**Type of Data**

1. **Structured Data**
   * Highly organized and easily searchable in databases (e.g., SQL databases, spreadsheets).
   * Examples: Tables, columns, rows.
2. **Unstructured Data**
   * Not organized in a predefined manner.
   * Examples: Text documents, images, videos, social media posts.
3. **Semi-Structured Data**
   * Contains elements of both structured and unstructured data.
   * Examples: XML, JSON, HTML.

Comparison between different type of data

| **Properties** | **Structured data** | **Semi-structured data** | **Unstructured data** |
| --- | --- | --- | --- |
| Technology | It is based on Relational database table | It is based on XML/RDF(Resource Description Framework). | It is based on character and binary data |
| Transaction management | Matured transaction and various concurrency techniques | Transaction is adapted from DBMS not matured | No transaction management and no concurrency |
| Version management | Versioning over tuples, row, tables | Versioning over tuples or graph is possible | Versioned as a whole |
| Flexibility | It is schema dependent and less flexible | It is more flexible than structured data but less flexible than unstructured data | It is more flexible and there is absence of schema |
| Scalability | It is very difficult to scale DB schema | It’s scaling is simpler than structured data | It is more scalable. |
| Robustness | Very robust | New technology, not very spread | — |
| Query performance | Structured query allow complex joining | Queries over anonymous nodes are possible | Only textual queries are possible |

Storage Needs for an Enterprise

# Application wise Usecases on storage

* WebApps: To store media (images, docs, videos)
* Backups and Archives: To store backups
* Applications:Disks
* IT:Shared disk

Cloud: think of infra as software (we use someone else's hardware and we pay according to usage)

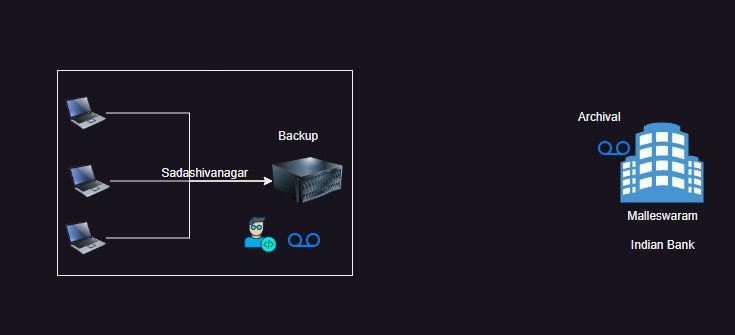
Backup and Archives

Backup:

* Used for quickly recovering from failures It should be quick
* prices should be moderate

Archive:

* Used for recovering from disasters
* It might take time in hours or days to recover Storage costs should be cheaper



# Media Storage

Media used by Websites, these are accessed from web This is ever-growing. Generally, companies use Content Delivery Networks.

CDN: A Content Delivery Network (CDN) is a network of distributed servers strategically placed around the globe to deliver web content and other digital assets to users more efficiently. The primary goal of a CDN is to reduce latency and improve the performance, speed, and reliability of web applications by bringing content closer to the end-users.

**How a CDN Works**

1. **Content Distribution**: A CDN caches (stores) content on servers in multiple geographic locations. These servers are known as edge servers or points of presence (PoPs).
2. **User Request**: When a user requests content (e.g., a web page, image, or video), the request is routed to the nearest edge server rather than the origin server.
3. **Content Delivery**: The edge server delivers the cached content to the user, resulting in faster load times and reduced latency.
4. **Cache Refresh**: If the requested content is not available on the edge server, the CDN fetches it from the origin server, delivers it to the user, and then caches it for future requests.

**Benefits of Using a CDN**

1. **Improved Performance**: By serving content from locations closer to the user, CDNs reduce latency and improve load times.
2. **Scalability**: CDNs can handle large volumes of traffic and sudden spikes in demand, ensuring that websites and applications remain available even under heavy load.
3. **Reliability and Redundancy**: With multiple servers distributed globally, CDNs provide redundancy and failover capabilities, enhancing the reliability of content delivery.
4. **Security**: CDNs offer enhanced security features, such as DDoS protection, web application firewalls (WAF), and secure sockets layer (SSL) encryption, to protect against cyber threats.
5. **Cost Efficiency**: By offloading traffic from origin servers, CDNs reduce bandwidth costs and server load, making content delivery more cost-effective.

