Open for Innovation
KNIME®

KNIME Server on AWS Marketplace

KNIME AG, Zurich, Switzerland
Version 4.8 (last updated on 2019-02-12)

Table of Contents

Introduction

KNIME Server is the enterprise software for team based collaboration, automation, management, and deployment of data science workflows, data, and guided analytics. Non experts are given access to data science via KNIME WebPortal or can use REST APIs to integrate workflows as analytic services to applications and IoT systems. A full overview is available here.

For an overview of use cases, see our solutions page. Presentations at KNIME Summits about usage of the KNIME Server can be found here.

Deployment on AWS

KNIME Server can be launched through the AWS Marketplace. There are several options:

- KNIME Server Medium
- KNIME Server Small
- KNIME Server (BYOL)

For a full list of product offerings including KNIME Analytics Platform, see here.

The KNIME Server (BYOL) instance requires you to Bring Your Own License file to use. To obtain a license file you should contact your KNIME representative, or sales@knime.com.

KNIME Server Small, Medium, and BYOL are single AMI instances, and are most easily launched via the AWS console. If you are familiar with the AWS CLI, you may also use this deployment method.

For self-build deployments using a custom base image, you should consult the KNIME Server Installation Guide.

Further reading

If you are looking for detailed explanations around the additional configuration options for KNIME Server, you can check the KNIME Server Administration Guide.

If you are looking to install KNIME Server, you should first consult the KNIME Server Installation Guide.

For guides on connecting to KNIME Server from KNIME Analytics Platform, or using KNIME WebPortal please refer to the following guides:

- KNIME Explorer User Guide
- KNIME WebPortal User Guide

There are additional resources such as the KNIME Server Advanced Setup Guide and KNIME Server Preview Functionality Guide.

Prerequisites

The person responsible for the deployment of KNIME Server should be familiar with basic AWS functionality surrounding configuring EC2 instances. KNIME Server administration requires basic Linux system administration skills, such as editing text files via the CLI, and starting/stopping systemd services.

KNIME Server Small, Medium, and BYOL are single AMI images and contain all software requirements.

For self-build instances, please consult the standard KNIME Server Installation Guide.

Pre-installed software

For convenience we have installed and pre-configured:

- OpenJDK 8 (required)
- Anaconda Python
- R
- Chrony
- AWS CLI
- Postfix
- iptables (redirects of requests on port 80, 443 to TomEE running on port 8080, 8443)

AWS resources

Launching an instance requires a VPC and subnet. The default security group will enable HTTP access on port 80, and HTTPS access on port 443. SSH access to administer the server on port 22.

