

PENGEMBANGAN MODUL MIGRASI DATA INDERAJA MENGUNAKAN METODE PEMROSESAN PARALEL DALAM LINGKUP *BIG DATA*

(*DEVELOPMENT OF REMOTE SENSING DATA MIGRATION MODULE USING PARALLEL PROCESSING METHODS IN BIG DATA ENVIRONMENT*)

Riyan Mahendra Saputra^{1,a}, Kartika Eka Yanindra¹, Emalia Efendy¹

¹Pusat Teknologi dan Data Penginderaan Jauh, LAPAN

^aKontributor Utama

e-mail: riyan.lpn@gmail.com

Diterima 16 November 2018; Direvisi 22 Maret 2019; Disetujui 5 Juni 2020

ABSTRACT

Nowadays, the era of information technology, data that used and produced by an organization in carrying out its business processes are getting big and continuously increasing so that it requires a good method in its management. Remote sensing data is data generated by several types of sensors (active and passive) satellites. Generate data stored in several media, formats, and sizes. Pustekdata LAPAN carries out one of its functions in managing remote sensing image data at a National Remote Sensing Data Bank, including data migration. The problem faced in data migration is slow to access, and the process using a folder-based sequential approach. The purpose of this study is to develop a data migration module using parallel processing methods in a big data environment to improve the performance of existing migration systems to be more efficient, effective, fast, easy, and safe. The data used for testing are medium resolution data of Landsat 8 with a size of 780 MB/scene, high-resolution SPOT 6 and 7 with a size of 300 MB - 5 GB/scene, and very high-resolution Pleiades with a size of 2 - 5 GB/scene. Migration testing does by using 10-100 samples data for each type of data. The experimental results showed an increase of 12,8 times faster in a 1Gbps network environment compared with the previous migration module process and 4,8 times faster in a 10 Gbps network environment compare with 1 Gbps network.

Keywords: *big data, data management, parallel processing, imagery data, remote sensing*

ABSTRAK

Era teknologi informasi sekarang ini, data yang digunakan dan dihasilkan oleh suatu organisasi dalam menjalankan proses bisnisnya semakin besar dan terus meningkat sehingga membutuhkan metode yang baik dalam pengelolaannya. Data penginderaan jauh merupakan data yang dihasilkan dari beberapa jenis sensor satelit aktif maupun pasif. Data tersebut disimpan dalam beragam media, format dan ukuran. Pustekdata LAPAN melakukan pengelolaan seluruh data citra penginderaan jauh yang dimilikinya di sistem Bank Data Penginderaan Jauh Nasional, termasuk migrasi data. Permasalahan yang dihadapi dalam migrasi data adalah akses yang lambat, terutama karena prosesnya masih menggunakan pendekatan sekuensial berbasis *folder*. Tujuan penelitian ini adalah mengembangkan modul migrasi data menggunakan metode pemrosesan paralel dalam lingkup *big data* untuk meningkatkan performa sistem migrasi yang ada menjadi lebih efisien, efektif, cepat, mudah dan aman. Pada pengujian ini digunakan data resolusi menengah Landsat 8 dengan ukuran 780 MB/*scene*, resolusi tinggi SOPT 6 dan 7 dengan ukuran 300 MB – 5 GB/*scene* dan resolusi sangat tinggi Pleiades dengan ukuran 2-5 GB/*scene*. Pengujian dilakukan dengan menggunakan 10-100 data sampel untuk setiap jenis data. Hasil percobaan menunjukkan peningkatan kecepatan 12,8 kali lebih cepat dari modul migrasi sebelumnya di lingkungan jaringan 1 Gbps dan 4,8 kali lebih cepat dalam lingkungan jaringan 10 Gbps dibandingkan 1 Gbps.

