THREDDS Clustering

Contents

1 Introduction **4**
- 1.1 Ansible - Automatic deployment and configuration 4
- 1.2 HTTP Load balancing 4
  - 1.2.1 HTTP Load balancing caveats 5
- 1.3 Direct Server Return 5

2 Deployment model **6**
- 2.1 The gateway 6
- 2.2 THREDDS instances 7
- 2.3 Collections and replicas 7

3 Roles and scenarios **8**
- 3.1 Overview of roles 8
- 3.2 Overview of scenarios 9

4 Roles reference **11**
- 4.1 Role virtualenv-conda 11
  - 4.1.1 Role dependencies 11
  - 4.1.2 Role usage 11
  - 4.1.3 Variables 11
  - 4.1.4 Documentation 11
- 4.2 Role virtualenv 11
  - 4.2.1 Role dependencies 11
  - 4.2.2 Role usage 12
  - 4.2.3 Variables 12
  - 4.2.4 Documentation 12
- 4.3 Role supervisord 12
  - 4.3.1 Role dependencies 12
  - 4.3.2 Role usage 12
  - 4.3.3 Variables 12
  - 4.3.4 Documentation 13
- 4.4 Role httpd 13
  - 4.4.1 Role dependencies 13
  - 4.4.2 Variables 13
  - 4.4.3 Documentation 14
- 4.5 Role httpd-bin 14
  - 4.5.1 Role dependencies 14
  - 4.5.2 Variables 14
  - 4.5.3 Documentation 15
- 4.6 Role jk-gateway 15
  - 4.6.1 Role dependencies 15
  - 4.6.2 Role usage 15
  - 4.6.3 Variables 15
  - 4.6.4 Documentation 16

4.7 Role tomcat . . . 16
4.7.1 Role dependencies . . . 16
4.7.2 Variables . . . 16
4.7.3 Documentation . . . 17
4.8 Role tds . . . 17
4.8.1 Role dependencies . . . 17
4.8.2 Role usage . . . 17
4.8.3 Variables . . . 18
4.8.4 Documentation . . . 18
4.9 Role tds-jk . . . 18
4.9.1 Role dependencies . . . 18
4.9.2 Role usage . . . 19
4.9.3 Variables . . . 19
4.9.4 Documentation . . . 19

5 Scenarios reference **20**
5.1 Scenario devel-jk . . . 20
5.1.1 Requirements . . . 20
5.1.2 Documentation . . . 20
5.1.3 Usage . . . 20
5.1.4 Infrastructure description . . . 21
5.1.5 Networking . . . 21
5.1.6 JPDA debug . . . 21
5.2 Scenario tds_standalone . . . 21
5.2.1 Requirements . . . 22
5.2.2 Usage in localhost . . . 22
5.2.3 Usage in vagrant . . . 22
5.2.4 Scenario's variables . . . 23

1 Introduction

The THREDDS project is developing middleware to bridge the gap between data providers and data users. The goal is to simplify the discovery and use of scientific data and to allow scientific publications and educational materials to reference scientific data. Due to THREDDS's lack of horizontal scalability and automatic configuration management and deployment, this service usually deals with service downtimes and time consuming configuration tasks, mainly when an intensive use is done as is usual within the scientific community (e.g. climate).

This project aims to improve the scalability of the THREDDS Data Server through the implementation of a cluster of TDS instances that are deployed in the backend and that are only visible from the outside through a reverse proxy in the frontend. This project also aims to improve the management of the TDS catalogs by partitioning the hierarchy of catalogs into multiple TDS instances that are deployed in the backend, allowing high availability of the datasets and tackling the current problem of large waiting times after performing a THREDDS reinit when publishing new catalogs.

Instead of the classic installation and configuration of a single or multiple independent THREDDS servers, manually configured, this work presents an automatic provisioning, deployment and orchestration cluster of THREDDS servers.

1.1 Ansible - Automatic deployment and configuration

Ansible is a tool for automating configuration management and deployment and is used in TDS Clustering in order to automate the creation of the infrastructure.

This solution is based on Ansible playbooks, used to automate the deployment and management of the agents that conform the cluster. The playbooks are based on modules (or roles) of different backends and frontends load balancing setups and solutions. This implementation allows to configure different infrastructure and deployment setups, as more workers are easily added to the cluster by simply declaring them as Ansible variables and executing the playbooks.

