

RANCANG BANGUN PENGOLAHAN DATA LANDSAT BERBASIS PEMROGRAMAN PARALEL

Marendra Eko Budiono^{*)}, Kustiyo^{*)}, Yudhi Prabowo^{*)}
^{*)} Pusat Teknologi dan Data Penginderaan Jauh, LAPAN
 e-mail: Marendra_lapan@yahoo.com

Abstract

Landsat 7 L1T processing in Data and technology centre in Indonesia Aeronautical and Space Agency consist of Top Of Atmosphere (radiometric correction), cloud masking and daily, monthly or yearly mosaicking. In order to improve the processing speed to achieve user demand of cloud free mosaick of Landsat 7 for the whole of Indonesia, the processing is run using High Performance Computing (HPC) in the form of blade server cluster with Linux operating system that has faster processing speed compared to desktop Personal Computer (PC) or even a single server. The Landsat 7 processing softwares are adapted from Landsat 7 processing methods that already been developed and modified to using Message Passing Interface (MPI) library to run under HPC. Finally a design for processing system is build. This system is able to automatically do the processing steps after executing only with single command using bash scripting language on Linux operating system.

Key Words: *Landsat 7 L1T, blade server, Message Passing Interface, Linux, bash scripting, design for processing system*

Abstrak

Proses pengolahan Landsat 7 L1T yang berjalan di pustekdata LAPAN terdiri atas koreksi *Top Of Atmosphere* (radiometrik), proses *cloud masking* dan proses mosaik harian, bulanan maupun tahunan. Agar proses berjalan lebih cepat untuk melayani kebutuhan user akan data Landsat 7 untuk seluruh indonesia yang bebas awan dan termosaik, proses pengolahan dijalankan pada *High Performance Computing* (HPC) berupa sebuah *cluster* yang terdiri atas beberapa *blade server* dengan *Linux operating sistem* yang memiliki kemampuan memproses data jauh lebih cepat dibandingkan komputer *desktop* maupun sebuah komputer server. Software - software pengolahan yang digunakan merupakan hasil adaptasi dari metode - metode pengolahan Landsat 7 yang telah dikembangkan sebelumnya, yang kemudian dimodifikasi agar dapat memanfaatkan library MPI (*Message Passing Interface*) agar dapat berjalan pada HPC. Kemudian dibuat rancang bangun sistem pengolahan pada HPC agar sistem pengolahan ini dapat dieksekusi hanya dengan 1 perintah untuk kemudian secara otomatis menjalankan langkah-langkah pemrosesan Landsat 7 tersebut dengan memanfaatkan bahasa pemrograman *bash scripting* pada Linux.

Kata Kunci: *Landsat 7 L1T, blade server, Message Passing Interface, Linux, bash scripting, rancang bangun sistem pengolahan*

1. Pendahuluan

Bidang Teknologi Pengolahan Data (Teklahta) adalah salah satu bidang di bawah Pusat Teknologi dan Data (Pustekdata) Lembaga Penerbangan dan Antariksa Nasional (LAPAN) yang memiliki tugas melakukan penelitian untuk pengembangan metode pengolahan dan mengolah data-data satelit penginderaan jauh yang dimiliki oleh Pustekdata. Data-data satelit yang diolah adalah untuk keperluan baik itu *request/order* dari *user* seperti instansi pemerintah dan akademi maupun pengolahan rutin untuk pembuatan data mosaik tahunan. Data satelit yang dapat dibuat mosaik tahunan ini adalah data yang telah diakuisisi sejak lama dan di *archive* pada sistem penyimpanan. Salah satu data satelit yang terdapat pada *archive* sistem penyimpanan di bidang teklahta adalah satelit Landsat 7 level L1T. Saat ini telah terdapat *archive* untuk data Landsat 7 L1T tahun 2010, 2011 dan 2012 yang perlu dibuat mosaik tahunan.

