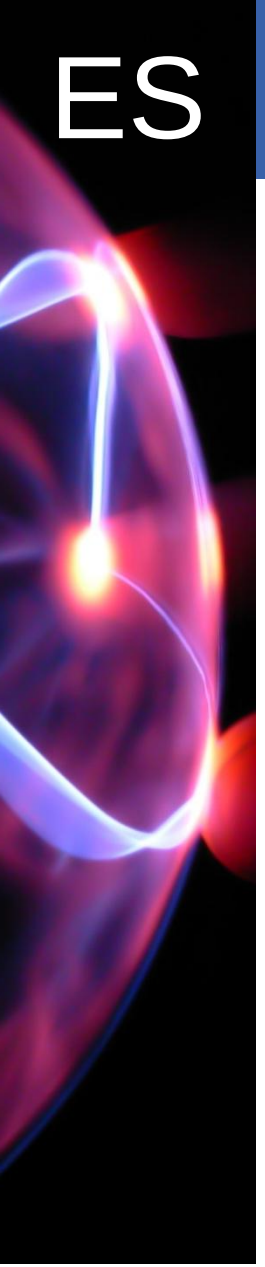
Web Redirector Tutorial for VOCs

Flavia Donno
CERN/IT-ES-VOS

Introduction



- ***Goal***: to provide enough background information to enable VOCs administer and operate the experiment specific Web Redirector

- Basic Concepts
- The web redirector service
- Web services configuration recommendations
- Rpms and quattor templates
- The user's perspective

CERN IT Department
CH-1211 Geneva 23
Switzerland
www.cern.ch/it

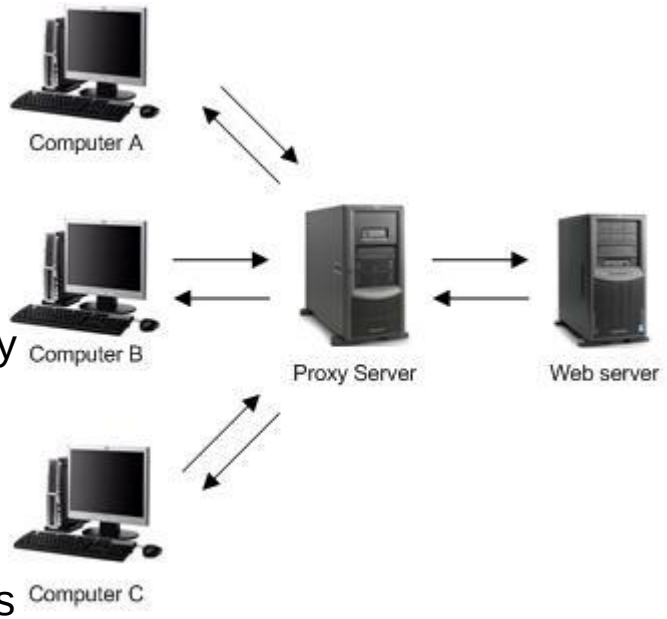
Basic Concepts

CERN IT Department
CH-1211 Geneva 23
Switzerland
www.cern.ch/it

Basic Concepts

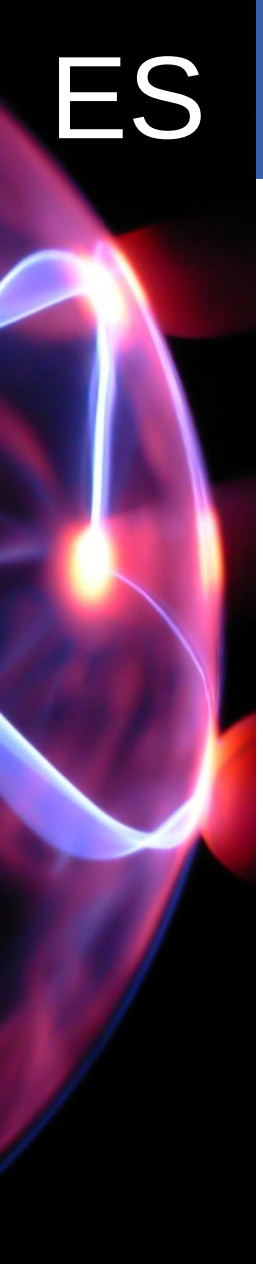
- ***Proxy Server***: is a server (a computer system or an application program) that acts as an intermediary for requests from clients seeking resources from other servers.

- A client connects to the proxy server, requesting some service, such as a file, connection, web page, or other resource, available from a different server.
- The proxy server evaluates the request according to its filtering rules.
- If the request is validated by the filter, the proxy provides the resource by connecting to the relevant server and requesting the service on behalf of the client.
- A proxy server may optionally alter the client's request or the server's response, and sometimes it may serve the request without contacting the specified server (cache).