Optional external dependencies

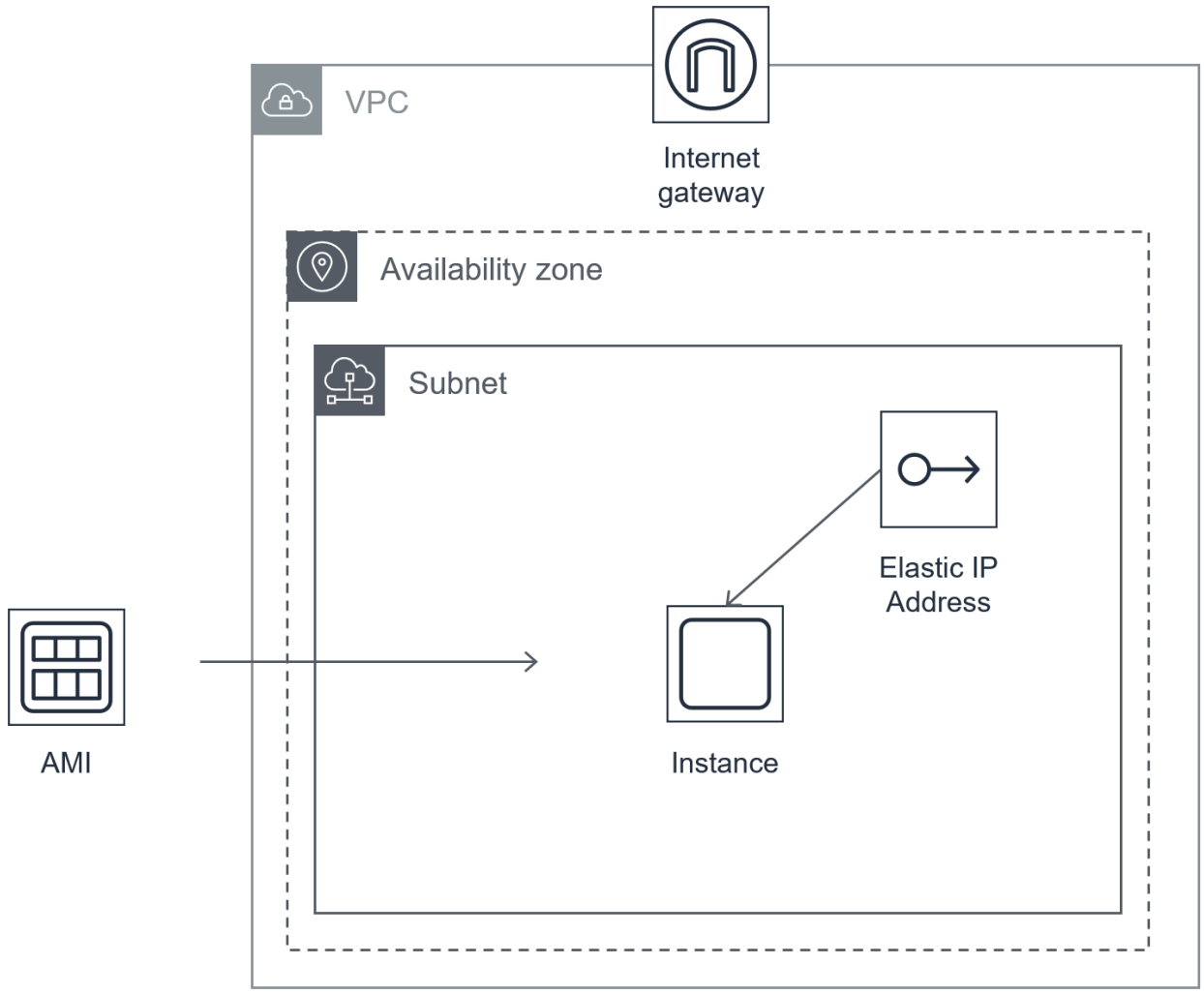
Optionally KNIME Server Large instances (currently available via BYOL license only) may choose to connect KNIME Server to an external LDAP/AD service. Full details are contained in KNIME Server Advanced Setup Guide.

Architecture

An overview of the general KNIME Server architecture is supplied. More detailed description of software architecture can be found in the KNIME Server Administration Guide.

KNIME Server Small/Medium

KNIME Server Small and KNIME Server Medium run as a single EC2 instance in a single subnet of a VPC. Use of an elastic IP is preferred since it simplifies the update/upgrade procedure.



Security

Detailed descriptions of general considerations for KNIME Server security configuration are described in the KNIME Server Administration Guide

Configurations specific to KNIME Server running on AWS are described below.

Access to KNIME Server instance

Root credentials are not required to access the KNIME Server instance.

IAM Roles and policies

IAM roles/policies that allow access to launch EC2 instances, and manage EBS volumes are required to launch the KNIME Server. It is assumed that a VPC with an internet gateway is configured and available.

Authenticating with AWS

KNIME Server does not require to authenticate with any AWS provided services.

Keys and rotation policies

An SSH key for access to the KNIME Server instance is created or chosen, via the AWS Management Console, or CLI at instance launch. You are responsible to manage this key as per the recommendations set within your organisation.

EC2 security groups and VPC access control lists

The default security group allows access to the KNIME Server via HTTP, and HTTPS on ports 80 and 443. Additionally advanced admin access via the SSH port 22 is enabled.

No VPC access control lists are defined.

Data encryption configuration

It is recommended that you enable EBS encryption and EBS snapshot encryption for all KNIME Server volumes. Full details available in the AWS documentation.

Audit trail

KNIME Server log files are accesible via the KNIME Server AdminPortal, or by accessing the files in their standard locations, as described in the KNIME Server Administration Guide

Tagging resources

You may wish to tag the EC2 instances and volumes for KNIME Server in order to identify e.g. owner, cost centre, etc. See the AWS Tagging Strategy document.

Costs

Software pricing

The software pricing for the KNIME Server is defined in the AWS Marketplace. See AWS Marketplace Pricing Questions regarding BYOL licensing should be directed to sales@knime.com.

