Calibration of low-frequency radio-astronomy data on cloud infrastructures

Jose Sabater Montes
Institute for Astronomy, University of Edinburgh

P. Best, S. Sanchez, J. Garrido, J. E. Ruiz, L. Verdes-Montenegro and the LOFAR collaboration

Outline

- LOFAR and ELAIS-N1
- Challenges
- Cloud solutions
  - Amazon Web services (SKA Astrocompute)
- Summary

LOFAR

- Low Frequency Array
- Software defined radio-interferometer working at low frequencies (30 to 240 MHz)
- One of the Square Kilometre Array pathfinders

LOFAR Stations

LOFAR Stations

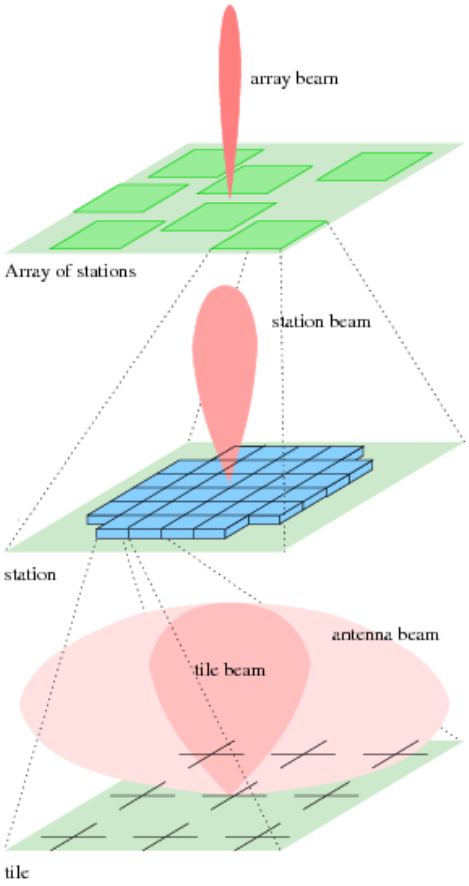
LOFAR frequencies

- LBA 30-80 MHz
- HBA 120-240 MHz

ELAIS-N1

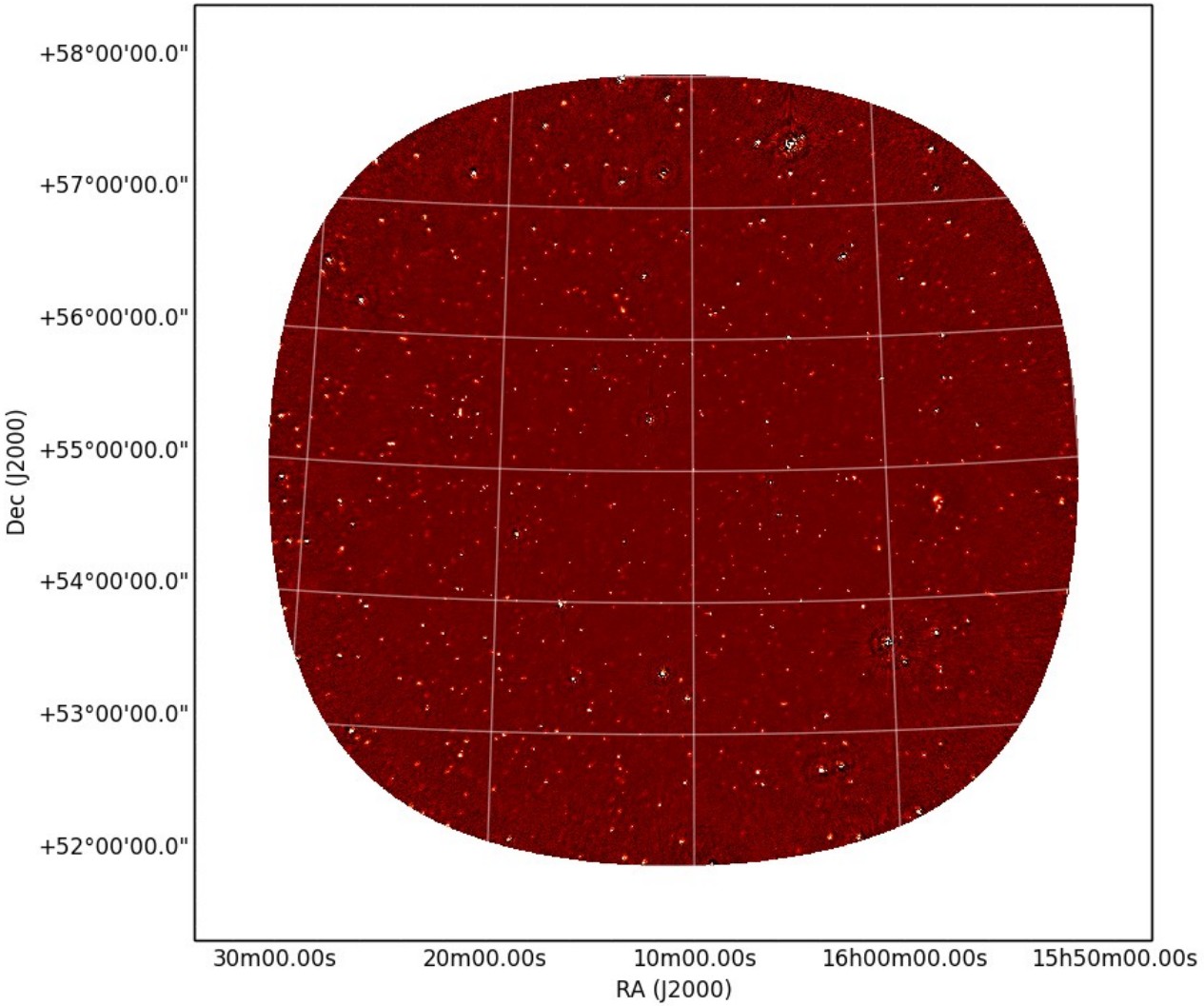
- Collaboration with Epoch of Reionization, magnetism and transients KSP.
- Public LOFAR deep field.
- Panchromatic data:
  - Spitzer, GMRT, suprime-cam, ALHAMBRA, etc.
  - SDSS BOSS spectrograph (4 plates).
- Observed: 260 hours observed so far → ~100 TB
  - Enough to reach 20 μJy rms. Aim 10 μJy.

LOFAR aperture synthesis



- field of view diameter of ~6 deg at 150 MHz
- resolution < 5 arcsec

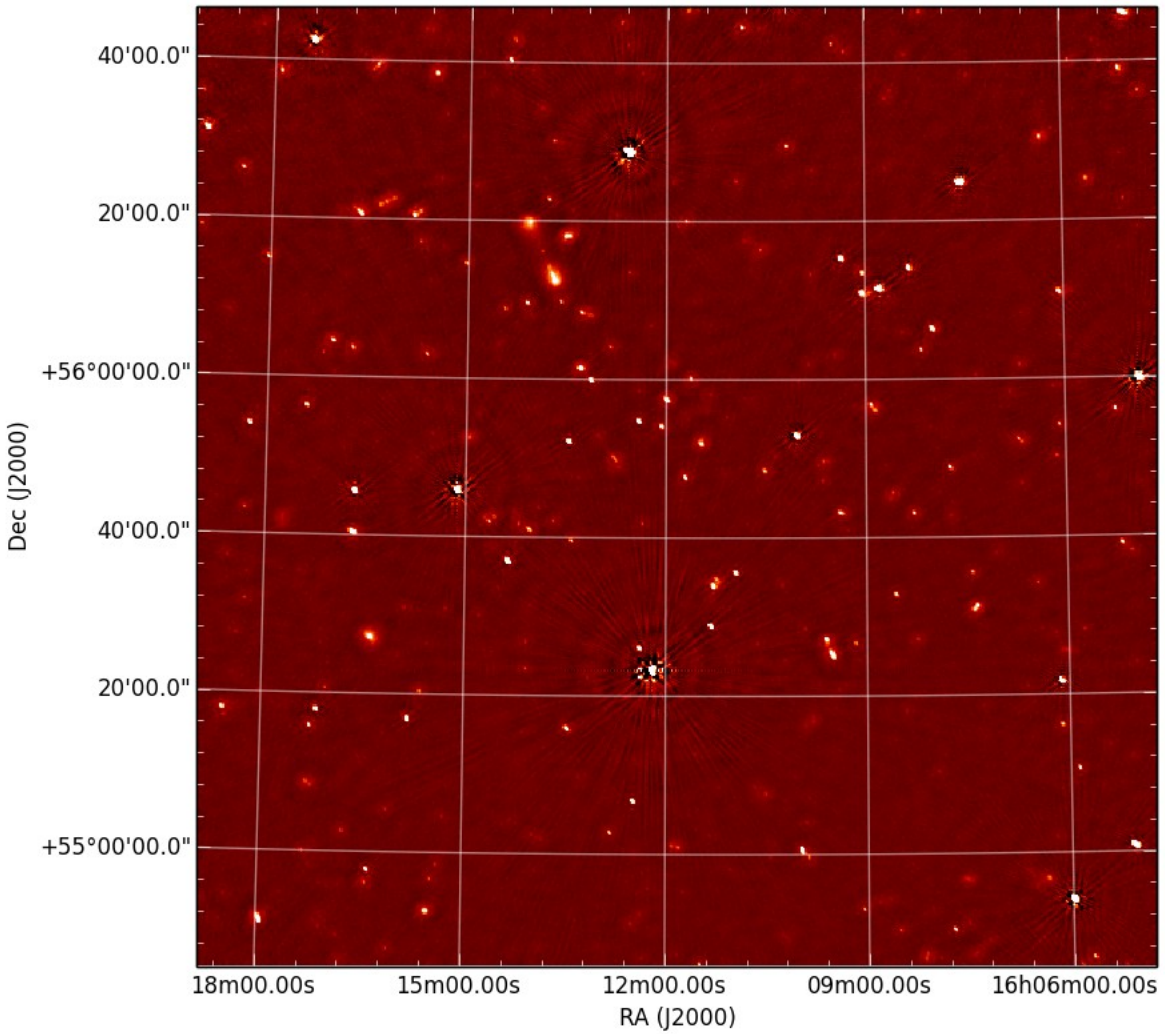
LOFAR imaging



r.m.s.
300 microJy
In 10 hours

Calibration on
IAA (Granada) cluster

LOFAR imaging



r.m.s.
300 microJy
In 10 hours

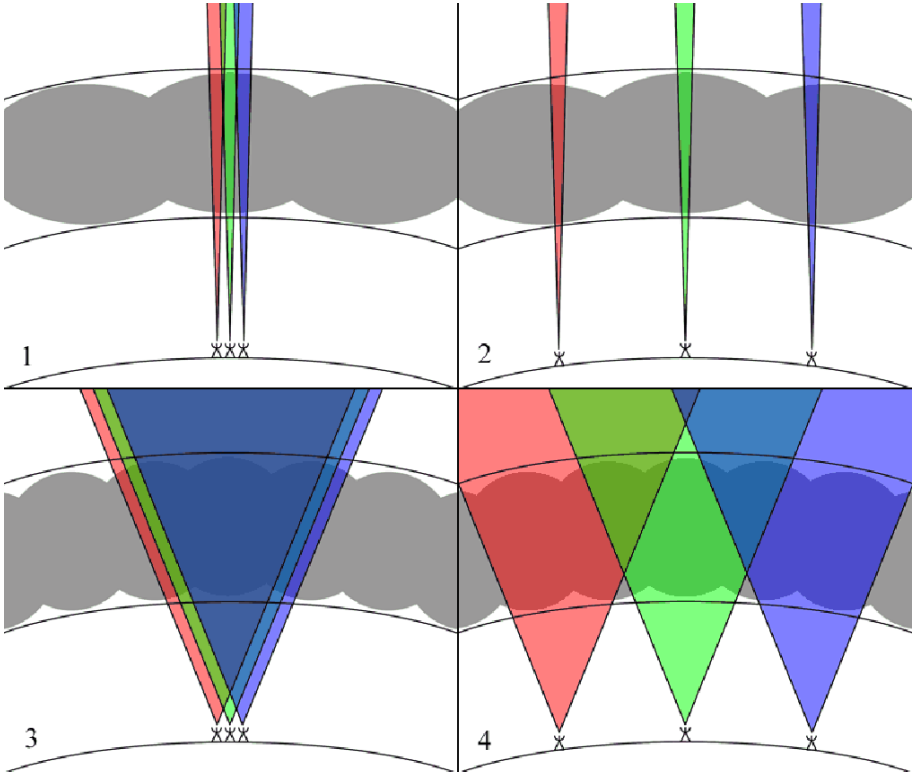
Calibration on
IAA (Granada) cluster

Challenges

- Effect of the ionosphere
- User data calibration
  - 10 hours full resolution → ~20 TB
  - Minimum of 2 CPU years to run the calibration
  - Experimental pipeline
- LOFAR calibration software
  - Difficult to install
  - Continuous development