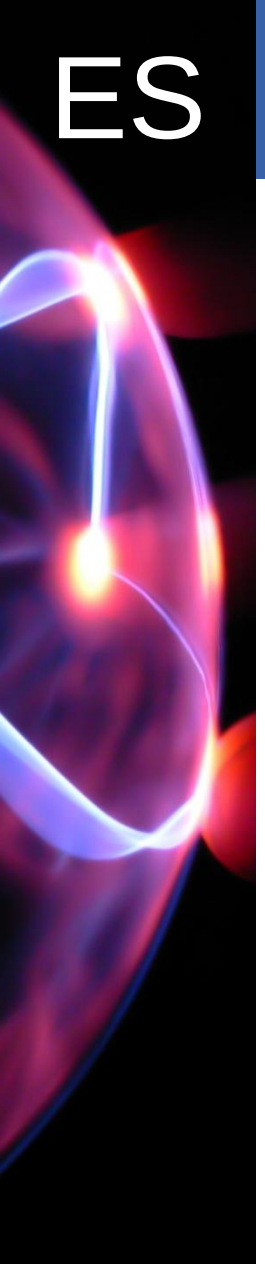
CERN IT Department
CH-1211 Geneva 23
Switzerland
www.cern.ch/it

Basic Concepts



- The purposes of a Proxy Server:
  - To keep machines behind it anonymous (mainly for security).
  - To speed up access to resources (using caching). Web proxies are commonly used to cache web pages from a web server.
  - To apply access policy to network services or content, e.g. to block undesired sites.
  - To log / audit usage, i.e. to provide company employee Internet usage reporting.
  - To scan transmitted content for malware before delivery.
  - To scan outbound content, e.g., for data leak protection.
  - To circumvent regional restrictions.

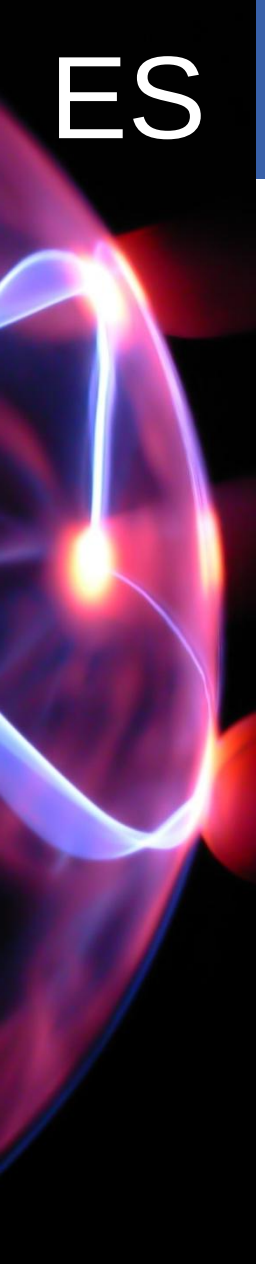
Basic Concepts



- ***Web Proxy***: A proxy that focuses on World Wide Web traffic.

- ***Reverse Proxy***: is (usually) an Internet-facing proxy used as a front-end to control and protect access to a server on a private network, commonly also performing tasks such as load-balancing, authentication, decryption or caching.

Basic Concepts



- ***Forward proxy*** acts as a proxy for out-bound traffic. For example, an ISP may use a proxy to forward HTTP traffic from its clients to external web servers.

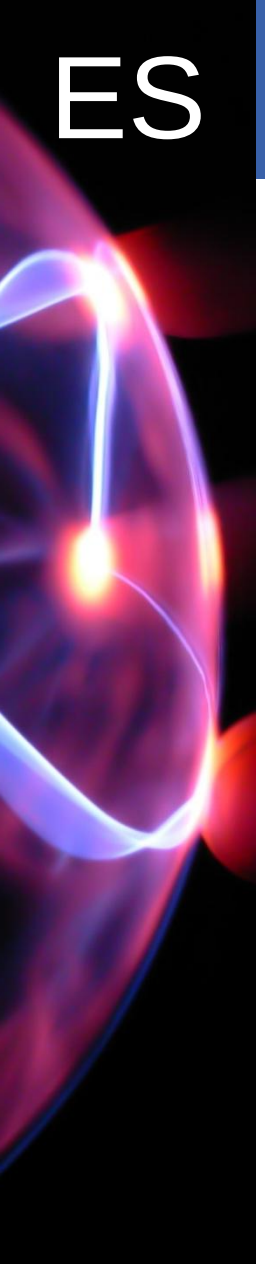
- ***Web Application Firewall*** *(WAF) is an appliance or software that provides customized protection for web applications against attacks.*

CERN IT Department
CH-1211 Geneva 23
Switzerland
www.cern.ch/it

The Web Redirector Service

CERN IT Department
CH-1211 Geneva 23
Switzerland
www.cern.ch/it

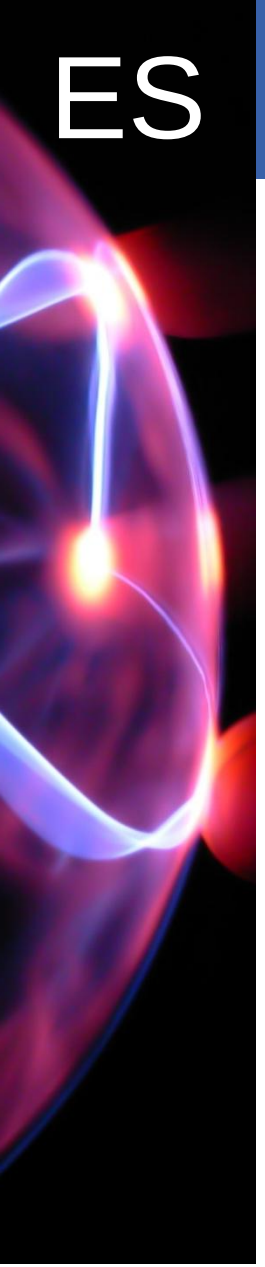
The Web Redirector: what is it?



