

# Documentação das Entidades - Sistema de Monitoramento ML V4

## Visão Geral do Sistema

O **Sistema de Monitoramento ML V4** foi projetado especificamente para suportar análises preditivas de falhas em máquinas industriais através de Machine Learning. A estrutura do banco de dados foi otimizada para capturar correlações entre sensores e facilitar o treinamento de modelos preditivos.

## ENTIDADE: FABRICANTE

### Propósito

Centraliza informações dos fabricantes das máquinas industriais, permitindo análises por marca, país de origem e facilitando contatos técnicos quando necessário.

### Campos

Campo	Tipo	Motivo da Inclusão
id_fabricante	<div>INT IDENTITY(1,1)</div> <div>PRIMARY KEY</div>	<b>Chave primária artificial:</b> Garante identificação única e eficiência em joins. IDENTITY elimina necessidade de controle manual de sequência.
nome_fabricante	<div>NVARCHAR(100) NOT NULL</div>	<b>Identificação do fabricante:</b> Campo obrigatório para rastreabilidade. NVARCHAR suporta caracteres especiais internacionais. Tamanho 100 suficiente para nomes corporativos.
pais_origem	<div>NVARCHAR(50)</div>	<b>Análise geográfica:</b> Permite identificar padrões de qualidade por região. Opcional pois pode não estar disponível para equipamentos antigos.
contato	<div>NVARCHAR(100)</div>	<b>Suporte técnico:</b> Facilita comunicação direta em caso de falhas recorrentes. Flexível para armazenar email, telefone ou URL.

### Justificativa da Entidade

- **Normalização:** Evita repetição de dados de fabricante em cada máquina
- **Integridade:** Centraliza dados corporativos para manutenção consistente
- **Análise:** Permite correlações entre fabricante e padrões de falha

## ENTIDADE: MAQUINA

## Propósito

Registro central das máquinas monitoradas. Funciona como dimensão principal para todas as análises, conectando dados de sensores às características físicas dos equipamentos.

## Campos

Campo	Tipo	Motivo da Inclusão
id_maquina	INT IDENTITY(1,1) PRIMARY KEY	<b>Identificador único:</b> Chave artificial para performance otimizada em consultas temporais massivas de sensores.
nome_maquina	NVARCHAR(100) NOT NULL	<b>Identificação operacional:</b> Nome usado pelos operadores. Obrigatório para rastreabilidade em relatórios e alertas de ML.
modelo	NVARCHAR(50)	<b>Agrupamento técnico:</b> Permite análises comparativas entre máquinas do mesmo modelo. Útil para identificar defeitos de projeto.
numero_serie	NVARCHAR(50) NOT NULL UNIQUE	<b>Identificação única física:</b> Rastreabilidade absoluta do equipamento. UNIQUE previne duplicação acidental.
id_fabricante	INT FOREIGN KEY	<b>Relacionamento com fabricante:</b> Permite análises de confiabilidade por marca. Foreign Key garante integridade referencial.
data_instalacao	DATE	<b>Análise temporal:</b> Correlaciona idade do equipamento com frequência de falhas. Fundamental para modelos de degradação.
localizacao	NVARCHAR(100)	<b>Contexto ambiental:</b> Permite identificar se localização influencia padrões de falha (umidade, temperatura ambiente, etc.).
status_operacional	NVARCHAR(20) DEFAULT 'ATIVO'	<b>Controle operacional:</b> Filtra máquinas ativas para análises. CHECK constraint garante valores válidos.

## Constraints Implementadas

- **FK\_maquina\_fabricante:** Integridade referencial com FABRICANTE
- **CHK\_status\_operacional:** Valores controlados ('ATIVO', 'INATIVO', 'MANUTENCAO')

## Justificativa da Entidade

- **Hub central:** Conecta todos os dados de sensores e falhas a características físicas
- **Contexto para ML:** Fornece features categóricas importantes (modelo, fabricante, idade)
- **Controle operacional:** Permite filtrar análises por equipamentos ativos

# ENTIDADE: LEITURA\_SENSORES

## Propósito

**Coração do sistema de ML.** Armazena todas as métricas dos sensores em um registro desnormalizado otimizado para análises. Estrutura projetada especificamente para alimentar algoritmos de Machine Learning.

