

# ***Big Data dan Komputasi Skala Besar***

## ***Big Data and Large Scale Computing***

**Suryo Bramasto<sup>1,2</sup>, Sunarto<sup>1</sup>**

<sup>1</sup>Program Studi Informatika Institut Teknologi Indonesia, Jl. Raya Puspiptek Serpong, Tangerang Selatan, Banten, 15320

(Diterima: 03 Maret 2016; Disetujui: 16 Agustus 2016)

### **Abstrak**

*Big data analytics dan komputasi skala besar dengan performa tinggi merupakan kebutuhan yang mulai dirasakan baik untuk keperluan ilmiah maupun industri, seiring perkembangan teknologi perangkat keras, perangkat lunak, algoritma komputasi dan pertumbuhan data di berbagai sektor. Ekosistem teknologi informasi global sedang mengalami transisi ke kondisi yang semakin membutuhkan layanan cloud dan rich data analytics. Big data analytics dan komputasi skala besar merupakan dua hal yang berbeda dari sisi analisis kebutuhan, perancangan, implementasi dan evolusi, namun pada kenyataannya kebutuhan terhadap big data analytics dan komputasi skala besar memiliki tingkat kepentingan yang sama serta diperlukan kolaborasi antara keduanya. Pada kenyataannya memang terdapat banyak perbedaan pada perangkat-perangkat dan budaya-budaya dari big data analytics dan komputasi performa tinggi, dimana perbedaan-perbedaan tersebut mengakibatkan kerugian terkait implementasi big data analytics dan komputasi skala besar, sehingga unifikasi terhadap spektrum dari domain-domain penelitian menjadi penting. Tantangan yang harus dihadapi dewasa ini terkait komputasi performa tinggi dan big data analytics adalah kemampuan dalam transmisi data, komputasi fungsi kompleks terhadap data, serta penyimpanan bagian-bagian penting dari data. Diperlukan suatu pendekatan sehingga tantangan-tantangan terkait bagaimana kolaborasi big data analytics dengan komputasi skala besar memberikan manfaat seperti yang diharapkan. Big data analytics dan komputasi skala besar merupakan elemen-elemen penting dari penelitian dan pengembangan komputasi terintegrasi, sehingga keduanya memiliki prioritas yang sama. Standarisasi global terhadap pengembangan lebih jauh dari arsitektur komputer menjadi kebutuhan, berikut keterbukaan ilmiah terhadap proses pengembangan tersebut.*

**Kata Kunci :** Analytics, Big Data, Skala Komputasi

### **Abstract**

*Nowadays, the necessities for big data analytics and high performance large scala computing are emerging as hardware, software, and computational algorithms development, and also more importantly is the growth of data on various sectors. Global information technology ecosystem is undergoing transition towards conditions which increasingly requires cloud services and rich data analytics. Big data analytics and large scale computing are two different things in terms of requirements analysis, design, implementation, and evolution, but the needs for big data analytics and large scale computing has the same level of importance as well as the necessary collaboration between the two. The tools and cultures of high performance computing and big data analytics have diverged, to the detriment of both; unification is essential to address a spectrum of major research domains. The challenges of scale tax the ability to transmit data, compute complicated functions on that data, or store a substantial part of it; new approaches are required to meet these challenges. Big data analytics and large scale computing are major elements for research and development of integrated computing, so both have the same priority. The international nature of science demands further development of*

*advanced computer architectures and global standards for processing data, even as international competition complicates the openness of the scientific process.*

**Keyword :** Analytics, Big Data, Computational Scale

<sup>2</sup>Penulis Korespondensi. Telp: +62 21 7561092; fax: +62 21 7560542  
Alamat E-mail: suryo.bramasto@iti.ac.id

## 1. Pendahuluan

Komputasi lebih dari sekedar *augmenter of science*. Tidak seperti perangkat lain yang terbatas terhadap *domain* pengetahuan tertentu, pemodelan komputasional dan *data analytics* dapat diaplikasikan pada semua area pengetahuan dan kerekayasaan yang menghidupkan segala pemodelan matematis secara ilmiah. Pemodelan komputasional dan *data analytics* juga memungkinkan peneliti dalam melakukan prediksi dan melakukan penelitian dengan lebih efisien. Pemodelan komputasional akan menyederhanakan pemodelan matematis yang sangat kompleks serta memungkinkan pemodelan dalam ranah berbagai ilmu pengetahuan dan kerekayasaan yang relatif menghemat biaya sekaligus juga pemodelan dalam ranah komputasional relatif lebih aman dibanding proses penelitian secara fisik.

Contoh kongkretnya misal pemodelan komputasional dari fenomena astrofisik pada ranah temporal dan spatial yakni untuk berbagai formasi planet dan benda langit lainnya, dinamika perbintangan, perilaku lubang hitam, formasi galaksi, dan sebagainya. Dapat juga dilakukan pemodelan komputasional untuk model iklim yang menunjukkan efek rumah kaca dan penggundulan hutan, yang membantu pemahaman terhadap pengaruh perilaku manusia terhadap perubahan iklim.

Ilmu pengetahuan dan teknologi komputasional juga memungkinkan desain dan optimasi multidisiplin sekaligus mengurangi waktu dan biaya prototyping. Misalnya yang telah dilakukan oleh Boeing dalam mendesain pesawat yang hemat bahan bakar, Goodyear yang dapat melakukan desain ban yang lebih aman secara relatif lebih cepat, serta Procter & Gamble yang mampu menciptakan material yang lebih baik untuk produk-produk *consumer goods*.

Serupa dengan komputasi, *big data*, *machine learning*, dan *predictive data analytics* telah dianggap sebagai paradigma keempat dari pengetahuan [1], yang memungkinkan peneliti menggali lebih dalam pada ranah instrumen ilmiah sekaligus ranah simulasi komputasional. Misalnya *machine learning* telah memungkinkan pendekatan baru dalam penelusuran dan agregasi

data terkait resiko-resiko kesehatan atau penyebaran penyakit melalui analisis jaringan sosial, analisis query pencarian pada mesin pencari berbasis web, berikut data-data di rumah sakit. *Machine learning* juga menjadi kunci dalam identifikasi *event* pada ranah biologi molekuler.

Seiring perkembangan instrumen ilmiah berskala besar, setiap generasi baru *advanced computing* senantiasa membawa kemampuan baru sekaligus tantangan teknis dan *trade off* secara ekonomi. Secara umum, kemampuan *data generation* dalam domain ilmiah tumbuh lebih cepat dibandingkan kemampuan komputasi sehingga dalam domain ilmiah menjadi bersifat *data-intensive* [2]. Dengan demikian menjadi kebutuhan bahwa komputasi performa tinggi dan sistem *big data* terikat menjadi satu kesatuan dalam ekosistem komputasi yang lebih luas berikut desain dan pasar dari ekosistem tersebut.