- The ATLAS/CMS/LHCB Web Redirector is *proxy server* that act as a *reverse proxy* and as *web application firewall*.
- It is used in front of other ATLAS/CMS/LHCB Web services in order to mitigate potential threats coming from the underlying network, managed and unmanaged clients and hosts, potential untrustworthy users.
- It runs on VOBOXes as described by the CERN/IT VOBOX SLA. The required OS is SLC5.

CERN IT Department
CH-1211 Geneva 23
Switzerland
www.cern.ch/it

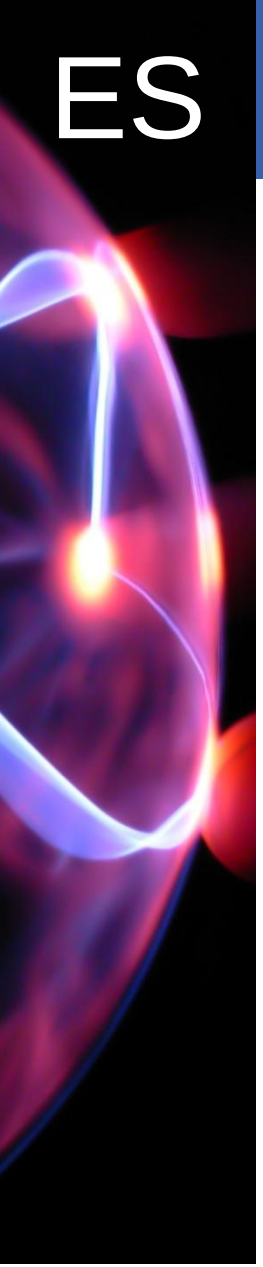
The Web Redirector: what is it?



- All connections coming from the Internet addressed to one of the ATLAS/CMS/LHCB Web servers are routed through the Web Redirector via a DNS alias (Virtual Host).

- The Web Redirector may either deal with the request itself or pass the request wholly or partially to the main ATLAS/CMS/LHCB web server/s dealing with the request.

The Web Redirector technologies



- The Web Redirector provides:
  - WAF through the *Apache ModSecurity* [3] module.
  - SSL based authentication. The CERN Single Sign On (SSO) *Shibboleth* service is used for this purpose.
  - Load distribution. Requests can be served by several ATLAS/CMS/LHCB web servers, each serving the same or its own application. Load distribution is achieved through the *mod_proxy_balancer*.
  - Caching support. The reverse proxy can offload the web servers behind it by caching static content through the Apache *mod_cache* and the *frontier-squid* server.
  - Support for special configurations: *AJP* protocol for Tomcat-based applications, customized redirection through *rewrite rules*; *session-aware forwarding*; *kerberos-aware sessions*.
  - Hardware sparing by supporting *virtualization*.
  - Web analytics through *awstats* and *webalizer*.

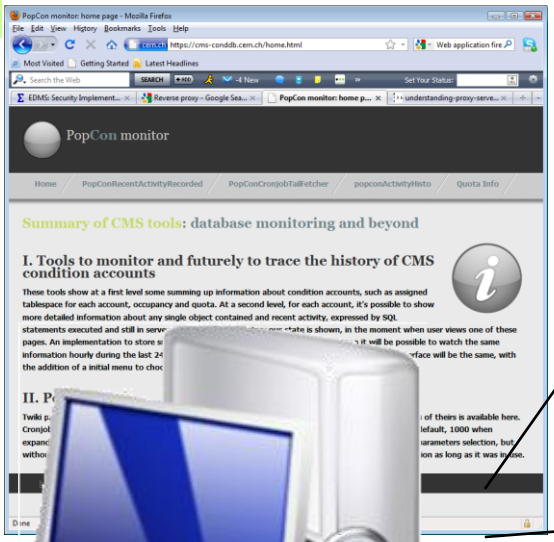
Supported configurations

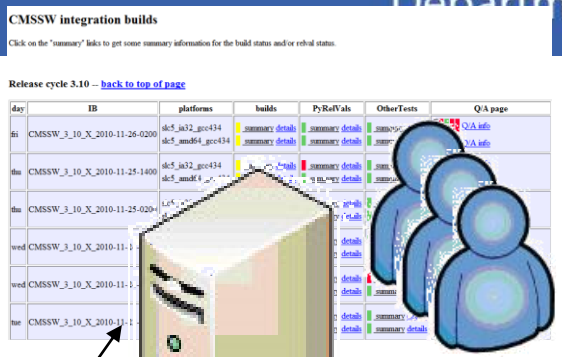
- Three configurations are at the moment supported:
  - The Web Redirector is hosted on the same machines running some of the experiment specific Web services.
  - The Web Redirector is hosted on a physical machine that hosts no other services.
  - The Web Redirector is hosted on a virtual machine as well as other experiment specific web services and they can run on the same or different hardware.

