

KAJIAN REKOMENDASI IMPLEMENTASI BIG DATA: END TO END APPROACH

(KONSEP, TEORI, PRAKTIK)

TA 2022

DIREKTORAT KETENAGALISTRIKAN,
TELEKOMUNIKASI DAN INFORMATIKA
KEMENTERIAN PERENCANAAN PEMBANGUNAN NASIONAL
2022

DAFTAR ISI

II	DAFTAR ISI	34	4. Hasil Pelaksanaan Kegiatan
III	DAFTAR GAMBAR		Kajian Implementasi Big Data :
IV	DAFTAR TABEL		End To End Approach
v	KATA PENGANTAR		
6	1. Pendahuluan		
6	Latar Belakang	34	Analisa SWOT
7	Tujuan	35	Tata Kelola Big Data
7	Ruang Lingkup	36	Sumber Daya Manusia Big Data
8	Keluaran	38	Arsitektur Big Data Nasional
9	2. Studi Literatur	40	Teknologi Big Data
9	Peraturan Pemerintahan Terkait	46	5. Strategi Pencapaian
17	Based Practices	46	Strategi Pengembangan
		48	Roadmap
31	3. Metode Pelaksanaan Kegiatan	54	Penutup
31	Metode Pelaksanaan		
32	Pengumpulan dan Analisa Data	55	Daftar Pustaka
33	Jadwal Pelaksanaan		

DAFTAR GAMBAR

Gambar 1. Pedoman Pelaksanaan SPBE

Gambar 2. Infrastruktur SPBE

Gambar 3. Penyelenggaraan SDI secara Terpadu

Gambar 4. Konsep dan Definisi dalam SDI

Gambar 5. *Data-Driven* Indonesia Framework

Gambar 6. *Data-Driven Maturity Level*

Gambar 7. Karakteristik *Big Data*

Gambar 8. *7 Big Data Tren in 2022*

Gambar 9. *7 Implementasi Big Data* di Berbagai Negara

Gambar 10. *7 Implementasi Big Data* di Sektor Industri

Gambar 11. *7 Metode Pelaksana*

Gambar 12. Analisa SWOT

Gambar 13. Tata Kelola *Big Data* Modifikasi dari Kim and Cho (2018)

Gambar 14. Karakteristik 5V pada *Big Data*

Gambar 15. Tantangan Praktik *Big Data*

Gambar 16. Sumber Daya Manusia Implementasi *Big Data*

Gambar 17. Arsitektur *Big Data* Nasional

Gambar 18. *Data Lake*

Gambar 19. *Data warehousing Architecture*

Gambar 20. Contoh Visualisasi Data

DAFTAR TABEL

- Tabel 1. Analisa Jurnal Implementasi Big Data
- Tabel 2. Jadwal Pelaksanaan
- Tabel 3. Roadmap Implementasi Big Data
- Tabel 4. Roadmap Tata Kelola Big Data
- Tabel 5. Roadmap Teknologi Big Data
- Tabel 6. Roadmap Infrastruktur dan Keamanan Big Data

KATA PENGANTAR

Puji syukur dan terima kasih kami panjatkan kepada Allah SWT atas rahmat, hidayah, kesehatan, perlindungan, keamanan, serta motivasi dan karunia-Nya sehingga dapat tersusunnya kajian *“Big Data Implementation: End to End Approach”*.

Penyusunan kajian ini dilakukan dalam rangka menyusun strategi terkait pemanfaatan Big Data Analytics. Hal ini agar selaras dengan regulasi seperti Peraturan Presiden Nomor 95 Tahun 2018 tentang Sistem Pemerintahan Berbasis Elektronik (SPBE), Peraturan Presiden Nomor 39 Tahun 2019 tentang Satu Data Indonesia, Peraturan Menteri PAN/RB Nomor 59 Tahun 2020 tentang Pemantauan dan Evaluasi SPBE, serta Peraturan Menteri PPN/Bappenas Nomor 16 Tahun 2020 tentang Manajemen Data Sistem Pemerintahan Berbasis Elektronik. Kajian ini telah disusun secara maksimal dan dengan bantuan berbagai pihak sehingga dalam prosesnya dapat berjalan dengan lancar. Dengan terbatasnya waktu, pengetahuan dan pengalaman, kami menyadari terdapat kekurangan dalam kajian ini. Oleh karena itu, saran dan kritik yang membangun sangat diharapkan dari pembaca dalam penyempurnaan kajian ini.

Dengan penuh kerendahan hati, kami menyampaikan rasa hormat dan menghaturkan terima kasih yang sebesar-besarnya kepada semua pihak yang telah membantu dalam penyusunan kajian *“Big Data Implementation: End to End Approach”* ini.

01

PENDAHULUAN

Latar Belakang

Gagasan dalam menggunakan data untuk pengambilan keputusan bukanlah sesuatu hal yang baru. Data memiliki peran strategis bagi setiap organisasi untuk tetap bertahan dan dapat bersaing di era digital ini. Saat ini perhatian terhadap data semakin meningkat, tren teknologi big data dalam beberapa waktu belakangan ini telah menarik perhatian semua kalangan untuk memanfaatkan data mereka. Dengan teknologi analitika big data, organisasi bisa merangkul pelanggan (customer) untuk menciptakan bisnis dan budaya baru, sehingga pada akhirnya menjadi pemenang dalam persaingan.

Pemerintah Indonesia saat ini telah menyadari pentingnya data dalam pengambilan keputusan yang cepat dan tepat. Hal ini dimulai dari sejak diterbitkannya Peraturan Presiden Nomor 95 Tahun 2018 tentang Sistem Pemerintahan Berbasis Elektronik (SPBE). Dengan adanya peraturan ini diharapkan terwujudnya tata kelola pemerintahan yang bersih, efektif, transparan, dan akuntabel serta pelayanan publik yang berkualitas dan terpercaya. Berdasarkan data dari Kementerian PANRB tahun 2020, hasil evaluasi indeks SPBE Nasional masih dalam predikat cukup, dimana pada tahun 2018 hanya sebesar 1.98 dari skala 5, pada tahun 2019 dengan nilai 2.18 serta pada tahun 2020 dengan nilai 2.26. Hal ini menjadi tantangan bagi pemerintah pusat dan daerah bagaimana meningkatkan layanan publik untuk menjadi lebih efektif dan efisien.

Pemanfaatan data dalam pengambilan keputusan pada sektor pemerintahan di Indonesia membutuhkan dukungan dari banyak pihak, baik dari pihak industri, akademisi dan sektor lainnya. Melalui Peraturan Presiden Nomor 39 Tahun 2019 tentang Satu Data Indonesia, pemerintah berharap dapat mewujudkan keterpaduan perencanaan, pelaksanaan, evaluasi

dan pengendalian pembangunan yang didukung dengan data yang akurat, mutakhir, terpadu, dapat dipertanggungjawabkan, mudah diakses dan dibagipakaikan serta dikelola secara seksama, terintegrasi dan berkelanjutan. Namun dalam penerapan dilapangan baik ditingkat pusat maupun tingkat daerah masih ditemui banyak kendala terkait permasalahan dasar seperti infrastruktur yang belum memadai, investasi anggaran yang terbatas, serta sumber daya manusia yang belum kompeten. Selain itu, kerangka kerja penerapan *Big Data Analytics* dan langkah-langkah strategis sebagai panduan dalam penerapan satu data Indonesia belum tergambarkan secara jelas sehingga pemerintah pusat dan daerah tidak memiliki arah pasti terkait kebutuhan dan peta jalan yang seharusnya dilakukan.

