

Chapter 24: Performance Optimization

24.1 React Performance

24.1.1 What is React Performance?

- * Performance in React refers to how **fast** and **efficiently** a React application **loads**, **renders** and **responds** to user interactions.
- * Good performance improves **user experience**, reduces load time and minimizes unnecessary **re-renders**.
- * Optimization ensures that the application uses **minimum resources** and updates only what is necessary.

Poor Performance



Slow, Unresponsive

Good Performance



Fast, Responsive

24.1.2 Why Optimize React Performance?

- * Improves **user satisfaction** and **engagement**.
- * Reduces **bounce rate** and increases **conversions**.
- * Saves **bandwidth**, **CPU** and **memory**.
- * Helps applications **scale better** with more users.
- * Follows **best practices** for clean and maintainable code.



Better Performance

24.1.3 Common Performance Issues in React



Unnecessary
re-renders



Large
bundle size



Inefficient
state management



Heavy
components



Too many
API calls

Key Points:

- ✓ React performance depends on how **components render** and **update**.
- ✓ Avoid unnecessary **re-renders** and **heavy computations**.
- ✓ Optimize **state**, **props**, **context** and **side-effects**.
- ✓ Use **code-splitting**, **lazy loading** and **memoization** for better performance.



24.1.4 Best Practices for Better Performance

- * Use **React.memo** to prevent unnecessary re-renders.
- * Use **useMemo** and **useCallback** for memoization.
- * Optimize **state updates** and avoid deep nesting.
- * Use **lazy loading** and **code splitting**.
- * Reduce **bundle size** and remove unused dependencies.
- * Use **production build** for deployment.



Remember:

Optimize **early**, measure **often** and focus on what truly affects performance.



Key Takeaway:



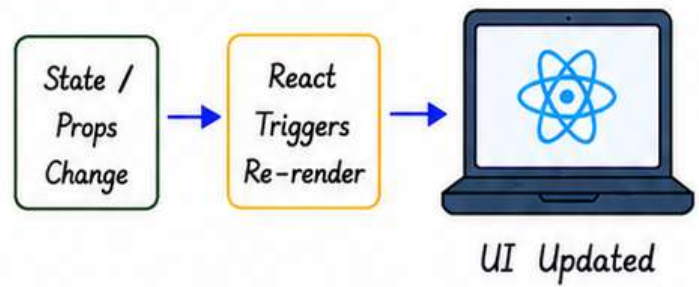
Optimizing React performance helps in building **fast**, **responsive** and **scalable** applications that provide a **smooth user experience**.



24.2 Component Re-rendering

24.2.1 What is Re-rendering?

- * In React, when a component's state or props change, it **re-renders**.
- * Re-rendering means React **calls the component function again** and updates the UI.
- * Re-rendering is normal, but **too many re-renders** can affect performance.



24.2.2 When Does a Component Re-render?

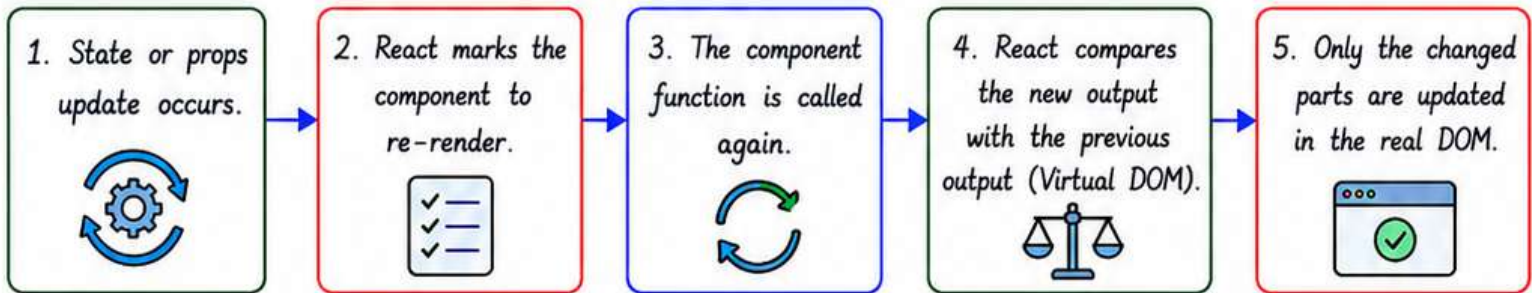
- * When its own **state** changes.
- * When its parent re-renders and passes new **props**.
- * When **context** value used by the component changes.
- * When a parent component re-renders for any reason.



Note:

Even if props or state values are the same, React **may still re-render** the component. This is called **unnecessary re-rendering**.

24.2.3 How Re-rendering Works in React



24.2.4 Causes of Unnecessary Re-renders



Inline functions created in JSX



Objects or arrays created inside the component



Passing props that change on every render



Not using memoization for child components



Unnecessary context updates

Key Points:

- ✓ Re-rendering is normal in React and keeps the UI in sync.
- ✓ Too many **unnecessary re-renders** can make the app **slow**.
- ✓ Understanding re-rendering is the first step to **optimize performance**.
- ✓ Use **memoization**, **stable props** and proper **state management** to reduce re-renders.



Key Takeaway:



Minimizing unnecessary re-renders helps React applications run **faster**, use fewer resources and provide a **better user experience**.

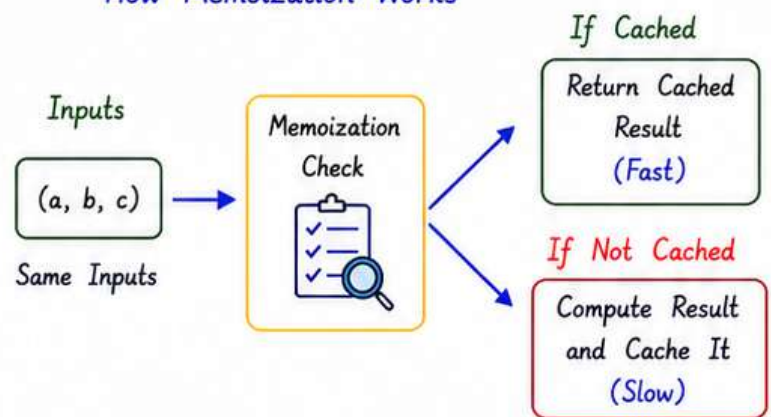


24.3 Memoization

24.3.1 What is Memoization?

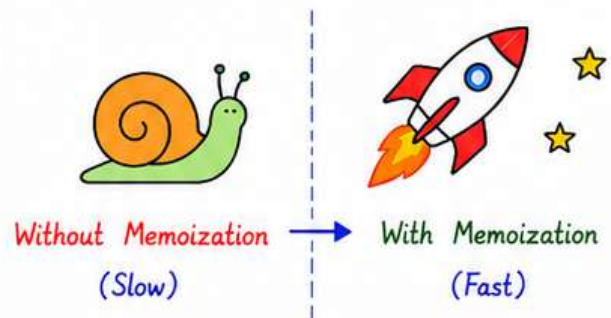
- * Memoization is a technique used to **store the result** of expensive function calls and return the **cached result** when the same inputs occur again.
- * It helps to avoid **unnecessary recomputations** and improves performance.
- * React provides built-in memoization utilities like **React.memo**, **useMemo** and **useCallback**.

