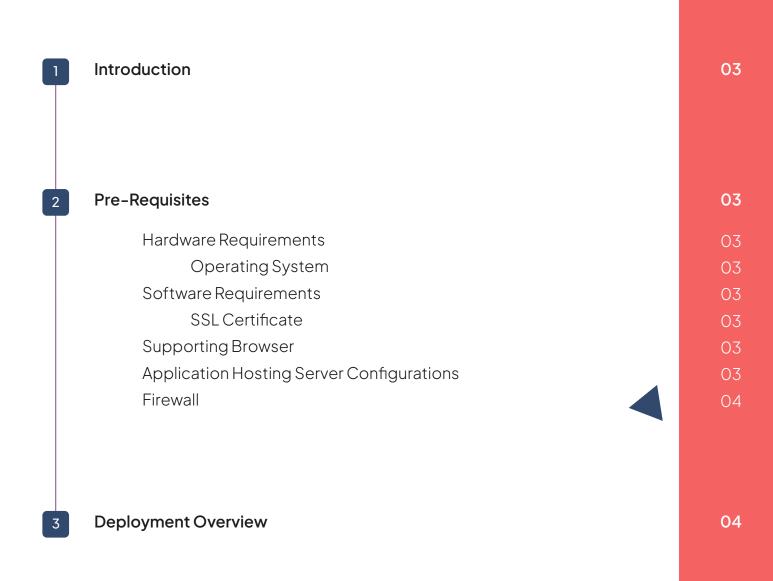![onix logo]

Prerequisites for Kingfisher Deployments on VMs

Table of Contents

Introduction

Kingfisher is an easy-to-use yet powerful and intuitive suite for generating Structured Synthetic Data. It is a web application that is deployed on a Tomcat server and uses the Postgres database as its metadata repository. It provides both a graphical interface as well as a RESTful API. This documentation outlines the prerequisite steps required for the successful setup of the Kingfisher application.

Pre-Requisites

Hardware Requirements

- Kingfisher application requires a baseline VM with a minimum of 8 cores and 32 GB RAM , with at least 500 GB persistent disk (N1 Std 8 machine or similar).

- Kingfisher Data Profiler requires a baseline VM with a minimum of 16 cores and 32 GB RAM , with at least 100 GB persistent disk (N1 Std 16 machine or similar).

Higher configuration will be required for heavy data generation needs. Talk to the Datametica team regarding your specific needs.

Operating System

Kingfisher is compatible with any of the following operating systems:

- Red Hat 7.7 and above
- CentOS 7 and 8
- Ubuntu 18.06 and above

Software Requirements

The following application software is required before installing Kingfisher:

- Docker (version 20.10.0 or higher)
- Docker Compose (version 1.0.29 or Higher)
- Kingfisher images (kingfisher-web, kingfisher-db, kingfisher-generator)

- Deployment package

SSL Certificate

Kingfisher can run either with 'http' or 'https' protocols. If you want to use Kingfisher with https, keep an SSL certificate ready with .pfx extension and in pkcs12 format. Passwords for the certificate will also be required.

To configure the SSL certificates, kindly request a Kingfisher HTTPS configuration document from Datametica

Supporting Browser

Kingfisher can be best experienced by using Google Chrome 5X or higher (Preferably the latest).

Application Hosting Server Configurations

The Kingfisher application consists of two components:

Core Kingfisher Web Application- This web-based application analyzes various artifacts and generates corresponding synthetic data. It is typically deployed in a non-production environment.

Kingfisher Data Profiler (Optional)- This command-line application requires SSH access to the container. It connects to the production environment to create a data profile based on actual production data. The Data Profiler must be deployed in the production environment, typically on a different virtual machine (VM) than the core application, using a separate Docker image.

The steps below apply to both instances:

1. The host user should have 'sudo' access to install Docker, docker-compose, and the GCloud utility.

2. A new Linux user will be created on the Kingfisher VM.

3. The Linux user will be added to the Docker group.

4. The Linux user needs permission to read and

write in the /opt directory and the extracted deployment directory.

5. Access to GCR will be required via the GCloud utility.

Firewall

- Kingfisher uses 8080 and 8081. These ports should be open between the user subnet and the Kingfisher VM.

- JDBC ports used on Target databases should be open between

  ◦ Kingfisher Application VM & Target

Note – Additional details are included in separate documents.

Deployment Overview