

## O que é Big Data na Saúde?

- **Definição:** Extração de informações relevantes a partir de grandes volumes de dados.
- **Importância:** Transformação de grandes volumes de dados em informações estratégicas.
- **Aplicações:** Estudos, análises e tomadas de decisões em clínicas e hospitais.

## Papel do Big Data na Medicina:

- **Análise Rápida e Eficiente:** Processamento de grandes quantidades de dados rapidamente.
- **Projeções Futuras:** Crescimento exponencial de dados no setor de saúde.
- **Ferramenta Necessária:** Big Data é essencial para lidar com o grande volume de dados previstos.

## Funcionamento do Big Data na Saúde:

- **Software e Programas Específicos:** Necessidade de ferramentas específicas para análise de dados.
- **Processo:**
  1. Definir estratégia.
  2. Gerenciar dados.
  3. Retornar dados específicos.
- **Etapas:** Integração, gerenciamento e análise.

## **Motivações para o Tema:**

- **Relevância Social:** Impacto positivo na qualidade de vida dos pacientes.
- **Benefícios Tangíveis:** Diagnósticos precisos, tratamentos eficazes, prevenção de doenças.

## **Problemas e Benefícios do Diagnóstico Preventivo:**

### **1. Identificação de Padrões de Saúde:**

- **Objetivo:** Detectar padrões e tendências em dados de saúde.
- **Ganhos:** Detecção precoce de doenças, melhor prognóstico.

### **2. Personalização de Tratamentos:**

- **Objetivo:** Desenvolver modelos de dados para tratamentos personalizados.
- **Ganhos:** Maior eficácia dos tratamentos, satisfação dos pacientes, redução de custos.

### **3. Análise de Efetividade de Tratamentos:**

- **Objetivo:** Avaliar a efetividade de tratamentos com base em dados.
- **Ganhos:** Melhoria na seleção de tratamentos, aumento das taxas de sucesso.

## **Problemas a Serem Resolvidos:**

- **Detecção Precoce de Doenças:** Identificação de sinais precoces nos dados.
- **Monitoramento Contínuo:** Monitoramento em tempo real de anomalias.
- **Personalização de Tratamentos:** Desenvolvimento de modelos preditivos.

- **Efetividade de Tratamentos:** Utilização de dados históricos para melhor decisão clínica.

### **Identificação dos Dados Relevantes:**

- **Dados Estruturados:** EMRs, dados demográficos, sensores e dispositivos wearables, dados de utilização dos serviços de saúde.
- **Dados Não Estruturados:** Anotações clínicas, imagens médicas, dados de redes sociais, vídeos de exames.

### **Fontes de Dados Relevantes e Periodicidade:**

- **Hospitais e Clínicas:** EMRs, anotações clínicas.
- **Laboratórios de Diagnóstico:** Resultados de exames, imagens médicas.
- **Dispositivos Wearables:** Monitoramento contínuo.
- **Redes Sociais e Fóruns de Saúde:** Postagens e comentários.
- **Bases de Dados Governamentais:** Dados demográficos, registros de saúde pública.
- **Seguradoras:** Dados de utilização de serviços, históricos de tratamentos.

### **Infraestrutura e Justificativa dos Recursos:**

- **Processador:** Intel Xeon Platinum 8280.
- **Servidor:** Dell EMC Power Edge.
- **Clusters:** Configurados com Apache Spark.
- **Memória RAM:** 256 GB.
- **Armazenamento:** SSDs e discos rígidos de alto desempenho.
- **Rede:** Redes de fibra óptica.

- **Switches e Roteadores:** Alta capacidade.
- **Gerenciamento Remoto:** Recursos avançados de gerenciamento.
- **Dispositivos IoT e Sensores:** Coleta de dados em tempo real.

### **Controle Total sobre a Infraestrutura:**

- **Seleção de Hardware:**
  - A equipe de TI escolhe componentes específicos.
  - Possibilidade de selecionar hardware personalizado para necessidades específicas.
- **Configuração e Manutenção:**
  - Controle completo sobre a configuração do sistema.
  - Manutenção gerenciada internamente pela equipe de TI.
- **Alta Personalização:**
  - Capacidade de personalizar o servidor para projetos únicos.
  - Ajustes específicos para atender requisitos de segurança e conformidade.
- **Vantagens para Projetos Especiais:**
  - Ideal para projetos com requisitos de segurança rigorosos.
  - Adequado para ambientes que demandam conformidade regulamentar específica.

### **Ferramentas Necessárias:**

- **Sistema Operacional:** Linux.
- **Plataforma de Big Data:** Apache Spark.
- **Banco de Dados:** Apache Cassandra (NoSQL).
- **Ferramentas de Integração e ETL:** Talend.

- **Visualização de Dados:** Grafana.

### **Preparação dos Dados:**

1. Entendimento dos dados.
2. Limpeza de dados inconsistentes.
3. Integração de dados.
4. Transformação de dados.
5. Validação e verificação de dados.
6. Documentação e rastreabilidade.
7. Automatização e escalabilidade.

### **Segurança e Privacidade dos Dados:**

- **Criptografia:**
  - Dados em trânsito: TLS.
  - Dados em repouso: AES.
- **Controle de Acesso:** Autenticação multifator.
- **Treinamento e Conscientização:** Realizar treinamento com todos os usuários para integridade dos dados.