Tujuan

Tujuan pelaksanaan kegiatan Penyusunan Implementasi Big Data menggunakan pendekatan *end to end* ini adalah sebagai berikut:

1. Melakukan assessment terhadap pelaksanaan penerapan regulasi nasional terkait penerapan Sistem Pemerintahan Berbasis Elektronik dan Satu Data Indonesia.
2. Melakukan studi literatur terhadap jurnal dan best practices penerapan Big Data Analytics.
3. Melakukan benchmarking analysis terhadap inisiatif yang telah dilakukan dalam penerapan Big Data Analytics.
4. Menyusun Arsitektur Big Data Analytics melalui pendekatan *end to end*.
5. Menyusun General Roadmap Implementasi penerapan Big Data Analytics.

Ruang Lingkup

Ruang lingkup pelaksanaan kegiatan Penyusunan Implementasi Big Data menggunakan pendekatan *end to end* ini adalah sebagai berikut:

1. Menyusun Arsitektur Big Data Nasional dengan Pendekatan *End to End* namun tidak terbatas hanya pada sektor pemerintahan saja.
2. Menyusun Arsitektur Big Data Nasional yang selaras dengan Sistem Pemerintahan Berbasis Elektronik (SPBE) dan Satu Data Indonesia (SDI).
3. Memberikan konsep terkait penerapan Data Driven.
4. Memberikan landasan teori yang kuat dalam penerapan Big Data
5. Menawarkan strategi dan rekomendasi pencapaian penerapan *End to End big Data Implementation*

Keluaran

Pengembangan Roadmap *“Implementasi Big Data Analytics: End to End Approach”* menjadi fokus dalam kajian ini. Roadmap ini disusun agar pemanfaatan Big Data Analytics berjalan terarah dan memperhatikan pendekatan yang terkait. Hal ini untuk menjaga agar implementasi berjalan komprehensif, tidak hanya memperhatikan aspek teknis saja, karena terdapat aspek non-teknis lain yang perlu dipertimbangkan untuk dikelola. Selanjutnya secara rinci kontribusi dari kajian ini adalah:

1. Tersusunnya suatu kajian yang menjelaskan perkembangan ilmu pengetahuan di bidang Big Data Analytics.
2. Terciptanya Arsitektur Big Data dengan pendekatan End to End yang dilandasi oleh kajian teori dan best practices agar dapat diterapkan pada lingkungan instansi pemerintah maupun industri di Indonesia.
3. Tersusunnya roadmap penerapan big data secara end to end sebagai pedoman dalam pemanfaatan big data analytics baik bagi instansi pemerintahan maupun industri pada umumnya.

02 STUDI LITERATUR

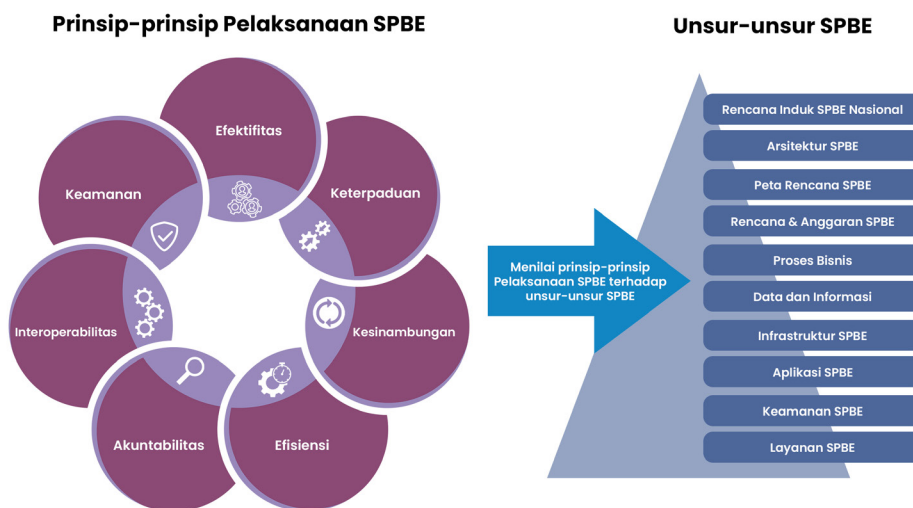
Peraturan Pemerintahan Terkait

Peraturan Presiden Nomor 95 Tahun 2018

Berdasarkan Peraturan Presiden Nomor 95 Tahun 2018, Tata kelola SPBE adalah kerangka kerja yang memastikan terlaksananya pengaturan, pengarahannya, dan pengendalian dalam penerapan SPBE secara terpadu. Arsitektur SPBE adalah kerangka dasar yang mendeskripsikan integrasi proses bisnis, data dan informasi, infrastruktur SPBE, aplikasi SPBE dan keamanan SPBE untuk menghasilkan layanan SPBE yang terintegrasi. Peta Rencana SPBE adalah dokumen yang mendeskripsikan arah dan langkah penyiapan dan pelaksanaan SPBE yang terintegrasi. Infrastruktur SPBE adalah semua perangkat keras, perangkat lunak dan fasilitas yang menjadi penunjang utama untuk menjalankan sistem, aplikasi, komunikasi data, pengolahan dan penyimpanan data, perangkat integrasi/penghubung dan perangkat elektronik lainnya.

Sedangkan, Sistem Pemerintahan Elektronik (SPBE) adalah penyelenggaraan pemerintahan yang memanfaatkan teknologi informasi dan komunikasi untuk memberikan layanan kepada pengguna SPBE. Layanan SPBE adalah keluaran yang dihasilkan oleh satu atau beberapa fungsi aplikasi SPBE secara nasional untuk jangka waktu 20 tahun. Sistem Penghubung layanan adalah perangkat integrasi/penghubung untuk melakukan pertukaran layanan SPBE. Aplikasi SPBE dikelompokkan menjadi 2 yaitu Aplikasi Umum dan aplikasi khusus. Aplikasi Umum adalah aplikasi SPBE yang sama, standar dan digunakan secara bberbagi pakai oleh instansi pusat maupun pemerintah daerah. Aplikasi Khusus adalah aplikasi SPBE yang dibangun, dikembangkan, digunakan dan dikelola oleh instansi pusat atau pemerintah daerah untuk memenuhi kebutuhan khusus yang bukan kebutuhan instansi pusat dan pemerintah daerah lain.

Gambar 1. Pedoman Pelaksanaan SPBE



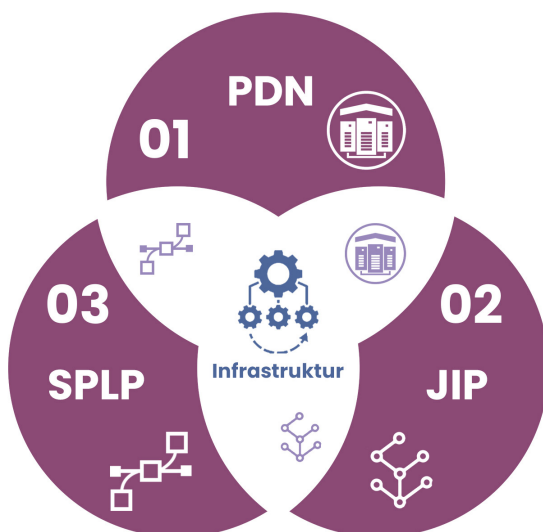
Pengaturan Sistem Pemerintah Berbasis Elektronik dimaksudkan untuk mengatur penyelenggaraan Tata Kelola SPBE yang bertujuan untuk memastikan penerapan unsur-unsur SPBE diantaranya :