Kata kunci: *big data, pengelolaan data, pemrosesan paralel, data citra, penginderaan jauh*

1 PENDAHULUAN

Era teknologi informasi sekarang ini, hampir setiap organisasi menghadapi permasalahan dalam pengelolaan data. Data tersebut menjadi besar dan bertambah diakibatkan oleh proses atau aktifitas organisasi dalam menjalankan tugas dan fungsinya, sehingga diperlukan pengelolaan data yang baik untuk menangani jumlah data yang besar (*big data management*) tersebut. Kegiatan penginderaan jauh terdiri dari kegiatan akuisisi data, pengolahan data dan pengelolaan data dari beberapa jenis sensor satelit aktif dan pasif (optis, radar dan lidar). Kegiatan tersebut menghasilkan data citra dengan format dan ukuran yang beragam serta terus bertambah sehingga memerlukan sistem pengelolaan yang baik. Pengelolaan data penginderaan jauh di PUSTEKDATA LAPAN dikelola dalam sistem Bank Data Penginderaan Jauh Nasional (BDPJN).

Data penginderaan jauh dapat dikategorikan berdasarkan resolusi spasial, yaitu resolusi rendah, menengah, tinggi dan sangat tinggi (Omali, 2018). Penyimpanan data penginderaan jauh berbasis *folder* disimpan dalam berbagai media penyimpanan. Sistem BDPJN terdiri dari beberapa subsistem yaitu, akuisisi data, pengolahan data, penyimpanan data dan distribusi data. Proses perpindahan data atau migrasi data antar subsistem dan media penyimpanan menjadi hal yang utama dan begitu juga dengan migrasi data antar media penyimpanan karena keterbatasan umur media penyimpanan.

Pengelolaan data penginderaan jauh di BDPJN telah dilakukan sejak tahun 2012 dan terus diperbaiki sampai saat ini. Permasalahan yang dihadapi dalam pengelolaan data di sistem BDPJN saat ini yaitu, pengaturan media penyimpanan, perpindahan file/data antar jenis media penyimpanan dan akses data.

Tabel 1-1 : PROSES MIGRASI DATA CITRA INDERAJA (Pustekdata, 2013)

No	Resolusi	Satelit	Kelas Data	Ukuran File	Waktu Migrasi
1	Resolusi Menengah	LANDSAT 8	H	780 MB	12 menit/ data
2	Resolusi Tinggi	SPOT 6	H	250 MB - 4 GB	14-35 menit/ data
3	Resolusi Tinggi	SPOT 7	H	250 MB - 5 GB	14-35 menit/ data
4	Resolusi Sangat Tinggi	PLEIADES	H	2 GB - 6 GB	27-50 menit/ data

Proses perpindahan atau migrasi data menggunakan aplikasi eksisting (*Migration Manager KIT* (PUSTEKDATA, 2013)) memerlukan waktu proses yang cukup lama ditunjukkan pada Tabel 1-1. Hal ini membutuhkan proses perbaikan pengelolaan penyimpanan data yang baik dalam migrasi data di sistem BDPJN.

Proses migrasi data eksisting dilakukan secara berurut (sekuensial). Proses yang dilakukan antara data yang satu dengan data lainnya tidak saling bergantung sehingga dapat diparalelkan. Pemanfaatan *multicore processor* (satu server) atau *multiworker* (beberapa server) dapat meningkatkan waktu proses dan memproses banyak data menggunakan kemampuan yang ada. Proses yang berjalan secara paralel bernilai optimal jika waktu yang diperlukan untuk menyelesaikan suatu masalah berbanding lurus dengan jumlah pemrosesnya. Upaya yang dilakukan untuk mencapai tingkat optimal pada pemrosesan secara paralel adalah sia-sia kecuali diikuti oleh pencapaian pada proses sekuensial yang hampir sama besarnya (Amdahl, 1967). Kamilah dan Saputra (2015), melakukan percepatan proses publikasi data menggunakan *parallel programming* dengan memaksimalkan jumlah *core processor* namun masih dapat ditingkatkan dan belum dalam lingkup *Big Data*.

Merujuk pada model 3V, definisi *big data* dibedakan berdasarkan ukuran (*volume*) terkait dengan pengelolaan penyimpanan, kecepatan (*velocity*) terkait akses atau operasional penyimpanan dan keragaman (*variety*) tentang pengelolaan data (Laney, 2001). Demchenko, *et al.* (2013) memperbaiki definisi *Big Data* dengan menambahkan 2 unsur yaitu nilai (*value*) dan ketelitian (*veracity*) menjadi 5V. Unsur tersebut didapatkan dari hasil inisiasi klasifikasi data dan pemrosesan pada model yang lebih spesifik.

