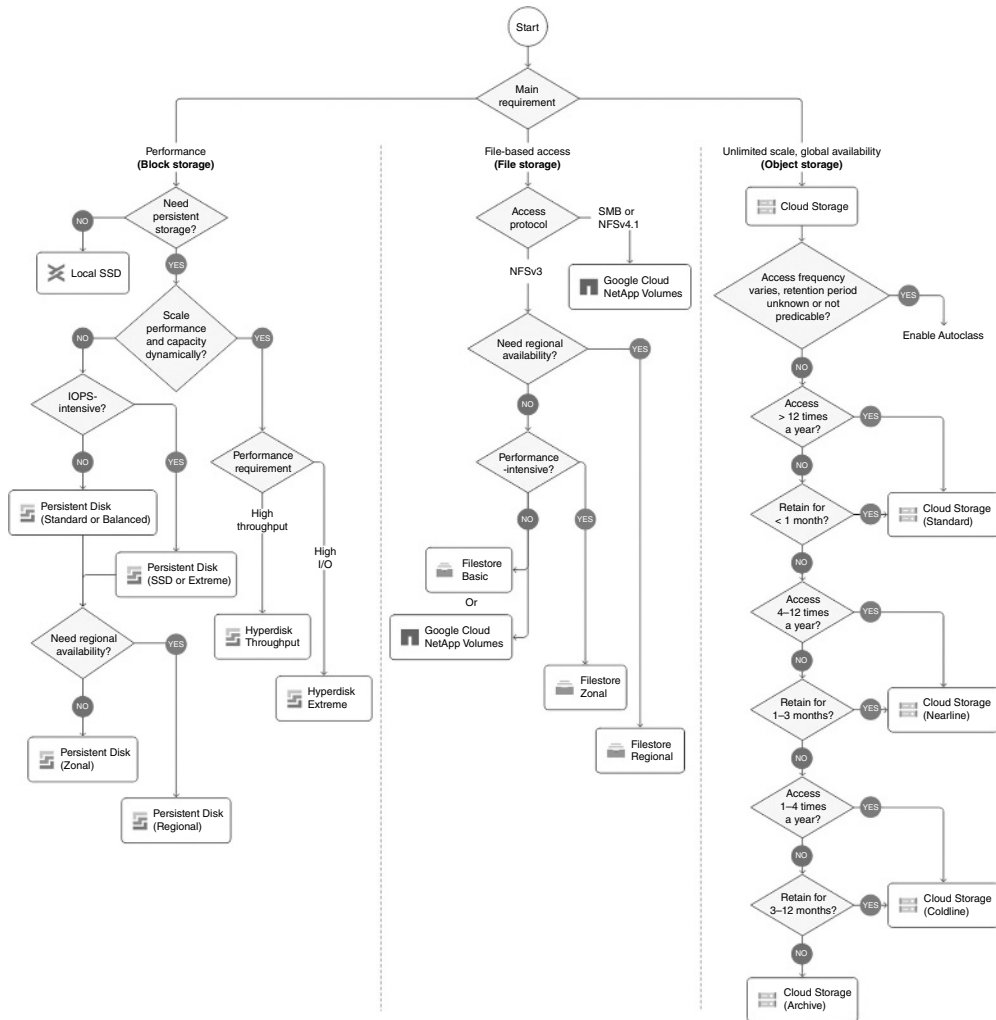
TABLE 1.8 Comparison of Google Cloud NetApp Volumes options

Feature	Flex	Standard	Premium	Extreme
Performance Tier	Entry level	General purpose	High performance	Ultra-high performance
Maximum Throughput	16 MiBps per TiB pool capacity	16 MiBps per TiB volume capacity	64 MiBps per TiB volume capacity	128 MiBps per TiB volume capacity
Workload Suitability	Development, testing, small file shares	File shares, VMs, databases	File shares, VMs, databases, high-performance computing	OLTP databases, low-latency applications
Price	Most economical	Balanced	Higher performance, higher cost	Highest performance, highest cost
Typical Use Cases	Noncritical workloads, cost-sensitive environments	General-purpose workloads, medium-sized file shares	Performance-sensitive workloads, large file shares	Mission-critical workloads, demanding applications
Recommended for	Smaller datasets, lower IOPS requirements	Medium-sized datasets, moderate IOPS requirements	Large datasets, high IOPS requirements	Very large datasets, extremely high IOPS requirements

Table 1.8 summarizes the four main NetApp Volumes confirmation options and compares the key features of each.