How Memoization Works



24.3.2 Why Use Memoization?

- * Reduces **unnecessary re-renders**.
- * Speeds up **expensive calculations**.
- * Improves the **performance** of React applications.
- * Optimizes rendering of **components** and **functions**.
- * Helps in handling **large lists** and **complex UIs** efficiently.



24.3.3 Types of Memoization in React

1. React.memo

- * Memoizes a functional component.
- * Prevents re-rendering if props remain the same.
- * Used for optimizing pure components.



`React.memo()`

2. useMemo

- * Memoizes the result of expensive calculations.
- * Returns the cached value unless dependencies change.
- * Used inside functional components.



`useMemo()`

3. useCallback

- * Memoizes a callback function.
- * Returns the same function instance unless dependencies change.
- * Useful for optimizing event handlers and child components.



`useCallback()`

24.3.4 When to Use Memoization?



When rendering large lists of data.



When a component re-renders frequently with same props.



When performing expensive calculations inside components.



When passing callbacks to child components.



When dealing with complex data and state updates.

Key Points:

- ✓ Memoization helps to **cache results** and avoid **unnecessary work**.
- ✓ Use **React.memo** to memoize components.
- ✓ Use **useMemo** to memoize values.
- ✓ Use **useCallback** to memoize functions.
- ✓ Use memoization **only when necessary**, as **overuse** can add **complexity**.



Key Takeaway:



Memoization is a powerful optimization technique that **improves the performance** of React applications by **caching results** and **preventing unnecessary re-renders**.

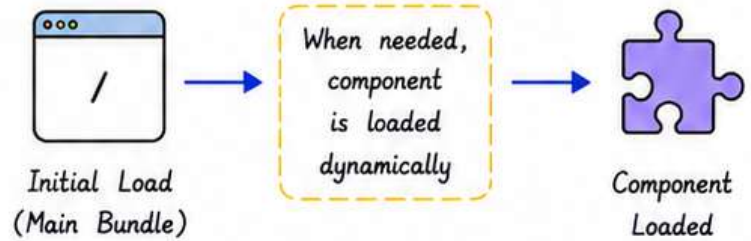


24.4 Lazy Loading

24.4.1 What is Lazy Loading?

- * Lazy loading is a technique that loads components **only when they are needed**, not at the initial load.
- * It helps to **reduce the initial bundle size** and improves performance.
- * React supports lazy loading using **React.lazy()** and **Suspense**.

How Lazy Loading Works



24.4.2 Why Use Lazy Loading?

- * Reduces the **initial load time**.
- * Improves **performance** and user experience.
- * Loads resources **only when users need them**.
- * Helps in building **scalable** and **efficient** applications.

24.4.3 How to Implement Lazy Loading in React?

```
import React, { Suspense } from 'react';
const About = React.lazy(() => import('./About'));

function App() {
  return (
    <div>
      <h1>My App</h1>
      <Suspense fallback= {<div>Loading...</div>} >
        <About />
      </Suspense>
    </div>
  );
}
```

- * **React.lazy()** is used to import a component dynamically.
- * **Suspense** is used to show a fallback UI while the component is loading.
- * The component is loaded **only when it** is rendered.

Note:
Lazy loading works only with **default exports**. **Named exports** require additional handling.

24.4.4 When to Use Lazy Loading?



When a component is used on a different route.



When the component is large and not immediately needed.



When loading admin panels or settings pages.



When building large scale applications with multiple features.



When you want to save bandwidth and improve initial load speed.

Key Points:

- ✓ Lazy loading **reduces the initial bundle size**.
- ✓ Use **React.lazy()** to load components dynamically.
- ✓ Use **Suspense** to handle loading states with a fallback UI.
- ✓ Lazy load components that are **not required at the initial render**.
- ✓ It improves **performance**, **user experience** and resource utilization.



Key Takeaway:



Lazy loading helps load components **only when needed**, reducing load time and **improving** the overall **performance** of React applications.

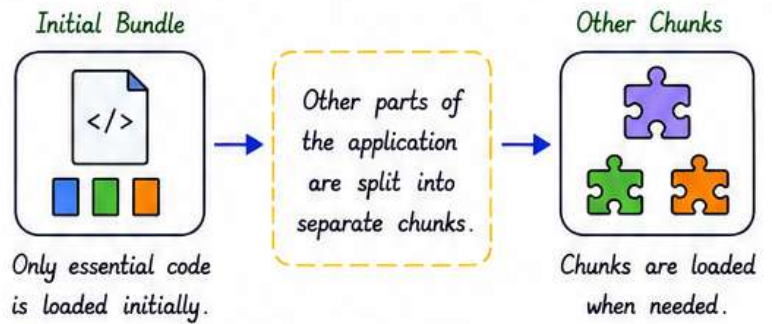


24.5 Code Splitting

24.5.1 What is Code Splitting?

- * Code splitting is a technique that breaks the application code into **smaller chunks**.
- * This allows the application to load only the **necessary code** for the current page.
- * It helps to **reduce the initial bundle size** and **improve performance**.
- * React supports code splitting using **React.lazy()**, **Suspense** and **dynamic import()**.

How Code Splitting Works



24.5.2 Why Use Code Splitting?

- * Reduces the **initial load time**.
- * Loads code **on demand** when required.
- * Improves **performance** and **user experience**.
- * Decreases **bandwidth** usage.
- * Helps in building **scalable** and **efficient** applications.



Note:

Every route or component can have its **own bundle**. These bundles are downloaded only when the user accesses that part.

24.5.3 How to Implement Code Splitting in React?

```
import React, { Suspense } from 'react';
const About = React.lazy(() => import('./About'));

function App() {
  return (
    <div>
      <h1>My App</h1>
      <Suspense fallback={<div>Loading...</div>}>
        <About />
      </Suspense>
    </div>
  );
}
```

- * **React.lazy()** is used to import a component dynamically.
- * **Suspense** is used to display a fallback UI while the component is loading.
- * The component is loaded **only when** it is rendered.



While Loading



After Loaded

24.5.4 When to Use Code Splitting?



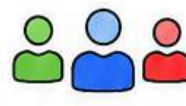
When different routes have different components.



When large components are not needed on every page.



When loading admin panels or settings pages.



When building large applications with multiple features.



When you want to improve performance and reduce load time.

Key Points:

- ✓ Code splitting breaks the code into **smaller chunks**.
- ✓ It **reduces the initial bundle size** and **improves performance**.
- ✓ Use **React.lazy()** and **Suspense** for dynamic imports.
- ✓ It loads components **only when** they are needed.
- ✓ It provides a **better user experience** and **saves bandwidth**.