Ionosphere

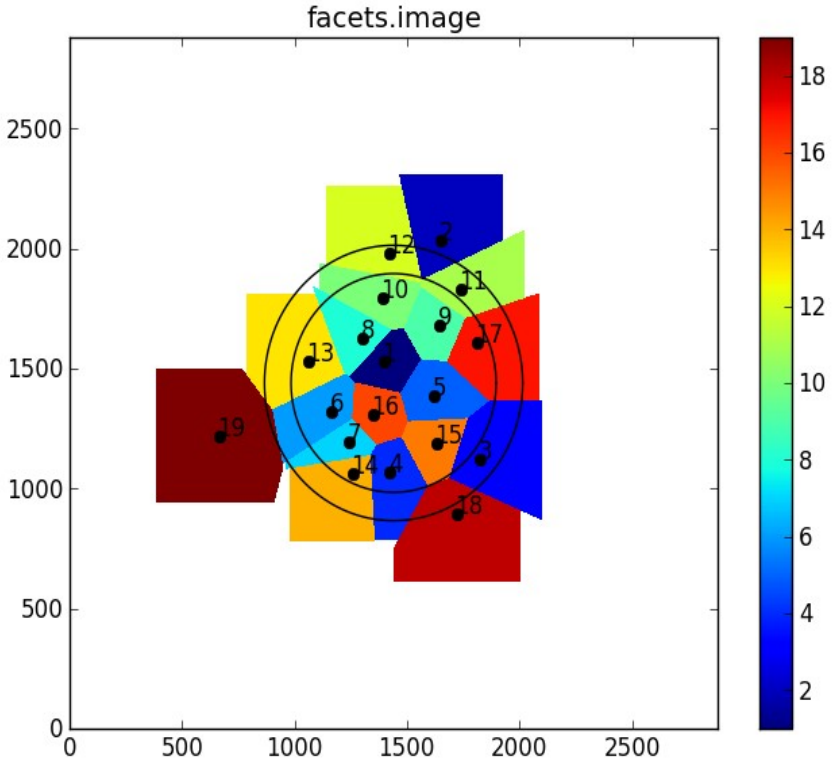
- Effect depends on frequency, length of the baselines and f.o.v.
- LOFAR, worst case:
  - Wide field of view
  - Long distance baselines
  - Low frequency



H. Intema

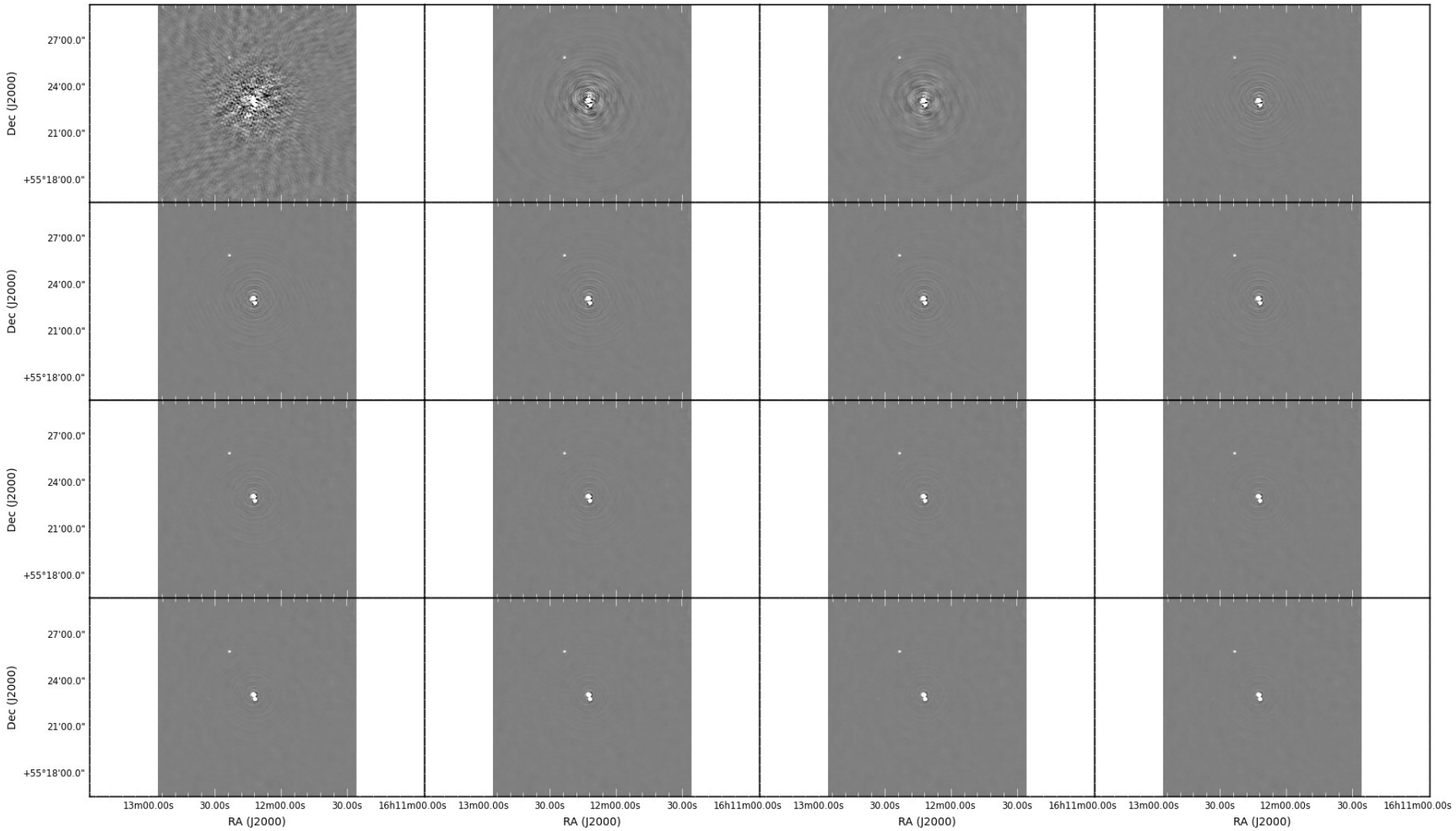
Direction dependent calibration

- Create facets centred in relatively bright sources.
- Iterate in each direction: shift, calibrate, image, shift back and remove artefacts.