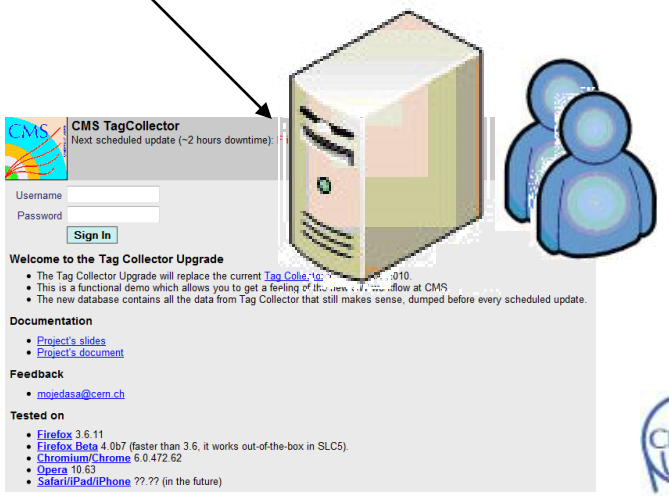
ES Supported Configurations

https://cms-conddb.cern.ch/

https://cmssdt.cern.ch/







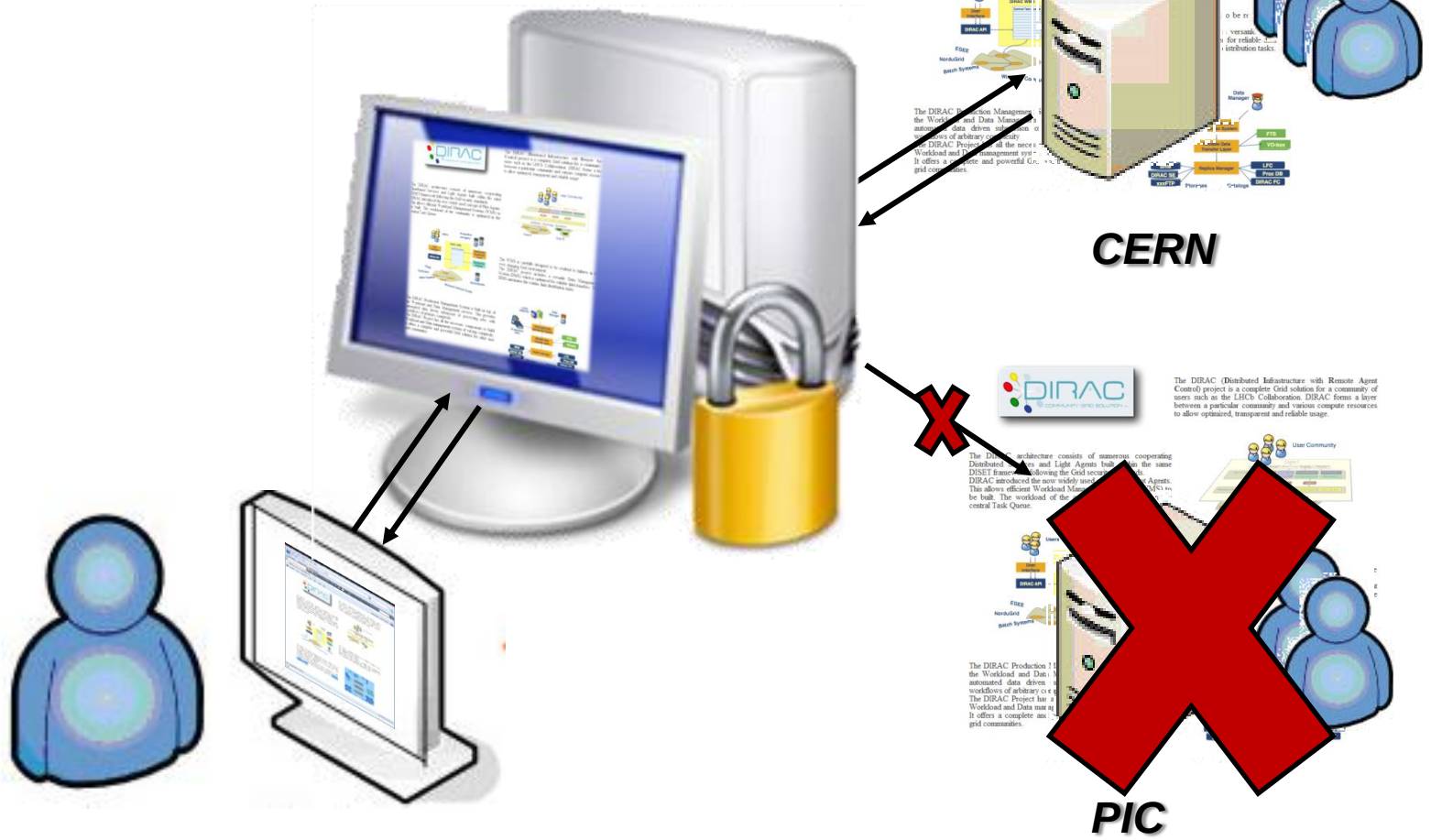
External clients only see the canonical web redirector URL

CERN IT Department
CH-1211 Geneva 23
Switzerland
www.cern.ch/it

ES Supported Configurations

CERN IT Department

http://lhcbweb-dev.cern.ch/



CERN IT Department
CH-1211 Geneva 23
Switzerland
www.cern.ch/it

Web Service configuration recommendations

CERN IT Department
CH-1211 Geneva 23
Switzerland
www.cern.ch/it

Web Services organization

- Each service running on or behind the web redirector is described by a service documentation card:
  - https://twiki.cern.ch/twiki/bin/view/LCG/WLCGWRSDC
- Services behind the web redirector run under a non-privileged account.
  - In particular 2 accounts per service should be used: an administrative account to customize the service and an account used to run the service.
  - Logins are disabled for the account running the service. Service managers can sudo to these accounts.

