BEAM / NEST Architectural Readiness for Sentinel 1, 2 and 3 Data

Technical Note
Version 1
October, 26th 2012

Document authors:

Luis Veci (Array Systems Computing Inc, Canada)
Norman Fomferra (Brockmann Consult GmbH, Germany)
Petar Marinkovic (PPO.labs, Netherlands)

Compiled from round table discussions during a meeting at ESA/ESRIN, October 5th, 2012 with:

Marcus Engdahl (ESA/ESRIN, Italy)
Peter Regner (ESA/ESRIN, Italy)
Andrea Minchella (ESA-RSAC/ESRIN, Italy)
Luis Veci (Array Systems Computing Inc, Canada)
Norman Fomferra (Brockmann Consult GmbH, Germany)
Petar Marinkovic (PPO.labs, Netherlands)
Executive Summary

This document identifies, at a high abstraction level, optimization and tailoring needed for an optimal processing of Sentinel data with ESA's open source NEST and BEAM toolboxes developed by Brockmann Consult, Array System Computing and PPO.labs. With the extensive experience of the development team in both software engineering and in the application domain and given the flexibility of the architecture, it is believed the NEST and BEAM toolboxes will be able to successfully handle the product data for Sentinel's 1, 2 and 3. However, some optimizations have been identified in order to improve the user experience. These optimizations, when implemented, should allow the toolboxes to better perform with extreme amounts of data on conventional hardware.

It is important to note, that only framework, memory and data model, etc. are addressed, while any of the algorithmic work required for a robust processing of Sentinel's data is not discussed. Dealing with algorithmic issues is outside the scope of this document. However, any uncertainty considering algorithmic work, poses little risk to the proposed optimization. What is being proposed is fairly well understood, and the development team is familiar with application area.

The Problem Definition and Challenges

The ESA Toolboxes BEAM and NEST currently handle several optical and SAR missions, and are used by several thousand scientific and industry users worldwide. The toolboxes are currently being extended to provide support for ESA’s Sentinel-1 (S-1), Sentinel-2 (S-2) and Sentinel-3 (S-3) missions. This brings new challenges which will need to be overcome in order to continue fulfilling the user experience expectations of the data users.

The Sentinel missions will not only have products that are much larger and acquired with higher temporal frequency than any of the previous ESA missions, but also users will have unrestricted access to significantly more data. Data users will now be able to readily create long time series of data, large stacks of products and global mosaics.

The data must be read, processed and visualized in an acceptable amount of time using laptop and desktop computers / workstations available in 2014.

Expected maximum product sizes for Sentinel data in Gb

<table>
<thead>
<tr>
<th></th>
<th>Sentinel-1</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>10 min data take</td>
<td>SM</td>
<td>IW</td>
<td>EW</td>
</tr>
<tr>
<td>S-1 Single Pol SLC</td>
<td>80</td>
<td>76</td>
<td>21</td>
</tr>
<tr>
<td>S-1 Dual Pol SLC</td>
<td>159</td>
<td>152</td>
<td>43</td>
</tr>
<tr>
<td>Sentinel-2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Data take</td>
<td></td>
<td>MSI</td>
<td></td>
</tr>
</tbody>
</table>
### Anticipated Data Use Cases

Two use cases, typical and extreme, are proposed and discussed to help ponder the following questions:

- Is the current BEAM/NEST architecture and data model adaptable for the use case? What adaptations would be needed to make working with Sentinel data possible? Or do we need a new paradigm?

- Is the user experience going to be acceptable? How can we increase the performance? What are the realistic minimum specifications of a machine needed for Sentinel processing?

### Typical

Typical use would be the continuity of applications from ENVISAT ASAR, RADARSAT-2, TerraSAR-X, etc. for SAR and MERIS, LandSat, etc for optical remote sensing data.

The toolboxes should be able to open, display and process single data sets of S-1, S-2, S-3 at a comparable user experience as with previous missions on NEST 4C and BEAM 4.10.

### Extreme

A use case to test the limitations of Sentinel data processing on a standard user computer would be:

- 10 data-takes over the same area with the same instrument,
- 20Gb per data-take, delivered in 2Gb slices for S-1.
- 20Gb of tiles for a data-take for S-2
- Format L1 IWS SLC for S-1 and L1C MSI for S-2.

The user would need to calibrate, co-register and orthorectify this dataset for S-1, and compute NDVI and band arithmetic for S-2.

### Brief overview of NEST/BEAM Architecture

#### Current Situation

BEAM and NEST use a tiled memory management to handle large products greater in size than can fit in memory. Image displaying is handled by image pyramids to allow rapid panning and zooming. A pyramid is composed of levels of images where each level image is a lower resolution version of its
Images of each level are tiled allowing to access only those parts of the resolution level that are currently being displayed.