Menurut Chi, *et al.* (2016) tantangan besar yang dihadapi metode *big data* dalam penginderaan jauh dibagi menjadi 3 yaitu, komputasi (meliputi perpindahan data, penyimpanan data dan distribusi data), metodologi (meliputi penyajian data, penggabungan data dan visualisasi data) dan aplikasi (identifikasi data, interpretasi data dan sebaran data).

Chi, *et al.* (2016) mengajukan kerangka kerja trinitas dalam memahami permasalahan *big data* di penginderaan jauh yaitu, kepemilikan data, metode pengolahan data dan eksploitasi *big data* untuk menyelesaikan permasalahan yang nyata.

Teknologi *big data* menurut Moorthy, *et al.* (2014) dapat diturunkan menjadi 2 komponen utama, yaitu perangkat keras dan perangkat lunak. Komponen perangkat keras merupakan komponen infrastruktur yaitu *server*, *storage* dan sistem jaringan, sedangkan komponen perangkat lunak berfungsi untuk melakukan pengelolaan dan pengaturan data serta pendukung pengambilan keputusan. Perangkat yang digunakan dapat dikategorikan berdasarkan jenis data yang digunakan untuk tujuan tertentu, seperti analisis data.

Penelitian ini bertujuan untuk melakukan perbaikan di sistem BDPJN dan dibatasi untuk pengembangan modul migrasi data menggunakan metode pemrosesan paralel dalam lingkup *big data* untuk meningkatkan performa sistem migrasi yang ada menjadi lebih efisien dan efektif.

2 METODOLOGI

Data menurut Moorthy, *et al.* (2014) dapat diklasifikasikan menjadi 3, yaitu data terstruktur (*structured*) yang dapat dikelompokkan menjadi suatu skema relasi kolom dan baris di dalam *database*; data semi struktur (*semi structured*) yang terbentuk dari data terstruktur namun tidak memiliki skema yang tetap seperti *weblog*; data tidak terstruktur (*unstructured*) adalah data yang tidak bisa dilakukan pembuatan indeks ke dalam tabel relasi secara mudah untuk analisa dan pencarian seperti data citra, audio dan video. Data citra penginderaan jauh terdiri dari file *image/raster* secara individual termasuk kategori data tidak terstruktur. Sedangkan metadata dan file informasi lainnya merupakan data semi struktur dan keduanya adalah bagian dari data spasial. Data citra penginderaan jauh di sistem BDPJN terdiri dari data citra optis (resolusi rendah sampai resolusi sangat tinggi) dan data radar. Data tersebut memiliki ukuran data yang beragam, ditunjukkan pada Tabel 2-1.

Data citra yang digunakan dalam penelitian ini dibatasi hanya 4 jenis satelit dari 22 jenis satelit yang ada, dengan kriteria data yang masih diakuisisi oleh stasiun bumi dan ukuran file yang besar. Data yang digunakan ditunjukkan pada Tabel 1-1.

Tabel 2-1: DATA CITRA PADA SISTEM BDPJN

No	Resolusi	Satelit	Ukuran File
01	Rendah	NOAA	300 MB
02	Rendah	MODIS TERRA	5,5 GB
03	Rendah	MODIS AQUA	5,5 GB
04	Rendah	NPP-NPOESS	10 GB
05	Menengah	LANDSAT 5	350MB
06	Menengah	LANDSAT 7	560MB
07	Menengah	LANDSAT 8	780 MB
08	Menengah	ALOS(AVNIR, PRISM)	280MB
09	Menengah	SPOT 2	60 MB
10	Menengah	SPOT 4	60 MB
11	Tinggi	SPOT 5	240 MB
12	Tinggi	SPOT 6	250 MB*
13	Tinggi	SPOT 7	250 MB *
14	Sangat Tinggi	IKONOS	600 MB *
15	Sangat Tinggi	QUICKBIRD	1,5 GB *
16	Sangat Tinggi	PLEIADES	2 GB *
17	Sangat Tinggi	WORLDVIEW	4 GB *
18	Sangat Tinggi	GEO EYE	4 GB *
19	Sangat Tinggi	RAPID EYE	600 MB*
20	Menengah	RADARSAT	410 MB
21	Tinggi	TERRASAR-X	1,1 GB
22	Menengah	ALOS PALSAR	750 MB