Key Takeaway:



Code splitting helps React applications load **faster** by loading only the code that is **necessary**, when it is **necessary**.

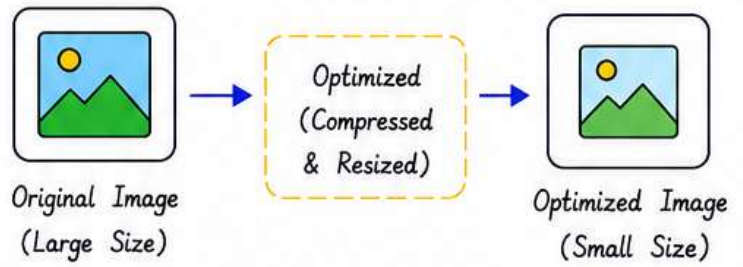


24.6 Image Optimization

24.6.1 What is Image Optimization?

- * Image optimization is the process of **reducing** the **file size** of images without losing quality.
- * It helps to **speed up** the website and **improve** **user experience**.
- * Optimized images load **faster** and **save** **bandwidth**.
- * It also helps in **better SEO** and performance.

How Image Optimization Helps



24.6.2 Why Optimize Images?

- * Reduces page **load time**.
- * Improves website **performance**.
- * Saves **bandwidth** and **server resources**.
- * Better **user experience** and engagement.
- * Helps in **better SEO** rankings.

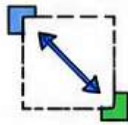
Impact of Image Optimization

	Before Optimization	After Optimization
Image Size	1.5 MB	120 KB
Load Time	5.2 sec	0.8 sec
Bandwidth Usage	High	Low
User Experience	Slow	Fast

24.6.3 Best Practices for Image Optimization



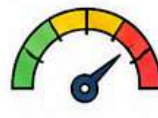
Use modern formats like WebP, AVIF.



Resize images to the required dimensions.



Compress images without losing quality.



Use lazy loading to load images when needed.



Use CDN to serve images from nearest location.



Avoid loading unused or hidden images.

24.6.4 Image Optimization in React

- * Use the **loading="lazy"** attribute for lazy loading.
- * Use responsive images with **srcSet** and **sizes**.
- * Use optimized image formats like **WebP**.
- * Use libraries like **next/image** (Next.js) for automatic optimization.
- * Store images in **CDN** or cloud storage.

```

```

Key Points:

- ✓ Optimized images load **faster** and **improve** performance.
- ✓ Use modern image formats like **WebP** and **AVIF**.
- ✓ **Compress** and **resize** images before uploading.
- ✓ Use **lazy loading** to load images only when needed.
- ✓ Use **CDN** to deliver images quickly to users.



Tip:
Smaller images =
Faster websites!



Key Takeaway:



Image optimization is essential for **fast**, **scalable** and **user-friendly** web applications. Always use optimized images for **better performance** and higher **user satisfaction**.

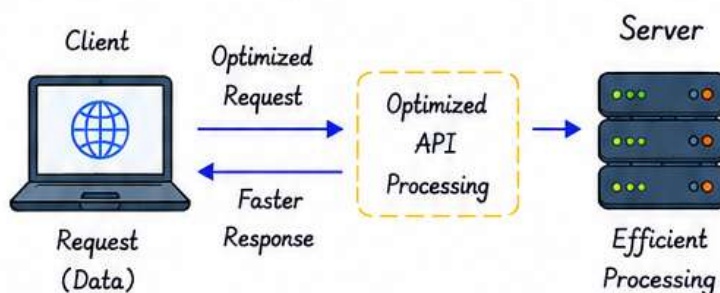


24.7 API Optimization

24.7.1 What is API Optimization?

- * API optimization is the process of improving the **performance** of APIs used in an application.
- * It reduces the **response time**, **server load** and **bandwidth usage**.
- * Optimized APIs help in building **faster**, **scalable** and **reliable** applications.
- * Good API optimization improves the overall **user experience**.

How API Optimization Works



24.7.2 Why Optimize APIs?

- * Reduces network **latency** and **response time**.
- * Decreases the **load on server** and **database**.
- * Saves **bandwidth** and reduces **data transfer**.
- * Improves **performance** and **user experience**.
- * Helps in handling **more users** and requests.

Impact of API Optimization		
	Before Optimization	After Optimization
Response Time	1200 ms	250 ms
Data Transfer	2.5 MB	400 KB
Server Load	High	Low
Bandwidth Usage	High	Low
User Experience	Slow	Fast

24.7.3 Best Practices for API Optimization



Use pagination to limit the amount of data returned.



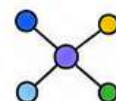
Use filtering to fetch only required data.



Use compression (Gzip/Brotli) to reduce payload size.



Implement caching to reduce unnecessary API calls.



Use field selection to request only needed fields.



Avoid multiple API calls by batching requests.

24.7.4 Techniques for API Optimization

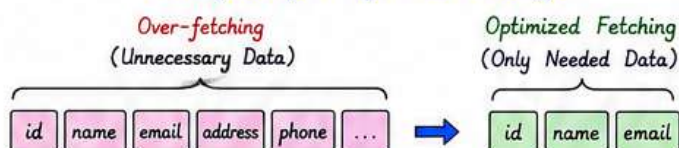
- * Use **pagination** (limit, offset) for large datasets.
- * Use **filtering** and **sorting** to get precise data.
- * Enable **response compression** to reduce data size.
- * **Cache responses** to avoid repeated requests.
- * Select only **required fields** (field selection).
- * Use **batch requests** to combine multiple calls.
- * Use **CDN** for static and semi-static data.



Note:

Always design APIs to return only the **necessary data**. Over-fetching and under-fetching can both affect performance.

Over-fetching vs Optimized Fetching



24.7.5 Example: API Call Optimization

Before Optimization

```
fetch('https://api.example.com/users')
  .then(res => res.json())
  .then(data => console.log(data));
// Returns all user fields (over-fetching)
```

After Optimization

```
fetch('https://api.example.com/users?fields=id,name,email&limit=10')
  .then(res => res.json())
  .then(data => console.log(data));
// Returns only required fields with limit (optimized)
```

Key Points:

- ✓ API optimization reduces **response time** and **server load**.
- ✓ Use **pagination**, **filtering**, **caching** and **compression**.
- ✓ Always request only the **required data**.
- ✓ Optimized APIs improve **performance**, **scalability** and **user experience**.



Key Takeaway:



Well-optimized APIs make applications **faster**, more **efficient** and **scalable** while providing a **better experience** to users.



24.8 Database Query Optimization

24.8.1 What is Database Query Optimization?

- * Query optimization is the process of improving database queries to make them **faster** and **more efficient**.
- * It reduces the **time taken to fetch data** from the database.
- * Well-optimized queries **reduce server load** and improve application performance.
- * It helps in **better resource utilization** and **faster response time**.

