Configure Memory Settings for the DLP Exact Data Matching Indexer

Contents

Introduction

Prerequisites

- Requirements
- Components Used

Overview

Problem

Solution

Introduction

This document describes how to increase available memory for the DLP Exact Data Matching Indexer to work with large data sources in Cisco Umbrella.

Prerequisites

Requirements

There are no specific requirements for this document.

Components Used

The information in this document is based on Cisco Umbrella.

The information in this document was created from the devices in a specific lab environment. All of the devices used in this document started with a cleared (default) configuration. If your network is live, ensure that you understand the potential impact of any command.

Overview

The Exact Data Match Indexer is part of the Exact Data Match feature in Umbrella DLP. The tool indexes a customer data source (CSV file) and generates fingerprints of critical records which are uploaded to Umbrella for use in DLP policies. This article explains how to increase the available memory for the indexer to work with large data sources.

Problem

When a large data source (CSV file) is indexed, this error displays:

```
ERROR: Out of heap space; please rerun with an increased size (-Xmx).
```

Solution

Run the indexing tool with `-Xmx` specifying the amount of memory to allocate to the indexing tool. The memory allocation can be specified in mebibytes (m) or gibibytes (g). For example:

- `-Xmx1000m` = 1000 mebibyte (1024 megabytes)
- `-Xmx1g` = 1 gibibyte (1074 megabytes)

The required memory depends on the file size of the source file (CSV file). Umbrella recommends allocating memory **at least twice the size of the source CSV file**.

For example, if the source data is 512 MB, the memory can be allocated like this:

```
java -X1g -jar edm-indexer.jar -i source_file.csv -e template-id
```

If the tool is being run in an automated way, then the memory allocation must be increased to account for changes in the source data size.