Storage Options Decision Tree

The decision tree in Figure 1.3 guides you through the Google Cloud storage recommendations.

FIGURE 1.3 Google storage options decision tree.

Summary

This chapter explored the fundamental concepts of data storage technologies within Google Cloud. It delved into the key differences between block, object, and file storage, highlighting their respective strengths and weaknesses. The chapter then examined the various storage options available within Google Cloud, including Persistent Disk, Hyperdisk, Local SSD, and Google Cloud Storage, providing a detailed analysis of their features, configuration options, and use cases.

Exam Essentials

Understand the evolving landscape of data. Recognize the trends driving the growth of unstructured data and its impact on storage needs. Explain the challenges of managing exploding data volumes and the advantages of cloud storage solutions.

Master Google Cloud storage options. Differentiate between block, object, and file storage formats in terms of their core characteristics, use cases, and access methods.

Describe the key features, configuration options, and use cases of Google Cloud's block storage options: Persistent Disk, Hyperdisk, and Local SSD.

Understand the different types of Persistent Disk (standard, performance, balanced, extreme) and their appropriate use cases.

Explain the advantages of Hyperdisk compared to Persistent Disk and identify the key benefits of each Hyperdisk type (balanced, high availability, extreme, ML, throughput).

Understand the strengths and limitations of Local SSDs and identify scenarios where they are most suitable.

Explore Google Cloud Storage (GCS). Understand the fundamentals of object storage and its advantages compared to block storage.

Describe the key features, configuration options, and use cases of Google Cloud Storage (GCS).

Analyze the different storage classes (Standard, Nearline, Coldline, Archive) and their appropriate use cases.

Explain the concept of data immutability in GCS and its implications for data integrity.

Differentiate between the different GCS location types (region, dual-region, multi-region) and their impact on availability, latency, and cost.

Apply a systematic approach to storage solution selection. Analyze workload characteristics such as data volume, growth rate, access patterns, durability requirements, retention policies, and compliance needs to inform your decision.

Consider factors like performance, scalability, cost optimization, data protection, and security to make informed storage design choices.

Review Questions

1. You are working on a project that involves storing large volumes of unstructured data, such as images, videos, and audio files. Which Google Cloud storage solution would be most appropriate for this scenario?
 - A. Persistent Disk
 - B. Google Cloud Storage
 - C. Cloud Filestore
 - D. Hyperdisk
2. You are tasked with designing a storage solution for a high-frequency trading platform. What is the most important consideration for this scenario?
 - A. Cost optimization
 - B. Data durability
 - C. Performance and scalability
 - D. Security and compliance
3. Which Google Cloud storage solution is directly attached to a virtual machine and is ideal for I/O-intensive workloads?
 - A. Persistent Disk
 - B. Hyperdisk
 - C. Local SSD
 - D. Google Cloud Storage
4. You are working on a project that requires high performance and low latency for data access. Which Google Cloud storage solution would you recommend?
 - A. Standard Persistent Disk
 - B. Balanced Persistent Disk
 - C. Extreme Persistent Disk
 - D. Hyperdisk
5. You are designing a storage solution for a large-scale data warehousing application. Which Google Cloud storage class would you recommend for older sales data that is accessed infrequently?
 - A. Standard
 - B. Nearline
 - C. Coldline
 - D. Archive

