

September 2023

Pengelolaan
*High-Performance
Computing* dan
Big Data di
Universitas
Indonesia

Policy Brief



UNIVERSITAS INDONESIA
SENAT AKADEMIK

Latar Belakang

Dengan meningkatnya kebutuhan penggunaan teknologi *Big Data* dalam penelitian di Universitas Indonesia (UI), maka pengelolaan *Big Data* di UI perlu mendapat perhatian serius oleh Eksekutif. *Big Data* adalah volume data yang sangat besar dan kompleks yang tidak dapat diolah dengan cara konvensional. Prinsip utama *Big Data* adalah kita dapat menemukan informasi dan wawasan yang berharga dari data yang sangat besar dan kompleks, melalui pemanfaatan teknologi penyimpanan data skala besar. *Big Data* melibatkan tiga hal, yaitu: volume, kecepatan, dan variasi. Komponen *Big Data* meliputi: *Data Sources*, *Data Storage*, *Data Processing*, *Data Analysis*, *Data Visualization*, *Data Security*, *Data Governance*, dan *Data Management*. Semua komponen bekerja bersama untuk memungkinkan organisasi mengumpulkan, menyimpan, memproses, menganalisis, dan memvisualisasikan data dalam skala besar dan kompleks secara efisien dan efektif.

Permasalahan *Big Data* tidak dapat diatasi hanya dengan komputer biasa yang dipakai di tingkat perkantoran. Secara umum infrastruktur kebutuhan ini disebut sebagai *High Performance Computing (HPC)*. HPC umumnya mengacu pada praktik menggabungkan daya komputasi dengan cara memberikan kinerja yang jauh lebih tinggi daripada komputer desktop atau workstation biasa untuk memecahkan masalah besar dalam sains, teknik, atau bisnis.

"Komputasi Kinerja Tinggi," atau HPC, adalah penerapan "superkomputer" untuk masalah komputasi yang terlalu besar untuk komputer standar atau akan memakan waktu terlalu lama. Komputer *desktop* umumnya memiliki satu chip pemrosesan, yang biasa disebut CPU. Sistem HPC, pada dasarnya adalah jaringan *node*, yang masing-masing berisi satu atau lebih *chip* pemrosesan, serta memorinya sendiri.

Menurut situs www.top500.org yang menampilkan 500 komputer tercepat atau terbaik di dunia, sepuluh negara industri menempati 10 tempat pertama. Hal tersebut mengindikasikan adanya hubungan yang kuat antara kematangan HPC suatu negara dengan kemajuan industrinya. Pada masa kemunculan COVID-19, komunitas HPC di seluruh dunia sangat berperan untuk mengatasi wabah ini dari sisi komputasi. Komunitas HPC di Amerika Serikat (<https://covid19-hpc-consortium.org/>) dan Eropa menyumbangkan waktu komputasi *super computer*-nya untuk penelitian melawan COVID-19. Konsorsium ini terdiri dari 43 anggota, 165K *nodes*, 87 *active projects*, 6.8 juta *cpu cores*, 50k *gpu*, 600 peta flops. Aktivitas komunitas ini antara lain, misalnya memahami bagaimana hubungan COVID-19 dengan wabah sebelumnya, Sars-Cov 2002 dan MersCov 2012; menargetkan-1 PRF untuk terapi COVID-19; desain obat berbasis struktur komputasi; mengidentifikasi antivirus dari produk alami yang menargetkan SARS-CoV-2; beberapa kemungkinan obat (*possible drugs*) untuk COVID-19 pun sudah ditemukan [Seo, 2020].

Policy Brief ini disusun untuk memberikan rekomendasi dalam pengelolaan HPC dan *Big Data* di UI agar sivitas akademika UI dapat memaksimalkan penggunaan teknologi *Big Data* dalam pelaksanaan tridharma dan memberikan dampak besar terhadap pengembangan UI. Kebijakan yang perlu dipertimbangkan UI terkait *Big Data* adalah mengenai *Data Management*, *Data Security*, *Data Governance*, Infrastruktur TI, Pelatihan, Kerja sama, dan Inovasi. Selain itu perlu ditambahkan: bagaimana teknologi *Big Data* digunakan dalam berbagai sektor, keuntungan dan risiko penggunaan teknologi *Big Data*, serta tantangan yang harus diatasi untuk memaksimalkan manfaatnya.