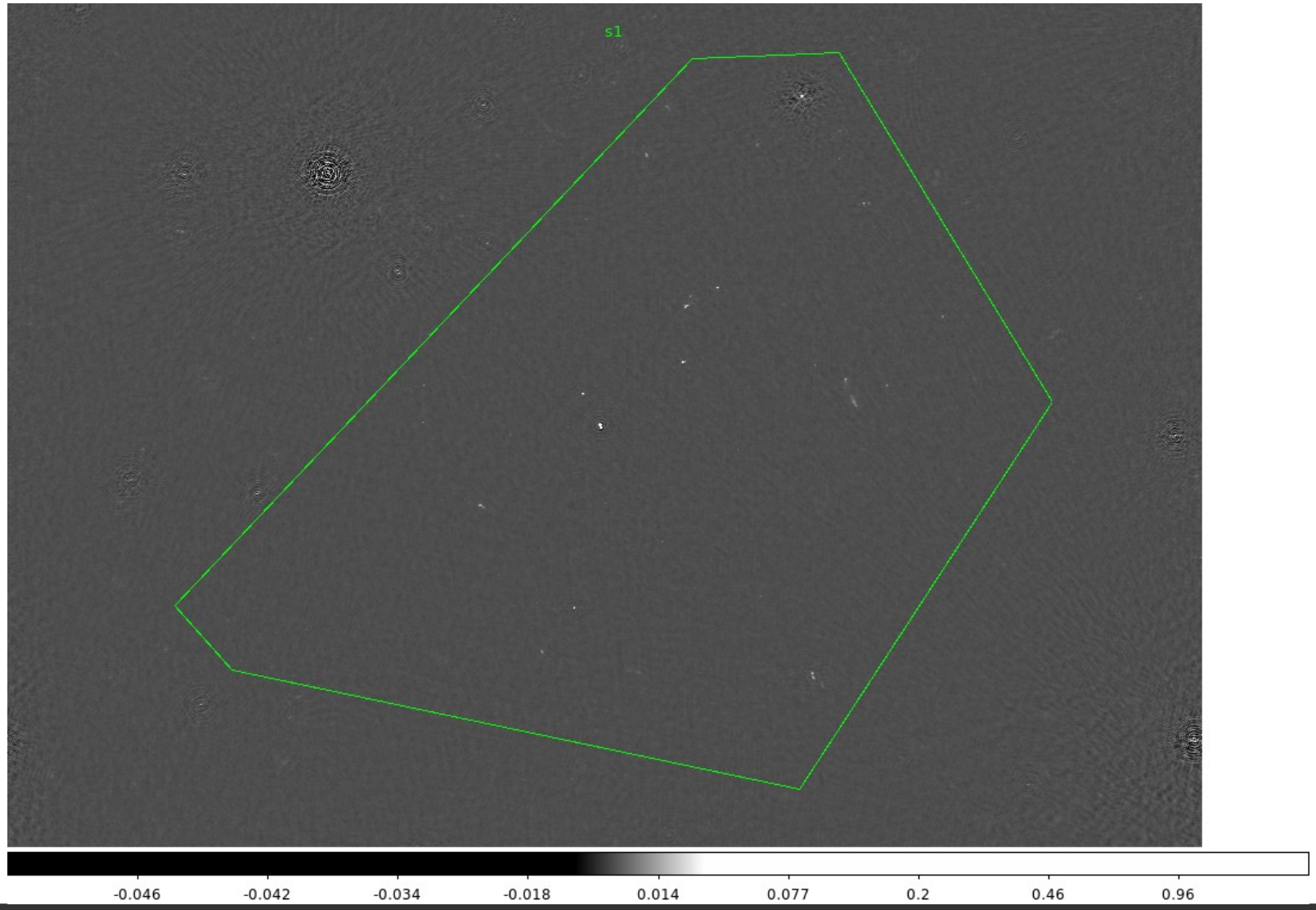


see Wendy Williams' talk

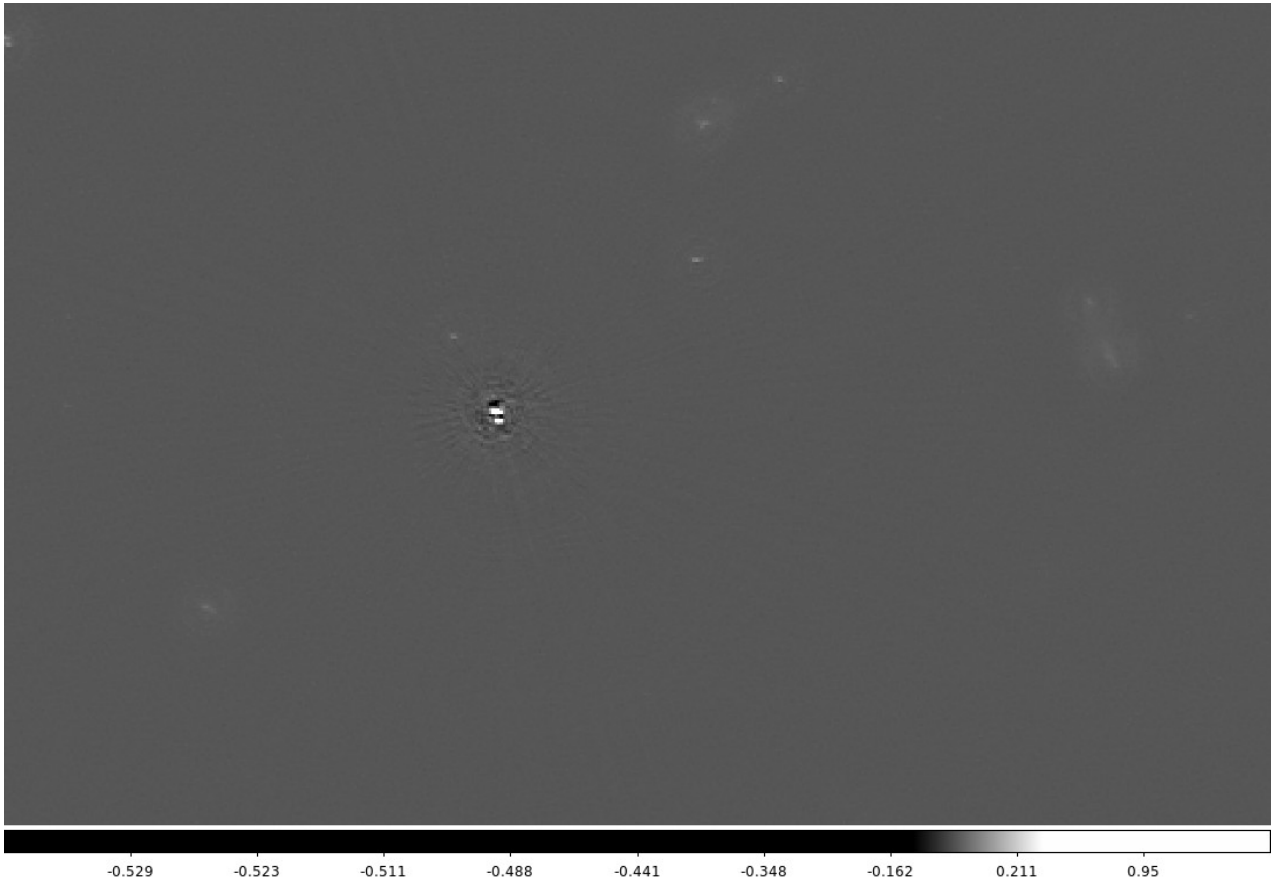
Facet self-calibration



Facet imaging

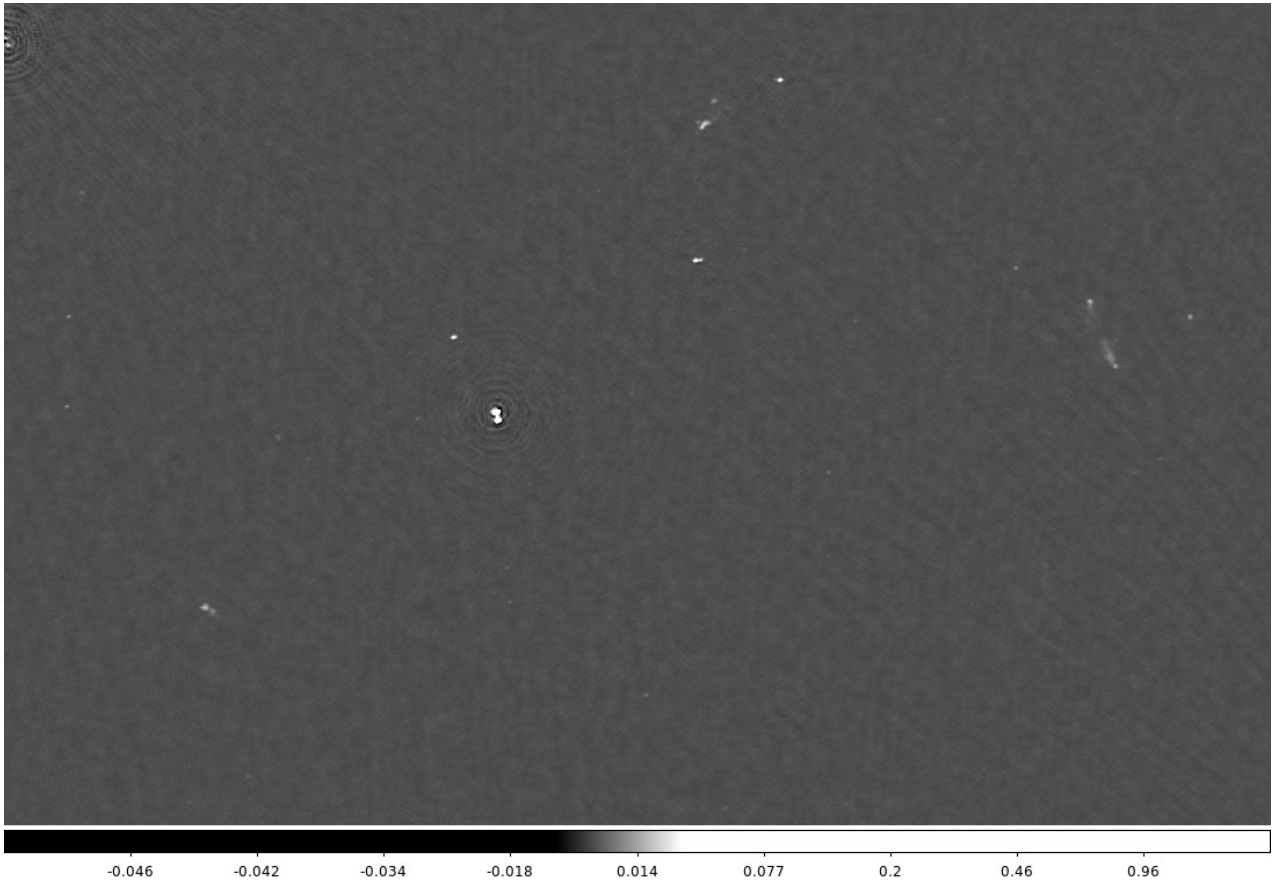


Comparison



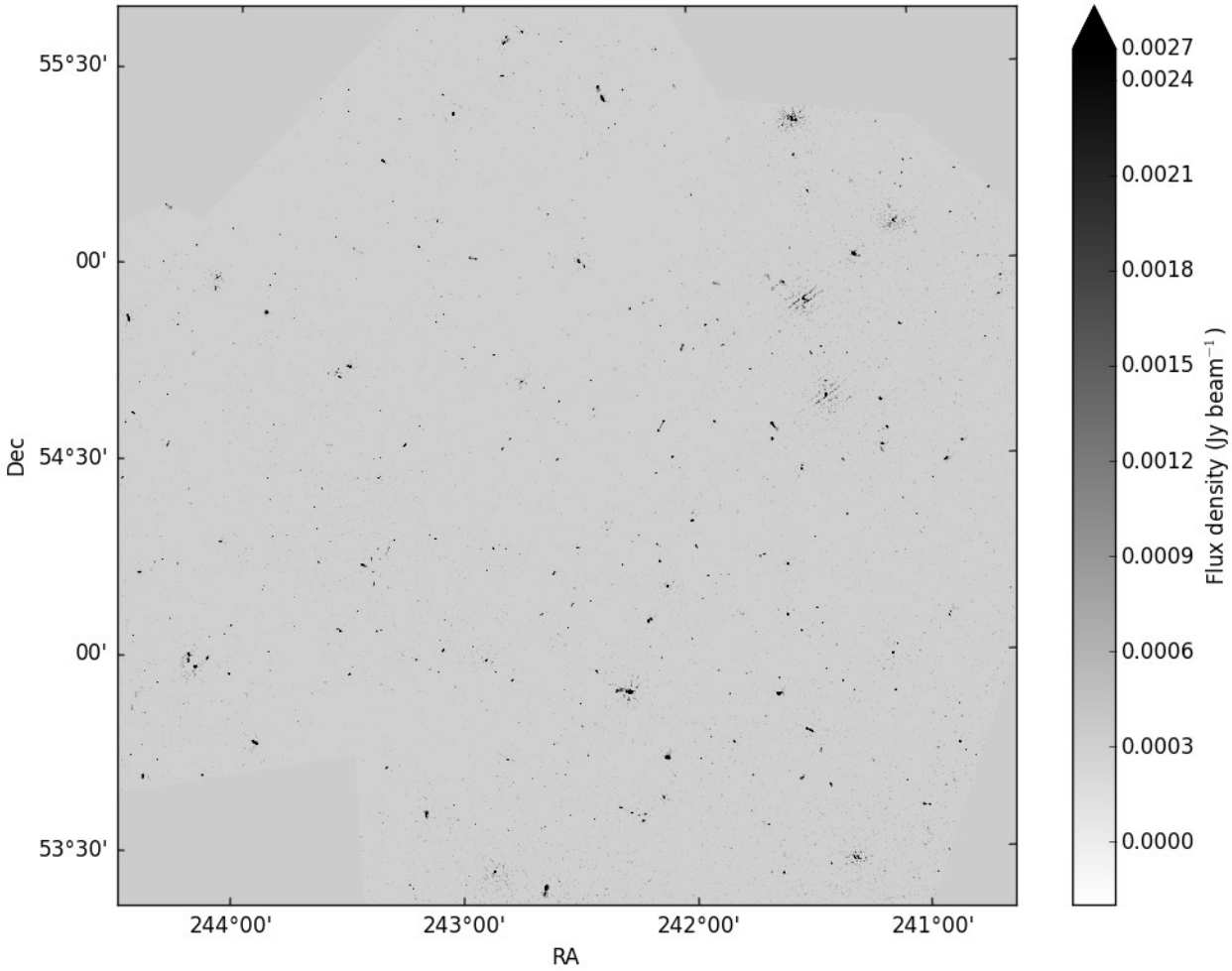
370 sb

Comparison



40 sb

Combined image

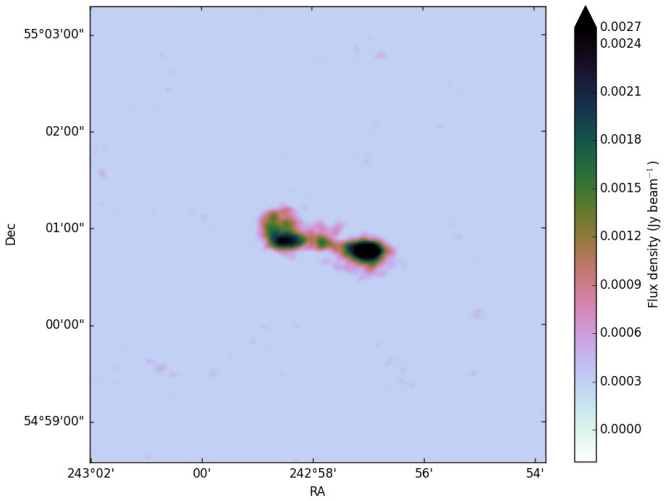


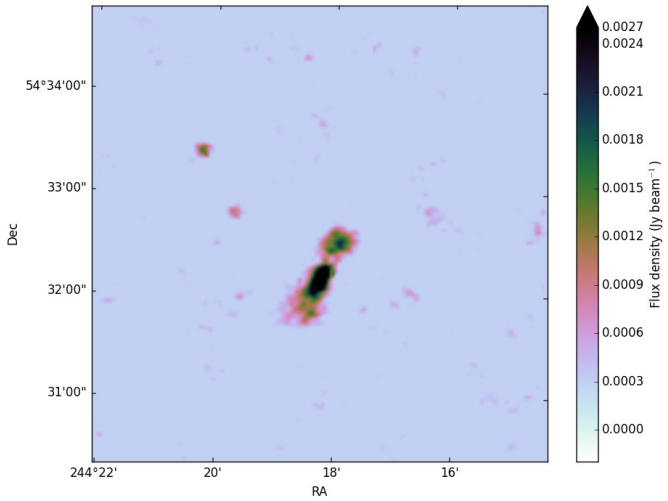
240 sb

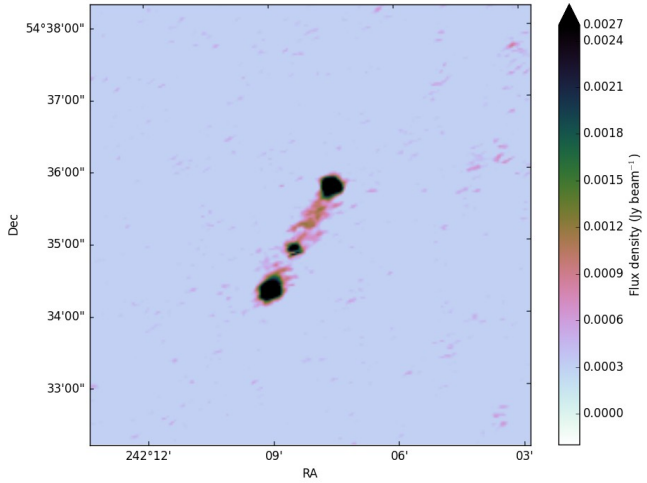
Combined image

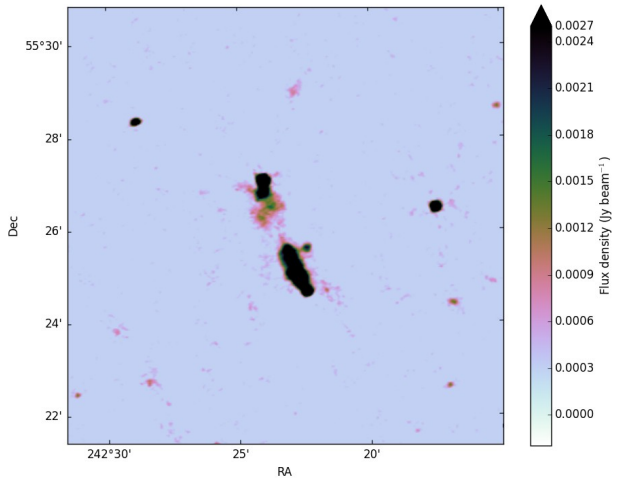
- First dataset of 10 hours:
  - Noise level ~140 microJy/beam → close to thermal noise
  - ~ 5000 sources detected

Extended sources







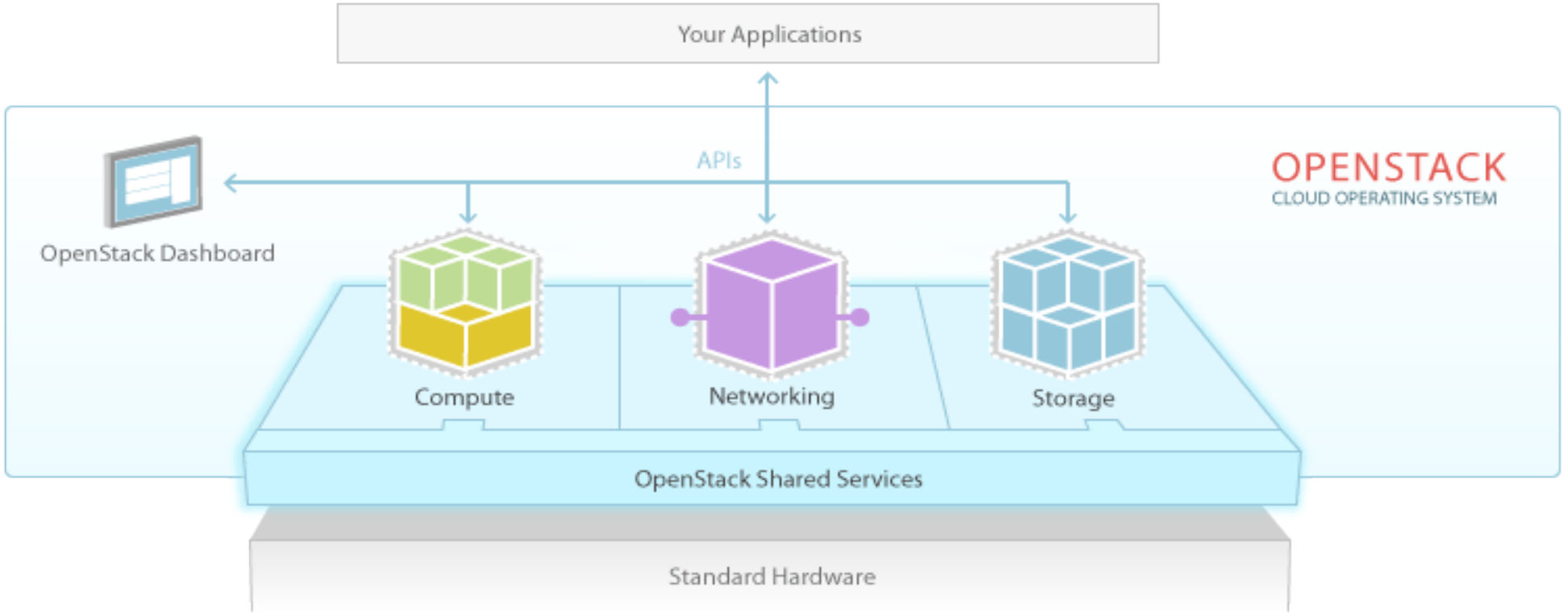


Computational solution needed

- Parallelizable:
  - Deal with a large amount of data in a reasonable time.
- Flexible:
  - Adapt the infrastructure (“hardware”) to different calibration strategies
  - Deal with quickly changing temperamental software
  - On-demand (optional but very useful)

