IBM QRadar: Uso de Coleções de Dados de Referência (Reference Data Collections)

Material técnico expandido e traduzido — Segurança da Informação e Administração de SIEM

Contexto

As coleções de dados de referência (Reference Data Collections) no IBM QRadar são componentes essenciais da infraestrutura de correlação e automação do SIEM. Elas permitem armazenar e manipular dados auxiliares — como listas de IPs, tabelas de usuários ou mapeamentos de ativos — que podem ser utilizados em regras, relatórios, painéis e APIs. Essas coleções funcionam como uma memória operacional do sistema, viabilizando análises mais contextuais e dinâmicas.

O curso ‘Using QRadar Reference Data Collections’ apresenta os principais tipos de estruturas de referência e suas finalidades, mostrando como usá-las de forma eficaz em ambientes corporativos para aprimorar a precisão das regras e o desempenho do SOC.

Desafios

O uso eficiente dessas coleções exige planejamento. Um dos desafios mais comuns é escolher o tipo correto de estrutura para o problema a ser resolvido: enquanto um ‘Reference Set’ é ideal para listas simples, um ‘Reference Map of Maps’ pode ser necessário para relacionamentos complexos. Outro desafio é a manutenção: coleções muito grandes ou sem expiração configurada podem degradar o desempenho do sistema.

Além disso, a consistência dos dados é fundamental. Quando múltiplas regras e integrações utilizam as mesmas coleções, é preciso implementar controles de acesso, auditoria e limpeza periódica (purge) para evitar dados redundantes ou obsoletos.

Conceitos Fundamentais

Propósito Geral das Coleções de Referência

As coleções de dados de referência têm como propósito fornecer um armazenamento temporário ou persistente de informações que são consultadas e manipuladas em tempo real pelas regras de correlação, pesquisas AQL, APIs REST e scripts do QRadar. Elas eliminam a necessidade de consultas externas e tornam o sistema mais responsivo e inteligente.

Por exemplo, uma regra pode verificar se um endereço IP que gerou um alerta pertence a um conjunto de IPs suspeitos (Reference Set) ou se o usuário está associado a um grupo de alto privilégio em um mapa de referência. Essa abordagem torna o SIEM capaz de correlacionar dados contextuais de forma rápida e confiável.

Reference Set

O ‘Reference Set’ é a estrutura mais simples, composta por uma lista de valores únicos. É amplamente usado para armazenar listas de IPs, domínios, usuários, hashes ou portas. Cada entrada pode ter um tempo de expiração (TTL – Time to Live) e pode ser alimentada por regras, APIs ou importações manuais. Ele é ideal para verificações binárias — presença ou ausência de um valor.

Exemplo: um analista pode criar um conjunto ‘Malicious_IPs’ e configurar uma regra para gerar um alerta sempre que um evento envolver um endereço presente nessa lista. É possível atualizar automaticamente o conjunto via integração com feeds de inteligência como MISP ou IBM X-Force.

Reference Map

O ‘Reference Map’ armazena pares de chave e valor, permitindo associar informações correlacionadas. Ele é útil quando é necessário relacionar um identificador a um atributo, como IP → País, Usuário → Departamento ou Domínio → Categoria. Cada chave é única e pode ser atualizada dinamicamente sem necessidade de recriar a estrutura.

Exemplo: um mapa pode conter IPs de origem como chaves e países de origem como valores. Uma regra pode então acionar um alerta se o país de um evento for classificado como de alto risco.

Reference Map of Sets

O ‘Reference Map of Sets’ expande o conceito de mapa, permitindo que cada chave aponte para um conjunto de valores. Isso é útil em cenários onde um elemento pode estar associado a múltiplos valores. Por exemplo, um usuário pode ter vários IPs ou dispositivos autorizados.

Exemplo: um mapa de conjuntos pode vincular cada usuário a todos os IPs usados nas últimas 24 horas. Regras podem detectar logins a partir de IPs fora desse conjunto, identificando possíveis comprometimentos de credenciais.

Reference Map of Maps

O ‘Reference Map of Maps’ é uma estrutura hierárquica que permite criar mapas dentro de mapas. Ele é ideal para armazenar dados multidimensionais e contextuais. Por exemplo, é possível mapear um nome de usuário para outro mapa contendo atributos como ‘Função’, ‘Departamento’ e ‘Último Login’.

Esse tipo de estrutura é frequentemente usado em automações e integrações avançadas, permitindo análises relacionais complexas e contextualização detalhada durante a correlação de eventos.

Reference Table

As ‘Reference Tables’ são estruturas tabulares semelhantes a bancos de dados relacionais, compostas por colunas nomeadas e registros. Elas são ideais para armazenar grandes volumes de dados estruturados, oferecendo suporte a consultas com AQL (Ariel Query Language). Cada linha contém uma chave primária e múltiplos atributos associados.

Exemplo: uma tabela pode registrar informações sobre logins de usuários, com colunas como ‘Usuário’, ‘IP’, ‘Horário’ e ‘Localização’. Regras podem consultar essa tabela para detectar atividades anômalas, como logins simultâneos em países diferentes.

Benefícios

As coleções de referência trazem flexibilidade, desempenho e escalabilidade para o QRadar. Elas permitem centralizar dados contextuais e evitam consultas externas, acelerando regras e buscas. Além disso, favorecem a automação e integração com feeds de inteligência, suportando ambientes de alta complexidade.

Recursos Avançados e Boas Práticas

1. **Automatize a atualização via API REST.** Utilize scripts ou integrações externas para popular as coleções. 2. **Defina prazos de expiração (TTL).** Evite crescimento indefinido e mantenha os dados atualizados. 3. **Monitore o tamanho e o desempenho.** Coleções muito grandes podem afetar o ECS e o CRE. 4. **Padronize nomes e descrições.** Use convenções claras para facilitar manutenção e auditoria. 5. **Combine tipos de coleções.** Use Tables para dados complexos e Sets para verificações rápidas.

Conclusão

As coleções de dados de referência são a base para um QRadar mais inteligente, eficiente e automatizado. Elas transformam dados brutos em conhecimento operacional, permitindo regras contextuais, integrações avançadas e detecção de ameaças em tempo real. Sua correta aplicação é um diferencial estratégico na maturidade de qualquer centro de operações de segurança (SOC).