Pada penelitian sebelumnya yang dilaksanakan di bidang Teklahta telah dikembangkan beberapa metode pengolahan data satelit Landsat 7 level L1T yang meliputi proses konversi dan koreksi radiometrik, mengubah proyeksi dari UTM ke *latitude longitude*, proses *cloud masking*, *mosaicking* dan pembuatan *quicklook*. Metode-metode tersebut telah diaplikasikan dalam bentuk modul. Modul – modul

tersebut dibuat dengan sistem operasi berbasis *Windows*, sehingga hanya dapat berjalan pada komputer dengan sistem operasi berbasis *Windows*. Pengolahan data Landsat di bidang Teklahta menggunakan *server* dengan sistem operasi *windows*, tetapi karena jumlah data yang sangat banyak, maka kecepatan pengolahan dengan sebuah mesin sekelas *server* dirasa belum cukup untuk memenuhi kebutuhan user akan data-data mosaick satelit Landsat L1T tersebut.

Dari sisi *hardware*, saat ini Pustekdata telah memiliki perangkat HPC (*High Performance Computing*) dengan sistem *cluster* yang berjalan dengan sistem operasi LINUX/UNIX. System HPC ini memiliki kelebihan untuk melakukan perhitungan yang kecepatannya melebihi komputer sekelas *server* sekalipun (*single machine*) karena fisiknya terdiri atas beberapa komputer yang saling berkomunikasi dalam melakukan proses perhitungan. Komunikasi antar CPU dalam HPC ini menggunakan *library message parsing interface* (*mpi*). Penggunaan HPC untuk melakukan proses koreksi dan mosaick diharapkan dapat meningkatkan kecepatan pemrosesan. Tetapi karena perbedaan sistem operasi yang digunakan, maka modul - modul pengolahan Landsat 7 L1T yang telah dikembangkan perlu dikonversi agar dapat berjalan pada platform HPC sekaligus diperlukan suatu desain otomatisasi agar sistem dapat melakukan proses koreksi sampai dengan mosaicking secara otomatis untuk pembuatan mosaick tahunan dari data - data yang telah diakuisisi selama ini.

2. Studi Literatur

Pengolahan data Landsat 7 L1T di Pustekdata terdiri atas beberapa langkah sebelum didapat citra hasil mosaick. Data input adalah data hasil download dari situs USGS yang sudah merupakan data dengan level L1T. Level L1T adalah data yang sudah terkoreksi geometrik sistematis menggunakan *Ground Control Point* (GCP) dan *Digital Elevation Model* (DEM). Data ini terdiri atas 7 band multispektral, 2 band thermal dan 1 band pankromatik dalam format GeoTiff yang dikompresi sehingga menjadi satu file. Kompresi biasanya dalam bentuk “.tar.gz”. Data input yang terkompresi ini harus diekstrak terlebih dahulu agar dapat diolah baru kemudian dilakukan proses konversi dari file yang terpisah dalam beberapa band menjadi satu file dengan format raster. Format ini terdiri atas 2 buah file, yaitu file data binari dan header yang disimpan dalam file berekstensi “.ers”. Alasan pemilihan format tersebut adalah karena format inilah yang dapat dibaca oleh modul – modul koreksi yang dikembangkan di Pustekdata sebagai input.

Proses koreksi selanjutnya adalah koreksi radiometrik dengan metode *Top Of Atmosphere correction* dengan formula sebagai berikut :

$$dnh_i = (\pi \times d_s^2 \times r_i) / (e \times \cos \theta_i) \times s1_i + s2_i \quad (2-1)$$

dengan :

dnh_i : digital number hasil koreksi radiometri

r_i : nilai radian dari band citra Landsat ke -i

d_s : jarak citra ke matahari

θ_i : sudut zenit dari tiap pixel

e : nilai exoatmospheric irradiance

$s1_i, s2_i$: Faktor skala untuk mengkonversi hasil koreksi menjadi citra 8-bit dari band ke-i

sedangkan,

$$r_i = dn_i \times g_i + o_i$$

dimana:

dn_i : nilai digital number dari citra landsat band ke-i

g_i : nilai gain dari citra landsat band ke-i

o_i : nilai offset dari citra landsat ke-i

Proses selanjutnya adalah mengubah proyeksi file citra Landsat tersebut yang semula menggunakan proyeksi UTM ke proyeksi *latitude longitude* dengan rumus sebagai berikut :

$$\theta = X' - (VII)q^2 + (VIII)q^4 - (D6)q^6$$

$$\Delta\lambda'' = (IX)q - (X)q^3 + (E5)q^5$$

$$\lambda = MT + \Delta\lambda \quad (\text{untuk titik di sebelah timur meridian tengah})$$

$$= MT - \Delta\lambda \quad (\text{untuk titik di sebelah barat meridian tengah})$$

Tujuan pengubahan proyeksi citra ke dalam *latitude-longitude* ini adalah karena daerah Indonesia dibagi menjadi beberapa zona UTM sehingga tidak dapat dimosaik menjadi satu file saja.

Kemudian langkah selanjutnya adalah melakukan koreksi *cloud masking* untuk melakukan *masking* pada awan yang terdapat pada citra. Untuk melakukan deteksi awan yang tebal digunakan nilai albedo (nilai kira-kira reflektan yang didapat dari band 1,2 dan 3 pada citra landsat) dan dengan melakukan deteksi suhu awan menggunakan kanal *thermal* dari citra Landsat. Untuk menghaluskan hasil cloud masking maka digunakan parameter – parameter seperti deteksi bayangan awan dengan nilai reflektan dari band 4 dan 5 citra Landsat, relasi spatial antara awan dan bayangan awan, *water indeks*, dan *bare land index*.

Pada penelitian sebelumnya di Pustekdata juga telah dinyatakan bahwa metode paralel pengolahan yang digunakan adalah dengan melakukan proses pembagian kerja pengolahan berdasarkan citra secara bersamaan karena proses paralel tersebut akan lebih efektif dibandingkan pembagian kerja pengolahan berdasarkan pembagian pengolahan berdasarkan kanal citra secara paralel. Karena metode-metode dan modul – modul yang diperlukan dalam pembuatan mosaik tahunan sudah tersedia, maka kemudian dibutuhkan suatu desain yang dapat merangkai modul – modul tersebut ke dalam suatu sistem dengan sistem operasi linux dan dalam lingkup komputasi paralel.

3. Alat dan bahan

Penelitian ini menggunakan 2 buah sistem sebagai perbandingan untuk menjalankan modul – modul koreksi. Yang pertama adalah server pengolahan dengan sistem operasi *windows server 2012* dengan sebuah storage yang dihubungkan dengan *fibre channel* dengan kecepatan *transfer data* hingga 8 Gbps. Berikut adalah spesifikasi dari *server* tersebut :

Tabel 3-1. Spesifikasi server

Kecepatan Processor	2 buah Intel Xeon E5 2690 2,9 GHz
Jumlah Inti processor	32 Inti
Sistem Operasi	Windows Server 2012 64 bit
Memori	48 GB
Storage	1,5 TB internal 7200 rpm
Direct Attach Storage	18 TB RAID 50 10000 rpm

Kemudian mesin kedua adalah HPC dengan sistem operasi *Rocks Cluster Distribution*, yaitu sistem operasi berbasis Linux khusus untuk *cluster*. *Cluster* sendiri adalah suatu kesatuan dari beberapa komputer yang mampu bekerja bersama melalui jaringan untuk mengatasi keterbatasan penggunaan satu buah mesin saja. Dalam hal ini cluster terdiri dari 16 buah blade server. Masing – masing blade server memiliki spesifikasi sebagai berikut:

Tabel 3-2. Spesifikasi HPC

Kecepatan Processor	2 buah Intel Xeon E5 2680 2,7 GHz
Jumlah Inti Processor	32 Inti
Sistem Operasi	Rocks Cluster Distribution 64 bit
Memori	16 GB
Storage	1 TB internal 7200 rpm
Direct Attach Storage	9 TB RAID 10 10000 rpm