Cloud computing

- Infrastructure as a Service (IaaS)



Tests on cloud infrastructures

- Ibercloud
- EGI Federated Cloud
- STFC RAL cloud
- Amazon Web Services

Ibercloud

- Based on OpenStack
- Very easy to use
- Discontinued and integrated on EGI Federated Cloud

EGI Federated Cloud

- Heterogeneous infrastructure (access using OCCI)
- Many resources and providers
- Good support

- Difficult to use:
  - Complex documentation
  - Site dependent issues
- **Blocker**: No block storage implemented

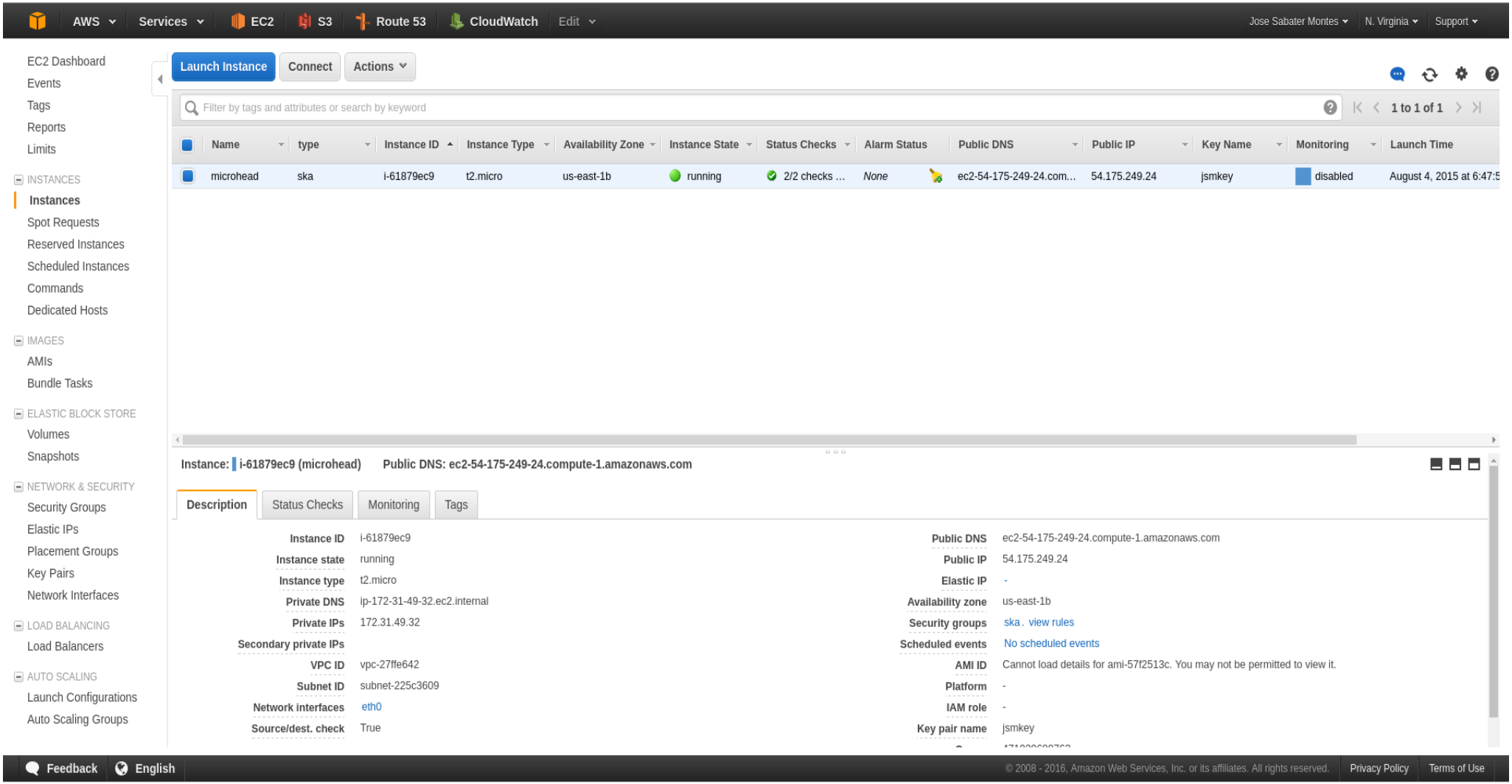
STFC RAL Cloud

- Based on OpenNebula (neither specially difficult nor easy to use)
- Good support
- Complex generation of the VM templates (human intervention, prone to errors)
- **Blocker**: Not enough block storage available

Amazon Web Services (AWS)

- Main services:
  - Elastic Compute Cloud; EC2 → Image templates and virtual machines or instances, including spot instances
  - Simple Storage Service; S3 → Long term storage of objects (files)
  - Elastic Block Storage; EBS → Data volumes
  - Route 53 → Dynamic DNS service

AWS console


AWS
Services
EC2
S3
Route 53
CloudWatch
Edit
Jose Sabater Montes
N. Virginia
Support
EC2 Dashboard
Events
Tags
Reports
Limits
INSTANCES
Instances
Spot Requests
Reserved Instances
Scheduled Instances
Commands
Dedicated Hosts
IMAGES
AMIs
Bundle Tasks
ELASTIC BLOCK STORE
Volumes
Snapshots
NETWORK & SECURITY
Security Groups
Elastic IPs
Placement Groups
Key Pairs
Network Interfaces
LOAD BALANCING
Load Balancers
AUTO SCALING
Launch Configurations
Auto Scaling Groups
Launch Instance
Connect
Actions
Filter by tags and attributes or search by keyword
1 to 1 of 1
Name
type
Instance ID
Instance Type
Availability Zone
Instance State
Status Checks
Alarm Status
Public DNS
Public IP
Key Name
Monitoring
Launch Time
microhead
ska
i-61879ec9
t2.micro
us-east-1b
running
2/2 checks ...
None
ec2-54-175-249-24.com...
54.175.249.24
jsmkey
disabled
August 4, 2015 at 6:47:5
Instance: i-61879ec9 (microhead)
Public DNS: ec2-54-175-249-24.compute-1.amazonaws.com
Description
Status Checks
Monitoring
Tags
Instance ID
i-61879ec9
Instance state
running
Instance type
t2.micro
Private DNS
ip-172-31-49-32.ec2.internal
Private IPs
172.31.49.32
Secondary private IPs
VPC ID
vpc-27ffe642
Subnet ID
subnet-225c3609
Network interfaces
eth0
Source/dest. check
True
Public DNS
ec2-54-175-249-24.compute-1.amazonaws.com
Public IP
54.175.249.24
Elastic IP
-
Availability zone
us-east-1b
Security groups
ska. view rules
Scheduled events
No scheduled events
AMI ID
Cannot load details for ami-57f2513c. You may not be permitted to view it.
Platform
-
IAM role
-
Key pair name
jsmkey
Feedback
English
© 2008 - 2016, Amazon Web Services, Inc. or its affiliates. All rights reserved.
Privacy Policy
Terms of Use

Orchestration with Ansible

- Ansible playbooks → Recipes to manage the infrastructure and the nodes
  - Written in YAML → easy to read and write by humans
  - Idempotent → Move the instances from state to state
  - Based on Python and easy to extend and integrate in programs

Example playbook

```
---
# tasks file for lofar

- name: install LSMTool
  pip: name='https://github.com/darafferty/LSMTool/archive/master.zip' extra_args='--allow-external --upgrade'
  sudo: True

- name: install LoSoTo
  pip: name='https://github.com/revoltek/losoto/archive/master.zip' extra_args='--allow-external --upgrade'
  sudo: True

- name: Download source of WSClean
  get_url: >
    url=http://sourceforge.net/projects/wsclean/files/wsclean-1.10/wsclean-1.10.tar.bz2/download
    dest=/tmp/wsclean-1.10.tar.bz2
# Alternative http://www.roe.ac.uk/~jsm/lofar_dist/src/wsclean-1.10.tar.bz2

- name: Compile and install WSClean
  script: wsclean.sh creates=/usr/bin/wsclean
  sudo: True

- name: download LOFAR package
  get_url: url=http://www.roe.ac.uk/~jsm/lofar_dist/trusty/lofar_2.15-1trusty_amd64.deb
  dest=/tmp/lofar_2.15-1trusty_amd64.deb

- name: Install LOFAR package
  apt: deb=/tmp/lofar_2.15-1trusty_amd64.deb
  sudo: True

- name: Remove temporary package
  file: path=/tmp/lofar_2.15-1trusty_amd64.deb state=absent

- name: copy LOFAR profile file
  copy: src=lofar_profile dest=/etc/profile.d/Z98-lofar.sh owner=root group=root mode=0755
  sudo: True

- name: install factor
  pip: name='https://github.com/revoltek/factor/archive/master.zip' extra_args='--allow-external --upgrade'
  sudo: True
```

see https://github.com/nudomarinero/Astrocompute-ELAIS-N1

AWS and ELAIS-N1

- SKA-AWS astrocompute proposal: See http://www.lofarcloud.uk
- Steps:
  - Preparation of the base infrastructure (virtual machine images, check provisioning of spot instances, etc)
  - Data transfer: 50 TB
  - Adapt calibration pipeline and run

http://www.lofarcloud.uk

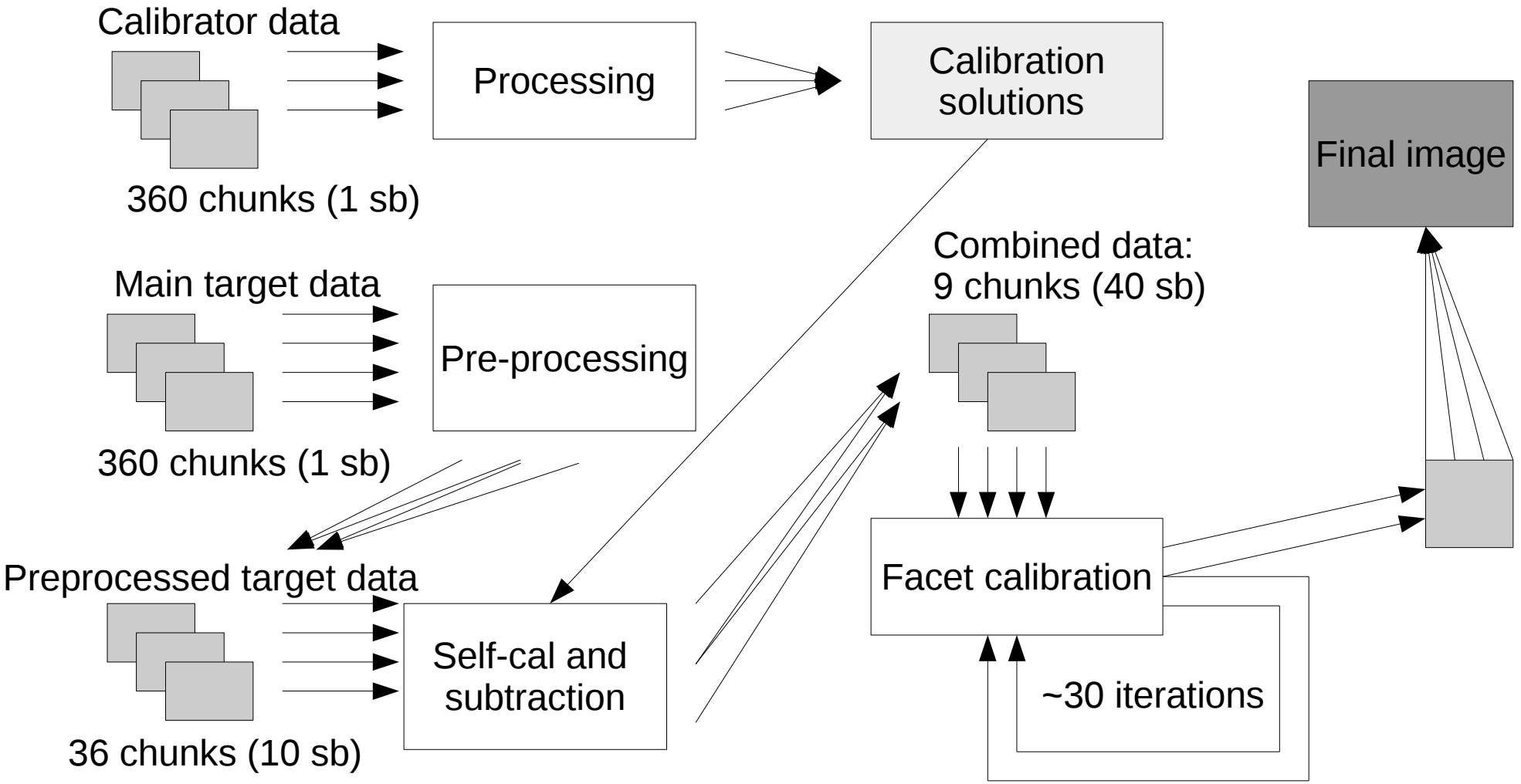
Creation of the images

- Creation of the template virtual machine images:
  - Ansible playbooks → recipe easy to update
  - LOFAR and extra software installed
  - Prepare once and use every time you need it
- Solves the problem with the installation of software in legacy or unsupported systems