Services organization

- Directory structure:

```
[root@vocms118 ~]# cd /var/vhost/
[root@vocms118 vhost]# ls -al
total 40
drwxr-xr-x  5 root     root 4096 Mar 31  2010 .
drwxr-xr-x 27 root     root 4096 Nov 26 13:39 ..
drwxr-xr-x 11 cmscdadm zp   4096 Jun  8 16:51 cms-conddb
drwxr-xr-x  9 cmsrqadm zp   4096 Jun 10 17:07 cms-reqman
drwxr-xr-x  3 root     root 4096 Apr  1  2010 cmssdt
[root@vocms118 cms-conddb]# cd ../cms-reqman/
[root@vocms118 cms-reqman]# ls -al
total 84
drwxr-xr-x 9 cmsrqadm zp       4096 Jun 10 17:07 .
drwxr-xr-x 5 root     root     4096 Mar 31  2010 ..
drwxr-xr-x 2 cmsrqadm zp       4096 Jun 10 16:48 cert
drwxr-xr-x 2 cmsrqadm zp       4096 Jun 11 14:54 conf
drwxr-xr-x 2 cmsrqadm zp       4096 Jun 11 15:27 conf.d
drwxr-xr-x 2 cmsrqadm zp       4096 Jun 10 16:30 html
drwxr-xr-x 2 cmsrqsrv cmsrqsrv 4096 Nov 21 04:02 logs
lrwxrwxrwx 1 root     root       18 Jun 10 16:48 modules -> /etc/httpd/modules
drwxr-xr-x 2 cmsrqsrv cmsrqsrv 4096 Sep 20 13:06 run
drwxr-xr-x 5 cmsrqadm zp       4096 Jun 11 15:50 secure
-rw-r--r-- 1 cmsrqadm zp        830 Jun 10 16:30 sysconfig-httpd
[root@vocms118 cmssdt]# pwd
/var/vhost/cmssdt
[root@vocms118 cmssdt]# ls -al cert
total 40
drwxr-xr-x 2 root root 4096 Apr  1  2010 .
drwxr-xr-x 3 root root 4096 Apr  1  2010 ..
-rwxr-xr-x 1 root root  670 Apr  1  2010 newkey.sh
-rw-r----- 1 root root 1326 Apr  1  2010 server.crt
-rw-r----- 1 root root  887 Apr  1  2010 server.key
```

CERN IT Department
CH-1211 Geneva 23
Switzerland
www.cern.ch/it

Services organization

- Store sensitive files in SINDES through the vobox_sindes script and quattor component
- Have pre/post configuration scripts
- Have clear stop/start/draining procedures
  - This is automatically provided by web redirector infrastructure in case of services hosted within the web redirector infrastructure (apache stop/start)

```
[voatlas47] ~ $ /etc/rc.d/init.d/httpd-atlas-runquery status
httpd-atlas-runquery (pid  23669) is running...
```

The Web Redirector Service configuration: quattor templates and rpms

CERN IT Department
CH-1211 Geneva 23
Switzerland
www.cern.ch/it

Default setup

- The default configuration (out of the box) provides:
  - Basic software configuration.
  - Classic redirection to one or multiple services.
  - SSO configuration.
  - Webalizer and AWSTATS configuration.

Recommendations

- DNS aliases should follow the convention <vo>-<service_name>.cern.ch.
- Web redirector software and service configuration files must come in rpms stored in experiment specific quattor repositories.
  - IT/ES provides tools to automate the generation of such rpms

Quattor templates

- At the moment, one web redirector template per VO
  - *prod/customization/lhcb/webredirector/software*
    - Example volhcb16
  - *prod/customization/cms/webredirector/software*
    - Example vocms118
  - *prod/customization/atlas/shared_web/software*
    - Example voatlas53
- In the future, common template served from prod/services tree as already done for frontier/squid

CERN IT Department
CH-1211 Geneva 23
Switzerland
www.cern.ch/it

Web Redirector rpms

- Check out the shared_web module from SVN in your own private directory (/<myworkingdirectory>)

```
cd /<myworkingdirectory>
svn checkout svn+ssh://svn.cern.ch/reps/adcops/shared_web
cd shared web
```

- Copy the file service-test.conf into the file <vo>-<servicename>.conf and define related variables
- Configure rpm macros so that rpms can be created in your working directory.

```
cat ~/.rpmmacros
%_topdir /<myworkingdirectory>/rpm

for d in _topdir _sourcedir _rpmdir _srcrpmdir _specdir _builddir _tmppath; do
  mkdir -p `rpm --eval %{$d}`
done
```

Web Redirector rpms

- Run the command:.
  - /new-service.sh <vo>-<servicename>.conf
- 2 rpms are created in /<myworkingdirectory>/rpm/RPMS/noarch.
  - vhost-vo-service-<version>.noarch.rpm
    - Only needed if service hosted on the same machine as the web redirector
  - vhost-wr-service-<version>.noarch.rpm
    - To be installed on the web redirector machine.
- Put these rpms into the experiment specific quattor rpm repository.

Service-test.conf


```
# Service_Test.conf

VHOST_SHORT_NAME=atst

#If you like to use existing usernames - please specify them below:
#VHOST_SERVER_USERNAME=stsesrv
#VHOST_ADMIN_USERNAME=stseadm
#

VHOST_SERVICE_NAME=vo-service

# Service port number used on VHOST_HOST_NAME
VHOST_HTTP_PORT=12921

#Exact DNS alias of service
#Default $VHOST_SERVICE_NAME.cern.ch
#VHOST_DNSALIAS_NAME=vo-service.cern.ch

VHOST_FULL_NAME="CERN VO Test Service"

#HTTPS port to be used for service
#Default $(( VHOST_HTTP_PORT + 1 ))
#VHOST_HTTPS_PORT=8243

#You may set folder where service files will be located
#Default is /var/vhost/$VHOST_SERVICE_NAME
#VHOST_PLACE=/var/serices/$VHOST_SERVICE_NAME
#The following example shows how it is possible to put both development
#and production versions of services into the same folder
#VHOST_PLACE=/var/vhost/`echo $VHOST_SERVICE_NAME|sed 's/-dev//g'`


#You may change location of secure and nonsecure files independently
#Defaultlocation is $VHOST_PLACE/html and $VHOST_PLACE/secure
#VHOST_HTTP_ROOT=/var/vhost/$VHOST_SERVICE_NAME/www
#VHOST_HTTPS_ROOT=$VHOST_HTTP_ROOT

#You may change version and release numbers for rpm
#VHOST_SERVICE_RELEASE=2
#VHOST_SERVICE_VERSION=0.3
```