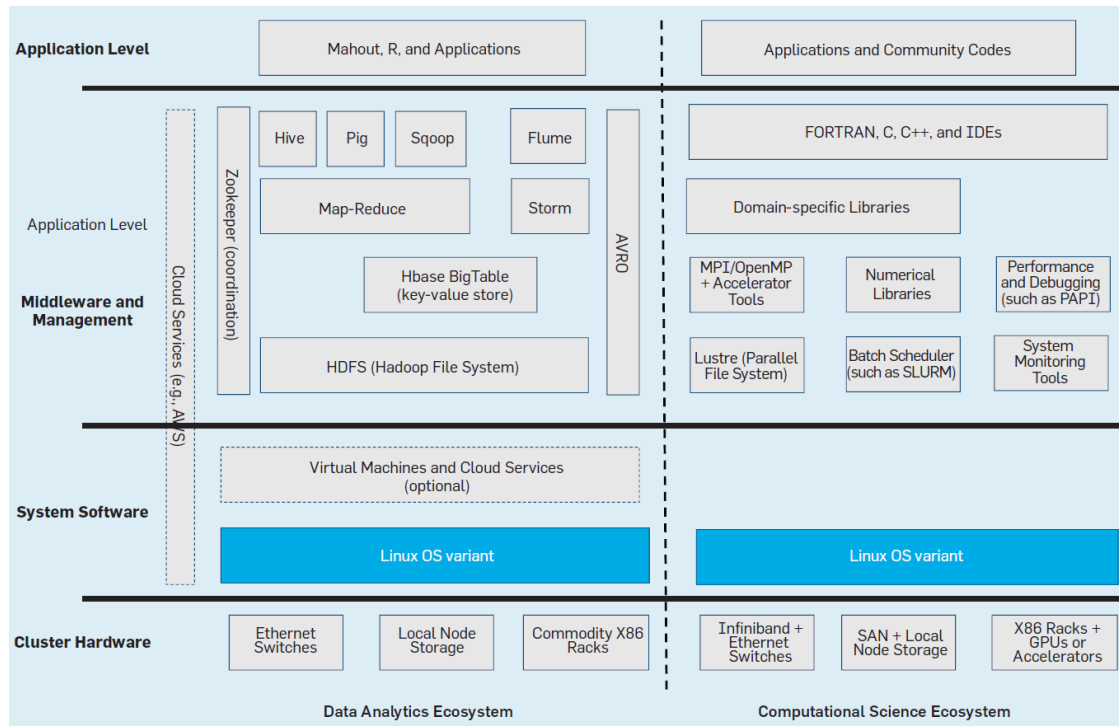
Penyatuan komputasi performa tinggi dan sistem *big data*, seiring dengan semakin tingginya biaya komputasi besar dan sistem analisis data, pastinya akan memunculkan pertanyaan terkait prioritas investasi, desain, dan model implementasi, sekaligus juga bagaimana kolaborasi dan kompetisi yang akan muncul. Makalah ini memaparkan beberapa tantangan teknis terkait penyatuan komputasi performa tinggi dan sistem *big data*, bagaimana saling ketergantungan pemodelan komputasional dengan *data analytics*, dan ekosistem global untuk *advanced computing*.

## 2. Teori Dasar

### 2.1 Advanced Computing Ecosystems

Sistem *advanced computing* mencakup perangkat keras, perangkat lunak, sekaligus algoritma-algoritma yang dibutuhkan guna menghantarkan kapabilitas komputasi tingkat tertinggi setiap saat. Gambar 1 menunjukkan ekosistem komputasi dan *data analytics*, dimana terlihat saling berbagi beberapa atribut antara lain ketergantungan terhadap piranti lunak *open source* dan ekosistem perangkat keras berbasis x86. Walau demikian, tetap terdapat perbedaan yang signifikan antara ekosistem komputasi dan *data analytics* terutama secara teknis. Dengan semakin tergantungnya penelitian-penelitian ilmiah terhadap

komputasi berkecepatan tinggi sekaligus juga terhadap *data analytics*, maka interoperabilitas dan konvergensi antara dua hal tersebut menjadi kebutuhan.



Gambar 1. Perbandingan ekosistem *data analytics* dengan komputasi

### 2.1.1 Komputasi Ilmiah

Teknologi komputasi ilmiah performa tinggi telah berkembang semenjak era *vector supercomputing* pada dekade 1980, *massively parallel processing* (MPPs) dan *shared memory multiprocessors* (SMPs) pada dekade 1990, dan komoditas *cluster* dan teknologi processor yang khusus dikembangkan untuk tujuan spesifik (misal IBM BlueGene) pada dekade 2000. Saat ini ekosistem komputasi performa tinggi telah berkembang dimana teknologi cluster diperkuat dengan akselerator komputasional dalam bentuk coprocessor dan *graphical processing units* (GPUs). Akselerator komputasional juga mendukung interkoneksi berkecepatan tinggi sekaligus *low latency* (Infiniband). Ekosistem komputasi performa tinggi saat ini menggunakan *storage area networks* (SANs) sebagai penyimpanan data yang handal (*persistent data storage*), sedangkan file-file temporer disimpan di penyimpanan lokal pada tiap *node*.

Seiring perkembangan teknologi *cluster*, Linux memberikan dukungan berupa sistem *file* paralel (Lustre) dan penjadwal *batch* (SLURM dan PBS) untuk pengelolaan tugas paralel. MPI dan OpenMP digunakan untuk mewujudkan paralelisme internode dan intranode, dengan diperkuat perangkat-perangkat yang digunakan oleh coprocessor. Masih dalam dukungan

terhadap komputasi performa tinggi, digunakan juga pustaka numerik (LAPACK dan PETSc) serta pustaka *domain-specific* pada lapisan aplikasi.

### 2.1.2 Data Analytics

Dewasa ini, sistem *cloud computing* untuk penelitian maupun komersial dapat terdiri atas penyimpanan sekunder dengan orde *petabyte*. Bahkan laboratorium penelitian individual dapat secara rutin mengolah data berukuran *terabyte*, yang dihasilkan oleh peralatan-peralatan ilmiah yang ada.

Seperti halnya pada ranah komputasi ilmiah, terjadi juga perkembangan yang pesat baik dari sisi perangkat keras maupun lunak guna *big-data analytics*. Namun tidak seperti pada ranah komputasi ilmiah, ranah *big-data analytics* pada umumnya berbasis pada komoditas jaringan *Ethernet* dan penyimpanan lokal, dengan biaya dan kapasitas sebagai prioritas utama.

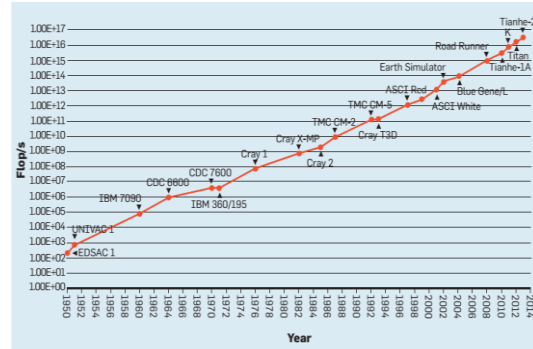
Guna kepentingan *big-data analytics*, pada tingkat middleware terdapat sistem Apache Hadoop yang menyediakan model untuk *data analytics* [3]. Hadoop memiliki sistem *file* terdistribusi guna mengelola file berukuran besar yang berjumlah sangat banyak (HDFS). Sistem *file* terdistribusi ini mengimplementasikan replikasi blok pada penyimpanan-penyimpanan

lokal dari *node-node cluster*. HDFS dan HBase, yang merupakan versi open source dari Google's BigTable, dapat menjadi *big-data analogs* untuk Lustre guna komputasi ilmiah. Hadoop memiliki perangkat yakni Pig yang menyediakan model pemrograman tingkat tinggi guna *data analytics* [4]. Sebagai ekosistem *data analytics*, Hadoop didukung oleh perangkat *big-data streaming* (Storm dan Flume); dukungan data relasional (Sqoop); perangkat penampil grafik; dan perangkat guna klasifikasi, rekomendasi, dan prediksi melalui pembelajaran *supervised* dan *unsupervised* (Mahout). Pada umumnya ekosistem *data analytics* dalam pengembangannya berbasis teknologi Java dan *web services* seperti Ruby on Rail.