The BEAM development platform provides a graph processing framework (GPF) to efficiently process large amounts of image data through directed, acyclic processing graphs whose nodes are pluggable processing operators. Tiles are processed in parallel according to the number of available cores.

**On-going Developments for Sentinel's**
Readers for S-1, S-2, S-3 are currently in development under the ESA SUHET project. Only the basic reading capabilities will be added with the data functioning with a subset of existing processing operators.

### Tools available for Sentinel data processing and analysis through the SUHET project

<table>
<thead>
<tr>
<th>BEAM Tools</th>
<th>NEST Tools</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Image Analysis</td>
<td>• Product Library local catalogue</td>
</tr>
<tr>
<td>• Layer and Mask Management</td>
<td>• Coregistration of Stripmap and Detected</td>
</tr>
<tr>
<td>• Band Arithmetic</td>
<td>• Multilook</td>
</tr>
<tr>
<td>• Spectrum View</td>
<td>• Terrain Correction</td>
</tr>
<tr>
<td>• Spectral Unmixing</td>
<td>• SAR Simulation Terrain Correction</td>
</tr>
<tr>
<td>• Import/export of var. raster data types</td>
<td>• Ellipsoid Correction</td>
</tr>
<tr>
<td>• Import/export of var. vector data types</td>
<td>• Slant to Ground Range</td>
</tr>
<tr>
<td>• Clustering</td>
<td>• Resampling</td>
</tr>
<tr>
<td>• Geo-corrections / -projections</td>
<td>• Mosaicking</td>
</tr>
<tr>
<td>• Region of interests</td>
<td>• Speckle Filtering</td>
</tr>
<tr>
<td>• Mosaicking</td>
<td>• Ship Detection</td>
</tr>
<tr>
<td>• Level-3 binning and aggregation</td>
<td>• Oil Spill Detection</td>
</tr>
<tr>
<td>• Time series analysis</td>
<td>• Principle Component Analysis</td>
</tr>
<tr>
<td>• Pixel extraction</td>
<td>•</td>
</tr>
<tr>
<td>• Match-up with point data sources</td>
<td>•</td>
</tr>
<tr>
<td>• Statistical analysis</td>
<td>•</td>
</tr>
<tr>
<td>• Histograms and percentiles</td>
<td>•</td>
</tr>
</tbody>
</table>

Current Sentinel readers are able to read, display and process single data sets for the typical data use case at acceptable performance levels.

**Known Issues**

**Tile caching strategy:** The default tile cache implementation in the GPF is currently (BEAM 4.10, NEST 4C) optimized for random access processing as required by highly interactive user interfaces. The default strategy may be ineligible for the efficient batch processing of large Sentinel data products and may be exchanged by a less memory-consuming one.
**Memory settings:** Although the GPF works on single tiles instead of entire images, its tile cache memory and the Java VM’s heap space can be exhausted if it is not properly configured for large data volumes. Tile sizes and tile cache sizes are static values defined in a configuration file. Maximum heap space is a VM setting. Most users are unaware how to properly set and adjust these settings.

**Bands of different sizes:** Handling of different size bands are required for processing S-1 IW and EW, S-2 L1 and S-3 SLSTR L1 products. The BEAM product model requires all bands in a product to have the same image dimensions and share the same spatial pixel resolution.

**Recommended Toolboxes Optimizations for Sentinel's**

Through preliminary tests involving simulated Sentinel data, and using the new readers, it is believed the NEST and BEAM toolboxes will be able to successfully handle the data. However, some optimizations have been identified in order to improve the user experience. These optimizations, when implemented, should allow the toolboxes to better perform with extreme amounts of data.

The following optimizations are recommended (ordered by priority/significance):

- Automatic memory management
- Bands of different size
- Deterministic GPF processing (tile caching) strategies
- Reader/writer optimizations
- Complex data optimizations
- Reduced data buffer copies
- Modification to existing product
- Processing previews

Implementation of these optimizations would be assisted through:

- Further benchmarking and optimizations are needed to determine and diminish bottlenecks for Sentinel data processing on GMES themed applications.
- Providing the development team with simulated and/or virtual data, both in terms of data size and spectral content of data is highly recommended.

The recommendations described would similarly benefit the processing and improved user experience when working with data from other missions.

**Summary**

The fundamental architecture that enables the processing of large Sentinel data is the tiled image pyramids and tiled graph processing framework. It is believed that with some minor modifications to the BEAM-NEST shared architecture, the toolboxes can be optimized to significantly improve performance, in particular for Sentinel products.

By using the BEAM and NEST toolboxes for exploitation of Sentinel data, several existing processing tools can be reused and made immediately available to the thousands of current BEAM and NEST users.