- Rencana induk SPBE Nasional yang bertujuan untuk memberikan arah SPBE yang terpadu dan bekersistinambungan secara nasional yang memuat visi, misi, tujuan, sasaran, arah kebijakan strategi dan peta rencana strategis SPBE. Rencana induk SPBE Nasional disusun berdasarkan Rencana Pembangunan Jangka Panjang Nasional dan Grand Design reformasi Birokrasi.
- Arsitektur SPBE Nasional yang bertujuan untuk memberikan panduan dalam pelaksanaan integrasi Proses Bisnis, data dan informasi, Infrastruktur SPBE, Aplikasi SPBE dan Keamanan SPBE untuk menghasilkan layanan SPBE yang terpadu secara nasional yang disusun untuk jangka waktu 5 tahun. Arsitektur SPBE disusun berdasarkan Rencana induk SPBE Nasional dan Rencana Pembangunan Jangka Menengah Nasional. Adapun dalam penyusunan Arsitektur SPBE harus memuat Referensi arsitektur yang mendeskripsikan komponen dasar arsitektur baku yang digunakan sebagai acuan untuk setiap domain arsitektur. Sedangkan domain arsitektur merupakan deskripsi dari substansi arsitektur yang memuat domain arsitektur Proses Bisnis, domain arsitektur data dan informasi, domain arsitektur infrastruktur SPBE, domain arsitektur aplikasi SPBE, domain Keamanan SPBE dan domain arsitektur layanan SPBE.
- Data dan Informasi mencakup semua jenis data dan informasi yang dimiliki oleh Instansi Pusat dan Pemerintah daerah dan/atau yang diperoleh dari masyarakat, pelaku usaha dan/atau pihak lain yang mengutamakan bagi pakai dengan berdasarkan tujuan dan cakupan, penyediaan akses data dan informasi dan sesuai dengan standar interoperabilitas data dan informasi. Sedangkan untuk tata kelola data dan informasi antar Instansi Pusat dan/atau Pemerintah daerah yang dikoordinasikan oleh Menteri yang menyelenggarakan urusan pemerintahan di bidang perencanaan pembangunan nasional.
- Infrastruktur SPBE, penggunaan infrastruktur SPBE nasional bertujuan untuk meningkatkan efisiensi, keamanan dan kemudahan integrasi dalam rangka memenuhi

kebutuhan infrastruktur SPBE bagi instansi Pusat dan Pemerintah Daerah yang terdiri dari infratraktur SPBE nasional seperti Pusat Data Nasional, Jaringan Intra Pemerintah dan Sistem Penghubung Layanan Pemerintah dan Infrastruktur Instansi pusat dan Pemerintah daerah seperti, Jaringan intra dan Sistem penghubung layanan instansi Pusat dan Pemerintah Daerah.

- Aplikasi SPBE digunakan oleh Instansi Pusat dan Pemerintah daerah untuk meberikan layanan secara elktronik yang dikelompokan berdasarkan aplikasi umum dan aplikasi khusus yang diman dalam pembangunan maupun pengembangan aplikasi menggunakan kode sumber terbuka dan dika menggunakan kode sumber tertutup instansi pusat maupun pemerintah daerah harus mencapatakan pertimbangan dari Menteri yang menyelenggarakan urusan pemerintahan di bidang komunikasi dan informatika.
- Keamanan SPBE mencakup penjaminan kerahasiaan, keutuhan, ketersediaan, keaslian, dan kenirsangkalan (nonrepudiation) sumber data terkait data dan informasi, infrastruktur SPBE dan aplikasi SPBE
- Layanan SPBE terdiri dari layanan administrasi pemerintahan berbasis elektronik dan layanan publik berbasis elektronik.

Gambar 2. Infrastruktur SPBE



Adapun penyelenggaraan Manajemen SPBE yang dilaksanakan oleh Instansi Pusat atau Pemeritah Daerah dilaksanakan sesuai dengan pedoman Standar Nasional Indonesia meliputi Manajemen Resiko yang bertujuan untuk menjamin keberlangsungan SPBE dengan meminimalkan dampak resiko dalam pelaksanaan SPBE yang dalam pelaksanaannya dipimpin oleh Instansi Pusat dan Kepala daerah berkoordinasi dan dapat melakukan konsultasi dengan Menteri yang menyelenggarakan urusan pemerintahan di bidah aparatur negara. Manajemen Keamanan Informasi dilkasanakan melalui serangkaian proses yang meliputi penetapan ruang lingkup, penetapan pengoperasian, evaluasi kinerja dan perbaikan berkelanjutan terhadap keamanan informasi SPBE. Manajemen Data dilaksnakan untuk menjamin terwujudnya data

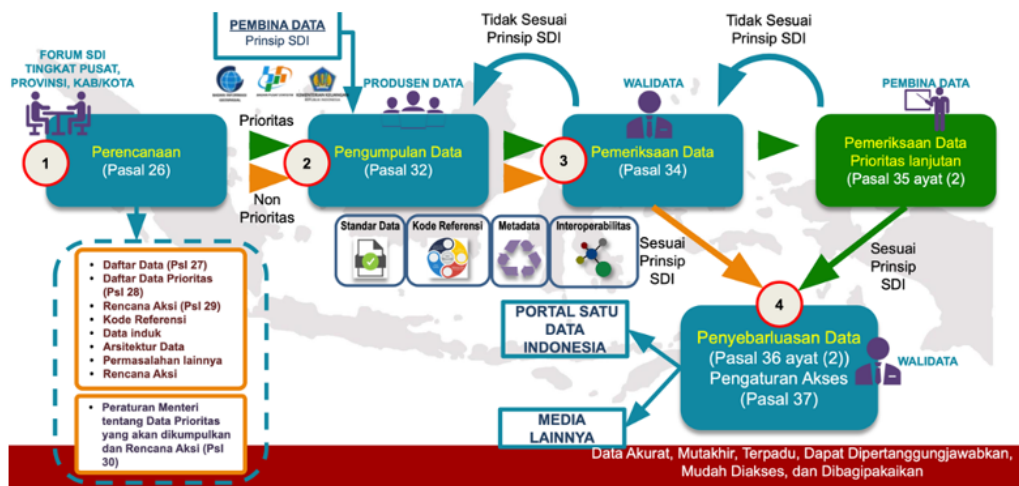
yang akurat, mutakhir, terintegrasi dan dapat diakses sebagai dasar perencanaan, pelaksanaan, evaluasi dan pengendalian pembangunan nasional. Manajemen Aset teknologi informasi dan komunikasi dilakukan melalui serangkaian proses perencanaan, pengadaan, penolakan dan penghapusan perangkat keras dan perangkat lunak yang digunakan dalam SPBE yang bertujuan untuk menjamin ketersediaan dan optimalisasi pemanfaatan asset teknologi informasi dan komunikasi dalam SPBE. Manajemen sumber daya manusia dilaksanakan untuk memastikan ketersediaan dan kompetensi sumber daya manusia yang bertujuan untuk menjamin keberlangsungan dan peningkatan layanan dalam SPBE. Manajemen pengetahuan dilaksanakan melalui serangkaian proses pengumpulan, pengolahan, penyimpanan, penggunaan dan alih pengetahuan dan teknologi yang dihasilkan dalam SPBE yang bertujuan untuk meningkatkan kualitas layanan SPBE dan mendukung proses pengambilan keputusan. Manajemen perubahan dilakukan melalui serangkaian proses perencanaan, analisis, pengembangan, implementasi, pemantauan dan evaluasi terhadap perubahan SPBE yang bertujuan untuk menjamin keberlangsungan dan meningkatkan kualitas layanan SPBE melalui pengendalian perubahan. Manajemen layanan SPBE dilaksanakan melalui serangkaian proses pekayaan pengguna SPBE, pengoperasian layanan SPBE dan pengelolaan Aplikasi SPBE yang bertujuan untuk menjamin keberlangsungan dan meningkatkan kualitas layanan SPBE kepada pengguna SPBE.