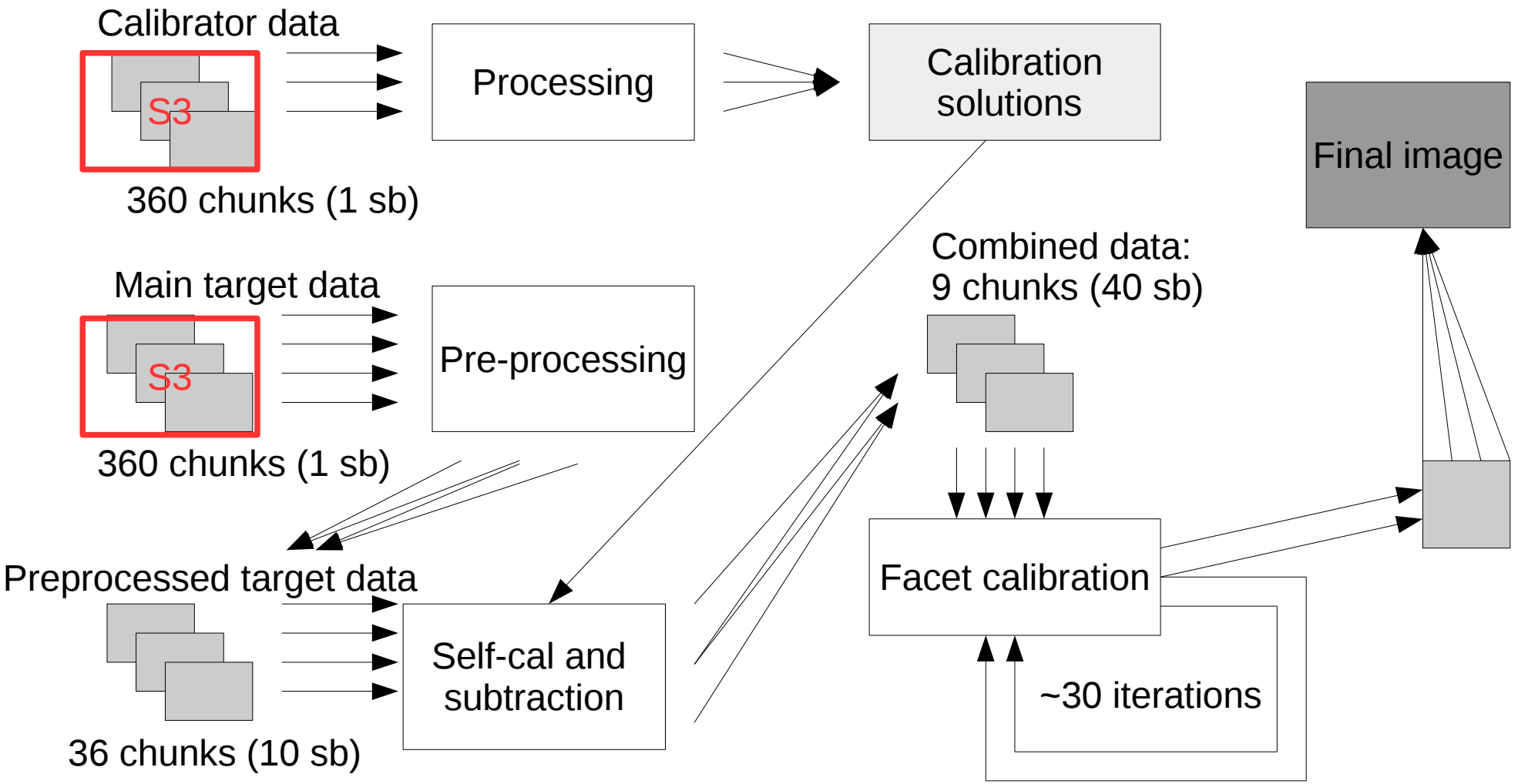
Data transfer

- ELAIS-N1 public data transferred to AWS us-east-1 region (Virginia)
- Publicly available
- 50 TB from the GRID in the Netherlands and Germany → 2 months; mainly due to the manual supervision of the transfer (renew proxy certificate and check errors)
- Consider Import/Export Snowball solution