How Query Optimization Works



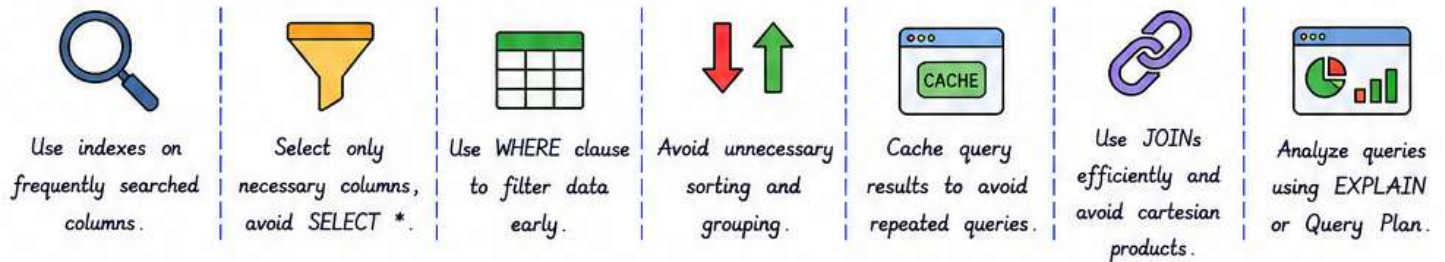
24.8.2 Why Optimize Database Queries?

- * Reduces query **execution time**.
- * Decreases **server load** and **CPU usage**.
- * Reduces unnecessary **data transfer**.
- * Improves **scalability** of the application.
- * Enhances overall **user experience**.
- * Helps in **better utilization** of database resources.

Impact of Query Optimization

	Before Optimization	After Optimization
Query Execution Time	1500 ms	180 ms
Database Load	High	Low
CPU Usage	High	Low
Data Fetched	Large	Required Only
User Experience	Slow	Fast

24.8.3 Best Practices for Query Optimization



24.8.4 Tips for Writing Optimized Queries

- * Use appropriate **indexes** for search and filter operations.
- * Use **LIMIT** clause to restrict the number of rows.
- * Avoid functions in **WHERE** clause on indexed columns.
- * Use **pagination** for large datasets.
- * Optimize **subqueries** and avoid nested subqueries.
- * Use **connection pooling** to handle database connections.
- * Keep database **statistics** updated.



Note:

EXPLAIN helps to **analyze** how a query will be executed by the database.

Example: Using EXPLAIN

```
EXPLAIN SELECT id, name FROM users
WHERE status = 'active' ORDER BY created_at DESC
LIMIT 10;
```

24.8.5 Example: Unoptimized vs Optimized Query

Unoptimized Query

```
SELECT * FROM orders
WHERE YEAR(order_date) = 2024
ORDER BY order_date DESC;
```

-- Fetches all columns, uses function in WHERE and may not use index.

Optimized Query

```
SELECT order_id, customer_id, total_amount, order_date
FROM orders
WHERE order_date >= '2024-01-01' AND order_date < '2025-01-01'
ORDER BY order_date DESC
LIMIT 50;
```

-- Retrieves only needed columns, uses index and limit.

Key Points:

- ✓ Query optimization reduces **execution time** and improves **performance**.
- ✓ Use **indexes**, **filtering**, and **pagination** for efficient queries.
- ✓ Avoid **unnecessary data** retrieval and complex subqueries.
- ✓ Regularly **analyze** and **optimize queries** for better efficiency.



Key Takeaway:



Optimizing database queries leads to **faster response times**, **lower server load** and **better user experience** in applications.

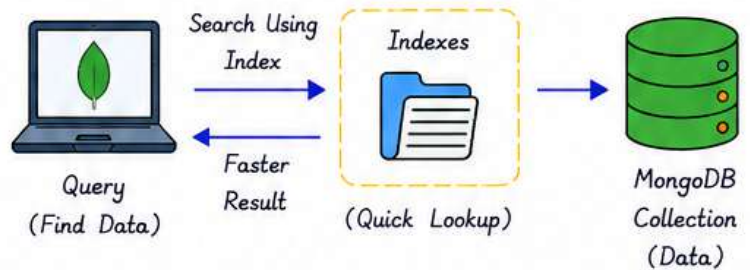


24.9 MongoDB Indexing

24.9.1 What is MongoDB Indexing?

- * Indexing in MongoDB improves the **speed** of data retrieval operations on **large collections**.
- * An index is a **special data structure** that stores a small portion of the collection's data in an easy-to-traverse form.
- * MongoDB uses indexes to **quickly locate documents** without scanning the entire collection.
- * Indexes improve **read performance** but slow down write operations slightly.

How MongoDB Indexing Works



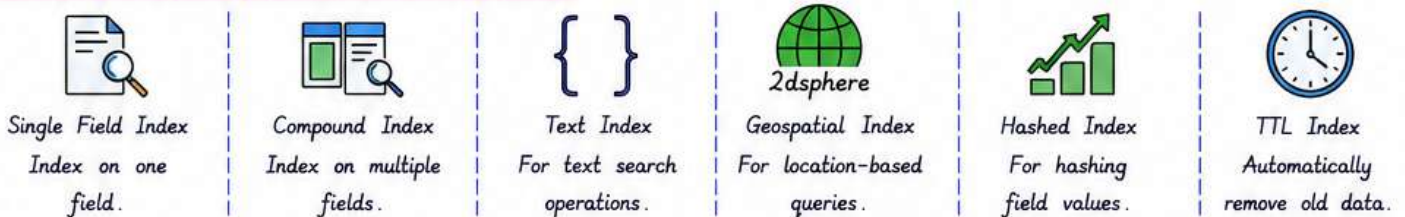
24.9.2 Why Use Indexes in MongoDB?

- * Speeds up **query performance**.
- * Reduces **response time**.
- * Improves **scalability** for large datasets.
- * Supports efficient **sorting** and **filtering**.
- * Helps in better **resource utilization**.
- * Essential for **high-performance** applications.

Impact of Indexing

Operation	Without Index	With Index
Query Performance	Slow (Collection Scan)	Fast (Index Scan)
Response Time	High	Low
CPU Usage	High	Low
Scalability	Poor	Good
Write Operations	Faster	Slightly Slower
Storage Usage	Low	Higher (Index Storage)

24.9.3 Types of Indexes in MongoDB



24.9.4 Create Index in MongoDB

```
// Create Index on Single Field
db.users.createIndex({ name: 1 });

// Create Compound Index
db.orders.createIndex({ customerId: 1, orderDate: -1 });

// Create Text Index
db.articles.createIndex({ title: "text", content: "text" });

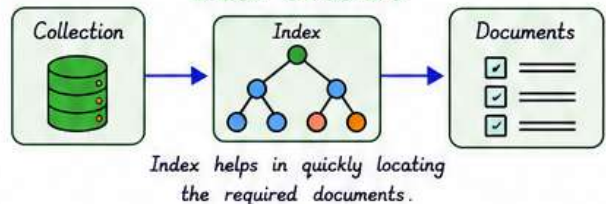
// Create TTL Index (Expired after 3600 seconds)
db.sessions.createIndex({ createdAt: 1 }, { expireAfterSeconds: 3600 });
```



Note:

Indexes improve **read operations** but consume extra storage and slow down **write operations** (insert, update, delete).