Dalam pelaksanaan Audit Teknologi Informasi dan Komunikasi SPBE dilakukan audit terhadap Aplikasi dan Keamanan SPBE yang meliputi penerapan tata kelola dan manajemen teknologi informasi, fungsionalitas teknologi informasi dan komunikasi, kinerja teknologi informasi dan komunikasi yang dihasilkan yang dilakukan oleh Lembaga pelaksana Audit TIK atau Lembaga pelaksana Audit TI yang terakreditasi sesuai dengan peraturan perundang-undangan. Pemantauan dan evaluasi SPBE bertujuan untuk mengukur kemajuan dan meningkatkan kualitas SPBE di instansi Pusat dan Pemerintah daerah yang dilakukan oleh Tim koordinasi SPBE nasional terhadap SPBE secara nasional dan oleh coordinator SPBE instansi pusat dan pemerintah daerah terhadap SPBE pada instansi pusat dan pemerintah daerah masing-masing.

Untuk meningkatkan keterpaduan pelaksanaan Tata Kelola SPBE, Manajemen SPBE dan Audit Teknologi Informasi dan Komunikasi, serta pemantauan dan evaluasi SPBE nasional maka dibentuk Tim Koordinasi SPBE Nasional yang bertugas melakukan koordinasi dan penerapan kebijakan SPBE pada Instansi Pusat dan Pemerintah Daerah yang di dalamnya terdiri dari Ketua yang dipilih dari Menteri yang menyelenggarakan urusan pemerintah di bidang aparatur negara dan Anggota yang terdiri dari Menteri yang menyelenggarakan urusan pemerintahan dalam negeri, bidang keuangan, bidang komunikasi dan informatika, bidang perencanaan dan pembangunan nasional, kepala lembaga yang menyelenggarakan tugas di bidang keamanan siber dan bidang pengkajian dan penerapan teknologi. Selain itu, percepatan penyelenggaraan SPBE diperlukan untuk meningkatkan kualitas penyelenggaraan pemerintah dan pelayanan publik dengan membangun Aplikasi Umum dan Infrastruktur SPBE nasional untuk memberikan Layanan SPBE yang mendukung kegiatan pemerintah di bidang perencanaan, penganggaran, pengadaan barang dan jasa pemerintah, akuntabilitas kinerja, pemantauan dan evaluasi, kearsipan, kepegawaian, dan pengaduan pelayanan publik.

Peraturan Presiden Nomor 39 Tahun 2019

Gambar 3. Penyelenggaraan SDI secara Terpadu



Sumber: Bappenas

Berdasarkan Peraturan Menteri Nomor 39 Tahun 2019 Data adalah catatan atas kumpulan fakta atau deskripsi berupa angka, karakter, simbol, gambar, peta, tanda, isyarat, tulisan, suara, dan/atau bunyi, yang merepresentasikan keadaan sebenarnya atau menunjukkan suatu ide, objek, kondisi, atau situasi. Data Statistik adalah data berupa angka tentang karakteristik atau ciri khusus suatu populasi yang diperoleh dengan cara pengumpulan, pengolahan, penyajian, dan analisis. Data Geospasial adalah data tentang lokasi geografis, dimensi atau ukuran, dan/atau karakteristik objek alam dan/atau buatan manusia yang berada di bawah, pada, atau di atas permukaan bumi.

Gambar 4. Konsep dan Definisi dalam SDI

Konsep dan Definisi dalam SDI



Sumber: BPS

Portal Satu Data Indonesia adalah media bagi-pakai Data di tingkat nasional yang dapat diakses melalui pemanfaatan teknologi informasi dan komunikasi. Pembina Data adalah Instansi Pusat yang diberi kewenangan melakukan pembinaan terkait Data atau Instansi Daerah yang diberikan penugasan untuk melakukan pembinaan terkait Data. Adapun walidata adalah unit pada Instansi Pusat dan Instansi Daerah yang melaksanakan kegiatan pengumpulan, pemeriksaan, dan pengelolaan Data yang disampaikan oleh Produsen Data, serta menyebarluaskan Data. Produsen Data adalah unit pada Instansi pusat dan Instansi Daerah yang menghasilkan Data berdasarkan kewenangan sesuai dengan ketentuan peraturan perundang-undangan. Pengguna Data adalah Instansi Pusat, Instansi Daerah, perseorangan, kelompok orang, atau badan hukum yang menggunakan Data.

1. Memberikan acuan pelaksanaan dan pedoman bagi Instansi Pusat dan instansi Daerah dalam rangka penyelenggaraan tata kelola data untuk mendukung perencanaan, pelaksanaan, evaluasi, dan pengendalian pembangunan;
2. Mewujudkan ketersediaan data yang akurat, mutakhir, terpadu, dapat dipertanggungjawabkan, serta mudah diakses dan dibagi pakaikan antar Instansi Pusat dan Instansi Daerah sebagai dasar perencanaan, pelaksanaan, evaluasi, dan pengendalian pembangunan;
3. Mendorong keterbukaan dan transparansi data sehingga tercipta perencanaan dan perumusan kebijakan pembangunan yang berbasis pada data; dan
4. Mendukung sistem statistik nasional peraturan perundang-undangan.

Satu Data Indonesia harus dilakukan berdasarkan prinsip sebagai berikut:

1. Data yang dihasilkan oleh Produsen Data harus memenuhi Standar Data;
2. Data yang dihasilkan oleh Produsen Data harus memiliki Metadata;
3. Data yang dihasilkan oleh Produsen Data harus memenuhi kaidah Interoperabilitas Data; dan
4. Data yang dihasilkan oleh Produsen Data harus menggunakan Kode Referensi dan/atau Data Induk.

Untuk memenuhi kaidah Interoperabilitas data tersebut, maka data harus:

1. Konsisten dalam sintak/bentuk, struktur/skema/komposisi penyajian, dan semantik/artikulasi keterbacaan; dan
2. Disimpan dalam format terbuka yang dapat dibaca sistem elektronik.

Selain didasarkan pada prinsip-prinsip diatas, penerapan satu data Indonesia dikoordinasikan oleh dewan pengarah yang terdiri atas ketua merangkap anggota, yaitu menteri yang menyelenggarakan urusan pemerintahan di bidang perencanaan pembangunan nasional. Adapun anggota terdiri dari menteri yang menyelenggarakan urusan pemerintahan di bidang pendayagunaan aparatur Negara, menteri yang menyelenggarakan urusan pemerintahan di bidang komunikasi dan informatika, menteri yang menyelenggarakan urusan pemerintahan dalam negeri, menteri yang menyelenggarakan urusan pemerintahan di bidang keuangan, kepala badan yang melaksanakan tugas pemerintahan di bidang kegiatan statistik dan kepala badan yang melaksanakan tugas pemerintahan di bidang informasi geospasial.

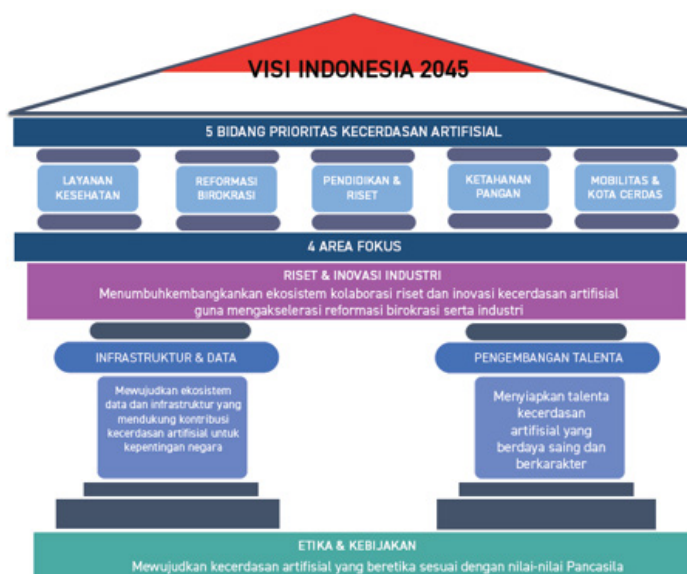
Dewan Pengarah mempunyai tugas adalah mengoordinasikan dan menetapkan kebijakan terkait Satu Data Indonesia, mengoordinasikan pelaksanaan Satu Data Indonesia, melakukan pemantauan dan evaluasi pelaksanaan Satu Data Indonesia, mengoordinasikan penyelesaian permasalahan dan hambatan pelaksanaan Satu Data Indonesia dan menyampaikan laporan penyelenggaraan Satu Data Indonesia tingkat pusat dan tingkat daerah kepada Presiden. Dalam melaksanakan tugas, Dewan Pengarah dapat melibatkan menteri atau kepala Instansi Pusat terkait lainnya.