Mahout berada pada level aplikasi. Perangkat lain pada level aplikasi yang biasa terdapat pada ekosistem *data analytics* adalah bahasa pemrograman R. R digunakan untuk komputasi statistik, analisis data, *data mining*, serta mengembangkan piranti lunak guna kebutuhan statistik spesifik.

## 2.2 Tantangan Dalam Peningkatan Skala

Perkembangan teknologi terjadi dengan sangat cepat. Sebagai contoh perbandingan, kekuatan komputasi pada iPhone 6 dan Samsung Galaxy S5 untuk aljabar linear standar telah melampaui Cray-1 (super komputer pertama, 1975). Contoh lain yakni standar Personal Computer (PC) yang ada sekarang dengan dilengkapi GPU NVidia atau akselerator Intel Xeon Phi dan penyimpanan lokal telah memiliki kekuatan komputasi pada orde *teraflops* dan *terabyte*, sehingga pada tingkatan *advanced computing* yang ada sekarang ini akan memiliki kekuatan pemrosesan pada orde *multiple petaflops* dan dukungan penyimpanan *cloud* dengan kapasitas penyimpanan sekunder hingga *multiple petabyte*. Gambar 2 menunjukkan fenomena peningkatan eksponensial dari perkembangan teknologi komputasi, terutama pada ranah *advanced computing* berdasarkan benchmark menggunakan *High-Performance LINPACK* (HPL) [5]. Dalam perkembangannya, komputasi performa tinggi juga memperoleh keuntungan dari kemajuan teknologi *semi-conductor* dan berkembangnya teknologi pemrosesan paralel.



21 juta. Teknologi memori (DRAM/*Dynamic Random Access Memori*) yang ada saat ini membutuhkan daya sebesar 260-450mW [9], sehingga memori perangkat *exascale* akan membutuhkan daya sebesar 5.5-9.5MW. Infrastruktur *exascale* merupakan sistem terdistribusi. Kebutuhan daya pada interkoneksi jaringan infrastruktur *exascale* diperkirakan akan membutuhkan daya sebesar 4-15.6 MW [6].

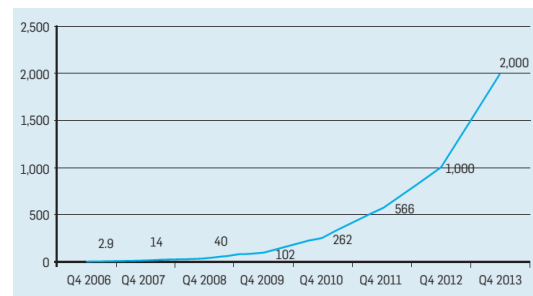
### 2.2.2 Tantangan Terkait Memori, Data, dan Perpindahan Data

Sejak dari awal terdapat perangkat komputer hingga era komputasi performa tinggi dewasa ini, perpindahan data senantiasa merupakan *performance bottleneck*. Perpindahan data senantiasa mendominasi pada semua tingkat hirarki memori sekaligus merupakan proses pada memori yang mengkonsumsi daya paling besar. Berbicara mengenai kebutuhan dari komputasi performa tinggi yakni pengolahan *scientific codes*, bahwa sebagian besar instruksi yang dieksekusi pada pengolahan *scientific codes* bukan *floating point* merupakan instruksi memori dan *integer*, yang mana komputasi instruksi *integer* membutuhkan sebagian besar alamat memori. Perangkat komputer yang ada sekarang menggunakan arsitektur *massively parallel processing* (MPP) yang hanya mampu melakukan akses data melalui *node* tunggal. Walau pada perangkat super-komputer yang ada saat ini tidak menggunakan MPP melainkan mekanisme *Partitioned Global Address Space* (PGAS) untuk perangkat keras dan lunaknya, mekanisme ini masih memiliki kelemahan yakni pengelolaan duplikat untuk setiap *node*. Ketidakseimbangan antara perkembangan teknologi *processor* dan memori, sekaligus dengan besarnya kebutuhan sumber daya memori pada piranti lunak-piranti lunak modern mengakibatkan rendahnya *Instruction Per Cycle* (IPC) pada *processor-processor* modern [10]. Pengembangan sistem yang lebih seimbang menjadi salah satu tantangan utama penghematan daya pada perangkat *exascale*.

Pertumbuhan data baik pribadi, bisnis, pemerintahan dan ranah ilmiah bahkan lebih cepat dibanding perkembangan komputasi performa tinggi. Saat ini penyedia layanan *cloud* komersial membangun jaringan *data center* yang tersebar di seluruh dunia, yang mana tiap *data center* dapat menghabiskan biaya hingga ratusan juta dolar guna mendukung layanan-layanan mesin pencarian, jejaring sosial, dan *cloud*.

Sebagai contoh, pada gambar 3 menunjukkan pertumbuhan eksponensial dari objek-objek yang tersimpan di Amazon S3, dimana Netflix berikut seluruh layanannya termasuk di dalamnya. Dewasa ini terjadi juga

peningkatan jumlah peneliti ilmiah yang menggunakan layanan-layanan *cloud* berikut teknik *machine learning* guna ekstraksi makna dari gambar ilmiah, grafik, serta data tekstual.



Gambar 3. Pertumbuhan Jumlah Object Tersimpan di Amazon S3

Dengan demikian secara alamiah akan muncul sinergi teknis dan ekonomi antara ilmu pengetahuan yang bersifat *data-intensive* dan komputasi performa tinggi seiring kemajuan keduanya. Ilmu pengetahuan yang bersifat *data-intensive* bergantung pada koleksi, analisis, dan pengelolaan data dengan volume sangat besar, yang diperoleh dari simulasi ilmiah atau fasilitas eksperimen.

