SANSA Catalogue Documentation

Tim Sutton, 2008

Contents

1 Introduction

1.1 The CSIR

This project has been carried out for the CSIR Satellite Applications Center, near Johannesburg, South Africa. The CSIR is the 'Council for Science and Industrial Research' - it is the main national science foundation of the country. They are a big organisation with many divisions of which SAC (Satellite Applications Center) is one.

1.2 SAC

SAC is a satellite ground station. This means they have a big campus with many antennas and collect information from satellites as they pass over our sky window they also do satellite tasking (telling satellites where to go and what to do) and satellite / space craft telemetry (tracking space vehicle orbit information etc).

SAC has two divisions:

1. Telemetry command and control where they do tracking, tasking etc.
2. EO (Earth Observation) where the focus is more software based to do remote sensing and generate products from imagery downloaded from satellites

SAC-EO is the client for this project.

1.3 SANSA Takover

South Africa is busy creating its own space agency - SANSA (South AfricAN National Space Agency). SANSA will aggregate space technology from various gov, parastatal, non-gov organisations to form a new organisation funded by the state. SAC-EO is scheduled to become part of SANSA as of 1 April 2011 and wil become SANSA-EO.

1.4 The project

SAC-EO has been building for the last 3 or 4 years an integrated system before this project (of which we form a small part), the processing of imagery was done manually and ad-hoc which is not very efficient and prone to difficulty if an expert leaves.

Thus they have started to build an integrated system called SAEOS (pronounced 'sigh-os'). The purpose of SAEOS is to create an automated processing environment through all the steps of the EO product workflow i.e.:

- satellite tasking ('please programme spot5 to take an image at footprint foo on date X')
- image processing (level 1a through 3a/b)
- image analysis (level 4)

- image ordering ('can I please get a copy of that SPOT image you took on dec 4 2008 of this area')
- product packaging ('bundle up the stuff that was ordered using a DVD robot, placing on an ftp site, writing to an external HD etc')

To achieve this goal they have a number of software components.

The first components are the 'terminal software'. Terminal software are provided by satellite operators such as SPOT5 (I will use SPOT as an example a lot as its the pilot sensor for their project, eventually to incorporate many more sensors) The terminal software is typically a linux box with the operators own proprietary software on top that lets the operators do the tasking of satellites (to collect an image at a given place and time) and also to extract archived images from their tape library

The second component is 'SARMES'. SARMES is a collection of EASI scripts / routines. EASI is a programming language that runs on top of PCI / Geomatica a proprietary GIS tool that runs on windows and linux. SAC are busy porting SARMES to SARMES II which has the same functionality but uses python language bindings of PCI/Geomatica instead of EASI script. SARMES has all the logic to do things like:

- take a raw image and convert it to a common GIS format e.g. pix, gtiff etc.
- collect GCP's automatically using a reference image
- orthorectify an image using a dem, gcps and other reference data
- reproject the image into different coord systems (typically UTM 33S - UTM 36S in our area but others may apply too)
- perform atmospheric correction to remove effects of the stratosphere interference between lens and ground target
- perform sensor specific correction to e.g. remove effects of lens distortion on a specific camera (using published sensor models)
- perform mosaicking of images to create one big seamless colour corrected dataset
- perform pan sharpening (make a colour image higher resolution by merging it with a pan-chromatic / grey scale band)
- chop up images in various tile schemes (e.g. degree squares, quarter degree sqauares etc)

These jobs are run by manual process - creating config files, placing input files in a specific dir heirachy etc.

The third component is DIMS. DIMS is a software system running on top of linux written in java, corba, and using oracle or postgresql as a backend (at SAC they are using PostgreSQL). DIMS is proprietary software written by a german company called

WERUM. The same software is used by the German Space Agency and others. DIMS provides automated tool chain processing. Basically you set up work flows and run them using an 'operating tool'. Although DIMS uses postgresql, there is no third party access to that db and the whole system should be considered a black box except for a few specific entry and exit points.

DIMS is being extended and customised for SAC-EO including modifications so it wll provide ogc interfaces. Before this had their own catalogue implementation and ordering system using very old standards or proprietary interfaces. So DIMS can process EO data and it builds up a catalogue of products that it has processed or 'knows about' - in its own silo. This catalogue is / will be accessible via CSW and for processing of ordering they are implementing the OGC

The OS4EO (ordering service for earth observation) is an ogc standard. The OS4EO standard is pretty simple and familiar. In essence it allows you to:

- get capabilities
- get quote
- place order

In DIMS it is implemented using SOAP rather than a RESTful service.