Strategi Nasional Kecerdasan Artifisial

Berdasarkan Strategi Nasional Kecerdasan Artifisial Indonesia yang merupakan negara yang strategis yang berbentuk kepulauan dengan jumlah penduduk yang banyak dan beragam. Maka Indonesia memiliki sejuta peluang dalam pemanfaatan Kecerdasan Artifisial (KA) karena teknologi ini dapat berpotensi memberikan peningkatan produktivitas bagi bisnis, efisiensi investasi pemanfaatan sumber daya manusia, dan inovasi di berbagai sektor seperti keuangan, kesehatan, Pendidikan, pertanian, hankam, transportasi dan kelautan. Kecerdasan Artifisial juga dapat memberikan solusi dalam mengatasi masalah infrastruktur yang hemat biaya, memberikan layanan sosial yang efektif, merencanakan sumber daya Pendidikan yang berkualitas dan membantu pemerintah menyusun kebijakan yang tepat berdasarkan data.

Namun, Indonesia sebagai negara yang memiliki kekuatan pasar ekonomi digital terbesar saat ini di Asia Tenggara akan memiliki banyak tantangan yang harus dihadapi dalam menerapkan teknologi ini. Tantangan-tantangan itu dapat dikelompokkan ke dalam empat hal penting, yakni kesiapan tenaga kerja terampil yang mengembangkan dan menggunakan kecerdasan artifisial, kesiapan regulasi yang mengatur etika penggunaan dan pemanfaatan kecerdasan artifisial yang bertanggung jawab, kesiapan infrastruktur komputasi dan data pendukung pemodelan kecerdasan artifisial, dan kesiapan industri dan sektor-sektor publik dalam mengadopsi inovasi-inovasi kecerdasan artifisial. Untuk menghadapi peluang dan tantangan tersebut di atas, pemerintah Indonesia memerlukan penyusunan sebuah strategi kecerdasan artifisial Indonesia dengan memperhatikan dan memperhitungkan isu-isu yang ada di strategi kecerdasan Artifisial negara-negara lain baik global maupun regional.

Pemerintah Indonesia berupaya untuk menyusun strategi nasional kecerdasan artifisial melalui pendekatan yang holistik, memiliki definisi kecerdasan artifisial yang jelas, memiliki sasaran yang terukur, dan mempertimbangkan solusi berupa ekosistem yang dapat mengorkestra seluruh kekuatan dan potensi yang dimiliki Indonesia. Strategi yang dibuat harus layak dan efektif untuk kemajuan bangsa. Selain itu, Indonesia akan mencanangkan sebuah slogan nasional kecerdasan artifisial yang dapat memberikan keyakinan kecerdasan artifisial untuk kesuksesan Visi Indonesia 2045.



Selain itu, Kecerdasan artifisial dikembangkan dan dimanfaatkan dengan tujuan untuk kepentingan positif bagi umat manusia. Dengan orientasi tujuan ini, diharapkan dapat menimbulkan kepercayaan publik (*public trust*) pada kecerdasan artifisial, berbagai sektor baik dari social dan ekonomi, yang dirasakan nyata oleh manusia. Untuk mendukung terwujudnya kecerdasan artifisial salah satunya berkaitan dengan tata kelola data dan privasi dimana Sistem Kecerdasan Artifisial harus menjamin privasi dan perlindungan data, disertai protocol data (untuk mengatur akses data). Dengan memperhatikan kemampuan Kecerdasan Artifisial untuk belajar secara mandiri, maka kualitas dan integritas data juga harus dijaga dan dilindungi. Adapun program inisiatif untuk terwujudnya keterbukaan untuk pertukaran data- data digital adalah setiap penyelenggara sistem elektronik pelayanan publik harus memperhatikan keamanan data dan kerahasiaan data pribadi melalui penguatan dan implementasi PP71/2019 tentang Penyelenggaraan Sistem dan Transaksi Elektronik(PSTE), melakukan kajian dan pendefinisian standarisasi skema data dan antarmuka sistem pertukaran data (SOA-Service Oriented Architecture atau Rest APIs). Selain itu tersedianya sistem penghubung data antara produsen dan consumer data juga perlu di dorong melalui sosialisasi dan penguatan implementasi Satu Data Indonesia dan melakukan kajian, pendampingan dan dukungan agar infrastruktur komunikasi dan penyimpanan data aman efisien dan Tangguh. Ketika KA sudah digunakan dengan menggunakan data-data yang diperlukan, maka sangat penting bahwa data-data tersebut akan selalu tersedia agar KA menjalankan fungsi yang dikehendaki. Data – data tersebut harus disimpan dalam suatu penyimpanan terpusat atau terdistribusi dengan sistem manajemen yang terorkestrasikan. Infrastruktur ini akan digunakan untuk persiapan, pemindahan, penyimpanan, pengolahan, dan analisa baik yang terpusat maupun yang terdistribusi dengan sistem manajemen dan akses yang aman, dan terkelola.

Sumber-sumber data pada umumnya terakumulasi secara sectoral dan di masing-masing instansi (data silos). Strategi yang tepat untuk mengatasi kendala data sektoral bukan dengan mengumpulkan semua data di satu tempat, tetapi dengan melakukan data integrasi. Secara teknologi saat ini, penyimpanan Big Data yang besar (*Volume*), beragam (*Variety*), dan cepat (*Velocity*) adalah dengan menggunakan Data Lake yang terimplementasi di infrastruktur Cloud (baik public maupun *private*). Dengan adanya data integrasi, data tetap terkumpul di masing masing instansi, tetapi ada penghubung dari setiap sumber data instansi ke pusat Data Lake di Cloud. Data akan lebih bermanfaat apabila basis-basis data sudah terhubung dan terintegrasi. Dengan demikian harus ada platform untuk jalur akses data yang terkendali dari Produsen data ke Data Lake yang dikelola oleh Data Broker, dan kemudian ke Konsumen data. Platform yang dimaksud adalah aplikasi di Cloud dengan penyediaan back-end services, frontend interface, disertakan dengan API untuk akses data.

Untuk mencapai sasaran strategis yang telah ditetapkan, dibutuhkan proses yang mengarahkan pada terciptanya produk , layanan dan solusi teknologi Kecerdasan artifisial yang tepat guna di sektor industri prioritas . Berikut merupakan beberapa teknologi Kecerdasan artifisial yang perlu dibangun atau dimanfaatkan untuk mendukung terciptanya produk,Layanan dan solusi berbasis kecerdasan buatan adalah Big data yang mejadi syarat mutlak bagi perkembangan teknologi Kecerdasan Artifisial modern yang di dominasi oleh machine learning.

Based Practices

Konsep Data Driven

Gagasan dalam menggunakan data untuk pengambilan keputusan bukanlah sesuatu hal yang baru. Data memiliki peran strategis bagi setiap organisasi untuk tetap bertahan dan dapat bersaing di era digital ini. Saat ini perhatian terhadap data semakin meningkat, trendnya teknologi big data dalam beberapa waktu belakangan ini telah menarik perhatian semua kalangan untuk memanfaatkan data mereka. Dengan teknologi analitika *big data*, organisasi bisa merangkul pelanggan (*customer*) untuk menciptakan bisnis dan budaya baru, sehingga pada akhirnya menjadi pemenang dalam persaingan.