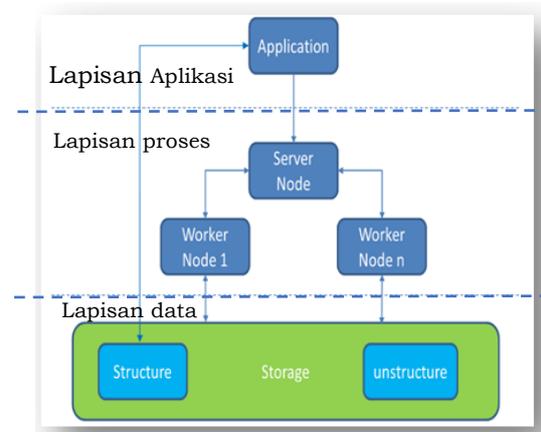
* Ukuran bervariasi tergantung luasan

Demchenko, et al. (2014) mengajukan *Big Data Architecture Framework (BDAF)* yang terdiri dari lima komponen dalam lingkungan sistem *big data*, yaitu, model, struktur dan jenis data (*format, non/relational, file system, etc*), pengaturan *big data* (siklus hidup, transformasi, klasifikasi *big data*), analisis dan perlengkapannya (aplikasi *big data*, target penggunaan, penyajian, visualisasi), infrastruktur *big data* (media penyimpanan, komputasi, jaringan,) dan keamanan *big data* (keamanan dalam proses, perpindahan, lingkungan pemrosesan yang dipercaya).

Lee dan Kang (2015) mengajukan desain arsitektur yang terdiri dari 3 lapisan yaitu, lapisan pengelolaan dan integrasi data spasial, lapisan analisa dan lapisan layanan dalam lingkup *big data*. Lapisan pengelolaan dan integrasi

data spasial mempunyai peran dalam penyimpanan, pengambilan, pembuatan indeks, dan pencarian data secara cepat. Lapisan analisa mempunyai tanggung jawab melakukan analisis data spasial (analisis statistik dan analisis citra). Lapisan layanan mempunyai tugas menyiapkan layanan untuk analisis, layanan berbagi pakai, pemantauan dan pengelolaan layanan yang bisa diakses oleh sistem lain.

Desain arsitektur aplikasi yang diusulkan menggunakan komponen-komponen dalam kerangka kerja arsitektur *big data* (Demchenko, et al., 2014) dan menggabungkan dengan pendekatan *three tier structure*. Lapisan data merepresentasikan tempat data (terstruktur dan tidak terstruktur); lapisan proses merupakan lapisan *middle ware* yang menghubungkan lapisan atas dengan bawah yang menjalankan eksekusi perintah dari lapisan atas; lapisan aplikasi menjadi *interface* dengan pengguna maupun sistem lainnya dalam pengaksesan data dalam aplikasi tertentu.



Gambar 2-1: Desain arsitektur aplikasi

Desain arsitektur sistem yang diajukan terlihat pada Gambar 2-1 mengadopsi desain arsitektur Lee, et al. (2017) dengan perbedaan dalam penanganan *file system*. Pengembangan modul migrasi data penginderaan jauh menggunakan metode pengembangan aplikasi (Kendall dan Kendall (2006), Jogyanto (2013)) dan mengadopsi metode pemrosesan paralel dalam lingkup *big data* yang dikembangkan. Percobaan dilakukan dengan menggunakan 3 buah server untuk pemrosesan dengan spesifikasi

processor 8 core, memory 16 GB dan 1 PC untuk modul aplikasi *client*.