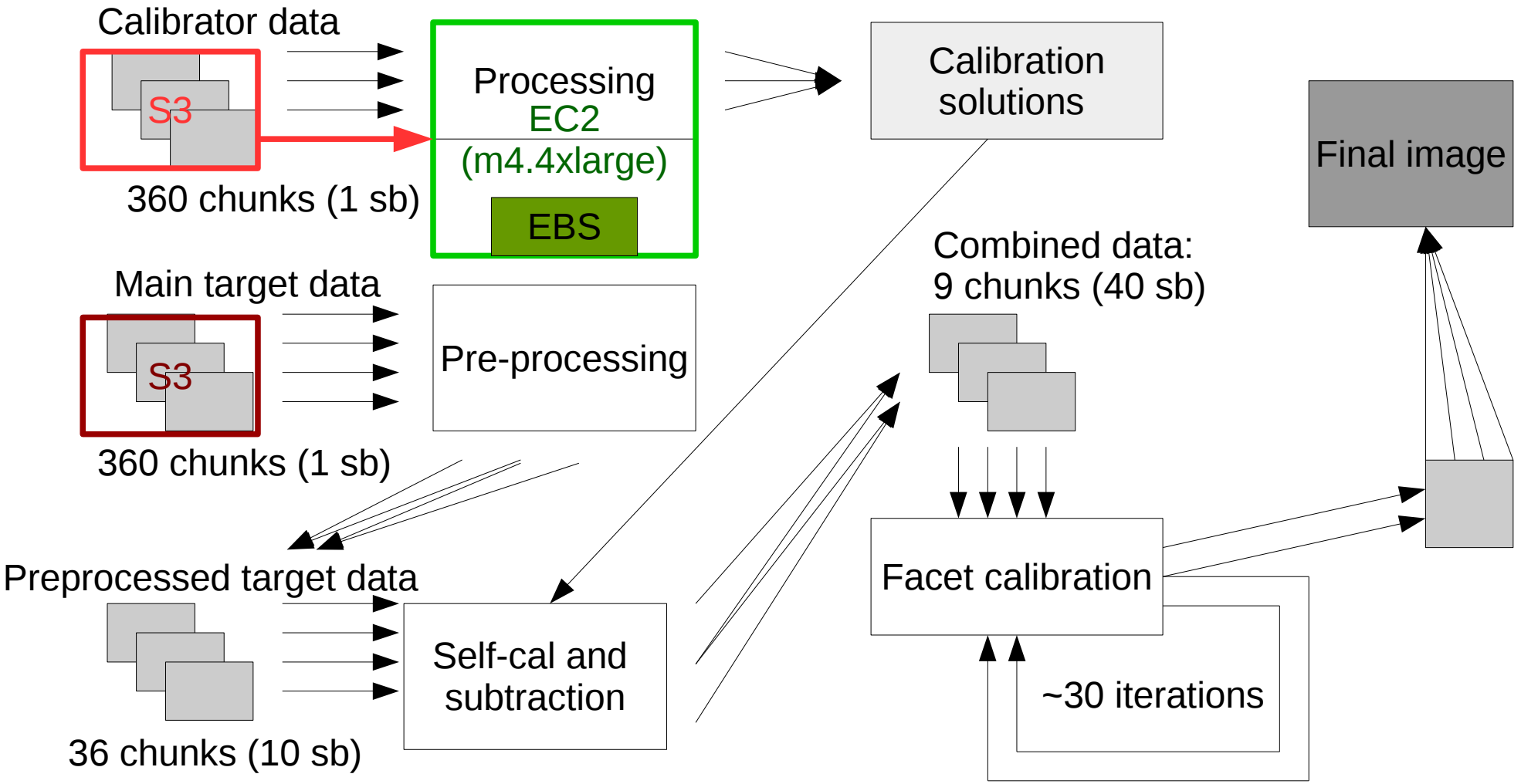
Adapt facet calibration pipeline



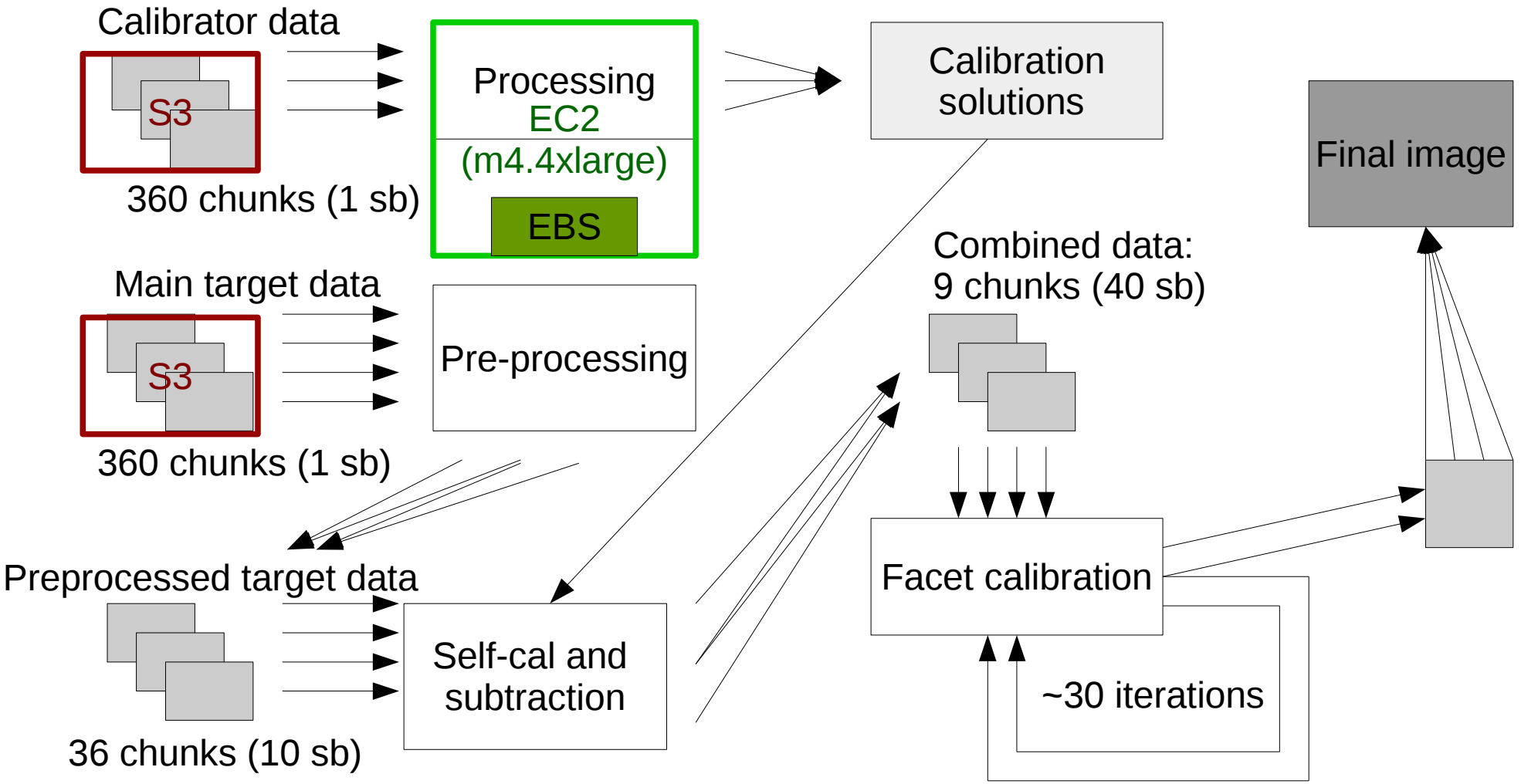
Data stored in S3



Process the calibrator

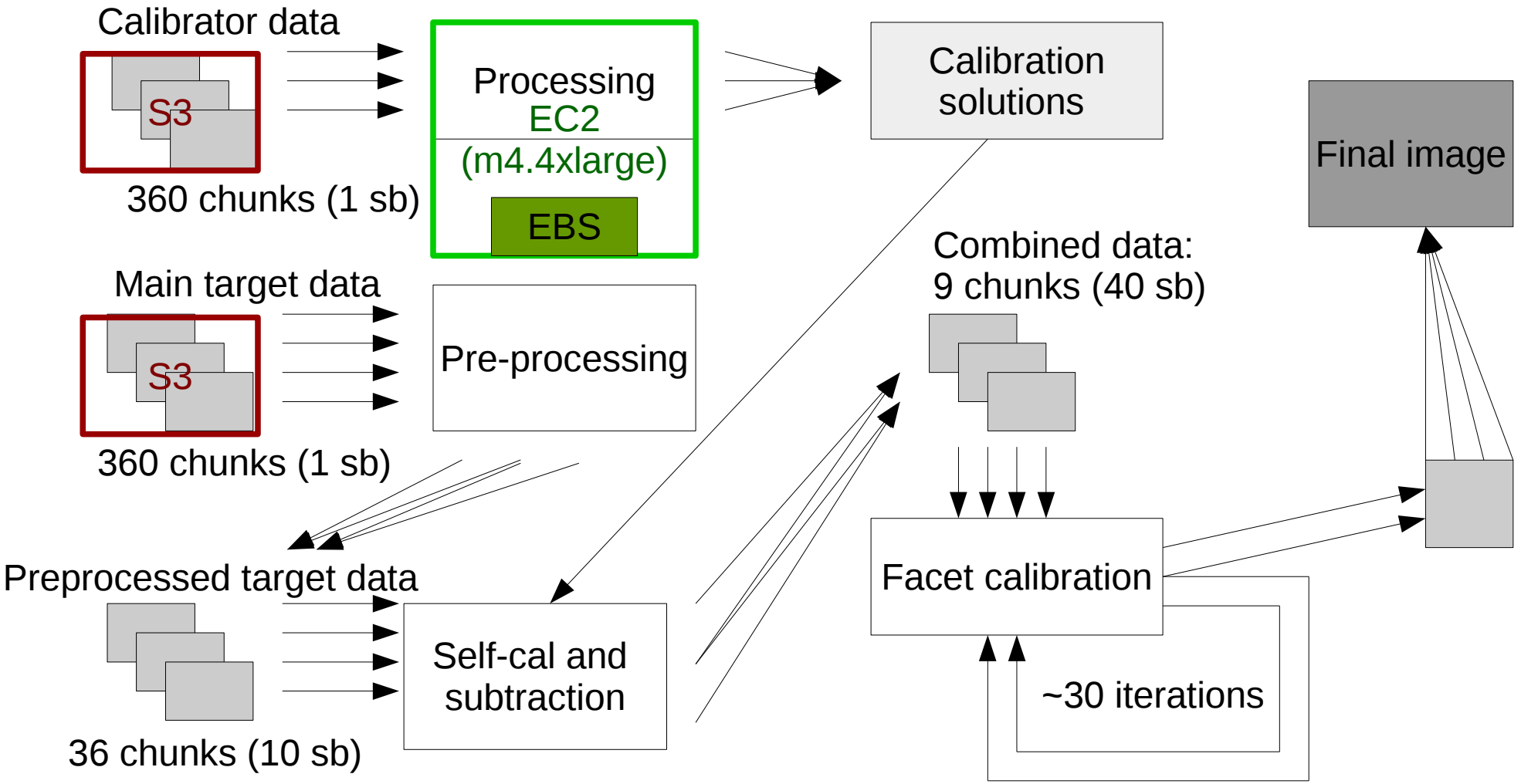


Process the calibrator

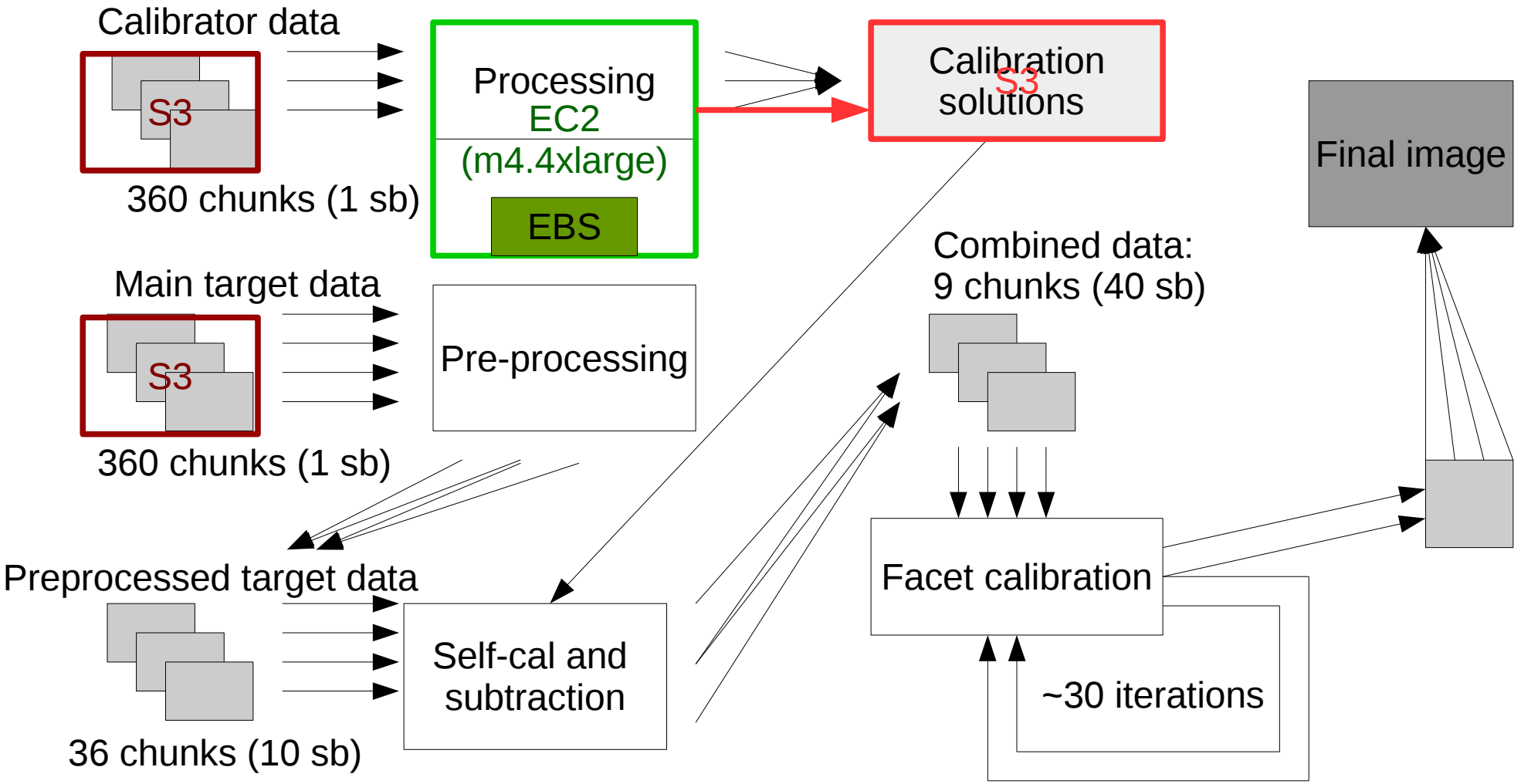


Process the calibrator

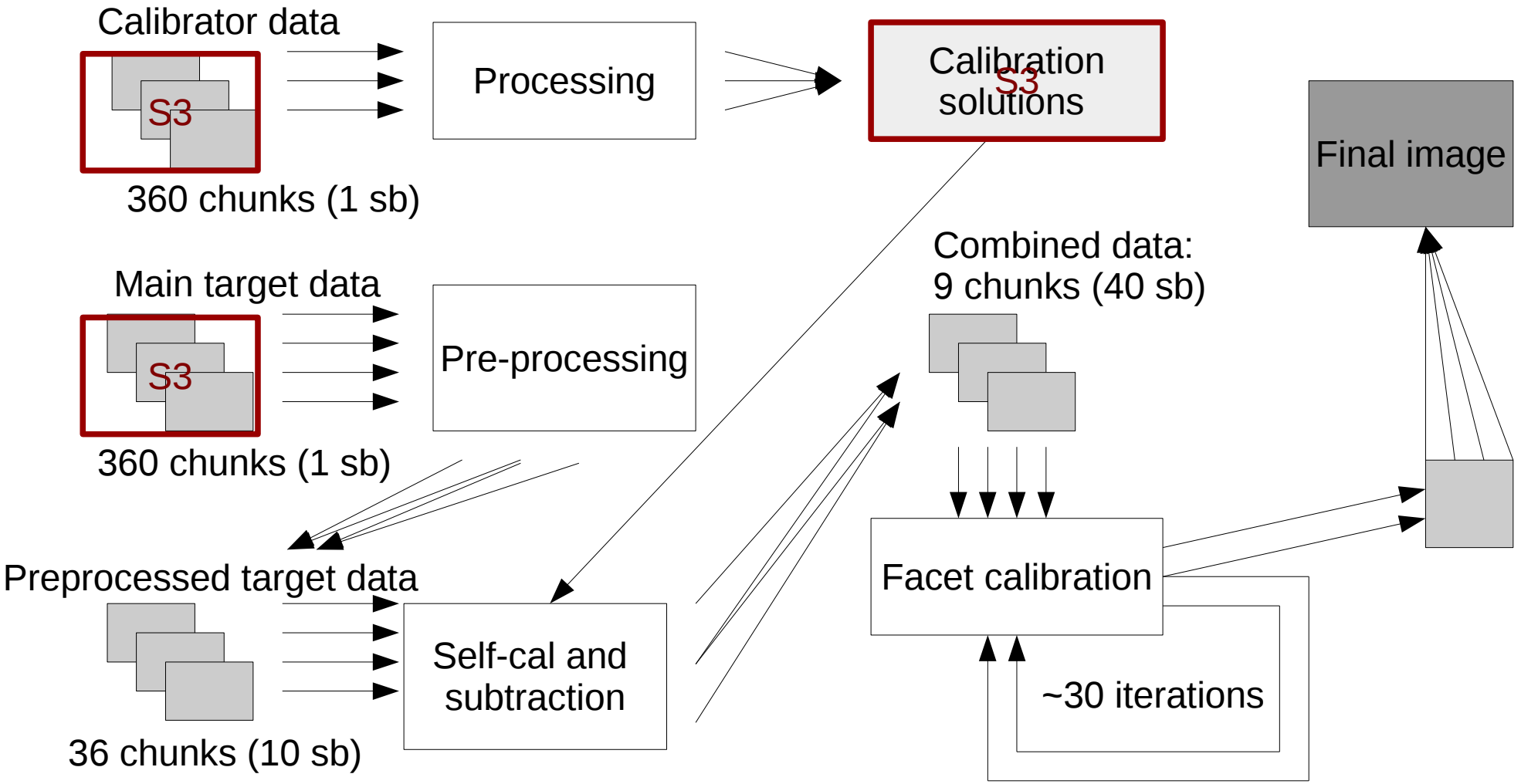
Resume after failure or shut-down is possible using EBS