Landasan Hukum dan Filosofis

Peraturan yang dipertimbangkan dalam pembuatan *policy brief* ini adalah sebagai berikut:

1

Undang-Undang Republik Indonesia Nomor 12 Tahun 2012 tentang Pendidikan Tinggi pasal 5 butir b menyebutkan bahwa salah satu tujuan Pendidikan Tinggi adalah: ‘dihasilkannya lulusan yang menguasai cabang Ilmu Pengetahuan dan/atau Teknologi untuk memenuhi kepentingan nasional dan peningkatan daya saing bangsa’;

2

Peraturan Pemerintah Republik Indonesia Nomor 75 Tahun 2021 tentang Statuta Universitas Indonesia Pasal 2 Ayat (2) butir c menyebutkan bahwa salah satu misi UI adalah ‘menciptakan lulusan yang berintelektualitas tinggi, berbudi pekerti luhur, dan mampu bersaing secara global, yang memiliki pola pikir, pola sikap, dan pola tindak dalam semangat kebangsaan’.

Peraturan pendukung terkait kebutuhan pengelolaan HPC dan *Big Data* pada tingkat UI maupun tingkat Nasional sejauh ini belum ada, sementara aktivitas dan kebutuhan pengelolaan HPC dan *Big Data* sudah semakin mendesak di kalangan pihak yang membutuhkan.



Analisis Situasi

1

Pengelolaan HPC dan *Big Data* di UI diawali dengan kegiatan penelitian dan pengajaran Dosen setelah selesai studi S3 di luar negeri. Sejak saat itu hingga kini kebutuhan HPC dan *Big Data* masih dilakukan atas inisiasi sendiri secara *self-supported*. Misalnya untuk keperluan penelitian dan pengajaran, peneliti menyisihkan anggarannya untuk membeli perangkat pendukung HPC dan *Big Data* yang masih minim. Sementara itu, kegiatan penelitian HPC dan *Big Data* di UI sudah semakin berkembang dengan melibatkan mahasiswa S1 sampai S3.

2

Beberapa tahun yang lalu Fakultas Ilmu Komputer (Fasilkom) UI mendapatkan donasi mesin NVIDIA DGX-1 dari Tokopedia untuk kegiatan penelitian di Tokopedia *AI Research Center* di UI. Fasilkom UI kemudian memperoleh Hibah dari Dikti untuk memperkuat infrastruktur dengan pengadaan mesin NVIDIA DGX A100. Peneliti dan mahasiswa yang memanfaatkan antara lain dari Fasilkom, FK, Farmasi, dan FMIPA UI.

3

Pada saat bersamaan, beberapa peneliti di perguruan tinggi lain juga memperoleh hibah dari Dikti untuk peralatan yang sama. Institusi lain seperti BRIN pada saat yang sama juga telah memiliki fasilitas HPC.

Model Struktur Organisasi

Kondisi di atas menggambarkan belum ada koordinasi atau pengelolaan dalam struktur organisasi di tingkat institusi maupun negara. *Best practices* di 10 negara urutan pertama di situs www.top500.org terkait model struktur organisasi adalah struktur organisasi yang dapat secara efektif mengelola HPC dan *Big Data* di institusi, yaitu:

Model Terpusat	Model Terdistribusi
<p>Pada model ini, departemen atau tim terpusat bertanggung jawab untuk mengelola HPC dan infrastruktur, serta layanan <i>Big Data</i>. Pada model terpusat, tim pusat menangani pengadaan, pemeliharaan, dan administrasi perangkat keras, perangkat lunak, dan sistem penyimpanan. Tim pusat juga menyediakan dukungan pengguna, pelatihan, dan alokasi sumber daya untuk peneliti dan pengguna lain dalam lembaga.</p>	<p>Dalam model ini, setiap departemen atau kelompok penelitian memiliki HPC dan sumber daya <i>Big Data</i> serta staf pendukung. Model ini memungkinkan kontrol dan fleksibilitas lebih lokal, karena setiap kelompok dapat menyesuaikan infrastruktur dengan kebutuhan spesifiknya. Namun, koordinasi dan kolaborasi antara kelompok yang berbeda memerlukan upaya tambahan.</p>
Model Hibrid	Model Federasi
<p>Model ini memadukan unsur sentralisasi dan pendekatan terdistribusi. Tim HPC pusat mengelola kumpulan inti dari sumber daya dan layanan berkinerja tinggi, sementara masing-masing departemen atau kelompok penelitian memelihara infrastruktur khusus mereka sendiri. Tim pusat memberikan dukungan, koordinasi, dan integrasi antara sumber daya yang berbeda.</p>	<p>Dalam model ini, banyak institusi atau organisasi datang bersama untuk membentuk federasi untuk HPC dan <i>Big Data</i>. Setiap institusi memelihara infrastruktur dan staf pendukungnya sendiri, tetapi mereka berkolaborasi dan berbagi sumber daya di dalam federasi. Model ini memungkinkan pengumpulan sumber daya, pembagian biaya, dan akses ke kumpulan sumber daya yang lebih besar dan lebih beragam.</p>

Pilihan struktur organisasi tergantung pada faktor-faktor seperti ukuran lembaga, sumber daya yang tersedia, kebutuhan penelitian, dan budaya lembaga. Penting untuk mempertimbangkan persyaratan dan tujuan spesifik institusi sebelum memutuskan struktur yang paling sesuai.

Kepemimpinan/Manajemen

Susunan organisasi struktural untuk mengelola HPC dan *Big Data* di institusi umumnya mencakup berbagai peran dan posisi. Latar belakang dan keahlian spesifik individu dalam perannya dapat bervariasi berdasarkan organisasi dan persyaratannya. Berikut adalah beberapa posisi umum dan latar belakang mereka:

A Direktur/ Manajer

Direktur/Manajer bertugas mengawasi HPC dan operasi *Big Data*, menetapkan sasaran strategis, mengelola anggaran, dan berkoordinasi dengan departemen atau pemangku kepentingan lain. Mereka biasanya memiliki latar belakang ilmu komputer, ilmu data, atau bidang terkait, bersama dengan pengalaman dalam mengelola proyek teknis yang rumit.

B Pakar Teknis

b.1. Administrator HPC: Para profesional ini bertanggung jawab atas administrasi harian dan pemeliharaan infrastruktur HPC, termasuk perangkat keras, perangkat lunak, jaringan, dan keamanan. Mereka memiliki latar belakang yang kuat dalam ilmu komputer, administrasi sistem, dan teknologi HPC.

b.2. Arsitek Data: bertugas merancang dan mengelola infrastruktur data organisasi, termasuk *database*, sistem penyimpanan data, dan proses integrasi data. Mereka memiliki keahlian dalam pemodelan data, sistem manajemen basis data, dan tata kelola data. Latar belakang mereka adalah ilmu komputer, sistem informasi, atau manajemen data.

b.3. Administrator Jaringan: Administrator jaringan menangani desain, implementasi, dan pemeliharaan infrastruktur jaringan lembaga, memastikan konektivitas yang andal untuk HPC dan sistem *Big Data*. Mereka memiliki latar belakang dalam teknologi jaringan, administrasi sistem, dan keamanan.

C Dukungan dan Pelatihan Pengguna

c.1. Spesialis Dukungan Pengguna: bertugas memberikan dukungan teknis dan bantuan untuk pengguna HPC dan sumber *Big Data*. Mereka membantu memecahkan masalah, menjawab pertanyaan pengguna, dan memberikan panduan tentang cara menggunakan

infrastruktur secara efektif. Latar belakang mereka kemungkinan besar adalah ilmu komputer, teknologi informasi, atau bidang terkait.

c.2. Koordinator Pelatihan: Koordinator pelatihan mengembangkan dan menyampaikan program pelatihan dan lokakarya untuk mengedukasi pengguna tentang HPC dan konsep *Big Data*, alat, dan praktik terbaik. Mereka memiliki pemahaman yang kuat tentang HPC dan teknologi *Big Data*, bersama dengan pengalaman dalam desain dan pengiriman instruksional.