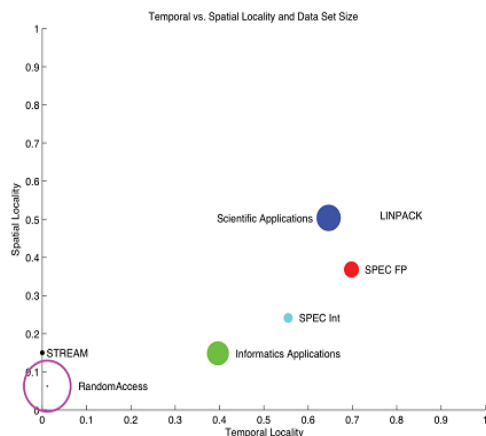
Pengukuran ketergantungan terhadap perpindahan data dan kebutuhan memori aplikasi yang telah dilakukan pada lingkungan *advance computing* ditunjukkan pada gambar 4, dimana menunjukkan pengukuran tiga macam perpindahan data utama sekaligus properti pola akses memori untuk semua aplikasi [11]. *Spatial Locality* mengukur utilitas memori sebagai hasil proses penjadwalan, sedangkan *Temporal Locality* mengukur reusability dari data yang telah digunakan. Aplikasi piranti lunak dalam hal ini diklasifikasikan menjadi:

- *Physical applications*, yakni aplikasi-aplikasi komputasional dan *analytics*, simulasi, dekomposisi spasial, pemodelan multidimensi, dan pengolahan *floating point*, dan cenderung menggunakan *data set* spesifik, dimana menjadi *core* dari komputasi performa tinggi.
- *Informatics applications* yang merepresentasikan kode-kode program yang berubah/berkembang dengan cepat seperti misalnya aplikasi *enterprise/transaksional*, yang cenderung tidak terstruktur, berorientasi *integer*, dan digunakan untuk membangun hipotesis dari *data set* yang besar.

Terdapat dua hal utama yang meningkatkan tantangan terkait perpindahan data pada perangkat *exascale*. Pertama, *informatics applications* sangat lebih berorientasi perpindahan data dibanding *physical applications*, yang ditunjukkan dengan rendahnya nilai *temporal* dan



*spatial locality*. Yang kedua adalah kecenderungan dimana aplikasi-aplikasi tidak *well tuned* sesuai performa mesin. Pada pengukuran ketergantungan terhadap perpindahan data tersebut juga digunakan *benchmark suites*. *Benchmark suites* tersebut yakni LINPACK yang merepresentasikan *physical applications* namun memiliki *data set* yang sangat kecil dan cenderung *temporal* dibanding *physical applications* pada umumnya, serta SPEC Integer (Int), dan SPEC Floating Point (FP). SPEC merepresentasikan baik *physical* maupun *informatics applications*. *Benchmark suites* digunakan sebagai *threshold representation* guna menemukan ambang kebutuhan memori yang dapat dipenuhi dengan teknologi yang ada sekarang.



**Gambar 4.** Hasil Pengukuran Ketergantungan Perpindahan Data dan Kebutuhan Memori Aplikasi pada Lingkungan *Advance Computing*

### 2.3 Tantangan Proteksi Pengguna

Terkait penerapan *big data*, perlu diperhatikan tentang aturan proteksi privasi, dimana jika menyesuaikan dengan prinsip keamanan informasi, paling tidak pengguna diinformasikan kapan *information retrieval* dilakukan, informasi apa yang dikumpulkan, serta juga tujuan *information retrieval* tersebut.

Pada umumnya penerapan *big data* lebih bergantung pada penggunaan informasi sekunder, yakni informasi yang dikumpulkan dari penggunaan mesin pencari, sensor data nirkabel, dan media sosial, dimana dalam hal ini pada umumnya secara teknis pengguna sulit untuk memberikan izin atau persetujuan terkait eksposisi datanya pada ranah *cyber*. Terkait dengan hal tersebut bahwa akan menjadi permasalahan pada era *big data* terkait anonimitas, atau dimana banyak pengguna yang beraktivitas di dunia *cyber* secara anonim, dimana identifikasi personal diperuntukkan pada *setting small data*, sedangkan untuk realita *big data* akan sangat dimungkinkan terjadi *re-identification* dikarenakan anonimitas

[12]. Contoh permasalahan *re-identification* (walau belum benar-benar pada ranah *big data*) yakni pada tahun 2007, AOL merilis *data set* yakni 20 juta *query* pencarian (dengan dua *search term*) oleh 657.000 pengguna. Selanjutnya dilakukan analisis terhadap *data set* tersebut dengan hasil ditemukannya identitas berikut profil detail seorang pengguna yang bocor/tersebar luas, padahal AOL telah menerapkan identifikasi pengguna dengan *numerical user name*.

Hal lain yang perlu diperhatikan dalam penerapan *big data* yakni kesulitan pemisahan antara *noise* dengan *signal*, atau dengan kata lain *big data* berarti lebih banyak informasi, tetapi juga berarti lebih banyak informasi yang salah [13]. Diperlukan keseimbangan antara potensi *big data* dengan peran manusia dalam *exploratory analytics*. Dengan *exploratory analytics*, seorang analis data dapat melakukan *data mining* guna mencari informasi yang penting dari data. Demikian juga masih diperlukan peran penting dari manusia guna melakukan koreksi kesalahan sekaligus membangun mekanisme pengambilan keputusan pada keseluruhan logika dari model *big data*.

*Big data* merupakan teknologi disruptif dengan proses *big data analytics* yang dapat dimanfaatkan baik pada ranah bisnis maupun ilmiah guna meningkatkan kualitas keluaran proses pengolahan data. Walau demikian perlu dipahami kesulitan-kesulitan tersembunyi dan kompleksitas penerapan *big data*. *Big data* yang secara fungsi merupakan otomatisasi dari penggabungan berbagai basis data, *number crunching*, serta penentuan korelasi, hanya dapat diwujudkan secara hati-hati, serta diperlukan sumber daya manusia dengan pengetahuan dan pengalaman tentang permasalahan yang sedang diinvestigasi dalam kaitannya dengan proses *big data analytics* terhadap *data set* yang layak mengimplementasikan teknologi disruptif tersebut.

### 2.4 Kesempatan Ilmiah dan Kerekayasaan

*Advanced computing* dan model komputasional telah lama digunakan pada bidang ilmiah dan kerekayasaan. Terlebih lagi dewasa ini bidang ilmu sosial, lingkungan, dan biologi juga telah menjadi bagian dari *big-data analytics*. Kemudian saat ini kegiatan kerekayasaan dan manufaktur telah sangat bergantung pada *advanced computing*. Dari ranah mesin turbo, kimia molekular, hingga pemodelan biomasa sel surya, *advanced computing* telah menjadi sinonim terhadap desain multidisipliner, optimasi, dan *advanced manufacturing*. Bidang-bidang yang memperoleh manfaat dari *advanced computing* antara lain:

- Biologi dan biomedicine

- Fisika *high-energy*
- Klimatologi
- Kosmologi dan astrofisika
- Astronomi
- Pengobatan kanker
- Ilmu bahan
- Produksi baja
- *Text* dan *data mining*

Terdapat dua kesamaan antara bidang-bidang tersebut. Yang pertama adalah rentang skala kompleksitas, temporal, dan spatial berikut interaksi non-linear antara berbagai proses fisik dan biologi. Hal ini sangat membutuhkan simulasi komputasional dengan dukungan performa komputer yang sangat tinggi. Tujuan simulasi komputasional dalam hal ini adalah simulasi prediktif, pengujian teori, identifikasi interaksi, dan memberikan arah penelitian. Yang kedua adalah yang berhubungan dengan data, yakni data ilmiah yang berjumlah atau berukuran sangat besar serta sangat beragam, asimilasi data ilmiah yang tidak dapat diprediksi, korelasi multidisiplin, analisis statistik. Baik pada ranah ilmu biologi atau fisika, rekayasa atau bisnis, *big data* menciptakan kebutuhan riset dan kesempatan baru.