Pada sektor pemerintah, hakikinya data merupakan sebuah asset strategis dalam penentuan kebijakan, peningkatan pelayanan dan pemecahan permasalahan masyarakat. Perumusan kebijakan berbasis data merupakan kebutuhan yang mendesak, akan tetapi cepatnya akselerasi dari volume data yang tersedia saat ini secara umum telah melampaui kemampuan pemerintah dalam mengelola dan menggunakan data tersebut untuk pengambilan keputusan. Setiap kebijakan yang dikeluarkan diperoleh melalui analisis data dari berbagai sumber, agar menghasilkan kebijakan yang tepat dan akurat serta dapat diperoleh perspektif ilmiah terhadap pendekatan penyelesaian masalah.

Pemerintah Indonesia saat ini telah menyadari pentingnya data dalam pengambilan keputusan yang cepat dan tepat. Melalui Peraturan Presiden Nomor 39 Tahun 2019 tentang Satu Data Indonesia, pemerintah berharap dapat mewujudkan keterpaduan perencanaan, pelaksanaan, evaluasi dan pengendalian pembangunan yang didukung dengan data yang akurat, mutakhir, terpadu, dapat dipertanggungjawabkan, mudah diakses dan dibagipakaikan serta dikelola secara seksama, terintegrasi dan berkelanjutan. Namun dalam penerapan dilapangan baik ditingkat pusat maupun tingkat daerah masih ditemui banyak kendala terkait permasalahan dasar seperti infrastruktur yang belum memadai, investasi anggaran yang terbatas, serta sumber daya manusia yang belum kompeten. Selain itu, langkah-langkah strategis sebagai panduan dalam penerapan satu data Indonesia belum tergambarkan secara jelas sehingga pemerintah pusat dan daerah tidak memiliki arah pasti terkait kebutuhan dan peta jalan yang seharusnya dilakukan.

Dewan TIK Nasional telah menyusun framework *Data-Driven* sebagai tindak lanjut dari Peraturan Presiden Nomor 39 Tahun 2019. *Data-Driven Framework* disusun berdasarkan praktik terbaik dari berbagai strategi teknologi di berbagai kota terkemuka di dunia. Adapun *framework* tersebut dapat dilihat pada gambar 5 berikut:

Gambar 5. Data-Driven Indonesia Framework



Berdasarkan gambar 5 diatas, *Framework Data-Driven* Indonesia terdiri dari 6 komponen sebagai berikut:

- *Leadership*
- *Capacity Building*
- *Community Engagement*
- *Data Supply*
- *Data Governance*
- *Data Analytics For Decision Making*

Selanjutnya, *framework data-driven* Indonesia diturunkan pada *maturity level* yang dipublikasikan oleh Jane Wiseman dengan judul *Data-Smart City Solution. Level* maturitas tersebut terbagi kedalam 4 tingkatan dengan mengacu pada prinsip data-driven Indonesia yaitu penggunaan data sebagai asset yang bernilai tinggi, keterlibatan masyarakat untuk pelayanan yang lebih baik, pembangunan kapasitas untuk meningkatkan performa dan terciptanya kepemimpinan data untuk keputusan yang lebih baik. Adapun tingkatan level maturitas tersebut adalah sebagai berikut:

- *Publish*
Open data dipublikasikan dan mulai dikembangkan *tools* dan keterampilan dalam menggunakan data terhadap pengambilan keputusan.
- *Polish*
Meningkatkan kualitas, kuantitas dan nilai data dengan membangun sistem yang dapat menganalisa data, melibatkan public dan mengembangkan budaya penggunaan data.
- *Analyze*
Pimpinan menunjukan komitmen terhadap data, tata kelola data terorganisir dengan baik, kapasitas dalam hal analisis meningkat. Adanya dukungan pelatihan dan platform berbagi dalam menumbuhkan budaya penggunaan data.
- *Optimize*
Pimpinan dan manajemen disemua level memiliki komitmen dalam penggunaan data. Terdapat dukungan dari komunitas data. Adanya dokumentasi dari setiap pencapaian serta adanya dukungan pelatihan atau pengembangan *skill*.

Gambar 6. Data-Driven Maturity Level



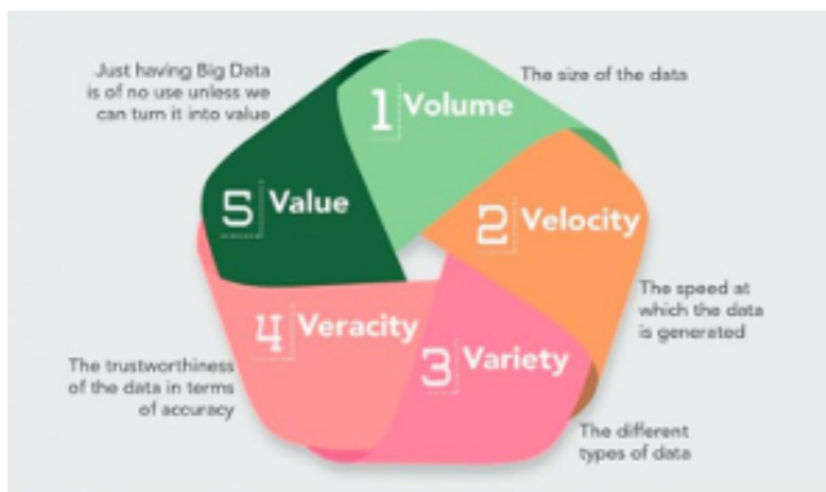
Source: Jane Wiseman, "Analytics Excellence Roadmap: A Four-Stage Maturity Model for Data-Driven Government," *Data-Smart City Solutions*, Ash Center at Harvard Kennedy School, July 2016.

Konsep Big Data

Big Data adalah istilah yang diberikan pada kumpulan data yang berukuran sangat besar dan kompleks, sehingga tidak memungkinkan untuk diproses menggunakan perangkat pengelola *database* konvensional ataupun aplikasi pemroses data lainnya. Dalam *Gartner IT Glossary*, *Big Data* didefinisikan sebagai berikut [1] : *Big Data is high-volume, high-velocity and/or high-variety information assets that demand cost-effective, innovative forms of information processing that enable enhanced insight, decision making, and process automation.*

Karakteristik utama *Big Data* meliputi 5 hal – biasa disingkat dengan 5V – yaitu *volume*, *velocity*, *variety*, *veracity* dan *value*. *Volume* terkait dengan besaran data yang harus dikelola berukuran super besar. *Velocity* berkenaan dengan kecepatan pemrosesan data yang harus mengimbangi pesatnya pertumbuhan jumlah data. *Variety* merujuk pada karakteristik sumber data yang sangat beragam, baik itu yang berasal dari basis data yang terstruktur maupun juga dari data-data yang tidak terstruktur. *Veracity* ini mengacu pada kualitas dan keakuratan data. Data yang dikumpulkan dapat memiliki bagian yang hilang, mungkin tidak akurat, atau mungkin tidak dapat memberikan wawasan yang nyata dan berharga. *Veracity*, secara keseluruhan, mengacu pada tingkat kepercayaan yang ada dalam data yang dikumpulkan. Sedangkan *value* mengacu pada nilai yang dapat diberikan oleh *Big Data*, dan ini terkait langsung dengan apa yang dapat dilakukan organisasi dengan data yang dikumpulkan tersebut. Mampu menarik nilai dari *Big Data* adalah suatu keharusan, karena nilai *Big Data* meningkat secara signifikan tergantung pada wawasan yang dapat diperoleh. Organisasi dapat menggunakan alat *Big Data* yang sama untuk mengumpulkan dan menganalisis data, tetapi bagaimana memperoleh nilai dari data itu harus unik.