1.2 HTTP Load balancing

The frontend load balancing system enables horizontal scalability by delegating requests to backend workers, consisting in a variable number of instances for the THREDDS server that are deployed inside Apache Tomcat containers. Through the clustering capacity of the Apache Tomcat server, in combination with the JK connector and the Apache Web Server, all HTTP features such as HTTP sessions, are available for the THREDDS cluster. This clustering also provides

fault-tolerance and better reliability since if any of the workers fail another instance of the cluster can take over it.

Additonally to the Apache httpd+mod_jk setup for the gateway, this project provides alternatives to perform the load balancing in the reverse proxy, such us HAproxy and nginx.

1.2.1 HTTP Load balancing caveats

In the context of HTTP load balancing, the load balancer becomes a bottleneck, since all the traffic goes through it. This is particularly problematic when large datasets are accessed by the clients, because the load balancer must perform heavy transfers of data at the kernel level, which involves a lot of overhead in data being copied from the network interface and the hard disk. This breaks the load balancer at the CPU level because of the excessive amount of work that it has to do.

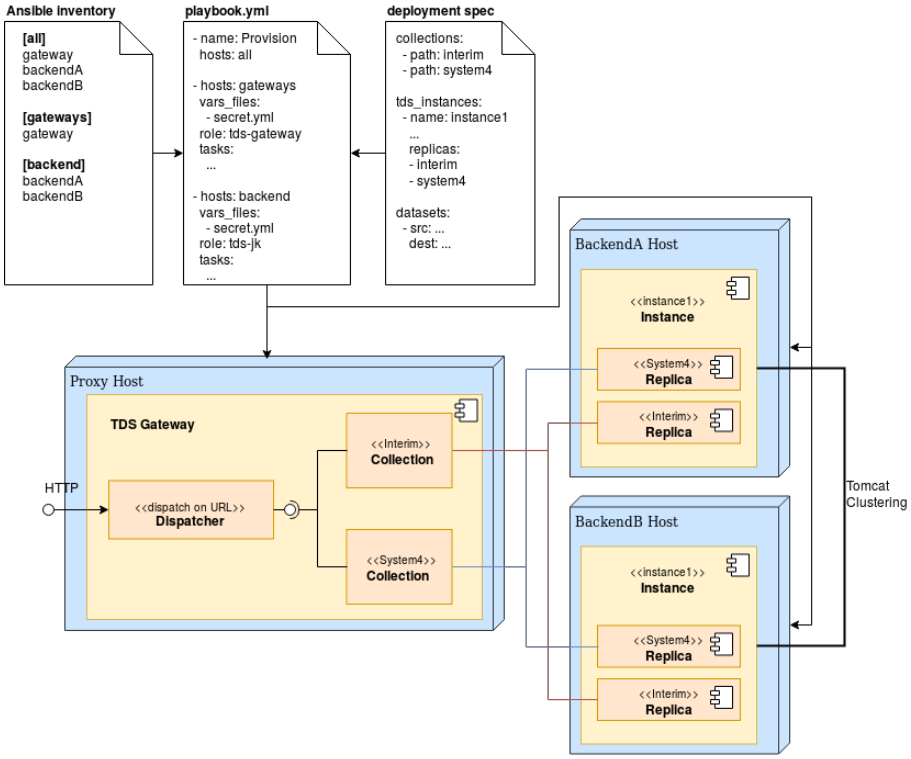
1.3 Direct Server Return

Direct Server Return is a type of load balancing in which the load balancer routes packets to the backends without changing anything in it but the destination MAC address. The backends process the requests and answer directly to the clients, without passing through the load-balancer.

This approach do away with the bottleneck created in the load balancer by the HTTP load balancing strategy, although it requires additional low level configuration and HTTP features are not available anymore.

2 Deployment model

The following image provides a first overview of the deployment model.



In the figure we can identify the following agents:

- Reverse proxy or gateway
- Instances (THREDDS instances)
- Collections
- Replicas

All these elements are described in detail in the following sections.

2.1 The gateway

The gateway is the software that works as a reverse proxy and it is in charge of performing the load balancing, forwarding requests to the backend TDS instances deployed in Tomcat application servers. Multiple options are considered in this project to act as a gateway: httpd+mod_jk, HAproxy, Linux Virtual Server and others.

2.2 THREDDS instances

Each TDS instance, or simply instance, corresponds to a Tomcat server proccess running the THREDDS web application in the backend hosts. Each host can run one or more instances.