Guna komputasi ilmiah dan rekayasa, peningkatan performa akan semakin ditingkatkan mencapai orde *exascale*. Seiring dengan hal tersebut pemeliharaan dan pemenuhan keberlanjutan data berskala besar, penciptaan metadata, fusi multidisiplin, serta pemenuhan privasi dan keamanan data menjadi tonggak pengembangan ranah *big data*. Definisi *advanced computing* tidak sekedar perhitungan kuantitatif terhadap laju operasi aritmatik, kapasitas penyimpanan, dan laju analisis, tetapi juga mencakup kapabilitas kemanfaatan untuk semua disiplin ilmu dengan skala apapun. Dengan demikian perlu dirumuskan standar penghantaran kemanfaatan tersebut.

Tantangan yang harus dihadapi pengembangan *advanced computing* meliputi konsumsi sumber daya, pendingin yang ramah lingkungan, paralelisme masif, kegagalan komponen, konsistensi data dan transaksi, pengelolaan metadata dan ontologi, presisi dan ketersediaan, serta pemeliharaan dan fusi data multidisiplin.

Yang harus diperhatikan juga bahwa ekosistem *advanced computing* kedepannya tidak kemudian menjadi terlalu kompleks sehingga hanya berguna di tangan beberapa ahli. Perangkat-perangkat *open source* seperti Hadoop, Mahout, dan Giraph berikut perangkat-perangkat pendukung masing-masing telah memungkinkan para peneliti dalam mengaplikasikan *machine learning* ke data ilmiah berskala besar tanpa

diperlukan pemahaman mendalam terhadap algoritma *machine learning*.

### 3. Hasil dan Pembahasan

#### 3.1 Menjawab Tantangan non-teknis

Perkembangan *advanced computing* dalam sejarah umat manusia sangat bergantung pada perkembangan algoritma, piranti lunak, dan arsitektur perangkat keras yang memungkinkan performa *floating-points* pada tingkat yang lebih tinggi untuk pemodelan komputasional. Kemajuan yang terjadi dewasa ini juga dibentuk oleh *data-analysis pipelines*, arsitektur data, dan perangkat *machine learning* yang mengelola data ilmiah dan rekayasa dalam jumlah besar. Tantangan yang dihadapi dalam kaitannya dengan perkembangan *advanced computing* secara non-teknis yakni dari sisi organisasi, dan ekonomi. Sistem komputasi berskala *exascale* tidak akan dapat diproduksi dengan biaya yang terjangkau jika dikorelasikan dengan batasan rekayasa yang realistis, baik dari sisi biaya operasi, kegunaan, dan keandalan. Walau demikian setiap kemajuan pada teknologi komputasi akan mendorong inovasi industri dan pertumbuhan ekonomi yang akan mempengaruhi keseluruhan spektrum komputasi, yakni dari *Internet of Things* hingga perangkat *mobile ubiquitous* hingga sistem komputasi terkuat dan penyimpanan data terbesar. Kemajuan dalam teknologi komputasi juga berpengaruh terhadap riset dasar dan terapan untuk setiap domain pengetahuan.

Tantangan teknis, politis, dan ekonomi bisa dihadapi hanya dengan koordinasi antara pemerintah, industri, dan dunia akademis, dengan berkomitmen terhadap keberlanjutan data serta untuk saling berbagi data, melakukan penelitian dan pengembangan bersama, serta bentuk kolaborasi lainnya.

#### 3.2 Menjawab Tantangan Teknis *Advanced Computing*

1. Sirkuit, sumber daya, dan teknologi pendingin yang hemat energi. Sistem komputasi *exascale* jika dikembangkan dengan teknologi *semi-conductor* yang ada saat ini akan membutuhkan daya hingga ratusan megawatt. Oleh karena itu diperlukan desain dan teknologi baru yang lebih hemat energi hingga ke tingkat yang lebih dikelola dan layak secara ekonomi (kisaran 20MW-40MW).

2. Teknologi interkoneksi performa tinggi. Pada ranah komputasi *exascale*, energi untuk memindahkan sebuah fakta akan melampaui energi untuk sebuah operasi *floating-point*, sehingga membutuhkan interkoneksi yang sangat hemat energi, *latency* rendah, dan *bandwidth* yang besar guna pertukaran data antar banyak

processor. Saat ini penyedia layanan cloud komersial telah mengimplementasikan komoditas interkoneksi Ethernet sekaligus model komputasi parallel sinkron, yang mana pada dasarnya terbukti efektif (berdasarkan adopsi perangkat MapReduce Hadoop). Namun interkoneksi konvergen dengan kebutuhan biaya lebih rendah akan memberikan keuntungan baik bagi platform komputasi maupun platform data-intensive, sekaligus membuka kesempatan baru untuk analisis yang lebih baik.

3. Teknologi memori dengan peningkatan kapasitas. Meminimalisir perpindahan data sekaligus penghematan energi (kapabilitas pengurangan jumlah memori per core pada system besar) merupakan tujuan teknologi memori baru seperti *processor-in-memory*, *stacked memory*, dan pendekatan memori *non-volatile*. Algoritma penentu kapasitas memori menjadi pendorong utama perkembangan teknologi memori.

4. Piranti lunak *scalable* yang mawas terhadap keamanan dan penggunaan energi. Piranti lunak komputasi performa tinggi yang ada sekarang dibangun dengan asumsi frekuensi kegagalan yang rendah. Dengan kemajuan yang semakin mendekati tingkat *exascale*, ketahanan sistemik terhadap kegagalan komponen menjadi semakin penting. Demikian juga pengelolaan energi yang adaptif harus menjadi bagian integral dari piranti lunak untuk alasan teknis dan ekonomi. Layanan *cloud* guna data-analytics telah memberikan *quality of services* (QoS) komersial dengan sejumlah besar teknik ketahanan yang mencakup *geo-distribution*, *automatic restart* dan *failover*, injeksi kegagalan, dan pengawasan introspektif. Contoh dari QoS ini adalah Simian Army pada Netflix.

5. Piranti lunak pengelolaan data yang mampu menangani volume, kecepatan, dan keberagaman data. Baik dihasilkan secara komputasional atau diperoleh dari perangkat spesifik, analisis data yang efisien membutuhkan aplikasi dan restrukturisasi alur kerja ilmiah, berikut teknik baru untuk koordinasi data, pembelajaran data, dan data mining. Analisis data yang efisien tidak menginginkan terjadinya kemacetan masukan-keluaran (input-output (I/O) *bottleneck*) yang akan membatasi kegunaan dan kemanfaatan aplikasi.

6. Model pemrograman yang mendukung paralelisme skala besar, lokalisasi data, dan ketahanan. Model proses communicating sequential atau pemrograman MPI yang mulai banyak digunakan senantiasa memeras otak pengembang piranti lunak guna mengatasi isu lokalisasi data dan paralelisme. Pada komputasi *exascale* akan terdapat banyak paralelisme, dimana pastinya akan banyak terjadi kesalahan,

sehingga dibutuhkan model pemrograman yang lebih ekspresif dimana mempermudah pengembang piranti lunak dalam mendukung paralelisme yang baik dan dinamis. Tonggak yang dapat diacu pada saat ini yakni pada layanan *web* dan *cloud* dimana lapisan abstraksi berikut perangkat domain-specific memungkinkan programmer untuk menggelar lingkungan eksekusi *custom* (mesin virtual) sekaligus memungkinkan reduksi data. Tantangan pada ranah komputasi ilmiah terkait model pemrograman adalah model pemrograman yang memungkinkan dalam mempertahankan ekspresivitas dan produktivitas, dengan tetap mampu menghantarkan performa tinggi.