Along the process of creating the SAEOS project, SAC-EO have also been investing in high end hardware - particularly storage. They have a petabyte capable heirachical storage system that in short works as follows:

- data is written to local hard drives
- after a certain period of inactivity moved down to slower sata drives (nearline storage)
- and after that its migrated down onto a tape library

The tape library (offline storage) is treated as part of the file system. It has a robot arm that loads tapes automatically. When you browse the file system, it appears that all data is local since all inodes are present in online storage. When you try to read a file that is offline, the robot fetches if from tape and puts it online - typically in under a minute, though that depends on system load.

DIMS is integrated with this file system (this file system is SGI's HFS - Heirachical File System). HFS is also proprietary software running on top of Linux. One of the things DIMS will be doing is de-archiving from old manually loaded tapes and moving them into HFS. De-archiving historically collected raw satellite imagery that is. When DIMS is finished going through that there will be hundreds of thousands (probably millions) of raw images stored in HFS and accessible via DIMS.

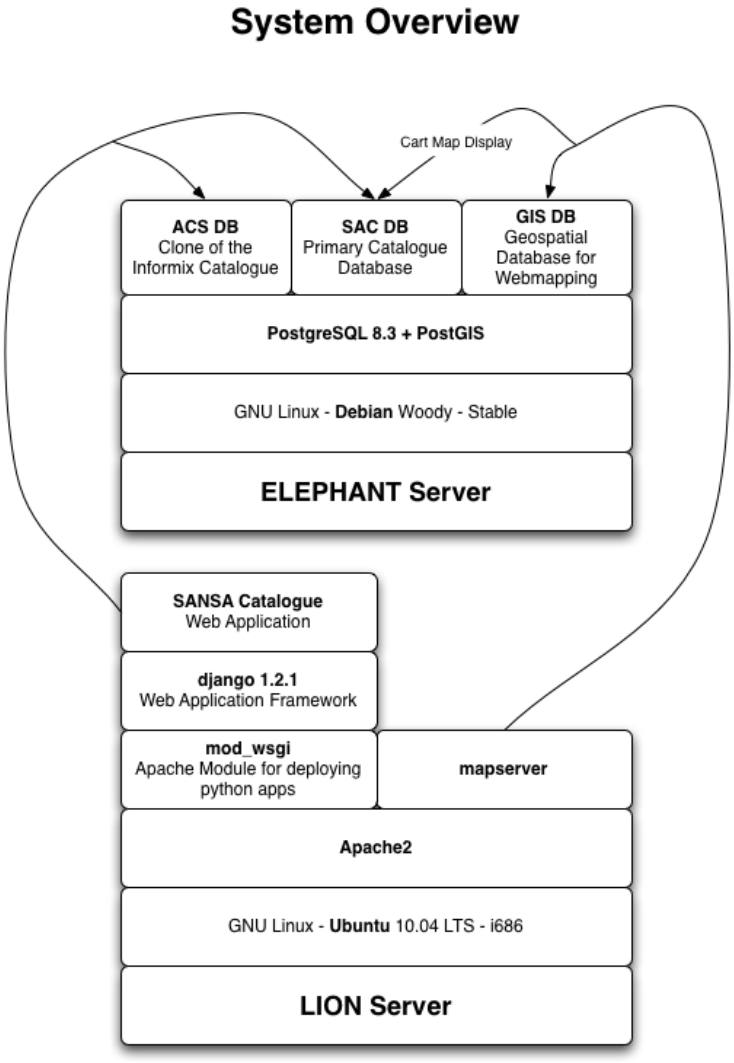
Since DIMS integrates with SARMES so you can do things like:

"Pull out that landsat 5 image from 2002, orthorectify it, correcto for atmospheric interference and lens distortions, reproject it to UTM 35S and clip it to this bounding box, then place the product on a dvd and write this label on it"

Thats the goal of the system - end to end automation with minimal operator intervention.

1.5 The Online Catalogue

Along side of these other packages, Linfiniti has been building a new web catalogue for SAC-EO. The catalogue is django + postgresql + all the other great FOSS tools we can use together to make a rich, interactive site.