6. A social media company needs to store a massive amount of user-generated content, including images, videos, and text posts. The data is expected to grow rapidly. Which storage type is most suitable for this scenario?
 - A. Block storage
 - B. File storage
 - C. Object storage
 - D. In-memory storage
7. A database architect needs a storage solution for a virtual machine running a high-performance database that requires low latency and high throughput. The data stored on this volume must persist even if the virtual machine is terminated. Which storage option is most appropriate?
 - A. Object storage
 - B. Persistent Disk
 - C. Local SSD
 - D. Cloud storage
8. Which Google Cloud storage option offers the best performance but does not guarantee data durability?
 - A. Persistent Disk
 - B. Hyperdisk
 - C. Local SSD
 - D. Object storage
9. What is a key difference between Google Cloud Filestore and Google Cloud NetApp Volumes?
 - A. Filestore is a Google-managed service, while NetApp Volumes is a partner-managed service.
 - B. Filestore supports NFS, while NetApp Volumes supports SMB.
 - C. Filestore is suitable for small-scale workloads, while NetApp Volumes is suitable for enterprise-grade workloads.
 - D. Filestore offers lower performance, while NetApp Volumes offers higher performance.
10. A genomics research company needs to store and analyze massive datasets containing DNA sequences. The data is expected to grow rapidly over time. Which storage type is most suitable for this scenario?
 - A. Block storage
 - B. Object storage
 - C. File storage
 - D. In-memory database

11. Which Google Cloud storage solution is a fully managed network file storage service that is compatible with standard NFS clients?
 - A. Cloud storage
 - B. Cloud Filestore
 - C. Hyperdisk
 - D. NetApp Volumes
12. A company is migrating an on-premises application to Google Cloud. The application requires a filesystem that can be accessed concurrently by multiple virtual machines. Which storage service best meets this requirement?
 - A. Cloud Filestore
 - B. Cloud storage
 - C. Persistent Disk
 - D. Local SSD
13. Your company is migrating an on-premises Oracle database to Google Cloud. The database requires high IOPS and throughput for optimal performance. Which Google Cloud storage option is most suitable for this scenario?
 - A. Cloud storage
 - B. Persistent Disk (pd-standard)
 - C. Persistent Disk (pd-extreme)
 - D. Local SSD
14. A startup is building a video editing application on Google Cloud and requires high-performance storage to handle large video files and processing tasks. The application also needs low latency access to temporary files during the editing process. Which storage options would you recommend for this use case?
 - A. Persistent Disk (pd-standard) for video files and Local SSD for temporary files
 - B. Hyperdisk Extreme for video files and Persistent Disk (pd-balanced) for temporary files
 - C. Cloud Filestore for video files and Local SSD for temporary files
 - D. Cloud Storage Nearline for video files and Persistent Disk (pd-ssd) for temporary files
15. You are designing a disaster recovery strategy for a mission-critical application running on Google Cloud. You need to ensure that data is replicated across multiple geographic regions to protect against regional outages. Which Google Cloud Storage location type should you choose?
 - A. Region
 - B. Dual-region
 - C. Multiregion
 - D. None of the above

16. Your team is developing a new application that requires a shared filesystem accessible from multiple compute engine instances. You need a storage solution that offers high performance and low latency, and that supports the NFS protocol. Which Google Cloud storage option would you recommend?
- A. Cloud storage
 - B. Cloud Filestore
 - C. Persistent Disk
 - D. Local SSD
17. You are working on a machine learning project that involves training a model on a massive dataset stored in Google Cloud Storage. The training process requires high throughput to read the data efficiently. Which storage class should you choose to optimize performance and minimize costs?
- A. Standard
 - B. Nearline
 - C. Coldline
 - D. Archive
18. You need to store a large volume of data that is infrequently accessed but requires high durability and low cost. This data will be used for disaster recovery purposes and long-term archival. Which Google Cloud storage class best fits these requirements?
- A. Standard
 - B. Nearline
 - C. Coldline
 - D. Archive
19. Your organization is subject to data residency regulations that require certain data to be stored within specific geographic regions. You need to choose a Google Cloud Storage location type that allows you to control data placement and ensure compliance. Which option should you select?
- A. Region
 - B. Dual-region
 - C. Multiregion
 - D. Global
20. You have a requirement to store frequently accessed data for a web application. This data needs to be highly available and deliver low latency to users across the globe. Which Google Cloud Storage location type is most suitable for this scenario?
- A. Region
 - B. Dual-region
 - C. Multiregion
 - D. None of the Above