7. Reformulasi permasalahan ilmiah dan refactoring algoritma solusi. Telah sangat banyak waktu yang diinvestasikan untuk kode-kode program guna kepentingan ilmiah dan rekayasa. Adaptasi kode-kode tersebut ke berbagai cara paralelisme akan membutuhkan desain ulang bahkan perumusan algoritma baru, atau bahkan juga reformulasi dari permasalahan-permasalahan ilmiahnya. Pemahaman tentang bagaimana integrasi piranti-piranti lunak data analytics, perangkat komputasi, bahasa pemrograman, dan berbagai model ilmiah secara efisien akan mempermudah penyelesaian permasalahan ilmiah kritis.

8. Mencegah kegagalan, mewujudkan *reproducibility*, dan verifikasi algoritma. Dengan seringnya terjadi kegagalan baik parsial maupun total, kurangnya kemampuan *reproducibility* dalam komunikasi kolektif, serta terdapatnya algoritma-algoritma minim verifikasi; maka validasi komputasi dan jaminan ketepatan akan menjadi sangat penting pada system parallel masif, baik untuk komputasi ilmiah, analisis data, atau keduanya.

9. Optimisasi matematis dan ketidakpastian kuantifikasi untuk penemuan, desain, dan keputusan. Komputasi skala besar pada dasarnya akan senantiasa eksperimental dimana ranah pembuktiannya ada pada ruang model numerik. Pemahaman terhadap sensitivitas prediksi komputasional terhadap model masukan dan asumsi, terutama saat melibatkan aplikasi kompleks multidisiplin; akan membutuhkan piranti dan teknik baru guna assessment dan validasi aplikasi. Hal yang tidak kalah pentingnya dalam kaitannya dengan *data analytics* skala besar dan machine learning adalah presisi (bagian dari data diperoleh yang relevan) dan *recall* (bagian relevan dari data yang diperoleh).

10. Rekayasa piranti lunak dan struktur pendukung yang mendukung produktivitas. Meskipun perangkat pemrograman, *compiler*, *debugger*, dan perangkat-perangkat pendukungnya telah mendukung produktivitas



penelitian sistem komputasi guna komputasi yang handal, efisien, dan tepat; tetapi juga dibutuhkan peningkatan produktivitas dari peneliti sehingga rumusan solusi tidak hanya didominasi rancang bangun aplikasi, tetapi juga dukungan komputasi.

11. Operasi hemat energi. Operasi hemat energi guna mencapai laju komputasi yang diinginkan, dalam kaitannya dengan adanya rugi-rugi daya.

12. Kapasitas Memori. Kapasitas memori primer dan sekunder yang dapat dicapai, dalam kaitannya dengan pembatasan sumber daya.

13. Konkurensi dan lokalitas. Konkurensi dan lokalitas guna mencapai target performa yang memungkinkan beberapa thread berhenti sementara saat operasi *long-latency*.

14. Ketahanan. Perwujudan ketahanan dengan menambah jumlah processor, mengurangi ukuran fitur *semi-conductor*, menggunakan *semi-conductor* hemat energi, serta penanganan kegagalan sementara dan permanen pada processor.

15. Penentuan skala aplikasi. Penentuan skala aplikasi dalam kaitannya dengan kapasitas memori dan batasan *latency* komunikasi.

16. Pengelolaan paralelisme. Pengelolaan paralelisme dan lokalitas dalam sistem piranti lunak dan model pemrograman portable.

17. Ketersediaan piranti lunak guna *performance tuning*, assessment ketepatan, dan pengelolaan energi.

18. Mendewasakan ekosistem R. R memang telah banyak digunakan di berbagai bidang ilmu, tetapi guna kepentingan *data analytics* untuk skala yang lebih besar, diperlukan *meta tool* yang memungkinkan eksplorasi interaktif terhadap data. Selain itu dibutuhkan juga perangkat pengembang R yang spesifik sesuai dengan kebutuhan peneliti. Pembangunan jejaring arsip komprehensif untuk R juga akan memberikan manfaat dalam kaitannya dengan *big-data analytics*.

19. *User friendly machine learning*. Pada ranah *advanced computing*, perangkat pengembang *machine learning* dikembangkan tidak hanya sebagai perangkat pemrograman, melainkan harus memprioritaskan kecepatan komputasi dan kemudahan penggunaan. Fitur-fitur yang dapat ditambahkan pada perangkat *machine learning* sehingga relatif lebih *user friendly* antara lain:

- Library yang terintegrasi dengan Application Programming Interface (API) sehingga mempersingkat data *pre-processing*, ekstraksi fitur, model fitting, dan tahapan validasi dengan urutan *transformed dataset* sebagai masukan tahapan berikutnya.

- *Dataframe* API untuk antarmuka ekosistem R, yang mendukung penerapan algoritma machine learning.
- Perangkat pemelihara primer yang memungkinkan pengguna untuk berpindah antara satu metode ke metode lain secara interaktif. Perangkat pemelihara primer ini juga dapat disesuaikan dengan kebutuhan pengguna melalui lingkungan *native C++*.

### 3.3 Menjawab Tantangan dari Sisi Arsitektur dan Perangkat Keras

1. *Post-dennard scaling*. Penerapan prinsip *Dennard scaling*, yakni memperkecil transistor dan sirkuit dengan kepadatan daya yang sama. Hal ini dilakukan dengan mengurangi ukuran linear transistor (hingga sekitar 22-nanometer, yang mulai banyak diterapkan saat ini) sekaligus memangkas sumber daya yang dibutuhkan, tanpa mengurangi laju *clock*. Hal lain yang perlu diperhatikan sehubungan dengan hal ini adalah dengan semakin kecilnya ukuran linear transistor dengan tetap mempertahankan laju *clock*, maka harus hati-hati dalam melakukan pemangkas daya selain juga akan meningkatkan emisi panas. Panas yang timbul pada *semi-conductor* sekaligus pemangkas sumber daya yang tidak tepat dapat mengakibatkan munculnya suatu efek negatif yang disebut *dark silicon*, yakni lebih dari satu transistor pada *chip* yang aktif secara simultan. Sehubungan dengan hal ini, dibutuhkan terobosan dalam desain chip terutama dari sisi arsitektur, yakni dengan *superpipelining*, *scoreboarding*, vektorisasi, dan paralelisasi, dengan tetap memperhatikan keseimbangan dengan performa dan konsumsi energi. Desain chip juga dilakukan dengan memperhatikan keseimbangan operasi floating-point, integer, dan percabangan sehingga dapat mendukung analisis data dan komputasi secara in-situ.

