**Alta Disponibilidade (High Availability - HA)**

**📌 Conceito**

Alta disponibilidade (HA) é a **capacidade de um sistema operar continuamente sem falhas** por um período de tempo predefinido. O objetivo é garantir que o sistema atenda a um **nível mínimo de desempenho operacional acordado**, mesmo diante de falhas ou imprevistos.

**Disponibilidade de "Cinco Noves" (99,999%)**: padrão de alta disponibilidade que significa que o sistema estará operacional por 99,999% do tempo.

**🏥 Onde a Alta Disponibilidade é Essencial**

Sistemas HA são críticos em ambientes onde a **falha pode afetar vidas humanas ou causar grandes perdas**. Exemplos:

* Controle militar
* Veículos autônomos
* Indústria e manufatura
* Sistemas hospitalares (EHR - Electronic Health Records)
* Infraestruturas críticas (energia, comunicação, transporte)

**🧱 Princípios de Projeto de Sistemas HA**

1. **Eliminação de Pontos Únicos de Falha (SPOF)**  
   Um componente que, se falhar, compromete todo o sistema.  
   Ex.: Um único servidor que roda um serviço crítico.
2. **Failover Confiável (Crossover)**  
   Substituição automática por um componente redundante sem perda de dados ou degradação de desempenho.
3. **Detectabilidade de Falhas**  
   As falhas devem ser **detectáveis automaticamente** e o sistema deve contar com mecanismos de **auto-recuperação**.
4. **Balanceamento de Carga (Load Balancing)**  
   Distribuição automática de cargas de trabalho entre múltiplos recursos. Exemplo: distribuição de requisições em uma arquitetura de nuvem híbrida.
5. **Clusterização de Servidores**  
   Agrupamento de servidores com replicação e failover automático entre eles, minimizando impactos.

**📏 Métricas de Disponibilidade**

A disponibilidade é normalmente expressa como uma **percentagem do tempo total de operação**:

plaintext

CopiarEditar

Disponibilidade = (Minutos do mês - Minutos de inatividade) / Minutos do mês

**Principais métricas:**

* **MTBF (Mean Time Between Failures)**  
  Tempo médio entre falhas do sistema.
* **MDT (Mean Downtime)**  
  Tempo médio em que o sistema fica indisponível.
* **MTTR (Mean Time to Repair)**  
  Tempo médio para recuperação ou reparo após falha.

**Comparação de SLA (Service Level Agreement):**

| **SLA (%)** | **Inatividade anual** | **Inatividade mensal** | **Inatividade diária** |
| --- | --- | --- | --- |
| 99% | ~3 dias e 15h | ~7h 18min | ~14min |
| 99,9% | ~8h 45min | ~43min | ~1min |
| 99,99% | ~52min | ~4min 23s | ~8,6s |
| 99,999% | ~5min 15s | ~26,3s | ~0,86s |

**🛠️ Etapas para Implementar Alta Disponibilidade**

1. **Projeto com foco em HA**  
   Minimize custos e complexidade, removendo SPOFs e usando redundância.
2. **Definição de Métricas de Sucesso**  
   Defina metas de disponibilidade e métricas para monitoramento (SLAs).
3. **Implantação de Hardware Resiliente**  
   Utilize equipamentos com suporte a *hot swap* e *hot plug* para manutenção sem desligamento.
4. **Testes de Failover**  
   Simule falhas regularmente para garantir que o sistema alternativo funcione corretamente.
5. **Monitoramento Contínuo**  
   Use ferramentas e métricas para rastrear o desempenho e detectar anomalias.
6. **Avaliação e Aprimoramento Contínuos**  
   Analise logs e indicadores para identificar oportunidades de melhoria no sistema.

**🔁 Recuperação de Desastres (Disaster Recovery - DR)**

* Foco em falhas catastróficas, como desastres naturais ou falhas de infraestrutura.
* Complementa estratégias HA, que tratam de falhas menores e localizadas.
* Deve haver backups e failover para:
  + Servidores
  + Armazenamento
  + Nós de rede
  + Data centers inteiros

**✅ Boas Práticas para Alta Disponibilidade**

* Eliminar **pontos únicos de falha**
* Manter **backups atualizados** de dados e sistemas
* Usar **balanceamento de carga** (ex.: HAProxy)
* **Monitorar continuamente** servidores e serviços críticos
* **Distribuir recursos** geograficamente (para tolerância a desastres)
* Utilizar **RAID/SAN** para tolerância a falhas de disco
* Implementar **detecção automática de falhas**
* **Testar e validar** os mecanismos de HA regularmente

**💬 Conclusão**

Alta disponibilidade é vital para sistemas que suportam **funções críticas** à vida, economia e serviços essenciais. Com uma arquitetura cuidadosamente projetada, baseada em redundância, monitoramento e recuperação rápida, é possível atingir altos padrões de confiabilidade e minimizar o impacto de falhas.