Hardware pricing

Hardware pricing is defined by AWS. See AWS Pricing

Required services

- EC2
- EBS volume
- Data transfer in/out
- Optional: EBS Snapshots

Estimated costs

Costs calculated using the AWS Cost Calculator.

1. Estimated annual costs for a KNIME Server Medium, launched in US-East-1 are calculated as follows:

Service	Cost
Software license fees	$ 29000.00
EC2 instance fee (R4.2xlarge, annual 1 year all upfront)	$ 2740.00
EBS volume charges (16Gb + 250Gb)	$ 360.00
EBS snapshots (+10% monthly)	$ 395.88
Data transfer out (100 Gb/month)	$ 106.92
Total (annual cost)	$ 32602.80

Sizing

There is no 'one size fits all' answer to questions around sizing of deployments. The answer will vary depending on your typical workload, number of concurrent users, desired calculation time, and so on. We provide some recommendations to help get started.

KNIME Server Small, and KNIME Server Medium are both sold via the AWS Marketplace with built in licenses for 5 named users, and a maximum of 8 cores for workflow execution. Additionally KNIME Server Medium allows 20 consumers to access the KNIME Server WebPortal via the web browser only. Please contact sales@knime.com if you require a larger number of users, consumers, or cores.

EC2 Instance selection

Typically workflow execution speed can benefit from additional available instance RAM. Therefore we recommend the 'R' instance types, since they provide the best value access to RAM.

The R4.2xlarge instance has 61 Gb RAM available, and also 8 CPU cores, thus is the largest instance that KNIME Server Small and KNIME Server Medium can make use of.

Full details of EC2 instance types can be found here.

Currently the R5 instance type is not supported for marketplace images.

EBS volume selection

The default root instance volume is 50Gb SSD (gp2), and in most cases it is not necessary to increase the volume size. The additional volume has a default size of 250Gb SSD (gp2) which should be appropriate for many new installations.

To help guide the volume size, you will need to consider:

- the size and number of workflows expected to be stored
- the number of jobs executed and job retention duration
- number and size of additional files stored in the workflow repository

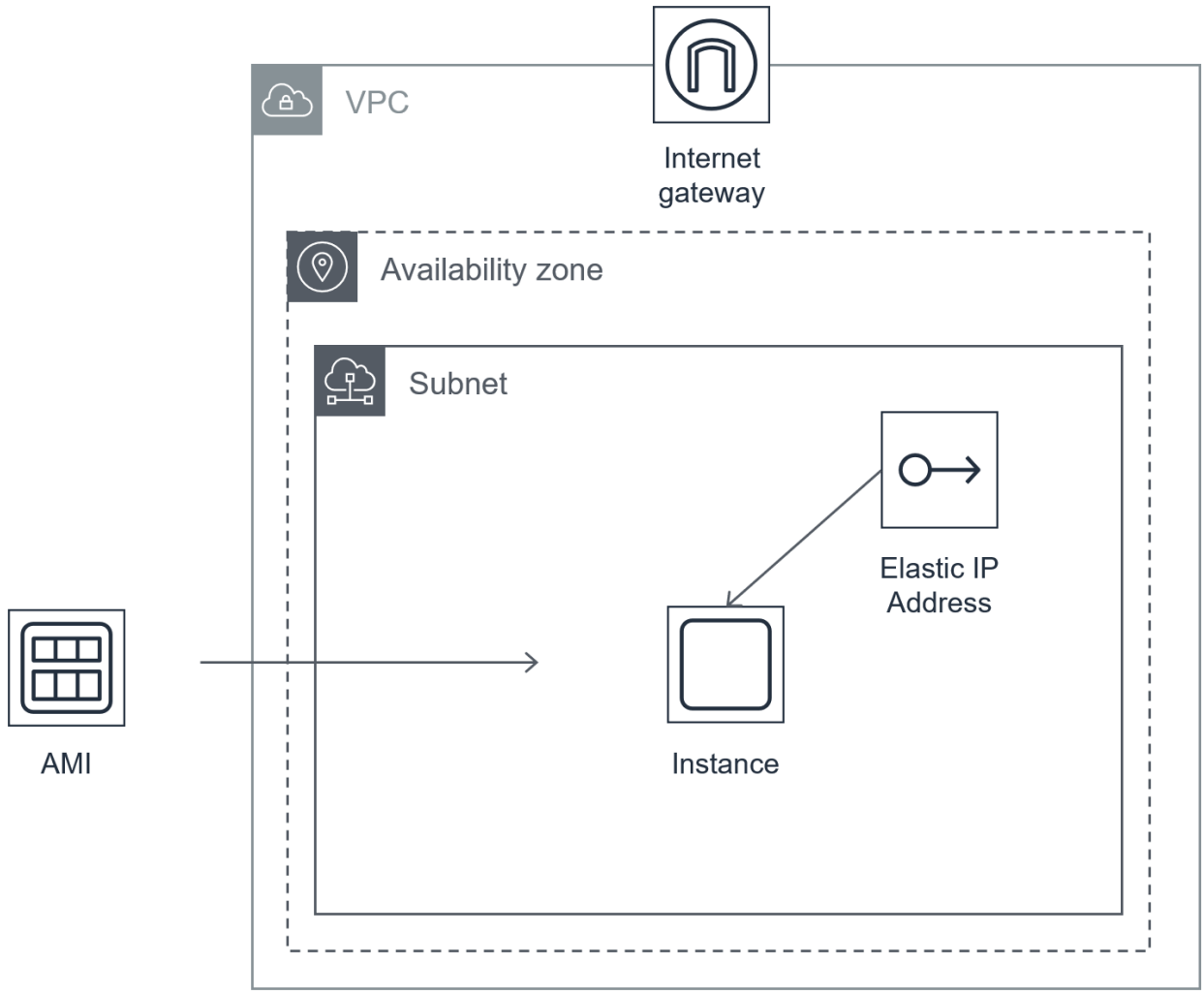
In case you need to add additional storage space later, please see the section Increasing EBS volume size.