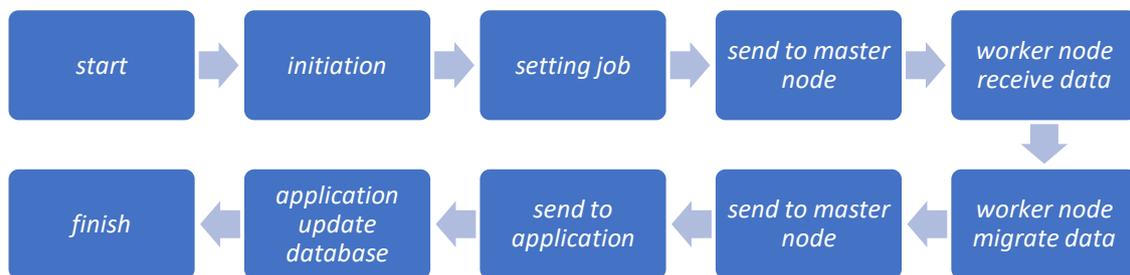
3 HASIL DAN PEMBAHASAN

Desain modul migrasi data terhubung dengan *database* BDPJN, aplikasi dibuat menggunakan bahasa pemrograman *Java* dan *Scala* untuk mengirimkan perintah ke lingkungan *server* pemrosesan (*Apache Spark*). Dalam lingkungan *server* pemrosesan yang telah dibuat memiliki satu *node Master* dan dua *node Worker*. *Node Master* menerima perintah dari aplikasi *client*, perintah tersebut didistribusikan ke *node Worker* untuk dieksekusi.

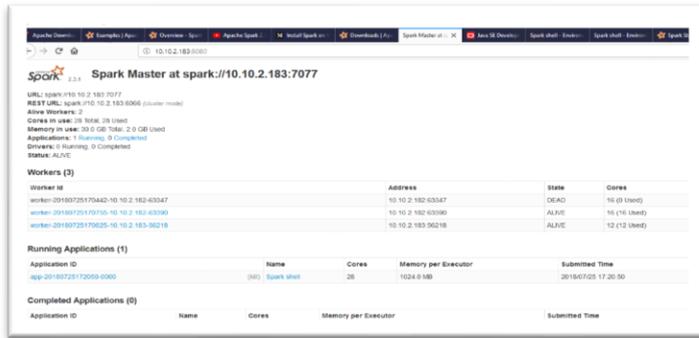
Identifikasi proses migrasi data yang ada di BDPJN diawali dengan pemilihan data berdasarkan *folder* data citra penginderaan jauh. Dilanjutkan pencarian secara manual oleh operator di dalam *folder* media penyimpanan. Akses langsung ke dalam media penyimpanan akan berdampak pada keamanan data terutama rentan terhadap ancaman virus dan ancaman lainnya, sehingga akses langsung perlu dibatasi dan digantikan melalui aplikasi. Setelah proses pemilihan *folder* data yang akan dimigrasi dilanjutkan dengan proses migrasi yang dilakukan secara berurut (sekuensial).

Alur Migrasi data dari sistem yang dikembangkan tampak pada Gambar 3-1. Tahap inisiasi merupakan proses awal untuk koneksi *database*. Tahapan selanjutnya adalah melakukan *setting job* untuk memilih jenis data, membuat

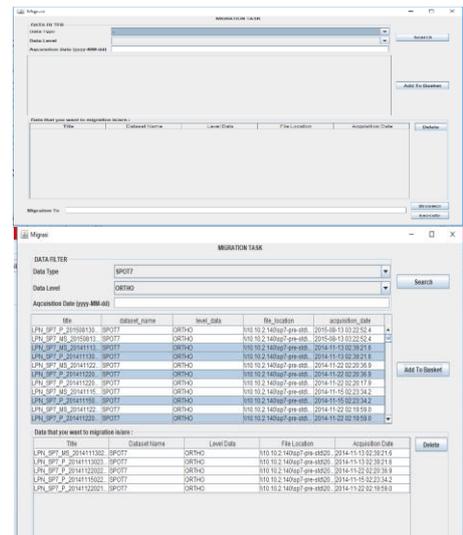
daftar data terpilih, dan menentukan tujuan migrasi. Dilanjutkan dengan *Send to Master Node*, setelah mendapatkan daftar data dalam bentuk *.txt* yang berisikan data dan tujuan migrasi. Daftar data dikirim ke *spark master* dan diubah dalam bentuk RDD (*resilient distributed dataset*) kemudian dibuatkan partisi di dalam RDD. Tahapan selanjutnya *Worker node receive data* menunjukkan bahwa *worker node* menerima RDD dari *master*. *Worker Node Migrate* data akan mengerjakan perintah berdasarkan RDD yang diberikan oleh *master*. Dilanjutkan dengan melakukan migrasi data ke folder tujuan sesuai dengan informasi yang diterima. Kemudian setiap *worker* akan membuat *log* keterangan lama waktu proses pekerjaan (berisi range waktu pengerjaan, status berhasil/gagal). Dilanjutkan dengan *Send to Master* untuk mengirimkan *log* yang sudah dibuat. *Master* juga menghitung berapa lama waktu yang dibutuhkan untuk menyelesaikan pekerjaan yang diberikan. Kemudian *Send to Application* merupakan tahap dimana *master* mengirimkan hasil *log* ke aplikasi, tahap selanjutnya aplikasi melakukan *update database* dari hasil *log* yang diterima. Gambar 3-2(a) menunjukkan halaman *Master* berisikan informasi kondisi lingkungan *Apache Spark* yang terdiri dari informasi *worker* dan aplikasi yang berjalan. Gambar 3-2(b) menunjukkan aplikasi yang dipakai oleh pengguna untuk melakukan pencarian data dan melakukan perintah migrasi.