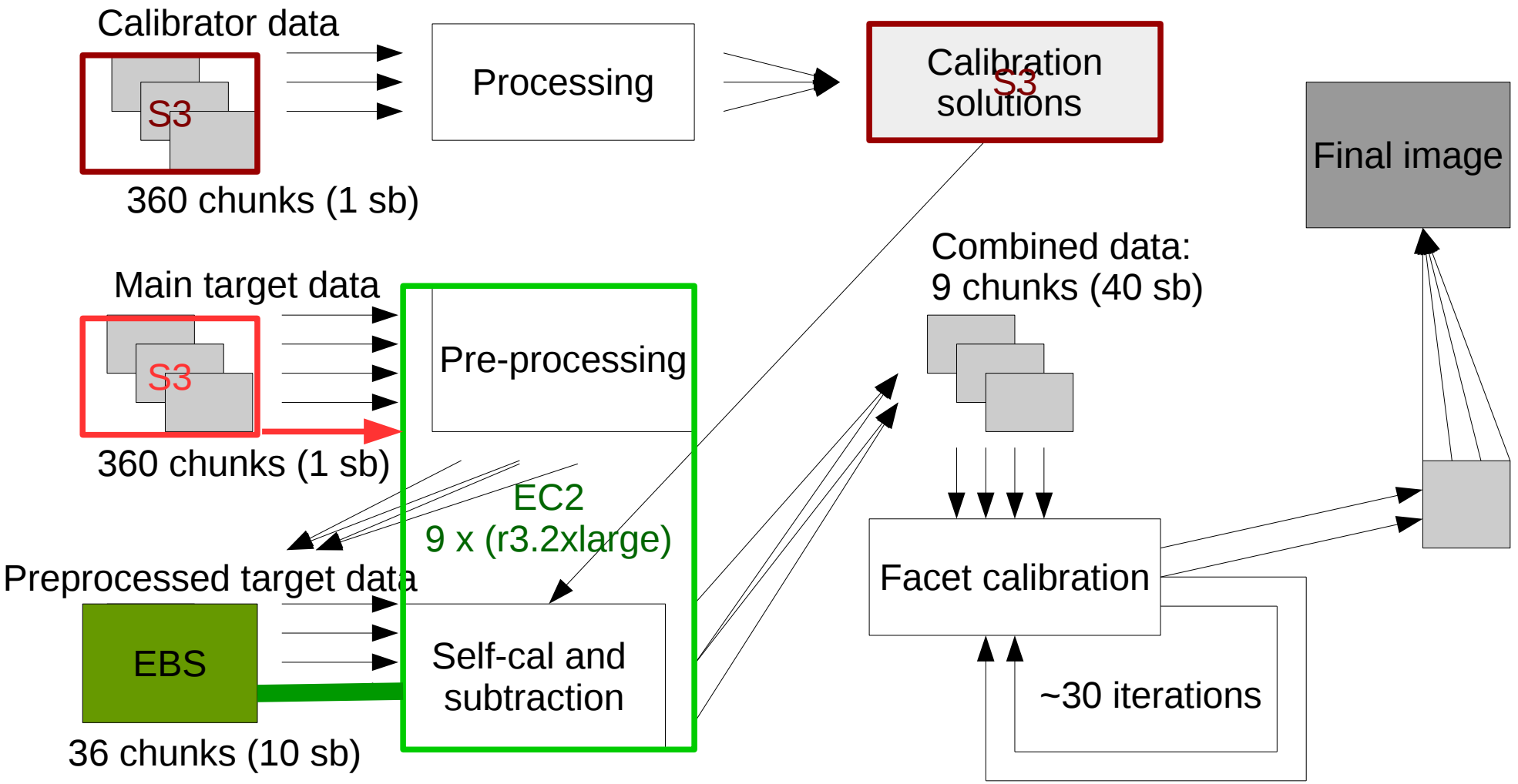
Process the calibrator



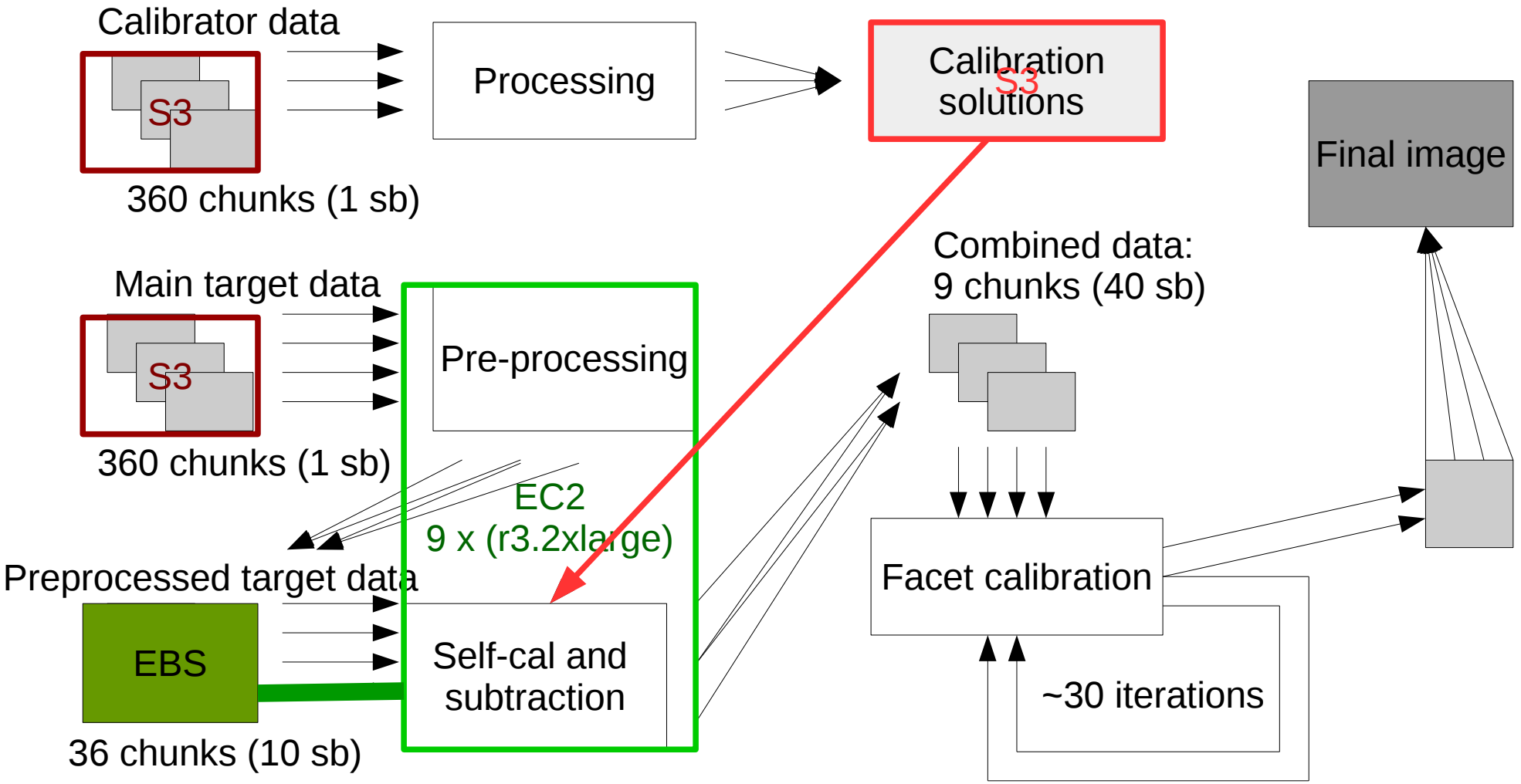
Calibrator data ready



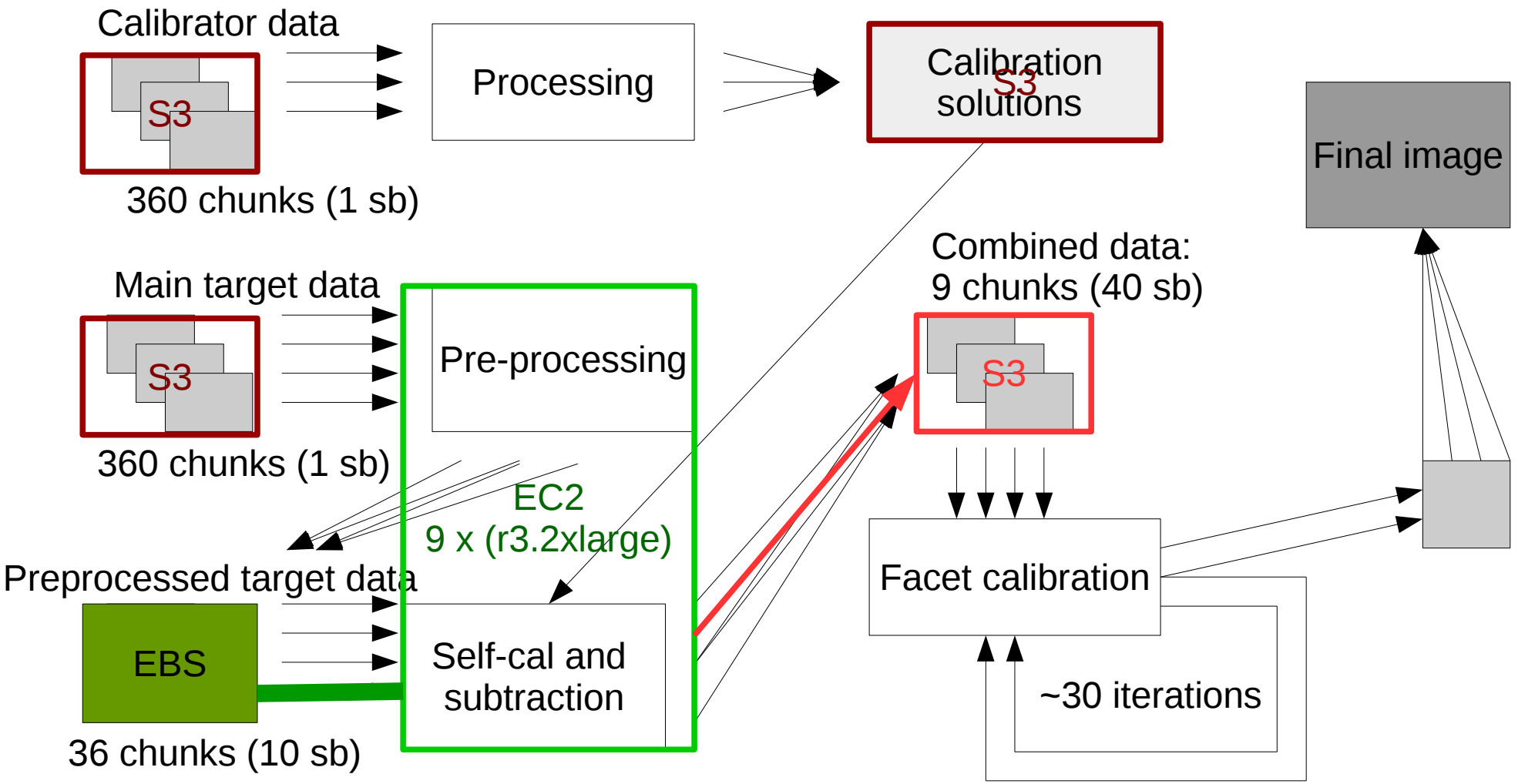
Pre-process the data



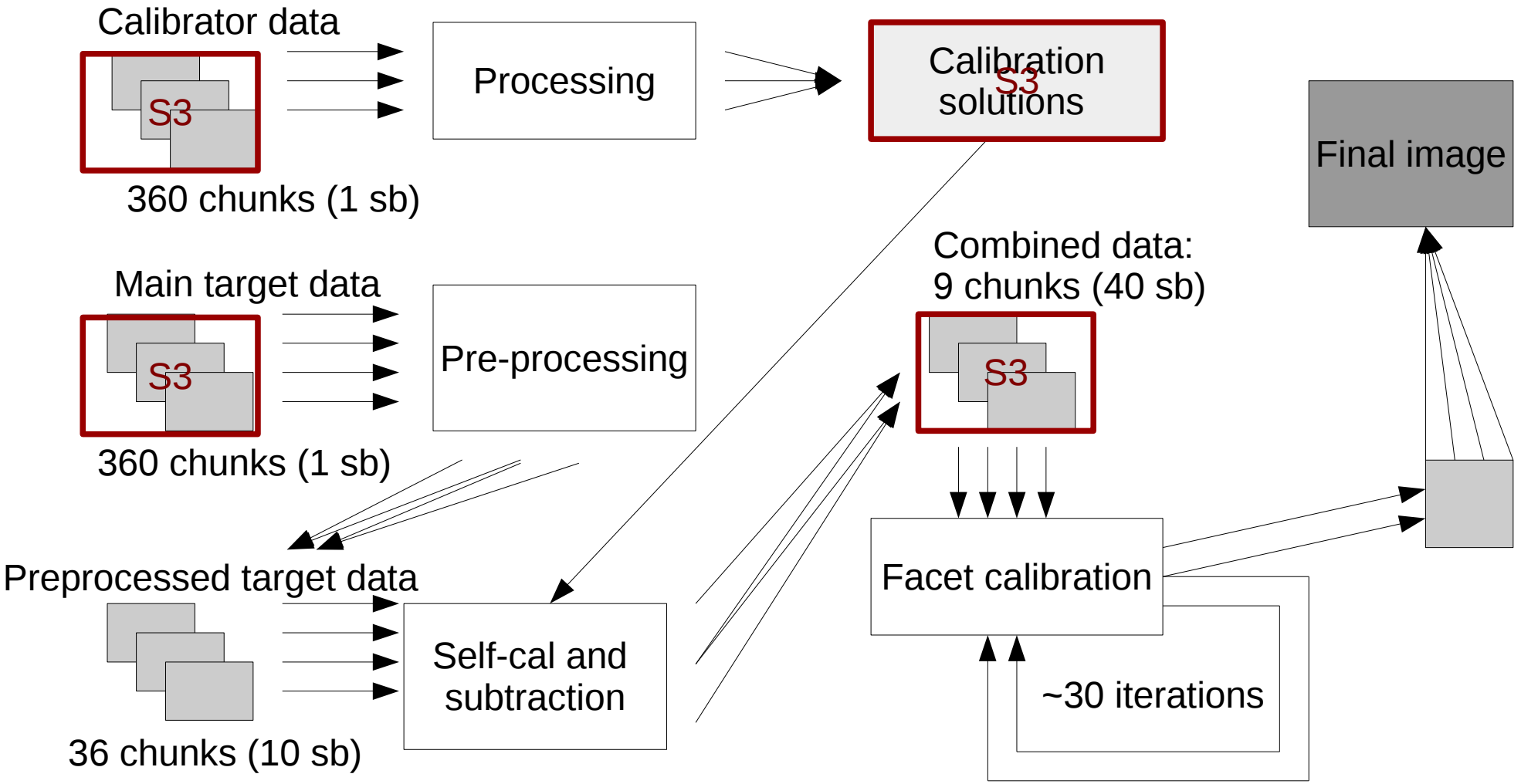
Pre-process the data



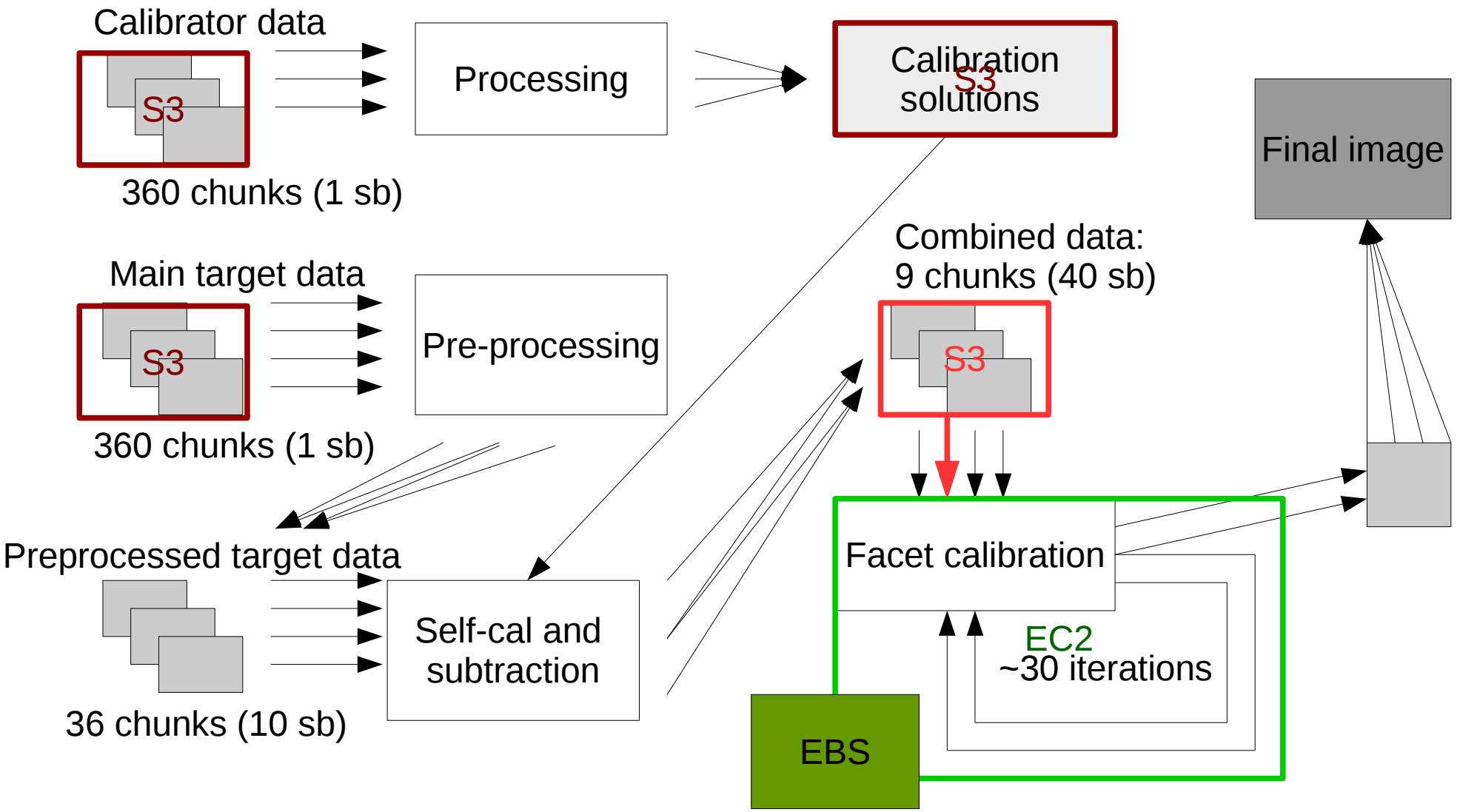
Pre-process the data



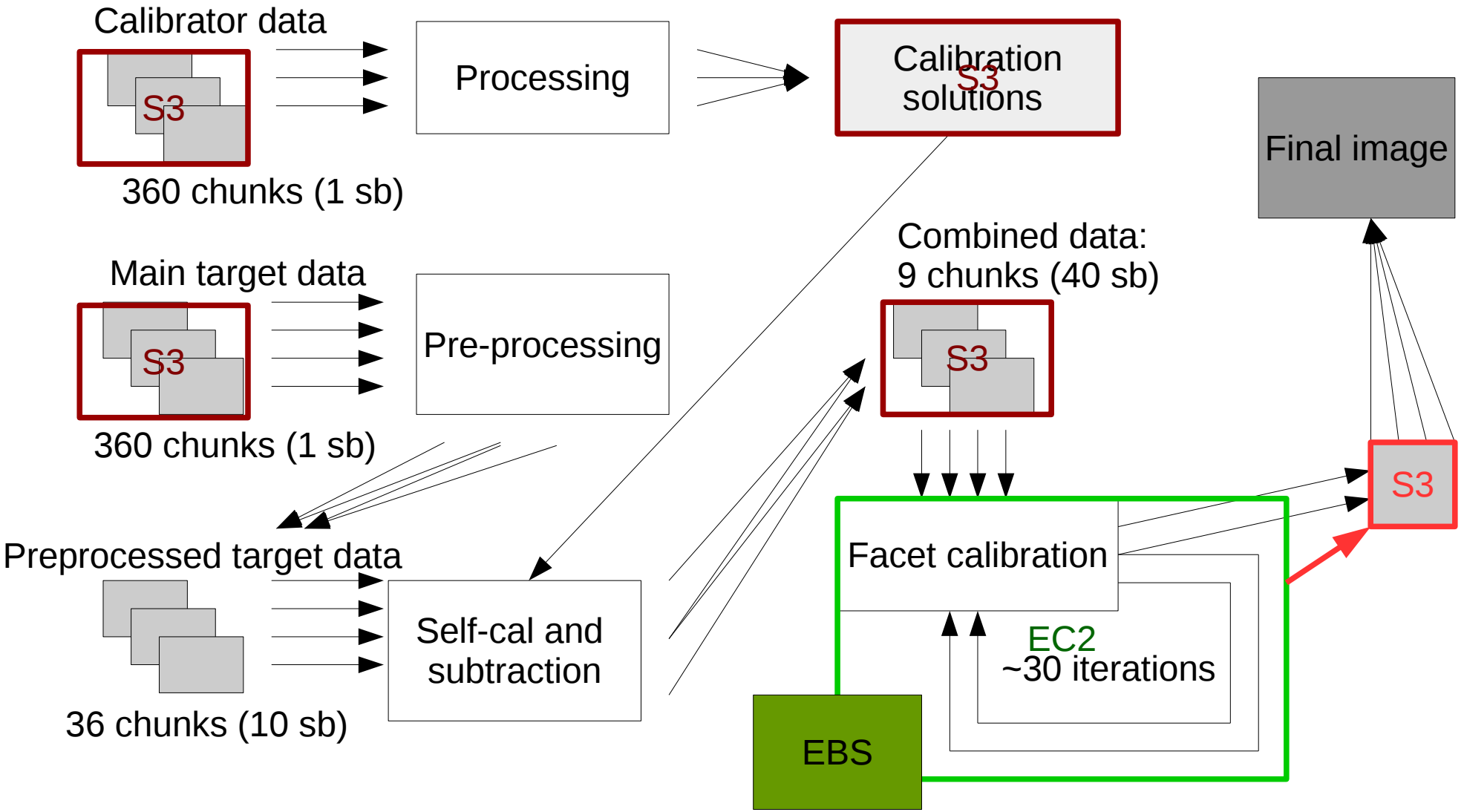
Combined base data



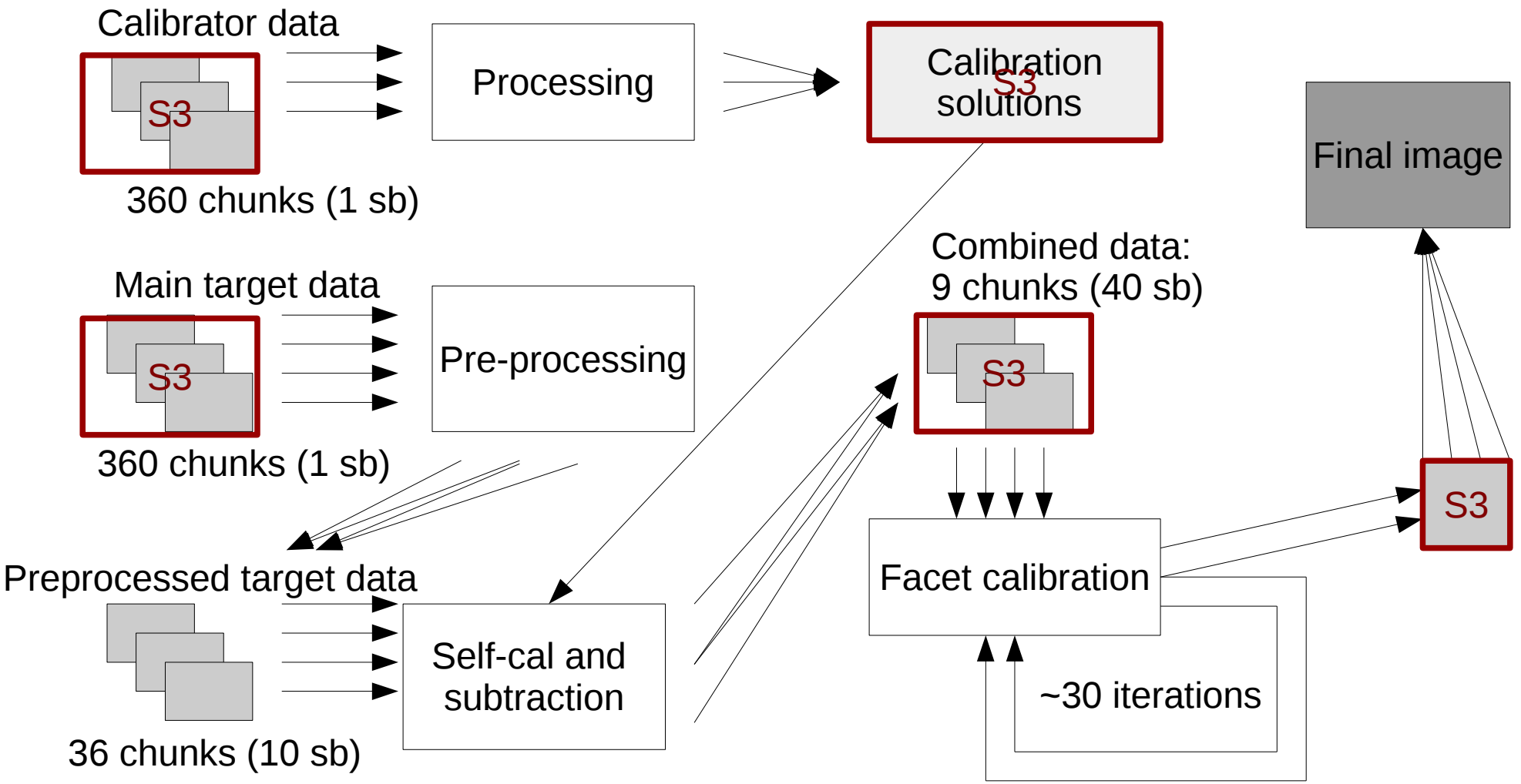
Facet calibration



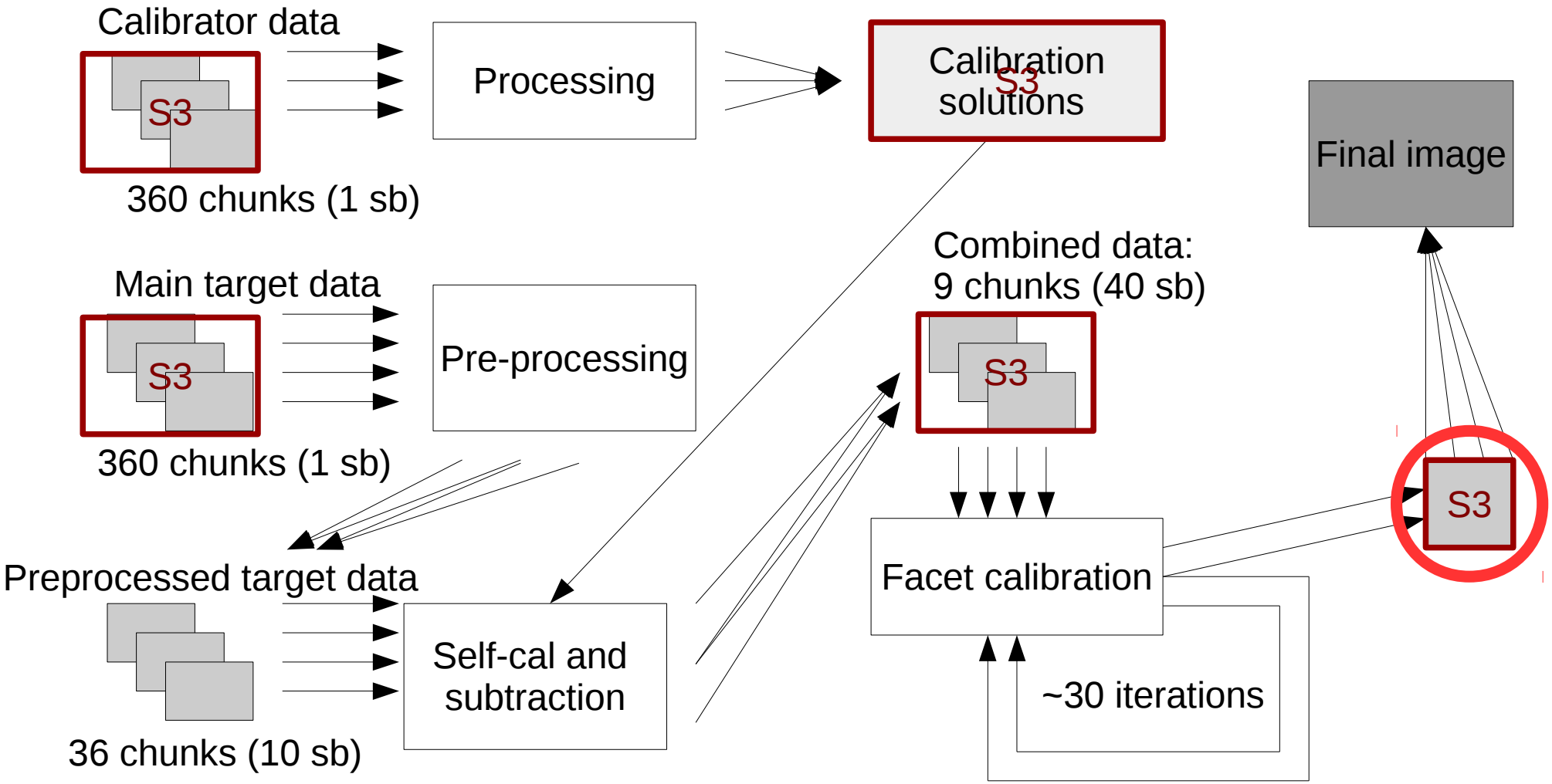
Facet calibration



Data products



Final data products



Highlights of AWS (or similar cloud infrastructures)

- Wide range of possibilities
- Size of the nodes optimised to account for limitations in the software (example, LOFAR BBS can only use one core)
- EBS:
  - scratch data storage
  - resume after failure (or shut-down) capability
- On demand consumption of resources → pay as you go with no special arrangements needed → ideal for a final user

Summary

- The new facet calibration strategy is being adapted to cloud infrastructures. Useful for SKA.
- Big software and data managing challenges associated to a software defined radio-telescope; even for final users.
- Cloud infrastructure to calibrate astronomical data:
  - Parallellization – Ability to deal with big data.
  - Flexibility – Quick development and testing of innovative strategies and on-demand consumption of (shared) resources.