2. Skala ketahanan dan efisiensi energi. Seiring perkembangan *advanced computing* dan sistem analisis data, maka frekuensi kegagalan operasional sistem juga semakin besar. Kegagalan sistem dapat dikelola dengan cara-cara deteksi dan pemahaman terhadap kegagalan komponen. Penelitian yang pernah dilakukan oleh Google pada tahun 2009 menyatakan bahwa terjadi peningkatan jumlah kesalahan pada *dynamic random access memory* (DRAM) sebesar *orders-of-magnitude* dibanding dengan tahun sebelumnya, dimana lebih dari 8% *dual in-line memory module* (DIMM) terpengaruh oleh kesalahan dalam satu tahun. Yang menjadi masalah adalah kesalahan tersebut merupakan kesalahan-kesalahan *hard*, bukan *soft* yang dapat diperbaiki dengan kode *error-correcting*. Selain ketahanan, efisiensi penggunaan energi dan pembuangan panas juga menjadi tantangan.

Dewasa ini sistem *advanced computing* dan data analisis mengkonsumsi daya hingga orde megawatt serta membutuhkan kapabilitas pendinginan yang mumpuni. Dengan tersebarnya sistem *advanced computing* dan data analisis secara geografis di seluruh dunia mengakibatkan keterbatasan dalam pemenuhan *peak power*. Penyedia layanan *cloud* komersial saat ini telah mulai membangun infrastruktur dengan pendekatan baru yang mencakup desain hemat energi, sistem pendingin yang semakin handal, akuntabilitas energi, dan efisiensi operasional.

### 3.4 Menjawab Tantangan dari Sisi Piranti Lunak dan Algoritma

Tantangan dari sisi piranti lunak dan algoritma pada ekosistem *advanced computing* dan *big-data analytics* merupakan konsekuensi dari sistem skala besar. *Advanced computing* berbagi permasalahan dengan layanan *cloud* dan *web*, tetapi berbeda pada optimasi keseimbangan dan pemenuhan performa tinggi baik guna komputasinya maupun analisis data. Dengan besarnya skala dan prediksi banyaknya kesalahan, perlu pendekatan baru terhadap desain dan implementasi algoritma untuk sistem komputasi *exascale*. Pendekatan baru tersebut misalnya eksplorasi algoritma yang bebas sinkronisasi secara global, algoritma yang *fault-oblivious* dan *error tolerant*, algoritma *architecture-aware* yang cocok untuk perangkat keras heterogen dan terorganisir secara hirarkis, algoritma yang mendukung aritmatika *mixed-precision*, dan algoritma untuk piranti lunak hemat energi.

### 3.5 Lokalitas dan Skala

Pada sistem komputasi *exascale*, mengelola keseimbangan beban pada semua level dari hirarki algoritma dan platform adalah kunci dari eksekusi yang efisien. Hal ini akan membutuhkan *runtime* yang dinamis dan adaptif, serta alokasi sumber daya yang toleran terhadap ketidakseimbangan algoritma, keberagaman performa dan kehandalan perangkat keras [14].

Energi dan *latency* komunikasi pada lokalitas komputasi *exascale* akan semakin dibutuhkan dibanding pada sistem komputasi yang ada sekarang. Saat ini operasi aritmatik kurang *energy intensive* dan lebih efisien dibandingkan komunikasi. Kompleksitas algoritma pada saat ini biasanya cenderung diekspresikan dengan performa jumlah operasi, bukan kuantitas perpindahan data ke memori. Pada komputasi skala besar, yang terjadi adalah sebaliknya (dari sisi biaya) yakni perpindahan ke memori akan menjadi sangat mahal tetapi operasional akan sangat murah. Hal ini sangat penting untuk diperhatikan dalam hubungannya dengan

algoritma *floating-point intensive* dan algoritma data analisis.

Biaya temporal perpindahan data membutuhkan terobosan pada desain algoritma dan optimasi komparatif, yang mana akan mengakibatkan komputasi *redundant* cenderung menjadi pilihan dibandingkan *data sharing*. Kemudian kompleksitas komunikasi juga akan meningkat seiring dengan peningkatan kompleksitas komputasi. Dengan demikian diperlukan desain algoritma yang sebisa mungkin meminimalisir komunikasi sekaligus model dan metode untuk meminimalisir dan mentolerir *latency*, optimasi perpindahan data, dan menghilangkan sinkronisasi global.

### 3.6 Sistem Piranti Lunak Adaptif

Pengelolaan sumber daya pada sistem komputasi performa tinggi yang ada saat ini adalah dengan penjadwalan terkoordinir dan sinkronisasi komunikasi secara ketat. Pada sistem komputasi skala besar keberagaman perangkat keras, sumber daya sistem, pembuangan panas, dan tingkat kegagalan komponen akan berpengaruh tidak hanya terhadap desain dan implementasi fungsi, tetapi juga terhadap desain piranti lunak terutama pada pengelolaan energi dan I/O. Demikian juga dengan laju pertumbuhan data ilmiah yang tinggi maka akan membutuhkan skala analisis data yang juga tinggi. Pada kondisi tersebut harus dipastikan abstraksi dan sistem *file* paralel pada komputasi teknis dapat mengimbangi skala analisis data.

Terobosan desain piranti lunak dan sistem operasi akan dibutuhkan guna mendukung pengelolaan sumber daya heterogen dan hirarki memori *non-cache-coherent*, menyediakan fungsi dan runtime dengan kendali lebih pada kebijakan penjadwalan tugas, serta kapabilitas pengelolaan *global namespaces*. Desain piranti lunak dan sistem operasi tersebut juga harus memungkinkan mekanisme pengukuran, prediksi, dan kendali pengelolaan sumber daya yang memungkinkan penjadwal memetakan komputasi ke akselerator fungsi spesifik, pengendali suhu, dan fungsi profil energi.

### 3.7 Dukungan Pemrograman Paralel

Dengan semakin meningkatnya keberagaman, kompleksitas, dan skala perangkat keras *advanced computing*, maka kompleksitas dan kesulitan pengembangan aplikasi juga semakin meningkat dengan banyaknya fungsi operasi yang harus didukung. Ditambah lagi adanya peningkatan kebutuhan aplikasi multidisiplin yang menggabungkan algoritma dan model yang menjangkau hingga skala spatiotemporal dengan pendekatan algoritmik.

Dapat diberikan contoh kompleksitas yakni misal sebuah program yang mengelola banyak data dengan pemrograman parallel atau model parallel sinkron secara *bulk*, dimana data tersebar di memori dan penyimpanan dari node-node komputasi, dan *node-node* tersebut berbagi data melalui jaringan message passing. Pada kondisi ini kode program pada setiap *node* mengelola hirarki komputasi *multilevel*, *multithreaded*, *core-core* heterogen, akselerator GPU, melakukan koordinasi I/O, melakukan *checkpointing*, mengelola sumber daya, dan mengatur pembuangan panas secara lokal. Kompleksitas seperti ini mengakibatkan pengembangan aplikasi menjadi tidak mudah dan harus dilakukan secara hati-hati.

Idealnya rancang bangun piranti lunak di masa depan akan meningkatkan level abstraksi, dan mempertimbangkan performa dan ketepatan pada output, bukan secara *ex-situ*. Lebih dari itu, perangkat-perangkat pemrograman harus terintegrasi dengan *compiler*, runtime system, lebih mendukung perangkat keras heterogen, mendukung model pemrograman *mixed*, serta kapabilitas pemrosesan data dan analisis yang lebih baik.

