Dicionário de Dados - microsoft_security_incident

Informações da Tabela

Nome da Tabela: `microsoft_security_incident`
Banco de Dados: `security_incident_prediction`
Charset: `utf8mb4_unicode_ci`
Objetivo: Armazenar dados processados de incidentes de segurança Microsoft para análise e machine learning

---

Estrutura das Colunas

Coluna	Tipo de Dado	Nulo	Descrição	Categoria
`id`	BIGINT	NOT NULL	Identificador único do registro (chave primária)	Identificador
`org_id`	INT	NOT NULL	Identificador da organização	Identificador
`incident_id`	INT	NOT NULL	Identificador único do incidente	Identificador
`alert_id`	INT	NOT NULL	Identificador único do alerta	Identificador
`timestamp`	DATETIME	NOT NULL	Data e hora do incidente	Temporal
`year`	INT	NOT NULL	Ano extraído do timestamp (coluna gerada)	Temporal
`month`	INT	NOT NULL	Mês extraído do timestamp (coluna gerada)	Temporal
`day`	INT	NOT NULL	Dia extraído do timestamp (coluna gerada)	Temporal
`hour`	INT	NOT NULL	Hora extraída do timestamp (coluna gerada)	Temporal

day_of_week	INT	NOT NULL	Dia da semana extraído do timestamp - 0=Segunda, 6=Domingo (coluna gerada)	Temporal
detector_id	INT	NOT NULL	Identificador do detector de segurança que gerou o alerta	Detecção
alert_title	INT	NOT NULL	Título do alerta (valor codificado)	Detecção
category	INT	NOT NULL	Categoria do incidente de segurança (valor codificado)	Classificação
mitre_techniques	INT	NOT NULL	Técnicas MITRE ATT&CK associadas ao incidente (valor codificado)	Classificação
incident_grade	INT	NOT NULL	Grau de severidade do incidente: 0=FalsePositive, 1=TruePositive, 2=BenignPositive	Classificação
entity_type	INT	NOT NULL	Tipo de entidade envolvida no incidente (valor codificado)	Entidade
evidence_role	INT	NOT NULL	Papel da evidência no contexto do incidente (valor codificado)	Entidade
device_id	BIGINT	NULL	Identificador único do dispositivo envolvido	Entidade
sha256	BIGINT	NULL	Hash SHA256 do arquivo envolvido (valor codificado)	Artefato
ip_address	BIGINT	NULL	Endereço IP envolvido no incidente (valor codificado)	Artefato
url	BIGINT	NULL	URL envolvida no incidente (valor codificado)	Artefato
account_sid	BIGINT	NULL	Security Identifier (SID) da conta Windows (valor codificado)	Artefato
account_upn	BIGINT	NULL	User Principal Name (UPN) da conta (valor codificado)	Artefato
os_family	INT	NOT NULL	Família do sistema operacional (valor codificado)	Sistema

`os_version`	INT	NOT NULL	Versão do sistema operacional (valor codificado)	Sistema
`country_code`	INT	NOT NULL	Código do país onde ocorreu o incidente (valor codificado)	Geolocalização
`state`	INT	NOT NULL	Estado ou província onde ocorreu o incidente (valor codificado)	Geolocalização
`city`	INT	NOT NULL	Cidade onde ocorreu o incidente (valor codificado)	Geolocalização
`last_verdict`	INT	NOT NULL	Último veredito atribuído ao incidente (valor codificado)	Classificação

---

Categorias de Informação

1. Identificadores

Colunas que identificam unicamente registros e relacionamentos:

- `id`, `org_id`, `incident_id`, `alert_id`

2. Informações Temporais

Colunas relacionadas ao momento do incidente:

- `timestamp`, `year`, `month`, `day`, `hour`, `day_of_week`

3. Detecção e Classificação

Colunas que descrevem como o incidente foi detectado e classificado:

- `detector_id`, `alert_title`, `category`, `mitre_techniques`, `incident_grade`, `last_verdict`

4. Entidades Envolvidas

Colunas que identificam as entidades relacionadas ao incidente:

- `entity_type`, `evidence_role`, `device_id`

5. Artefatos de Segurança

Colunas que armazenam indicadores técnicos do incidente:

- `sha256`, `ip_address`, `url`, `account_sid`, `account_upn`

6. Informações do Sistema

Colunas sobre o sistema operacional do dispositivo:

- `os_family`, `os_version`

7. Geolocalização

Colunas que indicam a localização geográfica do incidente:

- `country_code`, `state`, `city`

Observações Importantes

Colunas Geradas

As colunas `year`, `month`, `day`, `hour` e `day_of_week` são **colunas geradas automaticamente** pelo MySQL a partir do campo `timestamp`. Seus valores são calculados automaticamente e armazenados fisicamente (STORED) para melhor performance em consultas.

Valores Codificados (Encoded)

A maioria das colunas contém **valores numéricos codificados** que representam categorias originalmente textuais. Este processo de encoding foi aplicado na camada de transformação de dados para otimizar o armazenamento e processamento em modelos de machine learning.

Valores Nulos

Colunas relacionadas a artefatos de segurança (`device_id`, `sha256`, `ip_address`, `url`, `account_sid`, `account_upn`) **permitem valores nulos** (NULL), pois nem todos os incidentes possuem todos os tipos de evidências associadas.

Classificação de Incidentes

O campo `incident_grade` utiliza a seguinte codificação:

- **0:** False Positive (falso positivo - não é uma ameaça real)
- **1:** True Positive (verdadeiro positivo - ameaça confirmada)
- **2:** Benign Positive (positivo benigno - atividade legítima mas suspeita)

Framework MITRE ATT&CK

A coluna `mitre_techniques` referencia técnicas do framework MITRE ATT&CK, que é uma base de conhecimento de táticas e técnicas utilizadas por atacantes em ambientes corporativos.