Each instance has the option to support any number of collections and each instance will only respond to requests targeting data available in the supported collections. This is done through the appropriate filtering of requests in the gateway.

In case of using the httpd+mod_jk solution for the gateway, connectors of type AJP are used to connect the gateway with the instances. Usually, you only need one Connector element per gateway in the deployment, since Connector elements must reference their proxy. Each instance can contain one or more connectors (Connector xml elements in Tomcat's server.xml).

Instances are clustered in the backend using Tomcat's clustering capabilities, which allows user session replication and in case that one of the instances become unavailable, another can take over it without any disruption of the service.

2.3 Collections and replicas

From an user perspective, collections are an aggrupation of catalogs and datasets. For example, the collection Interim would contain all data that belongs to ERA-Interim. Collections have no more meaning to users. However, collections allow system administrators to partition and replicate the catalog and dataset infrastructure into replicas, isolating datasets from the Interim collection from datasets of other collections. Thus, changes to the Interim collection don't have side effects on the datasets of other collections.

From an administrator perspective, when using Apache httpd+mod_jk in the gateway, collections are mod_jk workers of type load balancing compunded of replicas, that is, mod_jk balance workers. Each replica points to a tomcat instance deployed in the backend hosts that holds a copy of the contents served by the collection. Each request is routed, through the use of urimaps, to the TDS instances that support the collection in which the requested data resides.

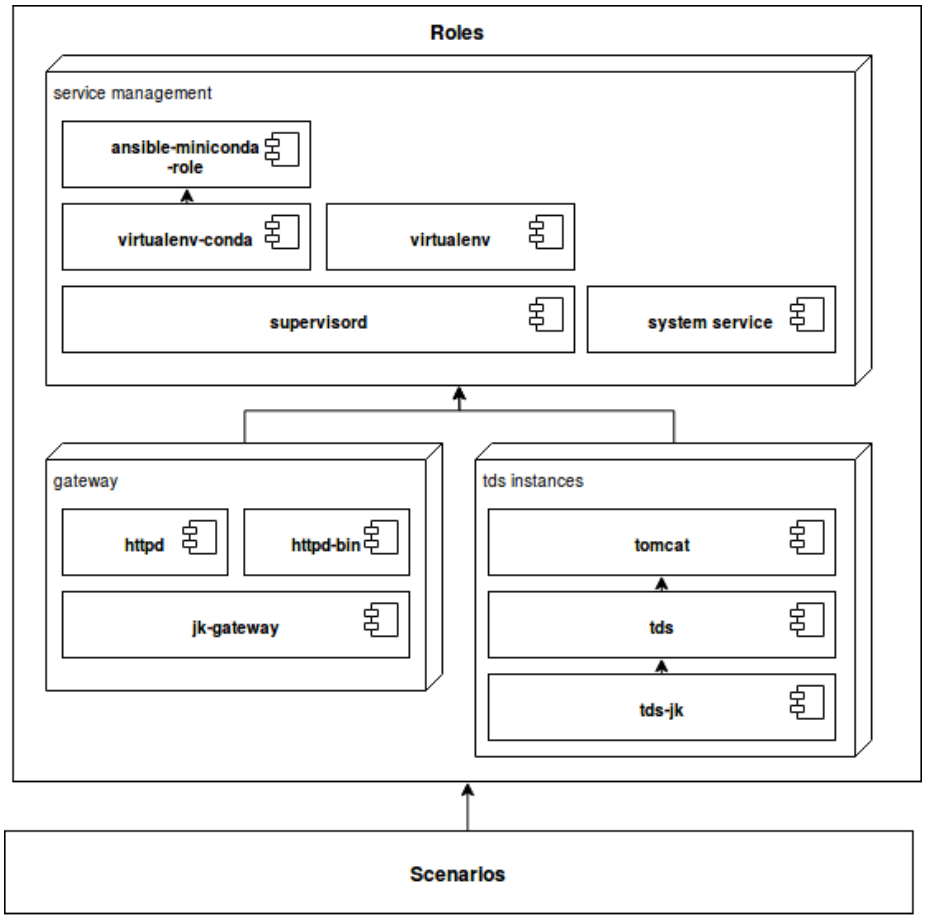
3 Roles and scenarios

Ansible roles have been created to split the deployment of TDS Clustering into reutilizable parts. A user who wants to implement his specific TDS Clustering scenario has to define an Ansible inventory and playbook, and also define the variables that the Ansible roles will use to do the deployment.