Model dan perangkat pemrograman merupakan perbedaan terbesar antara lingkungan komputasi ilmiah dengan ekosistem *big data*. Ekosistem *big data* menekankan pada abstraksi sederhana dengan format data *semantics-rich*, serta spesifikasi *high level*. Hal ini memungkinkan pengembang piranti lunak untuk membangun aplikasi *machine learning* kompleks hanya dengan sedikit pengetahuan tentang perangkat keras dan lunak pendukungnya. Sedangkan komputasi ilmiah sangat bergantung pada bahasa pemrograman dan *library*. Walau demikian, bahasa pemrograman generasi baru seperti OpenMP telah mendukung paralelisme, akselerator, dan *thread affinity*.

Bahasa *domain-specific* tertanam merupakan cara pragmatis dalam eksploitasi analisis mutakhir, sekaligus transformasi kapabilitas *compiler* pada bahasa pemrograman. Dengan bahasa *domain-specific* tertanam, pengembang aplikasi akan menulis *high-level primitives* yang ditransformasikan oleh *compiler* ke kode *low-level* yang efisien sedemikian rupa sehingga mewujudkan optimasi performa platform.

### 3.8 Menjawab Tantangan Algoritmis Matematis

Komputasi *exascale* akan banyak membutuhkan algoritma dalam dua hal yakni terhadap kebutuhan peningkatan lokalitas data guna melakukan komputasi secara efisien dan terhadap kebutuhan paralelisme yang mendukung

banyak *thread*. Dengan demikian akan dibutuhkan algoritma-algoritma baru yang menggali kapabilitas komputasional dari perangkat keras-perangkat keras baru.

Model pengembangan secara signifikan, desain algoritma, dan implementasi *scientific functions* yang didukung oleh model pemrograman yang tepat akan dibutuhkan oleh arsitektur *exascale*.

## 4. Kesimpulan

Konteks-konteks terkait komputasi *exascale* dan *big data analytics* di masa depan adalah:

- *post-dennard scaling* dan implikasinya terhadap kemajuan desain *semi conductor*,
- maraknya komputasi *mobile* dan *cloud*,
- pertumbuhan yang sangat pesat dari data-data ilmiah, bisnis, pemerintahan, dan pribadi,
- kesempatan bagi *data analytics* dan *machine learning*, dan
- peningkatan kebutuhan terhadap system computer yang semakin kuat guna ilmu pengetahuan dan rekayasa mutakhir.

Sehubungan dengan konteks-konteks tersebut, beberapa hal yang telah dapat dipastikan yakni:

1. *Big data* dan *exascale*. *High-end data analytics* (*big data analytics*) dan komputasi *high-end* (*exascale*) merupakan elemen-elemen penting dari penelitian dan pengembangan komputasi terintegrasi. Keduanya memiliki prioritas yang sama.
2. Algoritma, piranti lunak, dan aplikasi. Kedepannya penelitian dan pengembangan algoritma, piranti lunak, dan aplikasi *next generation* menjadi sama pentingnya dengan investasi pada perangkat keras dan *semi conductor*.
3. Ekosistem teknologi informasi. Ekosistem teknologi informasi global sedang mengalami transisi ke *next generation* yakni perangkat *mobile* berdaya rendah, layanan *cloud*, dan *rich data analytics*.
4. Riset swasta dan riset global. Kompetisi sektor swasta dan kolaborasi riset global diperlukan guna mewujudkan desain, pengujian, dan *deployment* dari komputasi *exascale* dan kapabilitas *big data analytics*.

## Daftar Pustaka

- [1] Hey, T., Tansley, S., and Tolle, K. 2009. *The Fourth Paradigm: Data-Intensive Scientific Discovery*, (Online), ([http://research.microsoft.com/en-us/UM/redmond/about/collaboration/fourth-paradigm/4th\\_PARADIGM\\_BOOK\\_compil\\_ete\\_HR.pdf](http://research.microsoft.com/en-us/UM/redmond/about/collaboration/fourth-paradigm/4th_PARADIGM_BOOK_compil_ete_HR.pdf), diakses 10 Desember 2015)

- [2] U.S. Department of Energy. Synergistic Challenges in Data-Intensive Science and Exascale Computing. Report of the Advanced Scientific Computing Advisory Committee Subcommittee, Mar. 30, 2013; (Online), ([http://science.energy.gov/~media/ascr/asca/c/pdf/reports/2013/ASCAC\\_Data\\_Intensive\\_Computing\\_report\\_final.pdf](http://science.energy.gov/~media/ascr/asca/c/pdf/reports/2013/ASCAC_Data_Intensive_Computing_report_final.pdf)), diakses 13 Februari 2016)
- [3] White, T. *Hadoop: The Definitive Guide*. O'Reilly Media, May 2012.
- [4] Olston, C., Reed, B., Strivastava, U., Kumar, R., and Tomkins, A. Pig Latin: A not-so-foreign language for data processing. In Proceedings of the 2008 ACM SIGMOD International Conference on Management of Data (Vancouver, BC, Canada, June 9–12). ACM Press, New York, 2008, 1099–1110.
- [5] Dongarra, J.J. The LINPACK benchmark: An explanation. In Proceedings of the *First International Conference on Supercomputing* (Athens, Greece, June 8–12). Springer-Verlag, New York, 1988, 456–474.
- [6] Kogge, P. (2008). *ExaScale Computing Study: Technology Challenges in Achieving Exascale Systems (Tech. Rep.)*. Notre Dame, IN: University of Notre Dame, Department of Computer Science.
- [7] Murphy, R., Rodrigues, A., Kogge, P., & Underwood, K. (2009). *The Implications of Working Set Analysis on Supercomputing Memory Hierarchy Design*. Paper presented at the International Conference on Supercomputing, Cambridge, UK.
- [8] ITRS International Roadmap Committee. International Technology Roadmap for Semiconductors, 2007.
- [9] Micron technology. (2007). *Calculating Memory System Power for DDR3*. Boies, ID.
- [10] Murphy, R. C. (2007, September 27-29). On the Effects of Memory Latency and Bandwidth on Supercomputer Application Performance. In *Proceedings Of the IEEE Internaional Symposium on Workload Characterization 2007 (IISWC07)*.
- [11] Murphy, R. C., & Kogge, P. M. (2007, July). *On the Memory Access Patterns of Supercomputer Applications: Benchmark Selection and Its Implications*. IEEE Transactions on Computers, 56(7), 937–945. doi:10.1109/TC.2007.1039
- [12] Mayer-Schonberger V, Cukier, K. *Big Data: A Revolution That Will Transform How We Live, Work, and Think*. Boston: Eamon Dolan Book/Houghton Mifflin Harcourt, 2013.
- [13] Taleb, Nassim N., Beware the Big Errors of 'Big Data', Aug. 02, 2013; (online), (<http://www.wired.com/opinion/2013/02/big-data-means-big-errors-people/>), diakses 26 Juli 2016)
- [14] Datta, K. et al. Stencil computation optimization and auto-tuning on state-of-the-art multicore architectures. In Proceedings of the 2008 ACM/IEEE Conference on Supercomputing (Austin, TX, Nov. 15–21). IEEE Press, Piscataway, NJ, 2008, 1–12.