Gambar 7. Karakteristik Big Data



Trend Big Data

Meningkatnya *volume* data informasi saat ini, yang menuntut pemrosesan yang inovatif dan efektif. Dengan menggunakan teknologi big data, sebuah organisasi data memperoleh wawasan dan dapat membuat keputusan yang lebih baik berdasarkan data. Berbagai sumber data yang tersedia saat ini memungkinkan sebuah organisasi untuk melacak tren dan memberikan perkiraan di berbagai sektor baik pemerintahan, kesehatan, bisnis dengan menganalisis data. Melihat hal tersebut big data dan analitik menjadi sangat penting saat ini.

Karena analitik *big data* semakin populer, sangat penting bagi sebuah organisasi untuk mengikuti prediksi analitik *big data* untuk menilai model dan tren pasar. Sebuah organisasi harus waspada terhadap tren masa depan dalam analitik *big data*. Ketika preferensi dan kebutuhan organisasi berubah seiring waktu, akan ada peningkatan penggunaan dan penerapan analitik *big data* oleh bisnis untuk menilai model, berikut merupakan beberapa tren *big data* diantaranya :

Gambar 8. 7 Big Data Tren in 2022



1. Data Akan dianalisis di *Cloud*
Penggunaan layanan *cloud public* dan pribadi untuk penyimpanan data dan analitik. Mengumpulkan, membersihkan, Menyusun dan menganalisis *volume* data yang besar adalah sumber kekhawatiran yang dapat diatasi melalui ilmu data di *cloud*.
2. Analisis Prediktif Akan Meningkat
Meskipun analitik prediktif bukan hal baru, akan ada lebih banyak perusahaan akan mengadopsinya, mengingat adanya pergeseran pola pembelian dan konsumsi.
3. Pengambilan keputusan berdasarkan Data
Investasi besar perangkat lunak untuk pemrosesan data yang mahal tidak akan membuahkan hasil apabila kita tidak menganalisis data untuk mendapatkan wawasan yang dapat ditindaklanjuti. Dengan data yang dapat ditindaklanjuti, dapat membuat keputusan yang lebih baik.
4. Analisis Klinis untuk Layanan Kesehatan
Pandemi telah menunjukkan kepada kita bahwa ada kebutuhan yang pasti untuk mendapatkan layanan kesehatan yang dapat didorong dengan melakukan analisa *big data*. Analitik klinis dan perawatan kesehatan akan membawa industri kesehatan yang baru dalam memperoleh layanan kesehatan.
5. *Artificial Intelligence* dan *Machine Learning*
Mengingat profelisasi data, kecerdasan buatan dan *machine learning* akan mendorong berbagai sektor pada tahun 2022. Mengintegrasikan kecerdasan buatan dan *machine learning* memungkinkan organisasi untuk meningkatkan proses pengambilan keputusan dan meningkatkan akurasi analisis data.

6. *IoT (Internet of Things)* pendorong Analisis *Streaming*

Sebagian besar perangkat yang kita gunakan saat ini diaktifkan dengan internet, mulai dari perangkat yang dapat dikenakan hingga peralatan rumah tangga. Seiring dengan meningkatnya jumlah perangkat ini, analitik streaming akan diperlukan untuk mengumpulkan wawasan yang bermakna dari data yang dihasilkan. Ini akan memungkinkan bisnis untuk tidak hanya memantau pergerakan dan penyimpanan data, tetapi juga menganalisis dan menguraikannya. Dengan analitik *streaming*, akan mendapatkan responsivitas dan *agility* yang lebih baik.

7. *Cyber Security*

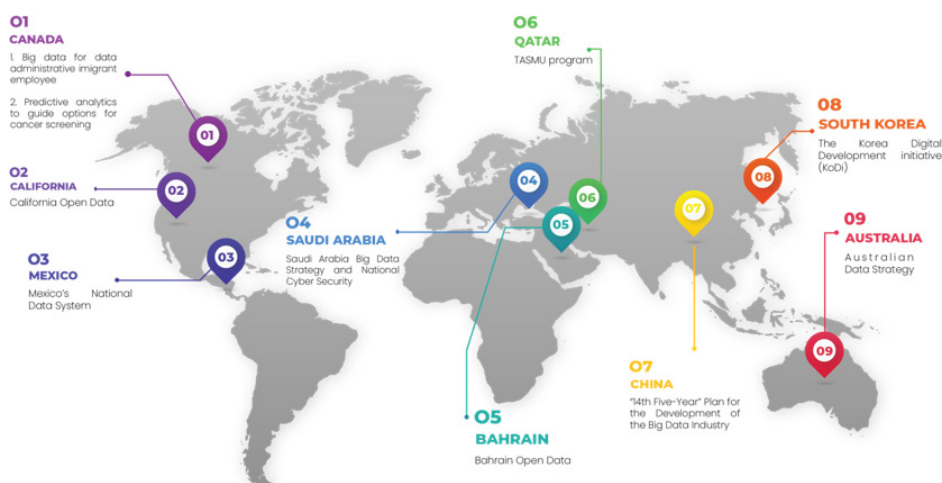
Serangan siber merupakan ancaman besar bagi pertumbuhan data. Pemodelan data tingkat lanjut akan mengurangi kemungkinan serangan siber untuk sebagian besar.

Analisa Benchmarking

Melihat tren *Big Data* dan pertumbuhan data saat ini, beberapa negara juga sudah mulai menggunakan teknologi *big data* untuk setiap proses bisnis yang dijalankan baik di sektor pemerintah maupun industri. Berikut merupakan beberapa negara yang sudah menerapkan *big data* dan inisiatif yang dilakukan.

Pemerintah Canada menerapkan teknologi analitik *Big Data* dalam melakukan proses administrasi pekerja imigran. Dengan analitik *Big Data* pemerintah Canada dapat mengetahui sebaran pekerja imigran yang ada di Canada. Selain itu, pemerintah Canada juga menggunakan teknologi *Big Data* untuk Analisa penyebaran penyakit kanker. Pemerintah Canada meyakini melalui Analitik *Big Data* mampu meningkatkan perekonomian Canada karena setiap pengambilan keputusan yang dilakukan berdasarkan pada data. Selain itu di Californasi dan Bahrain Pemerintah memanfaatkan *Open Data* sebagai penerapan *Big Data*. Masyarakat bisa menggunakan dan mengakses data melalui Portal *Open Data* yang sudah disediakan Pemerintah. Mexico sendiri mulai menerapkan *Big Data* dengan menyusun *Mexico's National Data System* sebagai acuan dalam penerepan *Big Data*. Hal yang sama juga dilakukan oleh Saudi Arabia melalui program Saudi Arabia *Big Data Strategy and National Cyber security*, Qatar dengan program Tasmu, China dengan program "14th – Five-year " *Plan fot the Development of the Big Data Industry*, Korea Selatan dengan program *The Korea Digital Development* dan Australia dengan Program *Australian Data Strategy*.

Gambar 9. 7 Implementasi Big Data di Berbagai Negara



Selain itu penerapan *Big Data* sangat berdampak bagi pertumbuhan ekonomi di sektor industri terbukti di beberapa perusahaan setelah menerapkan Analitik *Big Data* dalam strategi perusahaan pertumbuhan pendapatan dan pengguna dari setiap produk yang dimiliki meningkat pesat, seperti Netflix mengalami peningkatan pengguna sebanyak 213 juta pengguna dan mampu memberikan layanan yang jauh lebih baik dengan mengetahui tayangan apa yang sesuai dengan penggunaannya dari hasil Analisa interaksi dan respon pengguna terhadap setiap tayangan yang ada di Netflix. Berikut merupakan beberapa contoh penerapan big data di sektor industri komersial.