3.1 Overview of roles

The defined Ansible roles are:

- ansible-miniconda-role - Role to install miniconda
- virtualenv - Role to install a python virtualenv
- virtualenv-conda - Role to install a virtualenv using miniconda
- supervisord - Role to install supervisord in top of a virtualenv
- httpd - Role to install httpd from source code
- httpd-bin - Role to install httpd from system packages
- jk-gateway - Role to deploy the mod_jk gateway in top of httpd
- tomcat - Role to deploy a tomcat server that will contain multiple instances
- tds - Role to deploy the THREDDS Data Server in tomcat instances
- tds-jk - Brigde between jk-gateway and tds



Roles are further explained in the following chapters.

3.2 Overview of scenarios

Scenarios are concrete use cases that make use of the available Ansible roles to do deployments with specific settings. Roles do the deployment of the minimun infrastructure for the THREDDS cluster to run and any customization must be defined in the scenarios.

Any scenario will be composed of:

- inventory - Ansible inventory file, which defines the hosts involved in the deployment.
- main playbook - Ansible playbook that will contain two plays, one for the gateway hosts and other for the tds instances hosts, plus an optional thrird play that is in charge of the provisioning of the hosts.

- ansible.cfg - Ansible configuration file.
- user variables - Variables defined by the user that are required by the roles or that are defined to customize the deployment of the scenario.

4 Roles reference

4.1 Role virtualenv-conda

The purpose of this role is to deploy a python virtualenv in top of the miniconda installed by the role ansible-miniconda-role. The ansible-miniconda-role is installed in order to install libnetcdf for the tds instances to use it in their ncss service.

4.1.1 Role dependencies

ansible-miniconda-role

4.1.2 Role usage

This role enforces you to use the following variables in your playbook (variables required from the role `ansible-miniconda-role`):

- `miniconda_python` - Set it to `2` or `3`
- `miniconda_prefix` - Path where miniconda will be deployed, e.g. `"{{ ansible_env.HOME }}/miniconda2"`
- `miniconda_env.name` - Name of the virtualenv created inside miniconda
- `miniconda_env.dependencies` - Libs to install in the virtualenv

4.1.3 Variables

- `venv_home`: `"{{ miniconda_prefix }}/envs/{{ miniconda_env.name }}"`

4.1.4 Documentation

4.2 Role virtualenv

The purpose of this role is to deploy a python virtualenv.

4.2.1 Role dependencies

None.

4.2.2 Role usage

4.2.3 Variables

- `venv_home`: `"{{ ansible_env.HOME }}"` - Path where the virtualenv will be created.
- `venv_version`: `15.1.0`
- `venv_file_name`: `"virtualenv-{{ venv_version }}.tar.gz"`
- `venv_file_url`: `"https://pypi.python.org/packages/d4/0c/9840c08189e030873387a73b90ada981`
- `venv_sw`: `"{{ venv_home }}/sw"`
- `tmp_dir`: `"/tmp"`

4.2.4 Documentation

4.3 Role supervisord

This role install supervisord in the virtualenv created by one of the roles that installs virtualenv.

4.3.1 Role dependencies

None.

4.3.2 Role usage

If you want the supervisord to be proctected with password, you have to declare the following variables:

- `supervisord_user`: User name
- `supervisord_password`: User password

If they are not declared, supervisord is not protected.

4.3.3 Variables

- `supervisord_etc`: `"{{ venv_home }}/etc"`
- `supervisord_var`: `"{{ venv_home }}/var"`
- `supervisord_programs`: `"{{ supervisord_etc }}/supervisord.d"`
- `supervisord_port`: `9001`

4.3.4 Documentation

4.4 Role httpd

This role deploys an httpd server, downloading sources into the control machine and compiling it in every host. It is intented to allow the deployment by a non root user.

4.4.1 Role dependencies

None.