An example of a web redirector

- Let us look at one example of the quattor configuration for a (VM) machine that only hosts a web redirector

- voatlas46

Special Rewriting rules

```
#
# Disable TRACE
#
TraceEnable off
#
# This is the configuration file for the load balancer
#
<VirtualHost *:80>
   ServerName lhcbweb-dev.cern.ch

   ProxyRequests Off
   ProxyStatus On
   ProxyPreserveHost On
   ProxyVia On

   <Proxy lhcbweb-dev.cern.ch>
      Order deny,allow
      Allow from all
   </Proxy>

   ProxyPass /balancer-manager !
#   ProxyPass / balancer://lhcbweb/ stickysession=BALANCEID nofailover=On
   ProxyPass / balancer://lhcbweb/ nofailover=On
   ProxyPassReverse / balancer://lhcbweb/
#   ProxyPassReverse / http://volhcb06.cern.ch/
   ProxyPassReverse / http://lhcbweb.pic.es/
   ProxyPassReverse / http://volhcb12.cern.ch/
   <Proxy balancer://lhcbweb>
#      BalancerMember http://volhcb06.cern.ch route=pic loadfactor=1
      BalancerMember http://lhcbweb.pic.es route=pic loadfactor=8
      BalancerMember http://volhcb12.cern.ch route=cern loadfactor=1
      ProxySet lbmethod=byrequests
   </Proxy>

   DocumentRoot /var/www/lhcbweb

   <Location /balancer-manager>
      SetHandler balancer-manager
      Order deny,allow
      Allow from all
   </Location>

</VirtualHost>
```

volhcb16

/etc/httpd/conf.d

httpd-proxy-balancer.conf

CERN IT Department
CH-1211 Geneva 23
Switzerland
www.cern.ch/it

ES Special Rewriting rules

CERN IT Department

```
<VirtualHost *:443>
   ServerName cmssdt.cern.ch

   SSLEngine On
   SSLProxyEngine On
   SSLProtocol all -SSLv2

   #AB SSLCipherSuite ALL:!ADH:!EXPORT:!SSLv2:RC4+RSA:+HIGH:+MEDIUM:+LOW
   SSLCipherSuite HIGH:MEDIUM:-LOW:-SSLv2


   # This secures the server from being used as a third party proxy server
   ProxyRequests Off

   ProxyVia On

   SSLCertificateFile /var/vhost/cmssdt/cert/server.crt
   SSLCertificateKeyFile /var/vhost/cmssdt/cert/server.key

   ProxyPass /SDT http://vocms06.cern.ch:80/SDT
   ProxyPassReverse /SDT http://vocms06.cern.ch:80/SDT

   ProxyPass /dev http://vocms117.cern.ch:80/dev
   ProxyPassReverse /dev http://vocms117.cern.ch:80/dev

   ProxyPass /controllers http://cmsperfvm5.cern.ch:8085/controllers
   ProxyPassReverse /controllers http://cmsperfvm5.cern.ch:8085/controllers

   ProxyPass /qa/perfmondb http://cmsperfvm5.cern.ch:8085
   ProxyPassReverse /qa/perfmondb http://cmsperfvm5.cern.ch:8085

   ProxyPass /tcdev https://cmsntcdev.cern.ch
   ProxyPassReverse /tcdev https://cmsntcdev.cern.ch

   ProxyPreserveHost On
```

vocms118

/etc/httpd/conf.d

cmssdt.conf

```
   ### for Computer.Security @ CERN
   RewriteEngine on
   RewriteCond %{REQUEST_METHOD} ^(TRACE|TRACK)
   RewriteRule .* - [F]
   RewriteCond %{REQUEST_URI} ^/qa/perfmondb
   RewriteRule ^/qa/perfmondb$ /qa/perfmondb/ [NE,R,L]
   RewriteCond %{REQUEST_URI} ^/SDT
   RewriteRule ^/SDT$ /SDT/ [NE,R,L]
   RewriteCond %{REQUEST_URI} ^/dev
   RewriteRule ^/dev$ /dev/ [NE,R,L]
   RewriteCond %{REQUEST_URI} ^/controllers
   RewriteRule ^/controllers$ /controllers/ [NE,R,L]
   RewriteCond %{REQUEST_URI} ^/tcdev
   RewriteRule ^/tcdev$ /tcdev/ [NE,R,L]
   #RewriteRule ^.*$ https://cmsntcdev.cern.ch [NE,L]
   RewriteCond %{REQUEST_URI} !^/dev/.*
   RewriteCond %{REQUEST_URI} !^/SDT/.*
   RewriteCond %{REQUEST_URI} !^/tcdev/.*
   RewriteCond %{REQUEST_URI} !^/qa/perfmondb/.*
   RewriteCond %{REQUEST_URI} !^/controllers/.*
   RewriteCond %{REQUEST_URI} !/shibboleth-sp.*
   RewriteCond %{REQUEST_URI} !/Shib*
   RewriteRule ^/(.*)$ /SDT/$1 [NE,R,L]

   <Location /SDT>
       AuthType shibboleth
       ShibRequireSession On
       require valid-user
   </Location>

   <Location /qa/perfmondb>
       AuthType shibboleth
       ShibRequireSession On
       require valid-user
   </Location>

   <Location /controllers>
       AuthType shibboleth
       ShibRequireSession On
       require valid-user
   </Location>

   <Location /dev>
       AuthType shibboleth
       ShibRequireSession On
       require valid-user
   </Location>

</VirtualHost>
```