**Penghubung/
D Konsultan
Penelitian**

Konsultan Komputasi Riset: bekerja sama dengan peneliti dan staf akademik untuk memahami komputasi dan data mereka. Mereka memberikan panduan dalam memilih HPC yang sesuai dan besar sumber daya data, mengoptimalkan alur kerja, dan membantu analisis data dan penafsiran. Mereka memiliki latar belakang dalam disiplin ilmu yang relevan dan keahlian dalam metode komputasi.

Hal-hal penting untuk diperhatikan adalah peran dan latar belakang tertentu dapat bervariasi tergantung pada ukuran, sumber daya, dan kebutuhan spesifik lembaga. Beberapa organisasi mungkin memiliki posisi tambahan atau jabatan yang berbeda untuk memenuhi kebutuhan unik mereka.

Sedangkan *best practices* di level nasional di beberapa negara adalah sebagai berikut: Dalam konteks nasional, pengelolaan HPC dan *Big Data* umumnya melibatkan koordinasi berbagai organisasi, lembaga penelitian, dan badan pemerintah. Meskipun organisasi struktural tertentu dapat bervariasi antar negara, terdapat beberapa pendekatan umum:

**National
Supercomputing
Centers**

Banyak negara mendirikan pusat superkomputer nasional atau fasilitas komputasi berkinerja tinggi yang berfungsi sebagai hub pusat untuk mengelola HPC dan sumber daya *Big Data*. Pusat-pusat tersebut menyediakan infrastruktur canggih, layanan dukungan, dan keahlian bagi para peneliti dan institusi di seluruh negeri. Mereka umumnya memiliki tim khusus yang bertanggung jawab atas administrasi, dukungan pengguna, dan fasilitasi kolaborasi.

<p>National Research and Education Networks (NRENs)</p>	<p>NRENs adalah jaringan lanjutan yang menghubungkan lembaga penelitian dan pendidikan di suatu negara. Mereka sering memainkan peran penting dalam mengelola dan memfasilitasi HPC serta sumber daya dan layanan <i>Big Data</i>. NREN berkolaborasi dengan universitas, lembaga penelitian, dan lembaga pemerintah untuk menyediakan konektivitas berkecepatan tinggi, kemampuan transfer data, dan akses ke sumber daya komputasi terdistribusi.</p>
<p>Inisiatif Ilmu Data Nasional</p>	<p>Beberapa negara membentuk prakarsa atau lembaga ilmu data nasional untuk mengawasi pengelolaan dan pemanfaatan sumber daya <i>Big Data</i> di tingkat nasional. Inisiatif tersebut bekerja untuk mengembangkan strategi ilmu data, mempromosikan berbagi data dan kolaborasi, dan mendorong kemitraan antara akademisi, industri, dan pemerintah.</p>
<p>Pusat Penelitian Interdisipliner</p>	<p>Di negara tertentu, pusat penelitian interdisipliner didirikan untuk mengatasi tantangan ilmiah atau teknologi tertentu, termasuk HPC dan <i>Big Data</i>. Pusat-pusat tersebut menyatukan para peneliti dari berbagai disiplin ilmu dan memberi mereka infrastruktur, pendanaan, dan keahlian yang diperlukan untuk melakukan penelitian mutakhir menggunakan komputasi canggih dan metode analisis data.</p>

Hal-hal penting untuk diperhatikan adalah bahwa organisasi struktural tertentu bergantung pada sumber daya, prioritas, dan struktur tata kelola negara. Beberapa negara mungkin memiliki pendekatan terpusat dengan satu badan koordinasi, sementara yang lain mungkin mengadopsi model yang lebih terdesentralisasi dengan tanggung jawab terdistribusi di antara banyak organisasi. Tujuannya adalah untuk mendorong kolaborasi, berbagi sumber daya, dan inovasi dalam HPC dan *Big Data* di seluruh negeri.

Rekomendasi

Rekomendasi di bawah ini dapat dipertimbangkan sebagai asupan di tingkat perguruan tinggi, institusi lain dan di tingkat nasional.