Index Structure



24.9.5 Check and Remove Indexes

```
// List All Indexes
db.collection.getIndexes();

// Drop Single Index
db.collection.dropIndex("indexName");

// Drop All Indexes (Except _id Index)
db.collection.dropIndexes();
```

24.9.6 Best Practices for MongoDB Indexing

- * Create indexes on fields that are **frequently queried**.
- * Use **compound indexes** for multiple field queries.
- * Avoid **over-indexing** as it can impact write performance.
- * Monitor index usage and **remove unused indexes**.
- * Use **TTL indexes** for automatic data expiration.
- * Regularly review and **optimize indexes**.



Key Points:

- ✓ Indexes improve **query performance** and reduce **response time**.
- ✓ Use **appropriate indexes** based on query patterns.
- ✓ Indexes improve **read operations** but slow down **write operations**.
- ✓ Regularly **analyze** and **optimize indexes** for better performance.



Key Takeaway:



MongoDB indexing is crucial for building **fast** and **scalable** applications. Use indexes wisely to achieve the **best performance**.

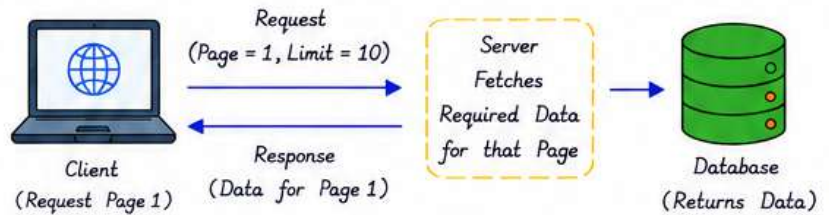


24.10 Pagination

24.10.1 What is Pagination?

- * Pagination is the process of **dividing** large data sets into **smaller chunks** (pages).
- * Only one page of data is loaded and displayed at a time.
- * It helps in **improving performance**, **reducing load time** and enhancing **user experience**.
- * It is widely used in tables, lists, search results, blogs, dashboards, etc.

How Pagination Works



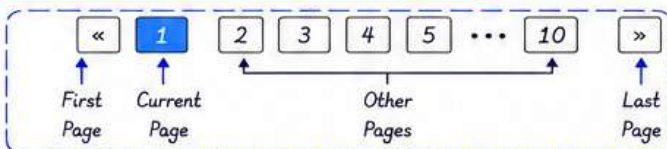
24.10.2 Why Use Pagination?

- * Reduces initial **load time**.
- * Lowers **server load** and database query cost.
- * Saves **bandwidth** and network resources.
- * Improves **user experience** and interface usability.
- * Easy to navigate large datasets.
- * Makes the application **scalable** and efficient.

Impact of Pagination

	Without Pagination	With Pagination
Data Loaded	All data at once	Small chunks (per page)
Load Time	High	Low
Server Load	High	Low
Bandwidth Usage	High	Low
User Experience	Slow	Fast
Scalability	Poor	Good

24.10.3 Example of Pagination UI



Note:

Pagination is better than loading all data at once. It improves **performance** and **user satisfaction**, especially for large datasets.

24.10.4 Pagination in Backend (Example in Node.js + MongoDB)

```
app.get('/api/users', async (req, res) => {
  let page = parseInt(req.query.page) || 1; // Default page = 1
  let limit = parseInt(req.query.limit) || 10; // Default limit = 10
  let skip = (page - 1) * limit; // Calculate records to skip

  let totalRecords = await User.countDocuments();
  let totalPages = Math.ceil(totalRecords / limit);

  let users = await User.find()
    .skip(skip) // Skip previous pages
    .limit(limit); // Limit to records per page

  res.json({
    data: users,
    currentPage: page,
    totalPages: totalPages,
    totalRecords: totalRecords
  });
});
```

- * **page**: Current page number.
- * **limit**: Number of records per page.
- * **skip**: Calculates the number of records to skip.
- * **countDocuments()**: Total number of records.
- * **find().skip().limit()**: Fetches only required data.

Example

page	limit	skip	Records Fetched
1	10	0	Records 1 - 10
2	10	10	Records 11 - 20
3	10	20	Records 21 - 30
...	...	...	...

24.10.5 Pagination in Frontend (Example in React)

```
function Pagination({ currentPage, totalPages, onPageChange }) {
  return (
    <div className="pagination">
      <button disabled={currentPage === 1} onClick={() => onPageChange(currentPage - 1)}>
        Prev</button>
      {[...Array(totalPages)].map((_, index) => (
        <button key={index + 1}
          className={currentPage === index + 1 ? "active" : ""}
          onClick={() => onPageChange(index + 1)}>
          {index + 1} </button>
      ))}
      <button disabled={currentPage === totalPages} onClick={() => onPageChange(currentPage + 1)}>
        Next</button>
    </div>
  );
}
```

Frontend Fetches Data for Selected Page



Tip:

Always **validate** page and limit values in the backend to avoid performance issues.

Key Points:

- ✓ Pagination divides large data into smaller pages.
- ✓ It reduces load time, server load and bandwidth usage.
- ✓ Use limit and skip (or offset) to fetch data.
- ✓ Always show total pages and current page to the user.
- ✓ Good pagination improves scalability and user experience.



Key Takeaway:



Pagination is a simple but powerful technique to handle large datasets efficiently. It ensures **fast performance**, **better scalability** and a **smooth user experience**.

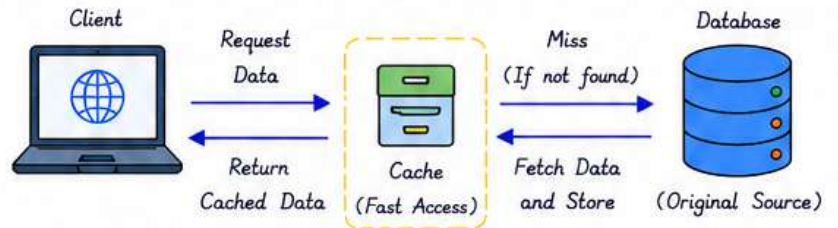


24.11 Caching Concepts

24.11.1 What is Caching?

- * Caching is the process of **storing frequently accessed data** in a **temporary storage** location.
- * It helps in **retrieving data faster** without hitting the original data source repeatedly.
- * Caching reduces **response time**, **server load** and **database calls**.
- * It improves application performance and **user experience**.