CERN IT Department
CH-1211 Geneva 23
Switzerland
www.cern.ch/it

ES Special Configuration support

CERN IT Department

```
<VirtualHost *:443>
   ServerName atlas-ddv.cern.ch
   Alias      atlas-ddv atlas-ddv.cern.ch

   SSLEngine On
   SSLProtocol all -SSLv2

   #AB SSLCipherSuite ALL:!ADH:!EXPORT:!SSLv2:RC4+RSA:+HIGH:+MEDIUM:+LOW
   SSLCipherSuite HIGH:MEDIUM:-LOW:-SSLv2


   # This secures the server from being used as a third party proxy server
   ProxyRequests Off

   ProxyVia On

   SSLCertificateFile    /etc/grid-security/hostcert.pem
   SSLCertificateKeyFile /etc/grid-security/hostkey.pem

   ProxyPass / ajp://atlas-ddv.cern.ch:8080/DDV/
   ProxyPassReverse / ajp://atlas-ddv.cern.ch:8080/DDV/

#SB    ProxyPass / http://atlas-ddv.cern.ch:12922/
#      ProxyPassReverse / http://atlas-ddv.cern.ch:12922/

   ProxyPreserveHost On
   ProxyTimeout 360
   Timeout 360

   ### for Computer.Security @ CERN
   RewriteEngine on
   RewriteCond %{REQUEST_METHOD} ^(TRACE|TRACK)
   RewriteRule .* - [F]
   RewriteCond %{HTTP_HOST} !^atlas-ddv.cern.ch$ [NC]
   RewriteCond %{HTTP_HOST}   !^$
   RewriteRule ^/(.*) https://atlas-ddv.cern.ch/$1 [L,R]
   RewriteCond %{REQUEST_URI} ^/$
   RewriteRule ^/$ /DDV.html [NE,L,R]

   <Location />
       AuthType shibboleth
       ShibRequireSession On
#       ShibRequireAll On
#       ShibExportAssertion Off
#       ShibUseHeaders On
       require valid-user
       #for e-groups:
       #Require adfs-group "atlas-readaccess-active-members", "atlas-external-operation"
   </Location>

</VirtualHost>
```

Support for AJP for Tomcat based applications

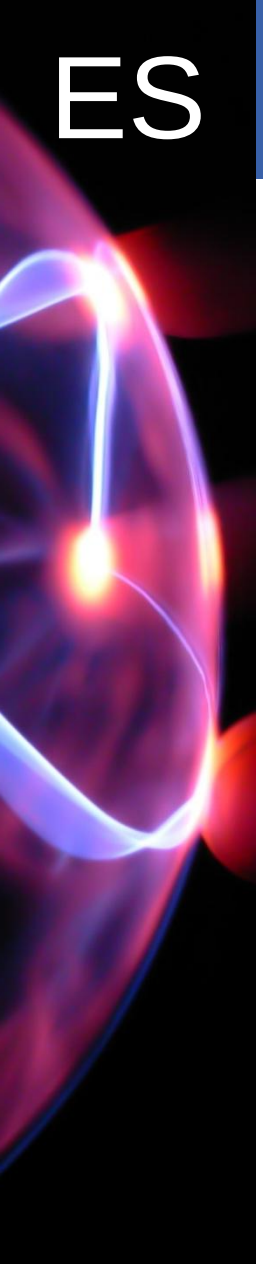
Authentication headers

voatlas53

/etc/httpd/conf.d

atlas-ddv.conf

CERN IT Department
CH-1211 Geneva 23
Switzerland
www.cern.ch/it



Configuring Shibboleth2

CERN IT Department

- Copy the file /etc/shibboleth/shibboleth2.xml from one of existing web redirectors (vocms118, volhcb16)
  - Remove unnecessary entries, change node names, add new entries for the new service.
- Copy the file /etc/shibboleth/ADFS-metadata.xml from one of existing web redirectors.
- Copy the file /etc/shibboleth/attribute-map.xml from one of existing web redirectors.
- Restart shibd service
  - /etc/rc.d/init.d/shibd restart

This process will be automated and part of the web redirector default configuration

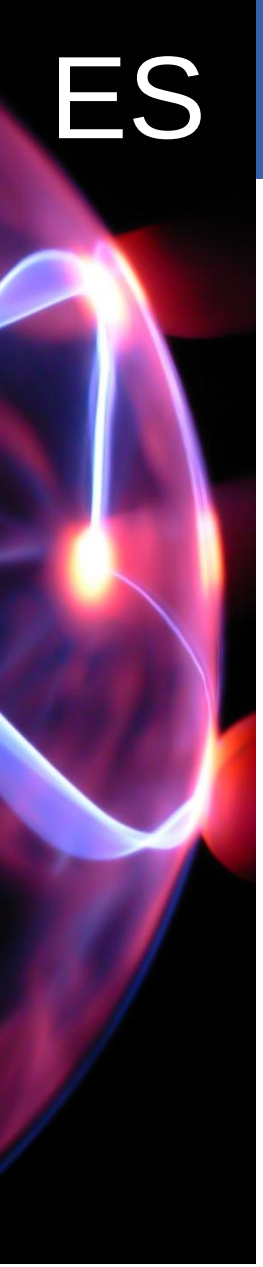
CERN IT Department
CH-1211 Geneva 23
Switzerland
www.cern.ch/it