**Common Use Cases**

1. **Websites and Web Applications**: CDNs accelerate the delivery of web pages, scripts, and stylesheets, improving the user experience.
2. **Video Streaming**: CDNs optimize the delivery of video content, reducing buffering and improving streaming quality.
3. **Software Distribution**: CDNs facilitate the fast and reliable distribution of software updates, patches, and downloads.
4. **E-commerce**: CDNs enhance the performance and reliability of online stores, ensuring a smooth shopping experience for customers.
5. **Gaming**: CDNs improve the delivery of game assets, updates, and patches, providing a better gaming experience.

**Major CDN Providers**

1. **Akamai**
2. **Cloudflare**
3. **Amazon CloudFront**
4. **Google Cloud CDN**
5. **Microsoft Azure CDN**
6. **Fastly**
7. **Limelight Networks**

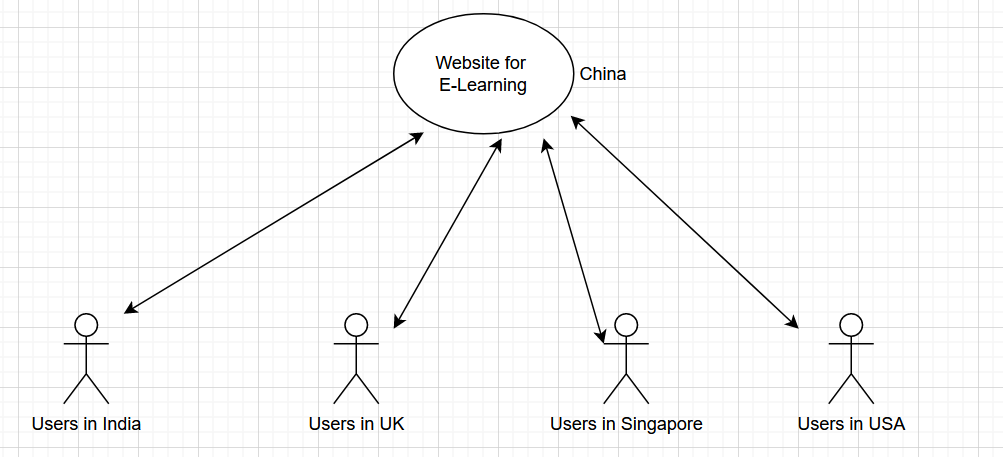
**Key CDN Components**

1. **Edge Servers**: Distributed servers that cache and deliver content to users.
2. **Origin Servers**: The original source of content that the CDN caches and delivers.
3. **PoPs (Points of Presence)**: Geographic locations where CDN servers are located.
4. **CDN Management Software**: Tools and interfaces for configuring, managing, and monitoring the CDN.

**Conclusion**

CDNs play a crucial role in the modern internet infrastructure, enhancing the performance, security, and scalability of web services. They are essential for businesses that require fast, reliable, and secure content delivery to a global audience.

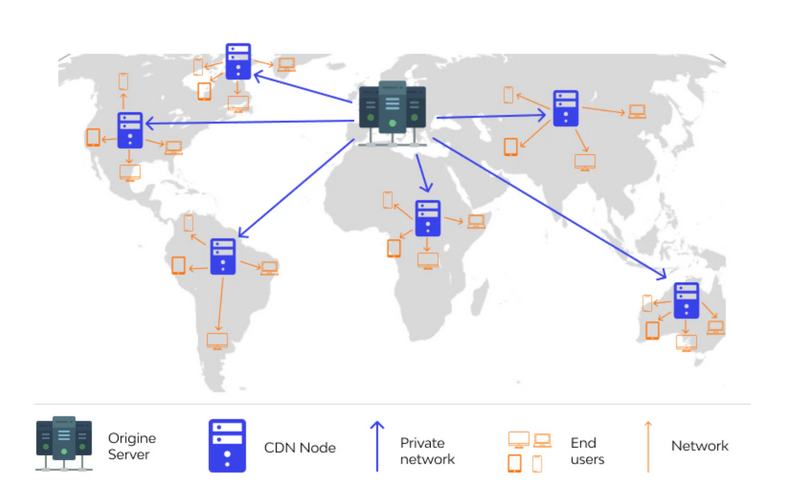
CDN by Diagram



Let say we have E-learning website hosting on China, and people across the glob are accessing it, so they need to feel latency.

Latency: Latency is the time delay experienced in a system, typically measured from the moment an action is initiated to the moment its effect is observed. In the context of networking and web services, latency refers to the time it takes for a data packet to travel from the source to the destination and back. Lower latency means faster response times, which is crucial for applications that require real-time interactions.

Now, CDN places a server near to user for improving latency.



Link: [https://www.shopify.com/ph/blog/what-is-cdn#](https://www.shopify.com/ph/blog/what-is-cdn)