Gambar 3-1: Alur proses migrasi data menggunakan metode paralel dalam lingkup *Big Data*



(a)



(b)

Gambar 3-2: (a) Master manajemen; (b) Aplikasi migrasi data

Percobaan yang dilakukan menggunakan data Landsat 8, SPOT 6, SPOT 7 dan Pleiades dengan jumlah sampel yang digunakan sebanyak 10, 20, 30, 40, 50, 60, 70, 80, 90 dan 100 sampel. Percobaan dilakukan dengan menggunakan koneksi jaringan 1 Gbps dan 10 Gbps. Jika menggunakan pendekatan *parallel programming*, peningkatan yang diperoleh tergantung

dari desain aplikasi dan jumlah pemrosesannya saja (Kamilah & Saputra, 2015). Jika menggunakan pendekatan *Big Data*, peningkatan tidak hanya diperoleh dari aplikasi dan jumlah pemrosesnya (jumlah *core* prosesor) saja, tetapi bisa ditingkatkan dari aplikasi dan *hardware* pemrosesnya (prosesor, *memory*, perangkat jaringan dan lain-lain).

Tabel 3-1 : HASIL PERCOBAAN MIGRASI DATA CITRA INDERAJA PADA (A) LINGKUNGAN 1GBPS; (B) LINGKUNGAN 10GBPS

No	Jumlah Data	Landsat 8		SPOT 6		SPOT 7		Pleiades	
		Ukuran (GB)	Waktu						
1	10	15,10	3m 42s	3,71	1m 42s	2,5	2m 48s	94,60	21m
2	20	27,20	12m	7,55	3m 24s	5,38	5m 54s	102,00	26m
3	30	39,90	18m	10,60	5m 24s	8,11	8m 42s	111,00	28m
4	40	51,90	22m	13,30	6m	10,9	11m	117,00	31m
5	50	66,00	30m	17,10	8m 18s	13,1	13m	124,00	35m
6	60	81,80	37m	19,40	9m 36s	15,5	16m	144,00	37m
7	70	98,30	44m	22,10	10m	18,3	19m	151,00	37m
8	80	113,00	57m	24,70	11m	21,2	21m	158,00	42m
9	90	127,00	1h	27,20	12m	23,4	24m	203,00	56m
10	100	141,00	44m	29,40	13m	26,2	26m	254,00	1h 20m

(a)

No	Jumlah Data	Landsat 8		SPOT 6		SPOT 7		Pleiades	
		Ukuran (GB)	Waktu						
1	10	15,10	36 s	3,71	30s	2,5	19s	94,60	5m
2	20	27,20	1m 18s	7,55	37s	5,38	25s	102,00	5m 12s
3	30	39,90	1m 30s	10,60	39s	8,11	36s	111,00	5m 54s
4	40	51,90	2m 12s	13,30	1m 18s	10,9	58s	117,00	6m 6s
5	50	66,00	3m	17,10	1m 18s	13,1	1m 24s	124,00	6m 30s
6	60	81,80	3m 24s	19,40	2m	15,5	1m 36s	144,00	7m 30s
7	70	98,30	4m	22,10	2m 6s	18,3	1m 48s	151,00	8m 36s
8	80	113,00	4m 30s	24,70	2m 18s	21,2	1m 54s	158,00	8m 48s
9	90	127,00	5m	27,20	2m 30s	23,4	2m	203,00	12m
10	100	141,00	5m 54s	29,40	2m 42s	26,2	2m 12s	254,00	16m