How Caching Works



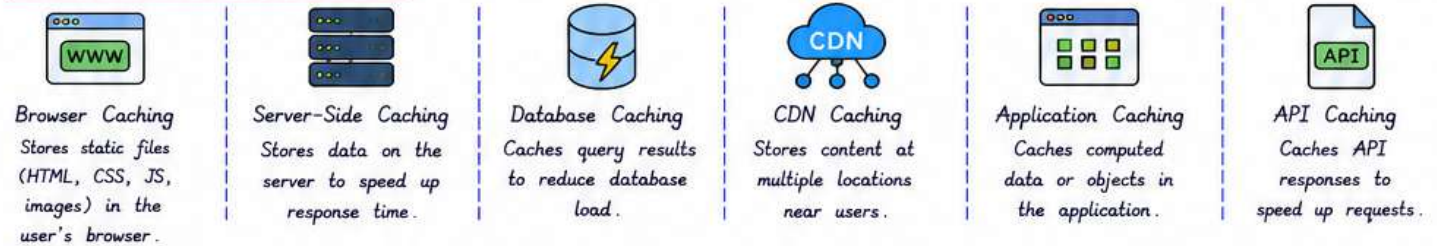
24.11.2 Why Use Caching?

- * Reduces data **retrieval time**.
- * Decreases **server** and **database load**.
- * Saves **bandwidth** and **network resources**.
- * Improves **scalability** and application **performance**.
- * Enhances the overall **user experience**.
- * Useful for **frequently accessed** or **expensive data**.

Impact of Caching

Parameter	Without Caching	With Caching
Response Time	High	Low
Server Load	High	Low
Database Calls	Many	Few
Bandwidth Usage	High	Low
Scalability	Limited	Better
User Experience	Slow	Fast

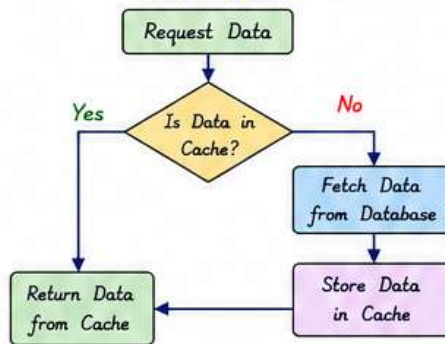
24.11.3 Types of Caching



24.11.4 Common Caching Strategies

- * **Cache-Aside (Lazy Loading)**: Fetch from DB only if data is not in cache.
- * **Write-Through**: Write data to cache and DB simultaneously.
- * **Write-Back (Write-Behind)**: Write data to cache first, then update DB later.
- * **Refresh-Ahead**: Refresh cache before data expires.
- * **Time-To-Live (TTL)**: Set expiry time for cached data.
- * **Eviction Policies**: Remove old data when cache is full (LRU, LFU, FIFO).

Cache Flow



Eviction Policies

- * **LRU (Least Recently Used)**: Remove least recently used data.
- * **LFU (Least Frequently Used)**: Remove least frequently used data.
- * **FIFO (First In First Out)**: Remove the oldest data.
- * **Random**: Remove random data.

24.11.5 Example: Caching in Node.js (Using Redis)

Without Caching (Always hit DB)

```
app.get('/users/:id', async (req, res) => {
  const user = await db.collection('users').findOne({ _id: req.params.id });
  res.json(user);
});
```

With Caching (Using Redis)

```
app.get('/users/:id', async (req, res) => {
  const cacheKey = `user:${req.params.id}`;
  const cachedUser = await redis.get(cacheKey);
  if (cachedUser) {
    return res.json(JSON.parse(cachedUser)); // Return from cache
  }
  const user = await db.collection('users').findOne({ _id: req.params.id });
  await redis.set(cacheKey, JSON.stringify(user), 'EX', 3600); // TTL = 1 hour
  res.json(user);
});
```

Key Points:

- ✓ Caching stores **frequently accessed data** for quick retrieval.
- ✓ It reduces **response time**, **server load** and **database calls**.
- ✓ Choose the right caching strategy and eviction policy.
- ✓ Always set appropriate **TTL** to keep data fresh.
- ✓ Caching is essential for building **fast**, **scalable** and **high-performance** applications.

Key Takeaway:



Caching is a powerful technique to improve **performance** and **scalability**. Use it wisely to deliver **faster** and **better user experiences**.



24.12 Backend Performance

24.12.1 What is Backend Performance?

- Backend performance refers to how **efficiently** the **server**, **database** and **application** handle requests and deliver responses.
- Good backend performance ensures **faster response time**, **high availability** and better **user experience**.
- It depends on **code efficiency**, **database queries**, **server resources** and **system architecture**.

24.12.2 Why is Backend Performance Important?

- Reduces **response time** and improves **user satisfaction**.
- Decreases **server load** and **infrastructure cost**.
- Improves system **scalability** and **reliability**.
- Prevents **timeouts** and **crashes** under high traffic.
- Helps in handling **more users** and **requests** efficiently.
- Enhances overall **application performance**.

24.12.3 Key Factors Affecting Backend Performance

Factor	Impact
 Efficient Code Optimized algorithms, clean and efficient code.	
 Database Queries Optimized queries, indexes and proper schema design.	
 Server Resources CPU, Memory, Disk and Network utilization.	
 Caching Use caching to reduce database and server load.	
 Load Balancing Distribute requests across multiple servers.	
 Asynchronous Processing Use queues and background jobs.	
 Monitoring & Logging Monitor performance and fix bottlenecks early.	

24.12.4 Best Practices for Better Performance

- Write **efficient** and **optimized** code.
- Optimize **database queries** and use **indexes**.
- Use **caching** (Redis, Memcached) for frequent data.
- Implement **pagination** to limit data transfer.
- Use **connection pooling** for database connections.
- Compress responses** to reduce data size.
- Use **asynchronous processing** for heavy tasks.
- Monitor performance and **fix bottlenecks**.
- Scale horizontally** using load balancers.
- Keep server and dependencies **updated**.



Note:

Always identify the **bottleneck** before optimizing. Optimize code, database, server and architecture for the **best performance**.

24.12.5 Example: Slow vs Optimized Query

Slow Query	Optimized Query
<pre>SELECT * FROM orders WHERE customer_id = 101 ORDER BY order_date DESC; -- Fetches all columns and sorts -- large data (slow)</pre>	<pre>SELECT order_id, total_amount, order_date FROM orders WHERE customer_id = 101 ORDER BY order_date DESC LIMIT 20; -- Fetches only required columns -- and limits results (faster)</pre>

24.12.6 Example: Caching in Backend (Node.js + Redis)

```
app.get('/products/:id', async (req, res) => {
  const id = req.params.id;
  const cacheKey = `product:${id}`;

  const cached = await redis.get(cacheKey);
  if (cached) { return res.json(JSON.parse(cached)); } // Return cached data

  const product = await Product.findById(id); // Fetch from DB
  await redis.set(cacheKey, JSON.stringify(product), 'EX', 3600); // Cache for 1 hour
  res.json(product);
});
```

Key Points:

- Backend performance depends on **code**, **database**, **server** and **architecture**.
- Optimize queries, use **caching**, **pagination** and **compression**.
- Monitor performance regularly and **fix bottlenecks**.
- Good backend performance ensures **scalability**, **reliability** and **better user experience**.