Deployment

Deployment of KNIME Server Small, and KNIME Server Medium is via a single AMI. High-availabilty options are available as part of KNIME Server Large which is not currently documented for AWS.

Recommended deployment

A typical deployment as per the sizing guidelines in the previous section looks something like:



Applying license file to KNIME Server (BYOL)

For KNIME Server (BYOL) you will need to apply your license file. This can be done by visiting `https://<public-hostname>/knime` from your web browser.

Logging in using the admin username: knimeadmin, and password: <instance-id>, will redirect you to the license upload page. Here you can apply your license file. A valid license file will be immediately applied, and you can begin to use all KNIME Server functionality.

You can find the <instance-id> in the AWS Console (EC2 page) for the required instance. Or you may login to the instance via SSH and issue the command:

```
curl http://169.254.169.254/latest/meta-data/instance-id
```

Testing the deployment

Simple testing of the deployment can be done by logging into KNIME Server WebPortal via the web browser. Certain functionality is only available to test via the KNIME Analytics Platform.

Connecting via the browser

Once you have launched the KNIME Server AMI, the resulting instance will automatically start the KNIME Server. The KNIME Server WebPortal is available in the browser at `https://<public-hostname>/knime`

Connecting via the Analytics Platform

Access to the KNIME Server from the KNIME Analytics Platform is via the KNIME Explorer. Full documentation is available in the KNIME Explorer User Guide Use the mountpoint address: `https://<public-hostname>/knime` Username is knimeadmin, and password is `<instance-id>` unless you already changed them.

Testing workflow execution

Click on any workflow from the WebPortal repository tree, and wait for the page to load. If the 'Start' button appears then workflow execution is working as expected. For automated testing strategies, see the section: Automated testing.

Operations

As part of any KNIME Server deployment you should consider monitoring your service for availability. KNIME Server has several endpoints that can be used to determine the system health.

AZ fault

Since KNIME Server Small/Medium runs in a single AZ an AZ fault will be detected by the application fault detection method described below.

Instance fault

An instance fault can be detected using the standard AWS techniques.

Application fault

A simple REST call to the deployed KNIME Server should always return a 200 response with a payload similar to:

```
curl https://<public-hostname>/knime/rest
```

rest_response

```
{
  "@controls" : {
    "self" : {
      "href" : "https://<public-hostname>/knime/rest/",
      "method" : "GET"
    },
    "knime:v4" : {
      "href" : "https://<public-hostname>/knime/rest/v4",
      "title" : "KNIME Server API v4",
      "method" : "GET"
    }
  },
  "version" : {
    "major" : 4,
    "minor" : 8,
    "revision" : 0,
    "qualifier" : ""
  },
  "mountId" : "<public-hostname>",
  "@namespaces" : {
    "knime" : {
      "name" : "http://www.knime.com/server/rels#"
    }
  }
}
```

A different response indicates a configuration issue, or application fault.