(b)

Hasil percobaan menunjukkan kecepatan proses untuk aplikasi menggunakan pendekatan paralel dalam lingkup *big data* ditunjukkan pada Tabel 3-1(a) dan 3-1(b). Waktu pengujian sangat dipengaruhi oleh kecepatan rata-rata transfer dari konfigurasi perangkat yang digunakan, akses ke media penyimpanan dan media penyimpanan. Untuk pengujian dalam lingkungan 1 Gbps, kecepatan rata-rata transfer maksimal 754 Mbps, kemudian dilakukan perbaikan konfigurasi perangkat dengan meningkatkan *buffer memory* pada konfigurasi kartu jaringan dan mengaktifkan mode *jumbo frame* sehingga didapatkan kecepatan rata-rata transfer maksimal 900 Mbps. Perbandingan migrasi data pada lingkungan 1 Gbps sebelum menggunakan metode *big data* (Tabel 1-1) dan sesudah menggunakan metode pemrograman paralel dalam lingkup *big data* (Tabel 3-1(a)) mengalami peningkatan yang signifikan. Singh (2013), mengatakan kecepatan digambarkan sebagai kecepatan yang diperoleh dari membandingkan waktu yang dibutuhkan menyelesaikan pekerjaan oleh satu prosesor dengan multi prosesor (n),

$$S(n) = T(1)/T(n) \dots \dots \dots (1)$$

dengan:

S adalah kecepatan

T adalah waktu pemrosesan

Jika memproses 10 data menggunakan metode sebelumnya, yaitu melalui aplikasi *Migration Management KIT* (MMK), waktu proses migrasi yang diperlukan untuk data Landsat 8 adalah 120 menit, untuk data SPOT 6 dan SPOT 7 adalah 140 menit, dan untuk data Pleiades adalah 270 menit (PUSTEKDATA, 2013). Waktu yang dibutuhkan untuk migrasi data Landsat 8 menggunakan pemrograman paralel dengan 16 *core* prosesor adalah 12 menit (Kamilah & Saputra, 2015). Jika menggunakan metode pemrograman paralel dalam lingkup *Big Data*, waktu yang dibutuhkan untuk migrasi data Landsat 8 adalah 3 menit 42 detik (kecepatan 32,4 kali lebih cepat dibandingkan dengan metode aplikasi yang ada dan 3,24 kali lebih cepat dibanding menggunakan pemrograman

paralel), 1 menit 42 detik untuk data SPOT 6 (kecepatan 82,3 kali lebih cepat dibandingkan dengan metode aplikasi yang ada dan 3,8 kali lebih cepat dibanding menggunakan pemrograman paralel), 2 menit 48 detik untuk data SPOT 7 (kecepatan 50 kali lebih cepat dibandingkan dengan metode aplikasi yang ada dan 3,5 kali lebih cepat dibanding menggunakan pemrograman paralel) dan 21 menit untuk data Pleiades (kecepatan 12,8 kali lebih cepat dibandingkan dengan metode aplikasi yang ada dan 1,28 kali lebih cepat dibanding menggunakan pemrograman paralel). Dalam percobaan menggunakan beberapa jumlah data di lingkungan 1 Gbps, peningkatan kecepatan tidak terlalu signifikan untuk data yang semakin besar ukurannya.

Kecepatan rata-rata transfer di lingkungan 10 Gbps adalah 5 Gbps. Perbandingan migrasi data pada lingkungan 1 Gbps dengan 10 Gbps untuk data Landsat 8 mengalami peningkatan kecepatan rata-rata 10,21 kali lebih cepat, untuk data SPOT 6 mengalami peningkatan kecepatan rata-rata 5,21 kali lebih cepat, untuk data SPOT 7 mengalami peningkatan kecepatan rata-rata 11,35 kali lebih cepat, dan untuk data Pleiades mengalami peningkatan kecepatan rata-rata 4,8 kali lebih cepat.