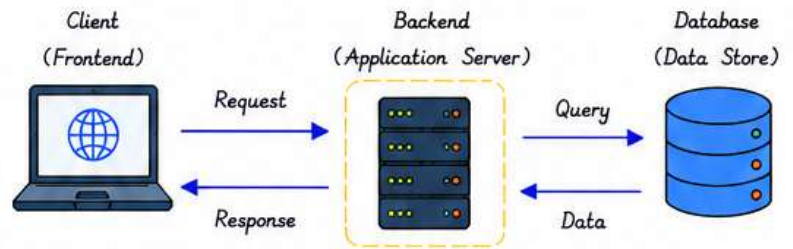
Key Takeaway:



A fast and optimized backend is the **backbone** of any successful application. Optimize smartly, monitor continuously and deliver the **best performance**!



How Backend Works



Impact of Backend Performance

Parameter	Poor Performance	Good Performance
Response Time	High	Low
Server Load	High	Low
Throughput	Low	High
Resource Usage	Inefficient	Efficient
Scalability	Poor	Good
User Experience	Slow & Frustrating	Fast & Smooth
Error Rate	High	Low

24.12.7 Tools for Monitoring Performance

- New Relic - Application performance monitoring.
- Datadog - Infrastructure and application monitoring.
- Prometheus + Grafana - Metrics and visualization.
- ELK Stack - Logs management and analysis.
- Google Cloud Monitoring - Cloud performance monitoring.
- Postman - API testing and performance check.



24.13 Network Optimization

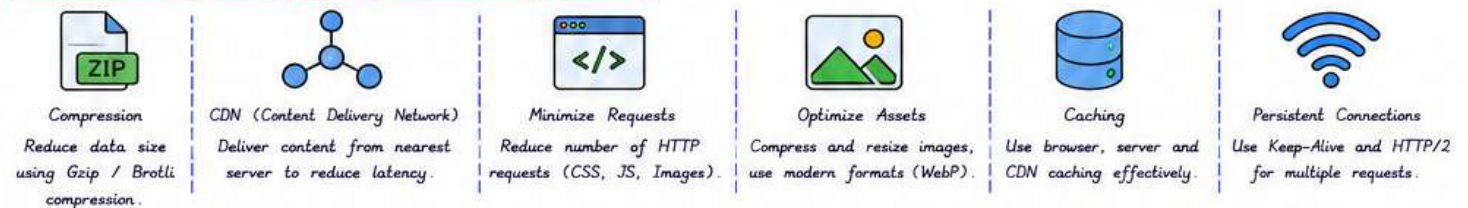
24.13.1 What is Network Optimization?

- * Network optimization is the process of improving network performance by **reducing latency**, **minimizing data transfer** and improving **bandwidth usage**.
- * It ensures **faster communication** between client, server and other network resources.
- * It helps in delivering web pages, APIs and media content **quickly and efficiently**.
- * Good network optimization **improves user experience** and reduces load on servers.

24.13.2 Why is Network Optimization Important?

- * Reduces **latency** and improves **response time**.
- * Saves **bandwidth** and reduces **data transfer**.
- * Speeds up **page load time** and **API responses**.
- * Improves **reliability** and reduces **packet loss**.
- * Enhances **user experience** and **satisfaction**.
- * Lowers **infrastructure** and **operational costs**.
- * Essential for **global users** and **mobile networks**.

24.13.3 Key Techniques for Network Optimization



24.13.4 Best Practices for Network Optimization

- * Enable **Gzip** or **Brotli** compression on the server.
- * Use **CDN** to serve static content globally.
- * Reduce the number of **redirects**.
- * Use **HTTP/2** or **HTTP/3** for faster connections.
- * Optimize **images** and use **lazy loading**.
- * Use caching headers (**Cache-Control**, **Expires**).
- * Minimize **CSS**, **JavaScript** and **HTML** files.
- * Remove **unused scripts**, **fonts** and **plugins**.
- * Monitor **network performance** regularly.

24.13.5 Example: Gzip Compression in Express.js

```
const express = require('express');
const compression = require('compression'); // npm install compression
const app = express();

// Enable Gzip Compression
app.use(compression());

app.get('/', (req, res) => {
  res.send('Hello! This response is compressed using Gzip.');
```

```
});

app.listen(3000, () => {
  console.log('Server running on http://localhost:3000');
```

```
});
```

24.13.7 Tools for Network Optimization



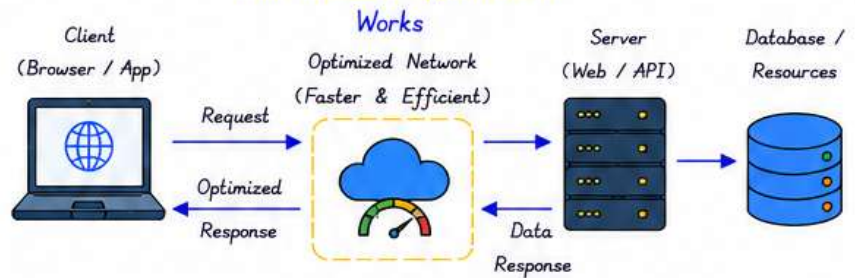
Key Points:

- ✓ Network optimization reduces **latency**, **bandwidth** usage and improves **reliability**.
- ✓ Use **compression**, **caching**, **CDN** and **optimized assets** for better performance.
- ✓ Minimize requests, use **HTTP/2** or **HTTP/3** and keep connections persistent.
- ✓ Regular **monitoring** helps in identifying and fixing network bottlenecks.

Key Takeaway:

 A well-optimized network ensures **faster delivery**, **lower latency** and **better scalability**.
Optimize the network, optimize the performance!

How Network Optimization Works



Impact of Network Optimization		
Parameter	Without Optimization	With Optimization
Latency	High	Low
Page Load Time	Slow	Fast
Bandwidth Usage	High	Optimized
Data Transfer Size	Large	Reduced
Server Load	High	Low
User Experience	Poor	Smooth & Fast
Reliability	Low	High



Note:

Small improvements in network can lead to huge **performance gains** and **better user experience**.