It is also possible to test for executor availablity. This requires authenticating against the KNIME Server and calling the following REST endpoint.

```
curl -X GET "https://<public-hostname>/knime/rest/v4/repository/Examples/Test Workflows
(add your own for databases)/01 - Test Basic Workflow - Data
Blending:execution?reset=true&timeout=300000" -H "accept:application/vnd.mason+json"
```

Storage capacity

You may monitor the storage capacity of the two EBS volumes (root and data) using standard techniques and services such as AWS CloudWatch. For more details see here.

We recommend triggering an alarm at <5% free space on either volume.

Security certificate expirations

Certificate expiration will be caught if the basic server check fails with an HTTP 400 status code.

Backup

KNIME Server can be backed up subject to the information available in the KNIME Server Administration Guide.

It is recommended to make use of the AWS EBS Snapshot functionality. See the AWS documentation section on taking EBS snapshots.

Recovery

To restore an AWS EBS Snapshot, see the AWS documentation section on restoring EBS volumes.

Routine Maintenance

Starting KNIME Server

KNIME Server starts automatically when the instance starts using standard systemd commands. Once the TomEE application has started successfully, it will automatically launch and executor. This means that in normal operation you will not need the below command.

In the case that you need to start a stopped KNIME Server, it may be started using the following command at the terminal:

```
sudo systemctl start knime-server.service
```

Stopping KNIME Server

Stop the KNIME Server by executing the command:

```
sudo systemctl stop knime-server.service
```

Restarting KNIME Server

Restart the KNIME Server by executing the command:

```
sudo systemctl restart knime-server.service
```

Bootnote, for versions older than KNIME Server 4.7

Note that starting, stopping and restarting differs from version 4.7 and older of KNIME Server, where knime-server.service was replaced with apache-tomee.service

Restarting the executor

It is possible to restart the executor by issuing the following command:

```
sudo -u knime touch /srv/knime_server/rmirestart
```

This will launch a new executor, leaving the existing executor running. All existing jobs will continue to run on the old executor, and all new jobs will be launched on the new executor. That is helpful when updating executor preference files without needing to interrupt existing running jobs. When the rmirestart file is automatically deleted, the new executor has been launched.

It is possible to perform a hard kill on a running instance, by issuing the command:

```
sudo -u knime kill -9 <PID>
```

where <PID> is the process ID of the running executor. You can find the <PID> by running:

```
ps aux | grep knime
```

and looking for the process(es) that are not the apache-tomee instance.

Key rotation

Managing SSH keys for accessing the KNIME Server is detailed here.