4 KESIMPULAN

Pembuatan modul migrasi data dalam sistem BDPJN menggunakan metode pemrograman paralel dalam lingkup *Big Data* dengan mengubah desain baik *hardware* maupun *software* aplikasi yang ada. Desain arsitektur aplikasi menggunakan komponen BDAF dengan membagi menjadi tiga lapisan, yaitu lapisan data, lapisan proses dan lapisan aplikasi. Hasil proses migrasi menggunakan pendekatan paralel dalam lingkup *Big Data* mengalami peningkatan kecepatan lebih dari 12 kali dibandingkan aplikasi yang ada dalam lingkungan jaringan 1 Gbps. Peningkatan proses migrasi lebih dari 4,8 kali kecepatan dalam jaringan 10 Gbps dibandingkan dengan aplikasi dengan 1 Gbps.

UCAPAN TERIMAKASIH

Penulis mengucapkan terima kasih kepada Kemenristekdikti yang telah membiayai penelitian ini melalui program Insinas No Kontrak 14/INS-1/PPK/E4/2019.

DAFTAR RUJUKAN

- Amdahl, G.M. (1967). *Validity of the Single Processor Approach to Achieving Large Scale Computing Capabilities*. American Federation of Information Processing Societies (AFIPS) Conference Proceedings (30): 483–485.
- Bernstein, A.J. (1966). *Analysis of Programs for Parallel Processing*. IEEE Transactions on Electronic Computers. EC15 (5): 757–763.
- Chi, M., Benediktsson, A., F., Sun, Z., Shen, J., & Zhu, Y. (2016). *BIG DATA for Remote Sensing: Challenges and Opportunities*, 104(11). <http://doi.org/10.1109/JPROC.2016.2598228>
- Demchenko, Y., De Laat, C., & Membrey, P. (2014). *Defining architecture components of the BIG DATA Ecosystem*. 2014 International Conference on Collaboration Technologies and Systems, CTS 2014, (May), 104–112. <http://doi.org/10.1109/CTS.2014.6867550>
- Jogiyanto. (2013). *Analisis & Desain Sistem Informasi : Pendekatan terstruktur teori dan praktik aplikasi bisnis*. Andi Offset. Yogyakarta
- Kamilah, Anis dan Saputra, R.M . (2015). *Percepatan Proses Publikasi Data di Katalog Bank Data Penginderaan Jauh Nasional (BDPJN) dengan Parallel Programming*. Seminar Nasional Penginderaan Jauh: 186–194.
- K. E. Kendall and J. E. Kendall. (2010). *Systems Analysis and Design, 8th ed*, S. Yagan, Ed. New Jersey:Prentice Hall,
- K. E. Kendall and J. E. Kendall . (2006). *Analisis dan Perancangan Sistem*. Edisi Keempat. Bandung : PT Index.
- Laney, D. (2001). *META Delta. Application Delivery Strategies*, 949(February 2001), 4. <http://doi.org/10.1016/j.infsof.2008.09.005>
- Lee, J., Kang, M., Lee, J., & Kang, M. (2015). *Geospatial BIG DATA: Challenges and Opportunities*. *Procedia Computer Science Geospatial BIG DATA: Challenges and Opportunities*. BIG DATA Research. <http://doi.org/10.1016/j.bdr.2015.01.003>
- Lee, et al. (2017). *Design of Educational BIG DATA application using SPARK*. ICACT 355–357. ISBN 978-89-968650-9-4
- Moorthy, M., Baby, R., & Senthamaraiselvi, S. (2014). *An Analysis for Big Data and its Technologies*. International Journal of Computer Science Engineering & Technology, 5(12), 412.
- Omali, Thomas U. 2018. *Impacts of Sensor Spatial Resolution on Remote Sensing Image Classification*. GSJ: Volume 6, Issue 1, January 2018, Online: ISSN 2320-9186.
- PUSTEKDATA (2013). *Laporan akhir Kegiatan Pengembangan sistem Bank Data Penginderaan Jauh Nasional di Pustekdata*, 2013
- Singh, I. (2013). *Review Article Review on Parallel and Distributed Computing*. Scholars Journal of Engineering and Technology (SJET), 218–225. ISSN 2321-435X