4.4.2 Variables

- `httpd_server_root`: `"{{ ansible_env.HOME }}/httpd"`
- `httpd_document_root`: `"{{ httpd_server_root }}/htdocs"`
- `httpd_version`: `2.4.25`
- `httpd_port`: `8000`
- `httpd_src`: `"{{ httpd_server_root }}/src"`
- `httpd_conf_path`: `"{{ httpd_server_root }}/conf"`
- `httpd_conf_file`: `"{{ httpd_conf_path }}/httpd.conf"`
- `httpd_version_major`: `"{{ httpd_versiontruncate(3, False, , 0) }}"`
- `httpd_mirror`: `https://archive.apache.org/dist/httpd/`
- `httpd_filename_unarchive`: `"httpd-{{ httpd_version }}"`
- `httpd_filename`: `"{{ httpd_filename_unarchive }}.tar.gz"`
- `httpd_download_url`: `"{{ httpd_mirror }}/{{ httpd_filename }}"`
- `httpd_to_be_removed`: `[build,build-1,icons,man,manual,src,include`
- `apr_version`: `"1.5.2"`
- `apr_filename_unarchive`: `"apr-{{ apr_version }}"`
- `apr_filename`: `"{{ apr_filename_unarchive }}.tar.gz"`
- `apr_mirror`: `http://archive.apache.org/dist/apr`
- `apr_download_url`: `"{{ apr_mirror }}/{{ apr_filename }}"`
- `apr_install_base`: `"{{ httpd_server_root }}"`
- `apr_util_version`: `"1.5.4"`

- apr_util_filename_unarchive: "apr-util-{{ apr_util_version }}"
- apr_util_filename: "{{ apr_util_filename_unarchive }}.tar.gz"
- apr_util_mirror: http://archive.apache.org/dist/apr
- apr_util_download_url: "{{ apr_util_mirror }}/{{ apr_util_filename }}"
- apr_util_install_base: "{{ httpd_server_root }}"
- pcre_version: "8.40"
- pcre_filename_unarchive: "pcre-{{ pcre_version }}"
- pcre_filename: "{{ pcre_filename_unarchive }}.tar.gz"
- pcre_mirror: http://ftp.cs.stanford.edu/pub/exim/pcre
- pcre_download_url: "{{ pcre_mirror }}/{{ pcre_filename }}"
- pcre_install_base: "{{ httpd_server_root }}"

4.4.3 Documentation

4.5 Role httpd-bin

This role deploys and httpd server using the system package manager. At the moment, it only works in yum based systems.

4.5.1 Role dependencies

None.

4.5.2 Variables

- httpd_server_root: /etc/httpd
- httpd_conf_file: /etc/httpd/conf/httpd.conf
- httpd_document_root: /var/www/html
- httpd_port: 80
- httpd_dependencies: ["@Development tools", "httpd-devel", "apr", "apr-devel", "apr-util'

4.5.3 Documentation

4.6 Role jk-gateway

Here follows the documentation of the role jk-gateway. This role deploys mod_jk in order to act as a reverse proxy for the TDS cluster. It can be used with an existing httpd installation or after executing the httpd role.

4.6.1 Role dependencies

Although this role does not define any dependencies in the `meta/` directory, it is supposed to be used on top of one of the `httpd` roles. It can also be used in top of an existing httpd installation that was previously carried out without any of the roles provided by this project, as it is the case of the ESGF scenario.

4.6.2 Role usage

If you want to deploy the status worker, you have to declare the following variables:

- `mod_jk_status_user`: User to login
- `mod_jk_status_passwd`: Password to login

If they are not declared, the `.htpasswd` file nor the status worker are created.

4.6.3 Variables

These variables control the deployment of the reverse proxy based on the underlying httpd installation. Default values assume that httpd has been installed using one of the provided httpd roles or that httpd has been installed from system packages.

- `mod_jk_version`: `"1.2.43"`
- `mod_jk_filename_unarchive`: `"tomcat-connectors-{{ mod_jk_version }}-src"`
- `mod_jk_filename`: `"{{ mod_jk_filename_unarchive }}.tar.gz"`
- `mod_jk_mirror`: `http://archive.apache.org/dist/tomcat/tomcat-connectors/jk`
- `mod_jk_download_url`: `"{{ mod_jk_mirror }}/{{ mod_jk_filename }}"`
- `mod_jk_src`: `"{{ ansible_env.HOME }}/mod_jk_src"`
- `mod_jk_conf_path`: `"{{ httpd_server_root  default(/etc/httpd) }}/conf.d"`
- `mod_jk_shm`: `"{{ httpd_server_root  default(/etc/httpd) }}/logs/mod_jk.shm"`

- `mod_jk_log`: `"{{ httpd_server_root  default(/etc/httpd) }}/logs/mod_jk.log"`
- `mod_jk_static_catalog`: `True`
- `mod_jk_status_path`: `"{{ mod_jk_conf_path }}"`

The `mod_jk_static_catalog` variable is used to indicate if the role should create static html files imitating THREDDS html catalog files in the reverse proxy. It is true by default and explicitly set to false in the ESGF scenario.