## Campos Principais

Campo	Tipo	Motivo da Inclusão
id_leitura	<div>BIGINT IDENTITY(1,1)</div> <div>PRIMARY KEY</div>	<b>Suporte a Big Data:</b> BIGINT suporta milhões de leituras. Essencial para dados temporais massivos.
id_maquina	<div>INT FOREIGN KEY NOT</div> <div>NULL</div>	<b>Ligação com equipamento:</b> Conecta métricas à máquina específica. NOT NULL garante integridade.
data_hora_leitura	<div>DATETIME2 DEFAULT</div> <div>GETDATE()</div>	<b>Série temporal:</b> Timestamp preciso para análises temporais. DATETIME2 oferece maior precisão que DATETIME.

## Métricas dos Sensores (Features para ML)

Campo	Tipo	Justificativa Técnica
corrente_eletrica	<div>DECIMAL(8,3)</div>	<b>Indicador de carga:</b> Correlaciona com esforço da máquina. Precisão de 3 casas para miliamperes. Range até 99.999A suficiente para equipamentos industriais.
pressao	<div>DECIMAL(8,3)</div>	<b>Estado hidráulico/pneumático:</b> Crítico para bombas e compressores. Detecta vazamentos e obstruções. Precisão necessária para variações sutis.
temperatura	<div>DECIMAL(6,2)</div>	<b>Indicador de atrito/sobrecarga:</b> Principal feature para predição de falhas. Correlaciona fortemente com desgaste. Range -99.99 a 999.99°C.
umidade	<div>DECIMAL(5,2)</div>	<b>Condições ambientais:</b> Afeta componentes eletrônicos e lubrificação. Percentual de 0.00 a 100.00%.
vibracao	<div>DECIMAL(8,4)</div>	<b>Deteção de desbalanceamento:</b> Indicador precoce de problemas mecânicos. 4 casas decimais para detectar micro-variações.

## Metadados Contextuais

Campo	Tipo	Valor para ML
qualidade_sinal	<div>TINYINT DEFAULT</div> <div>100</div>	<b>Confiabilidade da leitura:</b> Permite filtrar dados ruins. Feature adicional para ajustar peso das observações.

Campo	Tipo	Valor para ML
temperatura_ambiente	DECIMAL(5,2)	<b>Normalização contextual:</b> Distingue aquecimento interno de fatores externos. Melhora precisão dos modelos.
turno	NVARCHAR(10)	<b>Feature categórica temporal:</b> Captura padrões operacionais. Diferentes turnos podem ter perfis de uso distintos.

## Design para Machine Learning

### Estrutura Desnormalizada Intencional:

- Todos os sensores em um registro facilitam análises multivariadas
- Reduz complexidade de joins em consultas analíticas
- Otimizada para algoritmos que processam features simultaneamente

### Correlações Implementadas no Script:

- Temperatura ↔ Corrente: Equipamentos sob stress consomem mais energia
- Temperatura ↔ Vibração: Dilatação térmica causa desalinhamentos
- Corrente ↔ Pressão: Bombas/compressores consomem mais energia sob alta pressão

## ⚠ ENTIDADE: FALHA

### Propósito

**Target principal para ML.** Registra eventos de falha que os modelos devem aprender a prever. Estrutura otimizada para criar labels temporais para treinamento supervisionado.

