# Load Balancing و Load Balancer

**Load balancing** یا توزیع بار به معنای توزیع منابع و وظایف بین چندین سرور یا سیستم کامپیوتری است. هدف اصلی از استفاده از load balancing این است که بار کاری را بین منابع موجود توزیع کند تا عملکرد بهتر، کارایی بالاتر، و عدم وجود نقاط خرابی تک نقطه‌ای را فراهم کند.

**Load balancer** یا توزیع کننده بار، یک سیستم یا دستگاه است که وظیفه توزیع بار کاری بین منابع موجود را بر عهده دارد. وظایف اصلی یک load balancer شامل دریافت درخواست‌های کاربر، تحلیل بار کاری، تصمیم گیری برای توزیع بار، و ارسال درخواست به سرورهای مختلف است. Load balancer معمولاً در لایه بالاتر شبکه قرار می‌گیرد و به عنوان یک نقطه ورودی برای درخواست‌های کاربران عمل می‌کند.

## Workload یا Load

**Workload** یا بار کاری به میزان کار و فعالیتی اشاره دارد که بر روی یک سیستم یا منبع مشخص وجود دارد. بار کاری ممکن است شامل درخواست‌ها، تراکنش‌ها، پردازش داده‌ها و هر نوع فعالیت دیگری باشد که نیاز به منابع سیستمی دارد.

## Balance در شبکه

در شبکه، **balance** به توزیع منابع و بار کاری بین اجزای مختلف شبکه اشاره دارد. هدف اصلی توزیع بار کاری در شبکه، ایجاد تعادل و بهره‌وری بین سرورها، دستگاه‌ها و منابع موجود است. با توزیع منابع به طور متناسب، می‌توان از میزان استفاده بهینه از منابع برخوردار شد و عملکرد بهتری در شبکه به دست آورد.

## Redundancy در شبکه

**Redundancy** یا اضافی بودن به معنای تکرار و اضافی سازی اجزا و منابع در شبکه است. استفاده از اجزا اضافی و تکراری می‌تواند در کاهش خرابی‌ها و افزایش قابلیت اطمینان شبکه مؤثر باشد. با توزیع بار بین منابع تکراری، اگر یکی از آن‌ها خراب شود، بار کاری به منابع دیگر توزیع می‌شود و سرویس قابل اطمینان باقی می‌ماند.

## Distribute در شبکه

در شبکه، **distribute** به توزیع بار کاری بین منابع یا سرویس‌های مختلف اشاره دارد. توزیع بار کاری بین منابع می‌تواند به منظور بهبود کارایی و عملکرد شبکه، افزایش قابلیت اطمینان، و جلوگیری از بارزشی در یک سرور یا منبع خاص صورت بگیرد.

## Load Balancing در چه چیزهایی انجام می‌شود

Load balancing در موارد زیر انجام می‌شود:

1. **توزیع درخواست‌های وب**: وقتی یک سایت وب محبوب می‌شود و تعداد کاربران همزمان زیاد می‌شود، load balancing استفاده می‌شود تا درخواست‌های وب را بین سرورهای مختلف توزیع کند و بار کاری را به طور متناسب بین سرورها تقسیم کند.
2. **توزیع بار در سرورهای برنامه**: در محیط‌های برنامه‌نویسی و سرور، load balancing استفاده می‌شود تا بار کاری را بین سرورهای برنامه توزیع کند و منابع سیستمی را بهینه استفاده کند.
3. **توزیع بار در سرورهای پایگاه داده**: وقتی کاربران به سیستم پایگاه داده دسترسی دارند و تعداد درخواست‌های همزمان بالاست، load balancing می‌تواند به توزیع بار کاری در سرورهای پایگاه داده کمک کند و عملکرد و قابلیت اطمینان را بهبود بخشد.

## Load Balancing در لایه‌های OSI

Load balancing در معماری شبکه OSI در لایه‌های زیر کاربرد دارد:

1. **لایه لینک داده (Data Link Layer)**: Load balancing ممکن است در این لایه برای توزیع بار در شبکه‌های محلی (LAN) استفاده شود. به طور معمول، سوئیچ‌های لایه دوم قادر به توزیع درخواست‌های دریافتی بین دستگاه‌های متصل به آنها هستند.
2. **لایه شبکه (Network Layer)**: Load balancing ممکن است در این لایه برای توزیع بار در شبکه‌های گسترده (WAN) یا برای توزیع بار درون شبکه‌های خصوصی مجازی (VPN) استفاده شود. در این لایه، روترها می‌توانند بار کاری را بین مسیرهای مختلف تقسیم کنند.
3. **لایه انتقال (Transport Layer)**: Load balancing معمولاً در این لایه برای توزیع بار کاری بین سرورها در ارتباطات TCP/IP استفاده می‌شود. این لایه وظیفه توزیع و مدیریت ارتباطات بین نرم‌افزارهای مبدا و مقصد را بر عهده دارد.

## Single Point of Failure و Load Balancing

**Single Point of Failure (SPoF)** یا نقطه خرابی تک نقطه‌ای به یک عنصر در شبکه گفته می‌شود که در صورت خرابی آن، کل سیستم یا شبکه قادر به کارکرد صحیح نمی‌باشد. استفاده از load balancing می‌تواند از بروز SPoF جلوگیری کند؛ زیرا با توزیع بار کاری بین منابع مختلف، اگر یکی از آن‌ها خراب شود، منابع دیگر می‌توانند به طور خودکار کار را ادامه دهند و سرویس به کاربران فراهم باشد.

## Cluster در سرور

**Cluster** یا خوشه به گروهی از سرورها یا سیستم‌های کامپیوتری گفته می‌شود که به صورت مشترک عملکرد کرده و بار کاری را توزیع می‌کنند. در یک خوشه، هر عضو می‌تواند وظایف و بار کاری را بر عهده داشته باشد و همکاری کند تا بهبود عملکرد، قابلیت اطمینان و قابلیت انعطاف‌پذیری را فراهم کند.

## الگوریتم انجام Load Balancing

برای انجام load balancing، می‌توان از الگوریتم‌های مختلف استفاده کرد. برخی از الگوریتم‌های معروف شامل موارد زیر است:

1. **Round Robin**: در این الگوریتم، درخواست‌ها به ترتیب بین سرورهای مختلف توزیع می‌شوند. هر درخواست به یک سرور ارسال می‌شود و درخواست بعدی به سرور بعدی ارسال می‌شود، و این فرایند به صورت مکرر تکرار می‌شود.
2. **Least Connection**: در این الگوریتم، درخواست‌ها به سروری توزیع می‌شوند که در حال حاضر کمترین تعداد اتصال فعال را دارد. این الگوریتم بر اساس تعداد اتصالات فعال بین سرورها تصمیم می‌گیرد.
3. **IP Hash**: در این الگوریتم، بر اساس آدرس IP کاربر، تصمیم برای ارسال درخواست به یک سرور خاص می‌گیرد. به این ترتیب، درخواست‌های یک کاربر خاص به یک سرور خاص توزیع می‌شوند.
4. **Weighted Round Robin**: در این الگوریتم، وزن‌های مختلفی به سرورها اختصاص داده می‌شود و درخواست‌ها بر اساس این وزن‌ها توزیع می‌شوند. سرورهایی با وزن بالاتر بار کاری بیشتری دریافت می‌کنند.

همچنین، الگوریتم‌های دیگری مانند Least Response Time، Hash-based و Session-based نیز وجود دارند که بر اساس نیازها و محیط استفاده متناسب انتخاب می‌شوند.