Work on this started before any dims software was installed on site and the first major task was to migrate data and thumbnailss from a legacy, proprietary informix catalogue and get it into postgresql.

We also went through a few refactorings since the first version was almost a direct clone of the data model from the legacy informix system. Our current version uses the

concept of a 'generic product' for the data model.

XXXXXXXXX Insert image here XXXXXXXXXXXXXXX

On the left / center part of the diagram you will see an entity name 'Generic product'. This is a generic representation of any product (e.g. optical, radar, atmospheric, derived (like landcover map) and in the future we want to tune this to cater for vector data too.

There are five inherited models:

XXXXXXXXXXX Todo explain inherited models XXXXXXXXXXXXXXXXx

There are a bunch of small tables that provide foreign key dictionaries for the terms described in the models including:

- tables relating to ordering and tasking
- search and search record models are used to store user searches
- temporary tables used during product imports

The Online catalogue has the capability to deliver some products directly if they are held on local storage and also some basic capabilities for visitors to submit tasking requests.

Every time a search is made, its remembered, and the records the user is shown may be added to a cart and then assigned to an order To get started, first add an entry like this to your ssh config file in ~/.ssh/config:

```
Host linfiniti2
  Port 8697
  HostName 188.40.123.80
  FallBackToRsh no
```

1.6 Checkout Sources

Then setup a working dir and check out the sources (you can adapt dirs / paths as needed, using these paths will keep you consistent with all setup notes but its not required).

```
cd /home
sudo mkdir web
sudo chown -R <username>.<username> web
cd web
mkdir sac
git clone git@linfiniti2:sac_catalogue.git sac_catalogue
```

Then follow the instructions in README, skipping sections on informix, building gdal from source and source code checkout (you already checked it out if you have the readme :-)

1.7 Load a database dump

A recent database dump can be obtained from:

`http://196.35.94.243/sac_postgis_01February2011.dmp`

2 Working with Git

Each devloper works on a remote branch, others can track a specific branch locally and try out implemented features. After approving implementation, branch is merged with HEAD. (possibly closed/removed from tree)

This commands are based on http://www.eecs.harvard.edu/~cduan/technical/git/

2.1 Getting a list of branches

For local branches do:

```
git branch -v
```

For remote branches do:

```
git branch -r -v
```

2.2 To create remote branch

For current versions of git (at least git 1.7 or better). Say we want to create a new branch called 'docs-branch':

```
git branch docs-branch
git push --set-upstream origin docs-branch
git checkout docs-branch
```

2.3 Working with a remote branch

To be able to work with a remote branch locally (if it already exists remotely), we must create local branch and setup tracking of remote branch.

```
git pull #your local repo must be up to date first
git branch --track new-branch origin/new-branch
git checkout new-branch
```

Now you can go on to do your work in that branch.

To pull changes from remote repo do:

```
git pull origin
```

2.4 Deleting branches

Once you are done with a branch, you can delete it. For a local branch do:

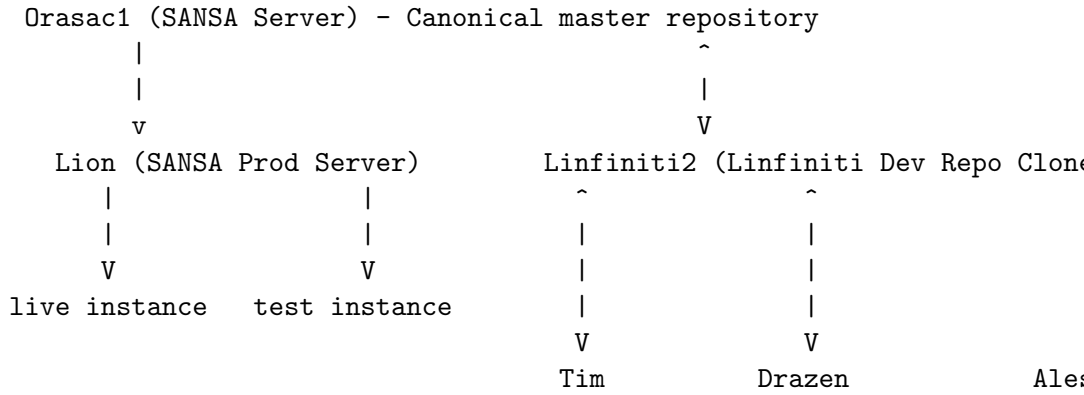
```
git branch -d new-branch
```

To delete a remote branch do (after first deleting it locally):

```
git push origin :new-branch
```