The `mod_jk_status_passwd` variable sets the path of the .htpasswd file that contains the user and password for the login in the jk status worker.

4.6.4 Documentation

4.7 Role tomcat

This role performs the deployment of a tomcat server by downloading files from the Internet and following a configuration that does not require the user to have roor privileges.

4.7.1 Role dependencies

The role does not include any dependency in it's `meta/` directory but it is supposed to be used on top of the `supervisord-conda` role.

4.7.2 Variables

- `jre_version`: `8u141`
- `jre_filename_unarchive`: `"jre-{{ jre_version }}-linux-x64"`
- `jre_filename`: `"{{ jre_filename_unarchive }}.tar.gz"`
- `jre_mirror`: `http://download.oracle.com/otn-pub/java/jdk/`
- `jre_download_url`: `"http://download.oracle.com/otn-pub/java/jdk/8u172-b11/a58eab1ec24421`
- `jre_header`: `"Cookie: oraclelicense=accept-securebackup-cookie"`
- `jre_install_base`: `"{{ tomcat_home }}/jre"`
- `tomcat_home`: `"{{ ansible_env.HOME }}/tomcat-home"`
- `tomcat_base`: `"{{ ansible_env.HOME }}/tomcat-base"`
- `tomcat_version`: `8.0.42`
- `tomcat_version_major`: `"{{ tomcat_versiontruncate(1, True, , 0) }}"`
- `tomcat_filename_unarchive`: `"apache-tomcat-{{ tomcat_version }}"`

- tomcat_filename: "{{ tomcat_filename_unarchive }}.tar.gz"
- tomcat_mirror: http://archive.apache.org/dist/tomcat
- tomcat_download_url: "{{ tomcat_mirror }}/tomcat-{{ tomcat_version_major }}/v{{ tomcat_v
- tomcat_to_be_removed: "[KEYS,LICENSE,NOTICE,RELEASE-NOTES,RELEASE-NOTES.html,RUNNING.txt

4.7.3 Documentation

4.8 Role tds

This role deploys a tds instance inside the tomcat server deployed by the 'tomcat' role.

4.8.1 Role dependencies

- tomcat

4.8.2 Role usage

This role requires that you define the following variables in your playbook: tds_instances, collections, datasets.

Every tds_instance corresponds to a tomcat process running in the $CATALINA_BASE directory defined in the tomcat role.

```
tds_instances:
  - name: Identifier for the instance, also used in the path to the instance (appended to to
    shutdown: Port for tomcat shutdown
    tds_content_root: Path to the content root of the TDS instance. Default is $CATALINA_BAS
    tds_debug:
      jpda_address: Default is localhost:8000
    connectors: Connectors for tomcat
      type: (http|ajp)
      port: Port number for the connector
      proxyName: proxyName attribute for the connector
      connectionTimeout: connectionTimeout attribute for the connector
    replicas: Collections that are supported by the instance
      name: Must reference the path attribute of the collection that this instance replicate
```

A collection is a tree of THREDDS catalogs. The root of the tree is a file, in xml format, whose name must be catalog.xml. This file follows the THREDDS xml catalog specification.

```
collections:
  - path: Identifier of the collection and path under content/thredds
    top: If True, a link will be added to the static catalog.html in the reverse proxy
    catalogs:
      src: Local path to the catalogs directory
```

The `datasets` variables allows the user of the role to copy files from the control host(in Ansible parlance), to the remote hosts. Usually, these files are datasets referenced by the catalogs.

```
datasets:
  - src: Local path to the datasets directory
    dest: Absolute remote path where datasets will be copied. If empty, datasets are copied
```

4.8.3 Variables

- `tds_version`: `ESGF-5.0.1`
- `tds_filename_unarchive`: `"tds-{{ tds_version }}"`
- `tds_filename`: `"{{ tds_filename_unarchive }}.war"`
- `tds_mirror`: `http://artifacts.unidata.ucar.edu/content/repositories/unidata-releases/edu/`
- `tds_download_url`: `"{{ tds_mirror }}/{{ tds_version }}/tds-{{ tds_version }}.war"`
- `tds_debug`: `False`

4.8.4 Documentation

TDS is downloaded from Unidata Nexus repository.