Kemudian data yang digunakan adalah data Landsat 7 level L1T yang terdapat di bidang Teklahta antara tahun 2010 – 2012 dengan rincian sebagai berikut :

Tabel 3-3. Jumlah data Landsat

Data Landsat tahun	Jumlah (scene)
2010	559
2011	875
2012	688

4. Hasil dan Pembahasan

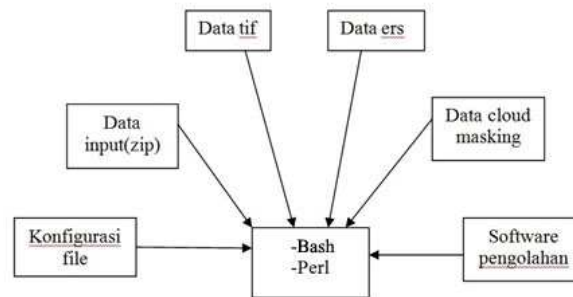
4.1 Desain otomatisasi

Proses koreksi citra Landsat L1T adalah proses bertingkat melalui beberapa modul yang merupakan aplikasi dari sebuah metode koreksi. Beberapa modul yang dibuat adalah sebagai berikut :

Tabel 4-1. Modul yang digunakan

Nama Modul	Fungsi
Mpi_tiff2ers	Mengubah file geotiff menjadi raster (.ers) sekaligus melakukan koreksi radiometrik
Mpi_utm2ll	Mengubah proyeksi UTM menjadi latitude longitude
Mpi_cs_ll	Melakukan cloud masking
Mpi_mosaik	Melakukan mosaicking
qmos	Menggabung semua file mosaick
Mpi_create_bmp	Membuat quicklook
Bmp_x	Membuat quicklook mosaick

Modul – modul tersebut adalah bagian dari satu kesatuan proses, dimulai dari tiff2ers yang input nya adalah *raw data* yang kemudian outputnya adalah file raster yang merupakan masukan bagi mpi_utm2ll artinya output dari modul sebelumnya akan menjadi input dari modul setelahnya sehingga proses ini adalah *sequential*. Karena itulah metode untuk melakukan kerja secara paralel adalah dengan melakukan salah satu modul untuk melakukan koreksi pada beberapa citra sekaligus. Kemudian karena langkah – langkah proses koreksi tersebut adalah *sequential* dan telah ditentukan file – file yang menjadi input dan output dari masing – masing modul, maka agar proses dapat berjalan secara otomatis, maka modul – modul tersebut dapat diotomatisasi dengan menggunakan bahasa pemrograman. Ide otomatisasi proses koreksi tersebut seperti gambar berikut:

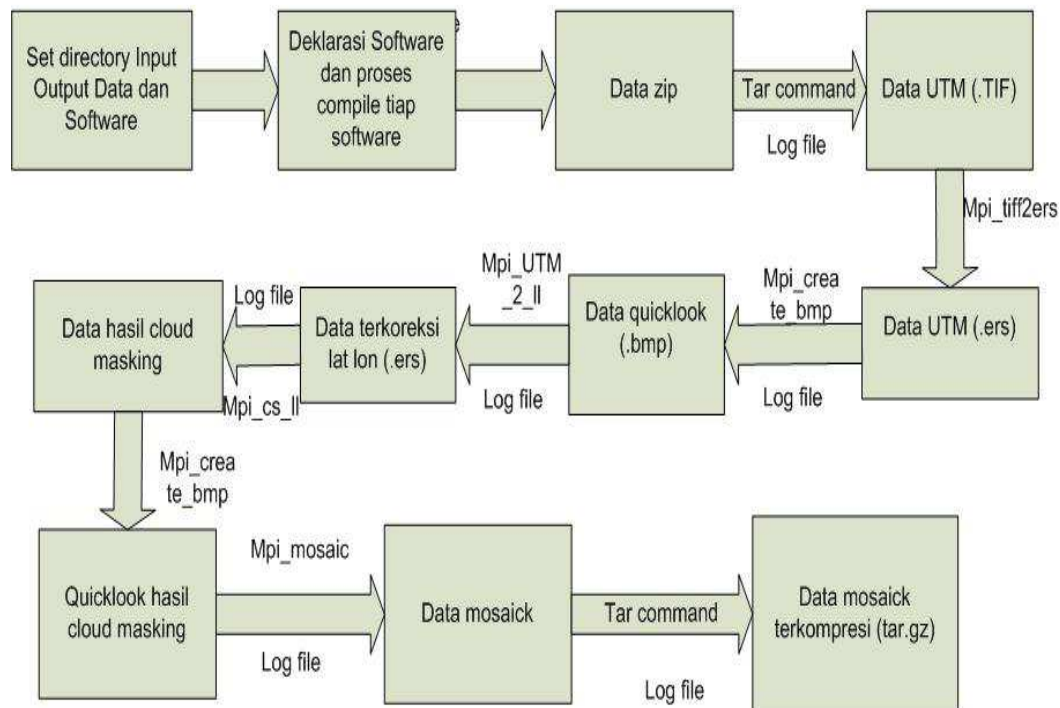


Gambar 4-1. Ide desain otomatisasi

Ide desain otomatisasi yang diusulkan adalah dengan membuat sebuah modul yang berisi informasi modul – modul pengolahan, input dan output serta lokasi nya pada hard disk. Konfigurasi file adalah bagaimana agar modul tersebut dapat dikonfigurasi untuk melakukan koreksi data – data Landsat tertentu saja dan dimanakah file input dan output dari proses tersebut. Bahasa pemrograman yang diusulkan adalah

bash dan perl, kedua bahasa pemrograman tersebut sudah terdapat pada mesin dengan sistem operasi Linux.

Kemudian dibuat implementasi pada diagram alir modul otomatisasi seperti tampak pada gambar :



Gambar 4-2. Diagram alir modul otomatisasi

Pada awal program, dilakukan inisialisasi direktori yang menjadi lokasi input dan output data serta modul – modul yang digunakan pada *hard disk*. Kemudian dilakukan proses *compile* tiap – tiap modul, hal ini bertujuan untuk menghindari adanya kesalahan modul yang digunakan. Kemudian input data yang berupa data terkompresi akan di ekstrak terlebih dulu dengan menggunakan command “tar” pada linux. Kemudian data GeoTiff yang masih menggunakan proyeksi UTM akan diubah ke dalam file raster sekaligus dilakukan koreksi radiometrik dengan modul “Mpi_tiff2ers”. File raster tersebut akan menjadi input dari modul “mpi_UTM_2_ll” yang akan mengubah proyeksi citra dari UTM ke *latitude longitude*. Diantara proses tersebut terdapat proses pembuatan quicklook berupa file bitmap dengan kombinasi band 453, pada modul otomatisasi ini, tiap proses akan dibuat quicklook nya. Selain quicklook, pada transisi antara modul koreksi satu dan yang lain dibuat file log. Tujuan pembuatan file log ini adalah sebagai mekanisme pemantauan jika terjadi kegagalan seperti masalah hardware atau jaringan selama proses pengolahan. Karena proses ini berjalan otomatis, maka file log ini akan sangat berguna jika terjadi kegagalan untuk melakukan penelusuran sampai dimana proses pada saat terjadi kegagalan yang disebabkan hardware atau karena sebab yang lain.

Kemudian modul selanjutnya adalah “mpi_cs_ll” yang akan melakukan cloud masking pada citra yang selanjutnya file citra yang sudah melalui proses cloud masking ini akan di mosaick, sehingga mosaick yang dihasilkan adalah yang bebas awan. Hal ini karena modul “mpi_mosaic” selain melakukan

mosaicking juga memiliki algoritma untuk mengisi data yang kosong (disebabkan masking yang dibuat untuk menutupi awan) dengan data citra yang sama dari tanggal yang berbeda. Tahap terakhir dari program ini adalah membuat quicklook dari mosaick yang dihasilkan dengan modul “Bmp_x” kemudian file mosaick akan dikompresi dengan “tar command” pada linux untuk memudahkan transfer data dari HPC ke storage server.