The Web Redirector: user's perspective

CERN IT Department
CH-1211 Geneva 23
Switzerland
www.cern.ch/it

The user's perspective

- The web redirector supports basic configuration for Apache httpd-based services
- The service managers providing the WS can login on lxvoadm.cern.ch and from there to the machine (HWS) hosting the WS.
- On the HWS, they can use sudo to login into an administriation account for their WS. Through this administration account, service managers can edit the files relative to their service.



The user's perspective

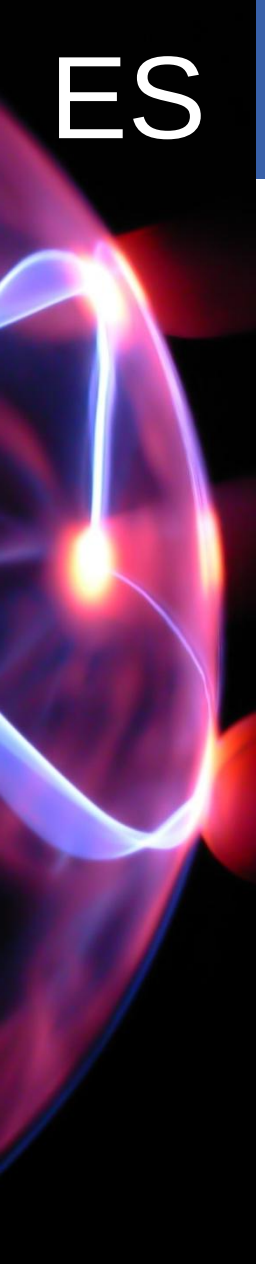
- The command to sudo into the administrative account is:
  - sudo -i -u <WSadminaccount>
- This account has its home directory in ***/data/<WSadminaccount>*** and can write into ***/var/vhost/<VO>-<WS>*** where the configuration, files, scripts of the WS are located.

CERN IT Department
CH-1211 Geneva 23
Switzerland
www.cern.ch/it

The user's perspective

- By default the WS has 2 document roots:
  - ***/var/vhost/atlas-<AWS>/html*** for anonymous access
  - ***/var/vhost/atlas-<AWS>/secure*** for authenticated (Shibboleth2 SSO) access
- The WS can be started through init.d under the special service account. To start the WS the following command can be used:
  - /sbin/service httpd-<VO>-<WS> start
- To stop the service or to check its status you need to use respectively the *stop* or *status* parameters instead of *start*.

The user's Apache modules

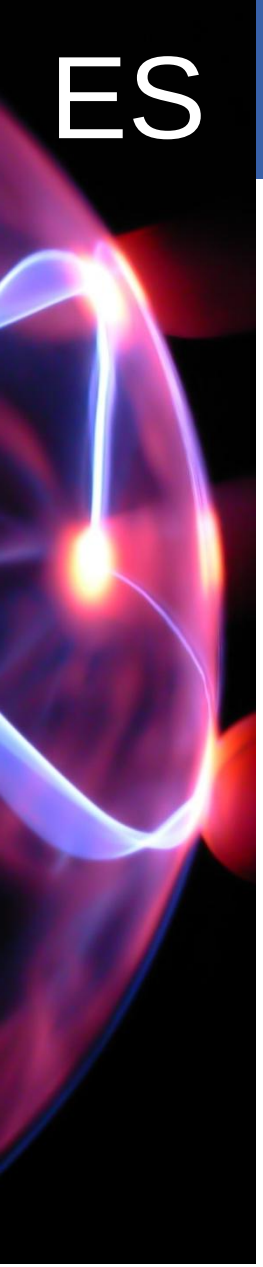


- If needed special apache modules like mod_python, mod_wsgi and others can be used for a specific WS. To do so the service manager must copy the corresponding .conf file from the main apache configuration directory:
  - cp /etc/httpd/conf.d/python.conf /var/vhost/<VO>-<WS>/conf.d/
- The AWS needs to be restarted as usual. After that the module is loaded and can be used.

Shibboleth2 registration

- Once the AWS is configured, it needs to be added to the allowed application list in CERN Single Sign On. To do so, simply go to this form and specify the following 3 items:
  1. Your **Application Name** (/not needed in Shib2 config ?: as declared in SessionInitiator property./). For our application we need to specify atlas-<AWS> even though for Shib2 this is not needed.
  2. Your application **URL**, as declared in saml:Audience property. (In our case this is: https://<VO>-<WS>.cern.ch/Shibboleth.sso/ADFS
  3. Your **name** and **email** for further contact.
- Once you do that, Shibboleth2 can be enabled for the service WS through the web redirector.

Conclusions



- We have introduced the basic concepts about proxy servers
- The web redirector is a proxy server that provides authentication, load distribution, caching, a web application firewall, web statistics, centralized support for special needs and better resources utilization.
- The installation and configuration of web redirector services has been automated providing for strong support for common practices and reducing management and operation effort.