2.5 Distributed Git Repository Topology

The repositories are arranged like this:



The orasac master repo must pull from the linfiniti2 server at regular (e.g. weekly) intervals using a command like this:

```
cd /opt/git/sac_catalogue
git pull git@linfiniti2:sac_catalogue.git
```

If changes have happened on the SAC side and committed to the repository on orasac1, those changes should be pushed over to the catalogue on linfiniti2 so that the two repos are in sync:

```
cd /opt/git/sac_catalogue
git push git@linfiniti2:sac_catalogue.git
```

Note that orasac1 also has an entry in /home/timlinux/.ssh/config like this:

```
Host linfiniti2
  HostName 188.40.123.80
  User timlinux
  Port 8697
```

The lion live and test instances are cloned from the orasac1 repo like this:

```
git clone timlinux@orasac1:/opt/git/sac_catalogue sac_live
git clone timlinux@orasac1:/opt/git/sac_catalogue sac_test
```

The instance on linfiniti2 gitosis was cloned in the same way into /opt/git/repositories/sac_catalogue.

For the Tim / Drazen / Alessandro clones, the clone was carried out as described in the first section of this doc.

2.6 Tracking branches from linfiniti with a master checkout from orasac

In this scenario, we want to be tracking master from orasac1 but occationally pulling down branches from linfiniti2 to test them under lion:/opt/sac_catalogue/sac_test. Make sure you have a linfiniti2 entry in your ~/.ssh/config as described further up in this document.

“ git remote add linfiniti2 git@linfiniti2:sac_catalogue.gi git fetch linfiniti2

```
You should see something like the output below showing you that the branches from the se
```

The authenticity of host ’[188.40.123.80]:8697 ([188.40.123.80]:8697)’ can’t be established. RSA key fingerprint is cd:86:2b:8c:45:61:ae:15:13:45:95:25:8e:9a:6f:c4. Are you sure you want to continue connecting (yes/no)? yes Warning: Permanently added ’[188.40.123.80]:8697’ (RSA) to the list of known hosts. _ _ __ _ _ _ _

(_)_ __ / _(_)_ __ (_) |_(_)
|
’_ \| |_|
’_ \|
__|
|
|
| _|
|
|
|_|

|_|_|_| |_|_| |_|_| |_|_|__|_|

– Authorized Access Only – Enter passphrase for key ’/home/timlinux/.ssh/id_dsa’: remote: Counting objects: 201, done. remote: Compressing objects: 100% (150/150), done. remote: Total 150 (delta 103), reused 0 (delta 0) Receiving objects: 100% (150/150), 1.10 MiB | 47 KiB/s, done. Resolving deltas: 100% (103/103), completed with 28 local objects. From linfiniti2:sac_catalogue * new ale -> linfiniti2/ale * new ale_test -> linfiniti2/ale_test * new map_resize -> linfiniti2/map_resize * new master -> linfiniti2/master * new tim-model-refactor-off-ale -> linfiniti2/tim-model-refactor-off-ale
‘

Now we are ready to check out the branch from there e.g.:

git branch map_resize linfiniti2/map_resize git pull #not sure if needed git checkout map_resize sudo /etc/init.d/apache2 reload

When you want to get back to the original again do:

git checkout origin/master

= System logic and rules =

== Computation of geometric_accuracy_mean ==

The geometric accuracy of a product is calculated as the mean of its geometric_resolutio

The values for geometric_resolution_x, geometric_resolution_y will vary per sensor and pe

(SAC to provide required detail here).

---------- TABLE PLACEHOLDER -------------------

== Sensor viewing angle ==

The sensor viewing angle

= Updates and imports of products =

vim /mnt/cataloguestorage/thumbnail_processing/thumb_blobs/lastblob.txt

```
Set desired blob no in above file.

python manage.py runscript --pythonpath=scripts -v 2 acs_importer
%!include: 004-installation-guide.t2t
%!include: 201-reporting-tools.t2t
%!include: 901-informix-acs-sync.t2t
%!include: 902-importing-other-products.t2t
```