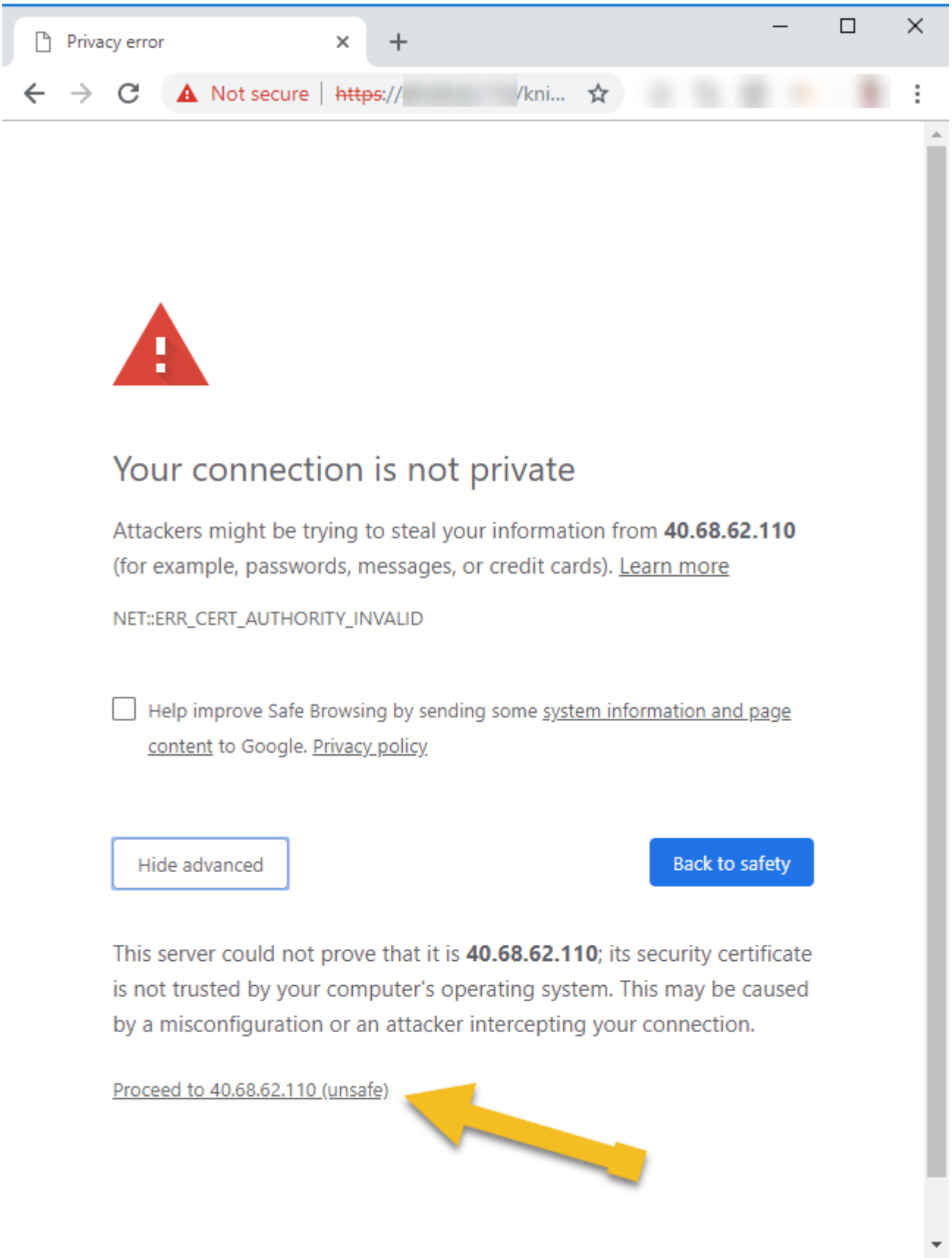
Managing Certificates

Detailed steps for managing the SSL certificate for KNIME Server can be found in the KNIME Server Administration Guide

Default Certificates

KNIME Server ships with a default SSL certificate. This allows for encrypted communication between client and server. However, since the certificate cannot be generated in advance for the server that you are running on, it will not be recognised as a valid certificate. Therefore, we recommend managing your own certificate as per the guidelines in the Managing Certificates section.

When testing with the default certificate, modern browsers will issue a warning as below. Choosing to ignore the warning, will allow you to access the KNIME WebPortal for testing.



Apply KNIME Server patches

Patches to the KNIME Server are announced on the KNIME Server Forum. You may subscribe to the topic. Details of update procedure are described in the KNIME Server Update Guide.

Update KNIME Server (feature version)

To make a feature update you have the option to follow the instructions in the KNIME Server Update Guide.

Increasing EBS volume size

It is possible to increase the size of the workflow repository EBS volume (default size: 250 Gb) after an instance has been launched. Follow the instructions here.

Emergency Maintenance

In case KNIME Server is not available due to degraded performance of an Availability Zone (AZ), EC2 instance fault, etc. It is possible to restore a snapshot and launch a new instance.

AZ recovery

AZ recovery is managed by launching a new instance into an unaffected AZ, using a recent snapshot.

Then attach the elastic IP from the affected instance to the new instance.

Region recovery

Region recovery is managed by launching a new instance into an unaffected region, using a recent snapshot.

Then attach the elastic IP from the affected instance to the new instance.

Support

KNIME Server Small support is provided by submitting questions in the KNIME Server forum.

KNIME Server Medium, and KNIME Server Large support is additionally supplied via contacting the support@knime.com email address.

When contacting KNIME Support you will need to include your Product Code, Instance ID, and AWS Account ID. We aim to respond to your question in under 48 hours.

Finding your product details.

Finding your Product Code, Instance ID and AWS Account ID:

The AWS documentation explains how to get access to the EC2 metadata that contains the information about your instance. You can also determine this information from the EC2 web management console.

Support costs

If you require additional support, please contact sales@knime.com for further information.

KNIME AG
Technoparkstrasse 1
8005 Zurich, Switzerland
www.knime.com
info@knime.com