1. Perlu adanya kebijakan atau peraturan di tingkat Perguruan Tinggi dan nasional agar kebutuhan HPC dan *Big Data* dapat dikelola secara profesional (termasuk *data governance*).
2. Perlu adanya struktur organisasi di tingkat Perguruan Tinggi dan Nasional agar pengelolaan HPC dan *Big Data* dapat dilakukan secara efisien.
3. Penting untuk memastikan bahwa data yang dikumpulkan dan digunakan sesuai dengan regulasi dan standar industri, seperti HIPAA, GDPR, dan lain-lain.
4. Universitas dan lembaga penelitian perlu terlibat dalam pengembangan teknologi *Big Data* dan pelatihan sumber daya manusia dalam penggunaan teknologi tersebut.
5. Perlu adanya investasi untuk memperkuat infrastruktur teknologi informasi yang mampu menangani volume data yang besar dan kompleks.
6. Perlu adanya kerja sama antara universitas, dan lembaga penelitian yang membutuhkan proses *Big Data* dalam pengumpulan data dan penggunaan teknologi *Big Data* untuk meningkatkan perawatan pasien dan efisiensi sistem kesehatan secara keseluruhan.
7. Perlu adanya literasi data untuk seluruh mahasiswa UI (termasuk pemanfaatan *Big Data*).

Penutup

Dengan meningkatnya kebutuhan penggunaan teknologi *Big Data* dalam penelitian di UI, maka pengelolaan *Big Data* di UI perlu mendapat perhatian serius oleh Eksekutif. Sementara ini kegiatan penelitian dan Pendidikan terkait HPC dan *Big Data* masih dilakukan secara individu baik di tingkat Perguruan Tinggi maupun Nasional. Agar UI dan Indonesia dapat bersaing di tingkat global, diperlukan kebijakan atau peraturan di tingkat Perguruan Tinggi maupun nasional agar kebutuhan HPC dan *Big Data* dapat dikelola secara profesional, termasuk struktur organisasi di tingkat Perguruan Tinggi dan nasional.



Photo Credit: Yugo K. Isal

Referensi

"The 5 Vs of Big Data", oleh IBM. Artikel ini menjelaskan tentang lima karakteristik utama dari Big Data, yaitu Volume, Variety, Velocity, Veracity, dan Value. Artikel ini dapat ditemukan di <https://www.ibm.com/analytics/hadoop/big-data-analytics>.

"Big Data: The 3 Vs Everyone Must Know", oleh Bernard Marr. Artikel ini membahas tentang Volume, Variety, dan Velocity sebagai karakteristik utama Big Data. Artikel ini dapat ditemukan di <https://www.forbes.com/sites/bernardmarr/2016/04/28/big-data-the-3-vs-everyone-must-know/#6c01f7d675c6>.

"The 7 Key Components of Big Data", oleh Bernard Marr. Artikel ini menjelaskan tentang tujuh komponen penting dari Big Data, yaitu Data Sources, Data Storage, Data Processing, Data Analysis, Data Visualization, Data Access, dan Data Security. Artikel ini dapat ditemukan di <https://www.datamation.com/big-data/the-7-key-components-of-big-data.html>.

"The 5 Key Components of Big Data Architecture", oleh Martin Brown. Artikel ini membahas tentang lima komponen penting dari arsitektur Big Data, yaitu Data Ingestion, Data Storage, Data Processing, Data Analysis, dan Data Visualization. Artikel ini dapat ditemukan di <https://www.datanami.com/2018/07/11/the-5-key-components-of-big-data-architecture/>.

"The 10 Key Components of a Big Data Solution", oleh Ali Hamza. Artikel ini membahas tentang 10 komponen penting dari solusi Big Data, yaitu Data Acquisition, Data Integration, Data Storage, Data Processing, Data Analysis, Data Visualization, Data Security, Data Quality, Data Governance, dan Data Privacy. Artikel ini dapat ditemukan di <https://www.3pillarglobal.com/insights/the-10-key-components-of-a-big-data-solution>.

Dr. Sangjae Seo, Korea Institute of Science and Technology., the Centra Webinar on KISTI Supercomputing for COVID-19 Challenges, Oct 21, 2020 09:00 AM Eastern Time (US and Canada)
<https://www.usgs.gov/core-science-systems/sas/arc/about/what-high-performance-computing>.
<https://www.nics.tennessee.edu/computing-resources/what-is-hpc>.