4.8.4.1 TdsClusterAuthorizer

ToDo

4.9 Role tds-jk

This role is the glue between the role `jk-gateway` and the role `tds`. For every `tds_instance` defined for the `tds` role, it configures the `mod_jk` reverse proxy or gateway to redirect requests to the appropriate `tds_instance`.

4.9.1 Role dependencies

- tds

4.9.2 Role usage

See documentation for the role `tds`.

4.9.3 Variables

4.9.4 Documentation

`mod_jk` sticky sessions are used.

5 Scenarios reference

5.1 Scenario devel-jk

This scenario serves as use case for development and testing of the TDS+`mod_jk` gateway infrastructure. It also serves as an example of different use cases that can be implemented.

5.1.1 Requirements

1. vagrant
2. ansible 2.5

5.1.2 Documentation

The main files in this directory are "source.yml" and "binary.yml". These playbooks deploy a httpd+mod_jk reverse proxy in the ansible host "proxy", that forward requests to a backend consisting on two hosts, "hostA" and "hostB", each running a tds instance. Configuration of variables can be found in the group_vars directory (see Ansible docs).

test-authentication.yml is a playbook that configures a simple setup in the THREDDS instances for testing of authentication and authorization. This playbooks creates a user named "alice" with password "1234" who has access to the datasets in the collection2 (see data/ directory).

The inventory file, named inventory, defines the three hosts involved in this scenario. The "test-simple" group is only used for quick testing avoiding the deployment of hostB.

The data directory contains sample THREDDS catalog files and NetCDF datasets.

5.1.3 Usage

1. vagrant up && vagrant reload
2. ansible-playbook (source.ymlbinary.yml) –ask-vault-pass [–limit test-simple]
3. (in web browser) localhost:9000/thredds, localhost:9000/status-jk

5.1.4 Infrastructure description

- CentOS 7 vagrant machines
- httpd 2.4
- tomcat 8
- tds 4.6

5.1.5 Networking

Virtual machines are configured to use an internal network (see Vagrantfile) and each virtual machine has the following configuration:

- proxy - eth1, ip: 192.168.50.10
- hostA - eth1, ip: 192.168.50.11
- hostB - eth1, ip: 192.168.50.12

The following ports are forwarded from the host to the guests:

- host 9000 -> proxy 8080
- host 9001 -> hostA 8080
- host 9002 -> hostB 8080

5.1.6 JPDA debug

Tomcat instances are configured to be debugged used JPDA. JPDA listens in the port 8000 in both hostA and hostB.

5.2 Scenario tds_standalone

This scenario deploys a sandbox, using the framework provided by TDS Collections, that contains:

- A THREDDS instance available at http://localhost:8080/thredds by default
- content/thredds is available in the root directory where you deployed the sandbox
- toolsUI jar
- ./toolsUI inits the toolsUI interface in the background
- ./ncjdump is a shorcut to `java -cp toolsUI.jar ucar.nc2.NCdumpW "$@"`

The THREDDS instance is populated with the following data:

- THREDDS catalogs contained in `data/catalogs`
- Datasets contained in `data/datasets`
- A <dataset> entry is created in the main catalog for each file found in data/datasets, although this is not recursive

5.2.1 Requirements

1. ssh key pair in your ~/.ssh directory and authorized to ssh to localhost
2. ansible 2.5

5.2.2 Usage in localhost

1. `git clone --recursive -b devel https://github.com/SantanderMetGroup/ansible-thredds-clu`
2. `cd ansible-thredds-cluster/scenarios/zequi/tds_standalone`
3. `./run.sh -h`
4. Deploy the TDS instance `./run.sh -r /tmp/sandbox deploy`
5. (in web browser) `localhost:8080/thredds`
6. After adding new content to the `data` directory and perform only the update of the catalogs: `./run.sh -r /tmp/sandbox -u deploy`
7. Stop all the processes: `./run.sh -r /tmp/sandbox stop`
8. Start all the processes: `./run.sh -r /tmp/sandbox boot`

5.2.3 Usage in vagrant

Vagrantfile is provided if you want to test the deployment without polluting your system. The default ip for the virtual machines are `192.168.50.10` and `192.168.50.11`.

1. `vagrant up`
2. `vagrant ssh ubuntu` or `vagrant ssh centos`
3. `./test.sh`

5.2.4 Scenario's variables

- root: Default is "/tmp/sandbox"
- catalogs_path: Default is "data/catalogs"
- datasets_path: Default is "data/datasets"