Network Optimization Flow



24.13.6 Example: Before vs After Optimization

Before Optimization	After Optimization
✗ Large image files	✓ Compressed & optimized images
✗ Multiple HTTP requests	✓ Minimized HTTP requests
✗ No compression	✓ Gzip / Brotli compression
✗ No caching	✓ Browser & CDN caching
✗ High latency	✓ Low latency
✗ Slow page load time	✓ Fast page load time



24.14 Production Optimization

24.14.1 What is Production Optimization?

- Production optimization is the process of **fine-tuning** an application and its infrastructure for maximum **performance, scalability, reliability** and **cost-efficiency** in a live environment.
- It includes **code optimization, server configuration, caching, monitoring, security** and **resource management**.
- The goal is to deliver **fast, stable** and **scalable** applications to real users.

How Production Optimization



24.14.2 Why is Production Optimization Important?

- Improves application **performance** and **response time**.
- Handles **more** users and traffic **efficiently**.
- Reduces **downtime** and prevents failures.
- Optimizes **resource usage** and reduces **operational cost**.
- Enhances **security, reliability** and **user experience**.
- Helps in **scaling** the application for future growth.
- Essential for maintaining a **competitive** and professional application.

Impact of Production Optimization		
Parameter	Without Optimization	With Optimization
Response Time	High	Low
Throughput	Low	High
Server Load	High	Low
Resource Utilization	Inefficient	Efficient
Downtime	Frequent	Minimal
Scalability	Poor	Excellent
Cost	High	Optimized
User Experience	Slow & Frustrating	Fast & Smooth
Reliability	Low	High

24.14.3 Key Areas of Production Optimization



24.14.4 Best Practices for Production Optimization

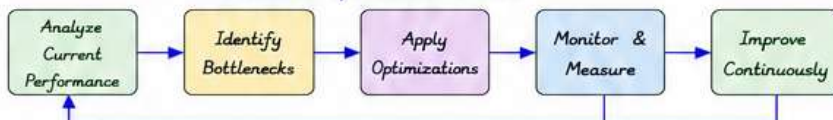
- Use **environment variables** for configuration.
- Minimize and **compress** assets (CSS, JS, Images).
- Enable **caching** at multiple levels.
- Optimize database queries and use **indexing**.
- Use **connection pooling** for database connections.
- Implement **pagination** for large datasets.
- Enable **Gzip / Brotli** compression.
- Use **CDN** for static assets.
- Configure **load balancers** and **autoscaling**.
- Monitor** application performance continuously.
- Regularly **update dependencies** and remove unused packages.
- Backup data** and test **disaster recovery plans**.



Note:

Production optimization is a continuous process. Always **monitor, analyze** and **improve**.

Production Optimization Process



24.14.5 Example: Enable Gzip Compression in Node.js

```
const express = require('express');
const compression = require('compression'); // npm install compression
const app = express();

// Use Gzip Compression
app.use(compression());

app.get('/', (req, res) => {
  res.send('Production Optimized Application!');
});

app.listen(3000, () => {
  console.log('Server running on http://localhost:3000');
});
```

24.14.6 Example: Database Connection Pooling (MongoDB)

```
const { MongoClient } = require('mongodb');
const uri = 'mongodb+srv://username:password@cluster.mongodb.net/dbname';

const client = new MongoClient(uri, {
  maxPoolSize: 20, // Maximum number of connections
  minPoolSize: 5, // Minimum number of connections
  serverSelectionTimeoutMS: 10000,
  socketTimeoutMS: 45000
});

async function connectDB() {
  await client.connect();
  console.log('MongoDB connected with connection pooling');
  return client.db('dbname');
}
```

24.14.7 Tools for Production Optimization



Key Points:

- Production optimization improves **performance, scalability, reliability** and reduces costs.
- Optimize **code, database, server, network** and **security** for best results.
- Use **caching, compression, CDN, indexing** and **connection pooling** for efficiency.
- Monitor** continuously and **fix bottlenecks**.
- Optimization is a **continuous process**, not a one-time task.

Key Takeaway:



A well-optimized production system delivers **fast performance, high availability, better scalability** and an **excellent user experience**.



24.15 Lighthouse Performance

24.15.1 What is Lighthouse?

- * Lighthouse is an **open-source, automated** tool built by Google to improve the quality of web pages.
- * It audits **performance, accessibility, best practices, SEO** and **PWA**.
- * It helps developers **identify issues** and **optimize** the website for a **better user experience**.

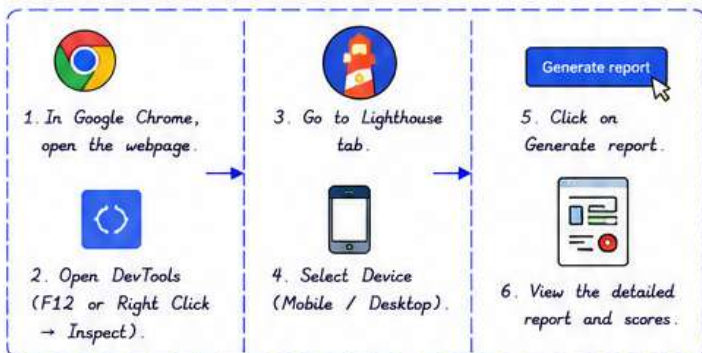
How Lighthouse Works



24.15.2 Why Use Lighthouse?

- * It provides a **detailed performance report**.
- * Helps find **performance bottlenecks**.
- * Improves **user experience** and **accessibility**.
- * Improves **SEO** and increases visibility.
- * Guides developers to follow **best practices**.
- * Useful in **development** and **production** environments.

24.15.3 How to Run Lighthouse?



24.15.4 Lighthouse Performance Score



Note:

The goal is to achieve a **score close to 100** in all categories.

24.15.5 Key Metrics in Performance Category

Metric	Meaning
First Contentful Paint (FCP)	Time taken to render the first content.
Largest Contentful Paint (LCP)	Time taken to render the largest content visible in the viewport.
Total Blocking Time (TBT)	Total time that the page is blocked from responding to user input.
Cumulative Layout Shift (CLS)	Measures visual stability while loading.
Speed Index (SI)	How quickly the contents of a page are visibly populated.

24.15.6 Example: Lighthouse Report (Sample)



24.15.7 Tips to Improve Lighthouse Score

- * **Optimize images** (use WebP, compress images).
- * **Minify** CSS, JS and HTML files.
- * Remove **unused** CSS and JavaScript.
- * Enable browser **caching**.
- * Use **CDN** to deliver static assets.
- * Reduce **render-blocking** resources.
- * Use **lazy loading** for images and iframes.
- * Ensure proper **contrast** for accessibility.
- * Follow **best practices** and fix warnings.

Key Points:

- ✓ Lighthouse audits **performance, accessibility, best practices, SEO** and **PWA**.
- ✓ It provides **actionable recommendations** to improve your website.
- ✓ Regularly running Lighthouse helps maintain **fast, accessible** and **SEO-friendly** websites.

Key Takeaway:



Lighthouse is a powerful tool to **measure** and **improve** website performance and quality. Use it regularly and follow the recommendations for the **best user experience**.



24.15.8 Lighthouse in Different Environments