4.2 Peningkatan kecepatan pengolahan

Berikut adalah tabel lama waktu pengolahan yang dilakukan oleh single server untuk tiap modul untuk melakukan proses pada satu citra :

Tabel 4-2. Waktu proses tiap modul

Nama Modul	Lama waktu (detik)
Mpi_tiff2ers	45
Mpi_utm2ll	60
Mpi_cs_ll	80
Mpi_create_bmp	30
Bmp_x	30

Karena pada single server modul yang dijalankan adalah sekuensial, maka waktu pengolahan akan semakin lama dengan makin banyaknya data yang diproses karena modul harus melakukan pemrosesan data satu demi satu. Untuk modul mosaick tahunan, lama waktu yang dibutuhkan cukup bervariasi tergantung data yang digunakan, tetapi paling tidak dibutuhkan 4 sampai dengan 6 jam untuk melakukan mosaick pada sekitar 500 citra. Sehingga waktu yang dibutuhkan untuk pengolahan data mosaick tahunan pada single server adalah seperti tampak pada tabel 6 berikut :

Tabel 4-3. Lama waktu proses data

Data Landsat tahun	Jumlah (scene)	Lama waktu (jam)
2010	559	47.7
2011	875	71.8
2012	688	56.5

Sedangkan lama waktu pengolahan dengan HPC tampak pada tabel 7 :

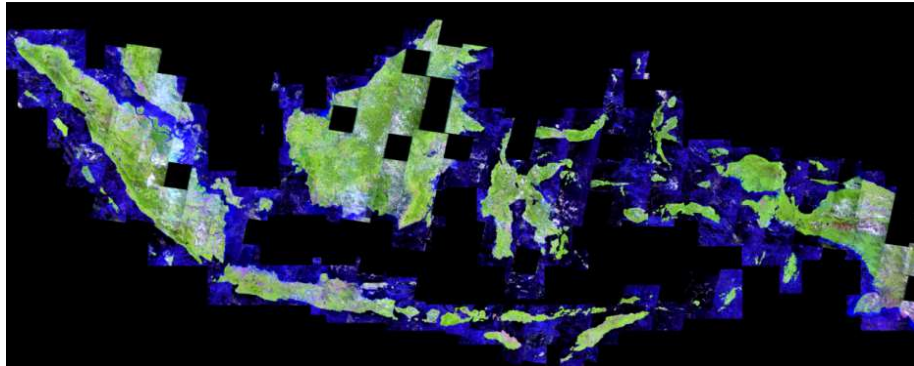
Tabel 4-4. Waktu proses tiap modul

Data Landsat tahun	Jumlah (scene)	Lama waktu (jam)
2010	559	3.5
2011	875	4
2012	688	3.5

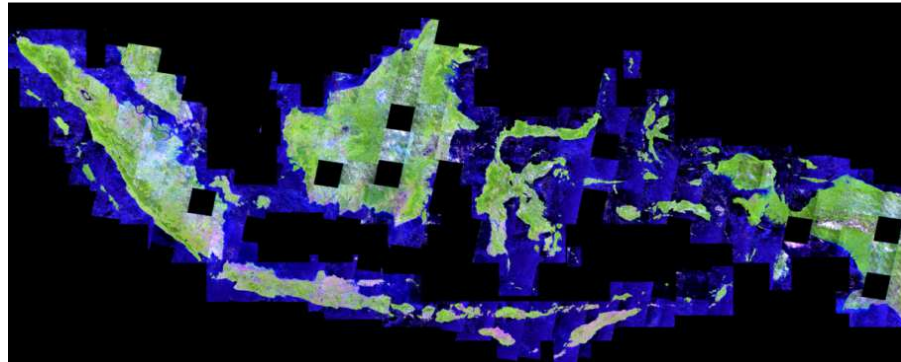
Dari tabel dapat dilihat peningkatan kecepatan yang didapat dengan penggunaan HPC menggunakan modul berbasis metode paralel programming per scene adalah 15x sampai 18x lebih cepat dibanding penggunaan single server.