### Campos

Campo	Tipo	Propósito no ML
id_falha	INT IDENTITY(1,1) PRIMARY KEY	<b>Identificação única:</b> Rastreamento individual de cada evento de falha.
id_maquina	INT FOREIGN KEY NOT NULL	<b>Conexão com contexto:</b> Liga falha à máquina específica e seus dados de sensores.
tipo_falha	NVARCHAR(50)	<b>Classificação multiclasse:</b> Permite modelos que predizem não apenas SE haverá falha, mas QUAL TIPO.
data_hora_falha	DATETIME2 NOT NULL	<b>Janela temporal para labels:</b> Define o momento exato para criar targets "falha nas próximas X horas".

Campo	Tipo	Propósito no ML
gravidade	NVARCHAR(20)	<b>Priorização:</b> Permite modelos focados em falhas críticas. Feature ordinal (BAIXA < MEDIA < ALTA < CRITICA).
tempo_inatividade_horas	DECIMAL(6,2)	<b>Impacto econômico:</b> Permite otimizar modelos para minimizar custos de parada, não apenas falhas.
descricao_falha	NVARCHAR(500)	<b>Análise qualitativa:</b> Texto livre para análise posterior com NLP se necessário.
resolvido	BIT DEFAULT 0	<b>Status operacional:</b> Filtra falhas resolvidas vs. ativas para relatórios.

## Lógica de Criação de Labels

A view `VW_DADOS_ML` implementa a lógica:

```
sql
CASE
  WHEN EXISTS (
    SELECT 1 FROM FALHA f
    WHERE f.id_maquina = ls.id_maquina
    AND f.data_hora_falha BETWEEN ls.data_hora_leitura
    AND DATEADD(hour, 2, ls.data_hora_leitura)
  ) THEN 1 ELSE 0
END AS falha_proximas_2h
```

**Janela de 2 horas** escolhida porque:

- Tempo suficiente para ação preventiva
- Evita labels muito esparsos (muito tempo) ou densos demais (pouco tempo)
- Equilibra precisão temporal com praticidade operacional

## VIEW: VW\_DADOS\_ML

### Propósito

**Interface otimizada para Machine Learning.** Combina dados de sensores com labels de falha em formato pronto para algoritmos de ML.

## Features Principais

- **falha\_proximas\_2h**: Target binário principal (0/1)
- **proxima\_gravidade**: Target categórico para classificação de severidade
- Todas as métricas de sensores como features numéricas
- Informações contextuais (turno, modelo, fabricante) como features categóricas

## Otimizações Implementadas

- Join único entre LEITURA\_SENSORES e MAQUINA
- Subconsulta otimizada para detectar falhas futuras
- Estrutura que evita necessidade de joins complexos em análises

---

## Decisões de Design para Machine Learning

### 1. Granularidade Temporal

- Leituras a cada ~21.6 minutos (4000 registros em 60 dias)
- Frequência equilibrada: captura tendências sem saturar o modelo

### 2. Balanceamento de Classes

- Script gera ~20% de condições de falha
- Evita datasets extremamente desbalanceados
- Probabilidades ajustadas por tipo de máquina e horário

### 3. Correlações Realistas

- Relações físicas implementadas via código
- Temperatura como driver principal
- Efeitos em cascata entre variáveis

### 4. Índices Estratégicos

```
sql
```

```
IX_leitura_data_maquina -- Consultas temporais por máquina
IX_leitura_maquina      -- Agregações por equipamento
IX_falha_data_maquina   -- Junção com janelas temporais
```

## 5. Tipos de Dados Otimizados

- DECIMAL para precisão numérica (evita erros de ponto flutuante)
  - BIGINT para suportar milhões de leituras
  - DATETIME2 para precisão temporal
  - CHECK constraints para qualidade de dados
- 

## Métricas de Qualidade Implementadas

O sistema inclui queries de validação que verificam:

- Diferenças estatísticas entre condições normais e de falha
- Correlações esperadas entre variáveis
- Distribuição balanceada de classes
- Qualidade temporal dos dados

Esta estrutura garante que o banco forneça dados limpos, correlacionados e balanceados para treinar modelos de Machine Learning eficazes na predição de falhas industriais.