4.3 Hasil pengolahan

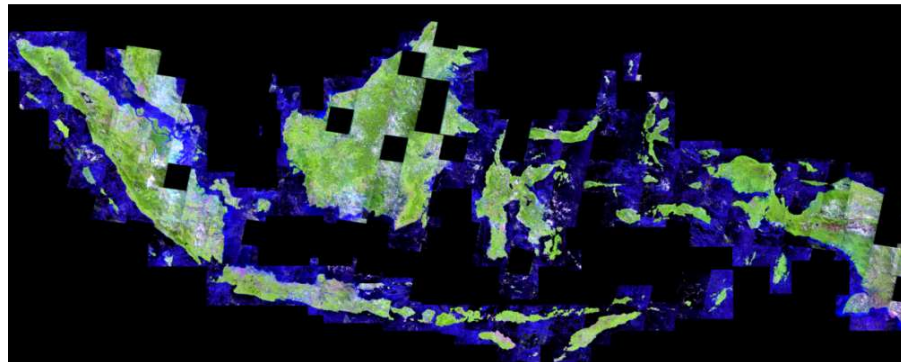
Berikut adalah quicklook hasil mosaick Landsat LIT tahun 2010, 2011 dan 2012



Gambar 4-3. Hasil mosaick Landsat 7 LIT tahun 2010



Gambar 4-4. Hasil mosaick Landsat 7 LIT tahun 2011



Gambar 4-5. Hasil mosaick Landsat 7 LIT tahun 2012

5. Kesimpulan

- Modul – modul pengolahan pada single server berbasis windows telah berhasil dikonversi ke dalam sistem berbasis Linux dan ke dalam lingkup pemrograman paralel.
- Telah dibuat desain otomatisasi untuk proses pengolahan Landsat 7 LIT.

- Peningkatan kecepatan pengolahan yang didapat dengan menggunakan HPC dan modul berbasis metode paralel programming per scene adalah 15x sampai 18x.

6. Daftar Rujukan

- Asanovic, Krste et al. (December 18, 2006). "The Landscape of Parallel Computing Research: A View from Berkeley" (PDF). University of California, Berkeley. Technical Report No. UCB/EECS-2006-183.
- Bader, David; Robert Pennington (June 1996). "Cluster Computing: Applications". Georgia Tech College of Computing. Retrieved 2007-07-13.
- Hennessy, John L.; Patterson, David A., Larus, James R. (1999). Computer organization and design : the hardware/perangkat lunak interface (2. ed., 3rd print. ed.). San Francisco: Kaufmann. ISBN 1-55860-428-6.
- Rocchio Laura.; Headley Rachel. (2008)About Landsat Data
- Dimiyati, Ratih Dewanti, 2010. The Use of Remote Sensing Data to Support Indonesian National Carbon Accounting System (INCAS). Makalah dalam seminar The Role of Space Technology in Addressing Climate Change. Melbourne : The Asia-Pacific Regional Space Agency Forum (APRSAP).
- Geometri Correction Of Remotely Sensed Imagery Remotely Sensed Imagery. Ramsey, R. Douglas. (http://www.gis.usu.edu/~doug/RS5750/lectures/L7_Geometriccorr.pdf, diakses 11 April 2011)
- Converting Digital Numbers to Top of Atmosphere (ToA) Reflectance (http://www.yale.edu/ceo/Documentation/Landsat_DN_to_Reflectance.pdf, diakses 11 April 2011)
- Sulyantara, Heri, Pengembangan Metode Cloud Masking dan Mosaicing Data Penginderaan Jauh Landsat Untuk Mendukung Indonesias National Carbon Accounting System (INCAS) (2011)