Gambar 10. 7 Implementasi Big Data di Sektor Industri



Teknologi Big Data

Perkembangan big data yang lahir karena semakin canggihnya teknologi menjadi begitu pesat. *Big Data* bahkan tidak mengenal kata berhenti, setiap hari akan menghasilkan data dalam jumlah yang sangat besar, dan akan terus bertambah setiap harinya seiring dengan bertambahnya pengguna teknologi. Berdasarkan data dari *International Data Corporation (IDC)*, memprediksi "*Global Datasphere*" data akan terus tumbuh dari 33 Zettabytes (ZB) pada tahun 2018 menjadi 175 ZB pada tahun 2025 artinya data yang dihasilkan setara dengan 1 triliun *gigabyte*.

Teknologi *big data* dirancang untuk menganalisis, memproses, dan mengekstrak informasi dari kumpulan data besar dengan struktur yang sangat kompleks yang tidak dapat ditangani oleh teknologi pemrosesan data secara manual. Teknologi *big data* dibagi menjadi 4 diantaranya :

1. Data Storage

Teknologi big data berhubungan dengan penyimpanan data yang memiliki kemampuan untuk mengambil, menyimpan dan mengelola data. Berikut merupakan beberapa teknologi big data yang banyak digunakan :

a. Apache Hadoop



Software Hadoop atau sebutan resminya adalah Apache Hadoop ini merupakan salah satu implementasi dari teknologi *Big Data*. Software yang bekerja lebih dari sekedar perangkat lunak ini, dapat diakses secara terbuka atau *open source*. Seperti yang sudah dijelaskan secara singkat di atas, Hadoop sendiri merupakan sekumpulan software yang mampu menyelesaikan permasalahan dari sekumpulan data dengan jumlah yang besar.

Dengan besarnya *volume* dan banyaknya variasi data yang diperoleh suatu perusahaan, menjadikan Hadoop bisa menjadi solusi dalam menyelesaikan masalah tersebut. Adapun beberapa perusahaan besar yang mengimplementasikan Hadoop dalam proses pengolahan Big Datanya, salah satunya adalah media sosial Facebook. Secara garis besar, Hadoop bekerja dan memiliki peran lebih dari sekedar perangkat lunak. Kumpulan beberapa software dalam Hadoop ini saling bekerja sama untuk meminimalisir waktu pengolahan dan memaksimalkan penyimpanan data yang ada. Terdapat inti atau core dari Hadoop, dimana *software* ini bisa didapat di website Apache Hadoop. Inti Hadoop ini ialah Hadoop Distributed File System, dimana HDFS ini sendiri merupakan tempat penyimpanan data-data dan file.

Selain itu terdapat *Data Mining*, dimana bagian ini dibuat untuk mempermudah dalam menjalankan *Map Reduce*. Contoh dari bagian ini adalah Apache Hive dan Apache Pig. Lalu, terdapat Database NoSQL (Not only SQL) yang bisa memberikan akses data dengan waktu yang cepat, contohnya adalah Apache Cassandra dan Apache HBase.

b. Mongo DB



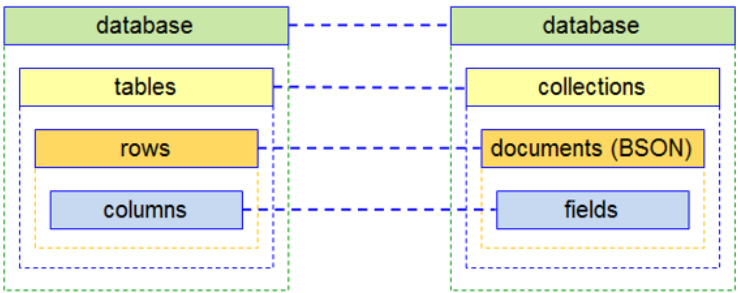
MongoDB adalah *database* NoSQL yang berbasis dokumen yang bersifat *cross-platform*, *open-source*, dan dapat digunakan secara gratis. MongoDB telah hadir sebagai salah satu database NoSQL yang sedang naik daun seiring dengan makin populernya penggunaan database NoSQL pada era *Big Data* sekarang ini. Adapun perusahaan-perusahaan besar yang telah menggunakan MongoDB diantaranya: Adobe, Facebook, eBay, video game FIFA, Foursquare, LinkedIn, McAfee, MetLife dan masih banyak lagi.

MongoDB adalah *database* multi-fungsi yang kuat, fleksibel, dan skalabel. MongoDB menggabungkan kemampuan bekerja dalam berbagai skala dengan fitur-fitur seperti *secondary indexes* (indeks tambahan selain indeks utama), *range queries* (penelitian dalam suatu rentang tertentu), *sorting* (pengurutan data), *aggregations* (penggabungan *dataset*), dan *geospatial indexes* (indeks geospasial).

Sebagai *database NoSQL* yang berbasis dokumen atau berorientasi pada dokumen (*document oriented*), MongoDB tidak menggunakan model relasional *database*. Alasan utama untuk meninggalkan *database* model relasional dan menggunakan database NoSQL layaknya MongoDB adalah demi kemudahan dalam meningkatkan skala *database* secara horisontal (*scale out*).

konsep database relasional

konsep database MongoDB



Model *document-oriented* ini secara alami sesuai dengan model data yang dipakai oleh para *developer* yang menggunakan bahasa program *object oriented modern*. Dalam *document-oriented database*, tidak ada keharusan untuk menentukan skema tertentu seperti halnya dalam *database* relasional. Data disimpan dalam bentuk dokumen yang terdiri atas *key* dan *value* dengan *type* dan ukuran yang bersifat variabel (tidak ditetapkan sebelumnya). Karena tidak ada skema yang harus ditetapkan, berdampak pada kemudahan dalam menghapus dan menambah *field* kapan saja diperlukan. Hal ini memungkinkan para *developer* untuk lebih mudah melakukan eksperimen-eksperimen dalam memilih model data yang akan digunakan hingga menemukan model yang terbaik.

2. *Data Mining*

Data mining adalah proses ekstraksi informasi dari data mentah. Biasanya, data yang di ekstrak dalam *volume* besar, dengan variabilitas tinggi dan *streaming* dengan kecepatan luar biasa. Di bawah ini adalah beberapa teknologi *big data* teratas yang digunakan untuk proses *data mining* :

a. Presto



Presto adalah mesin eksekusi *query paralel* yang terdistribusi yang dioptimalkan untuk *latency* rendah dan analisis *query* interaktif. Presto menjalankan *query* dengan mudah dan timbangan tanpa turun waktu bahkan dari *gigabyte* sampai *petabyte*.

Permintaan Presto tunggal dapat memproses data dari berbagai sumber seperti HDFS, MySQL, Cassandra, HIVE dan sumber data lebih banyak. Presto dibangun dengan JAVA dan mudah untuk mengintegrasikan dengan komponen infrastruktur data lainnya.

b. Rapidminer



Rapid Miner merupakan perangkat lunak yang dibuat oleh Dr. Markus Hofmann dari Institute of Teknologi Blanchardstown dan Ralf Klinkenberg dari rapid-i.com dengan tampilan GUI (*Graphical User Interface*) sehingga memudahkan pengguna dalam menggunakan perangkat lunak ini. Perangkat lunak ini bersifat *open source* dan dibuat dengan menggunakan program Java di bawah lisensi *GNU Public Licence* dan Rapid Miner dapat dijalankan di sistem operasi manapun. Dengan menggunakan Rapid Miner, tidak dibutuhkan kemampuan *coding* khusus, karena semua fasilitas sudah disediakan. Rapid Miner dikhususkan untuk penggunaan *data mining*. Model yang disediakan juga cukup banyak dan lengkap, seperti *Model Bayesian*, *Modelling*, *Tree Induction*, *Neural Network* dan lain-lain.

Banyak metode yang disediakan oleh Rapid Miner mulai dari klasifikasi, klustering, asosiasi dan lain-lain. Jika tidak ada model atau model algoritma yang tidak ada dalam Weka, pengguna boleh menambahkan modul lain, karena weka bersifat *open source*, jadi siapapun dapat ikut mengembangkan perangkat lunak ini.

3. *Data Analytic*

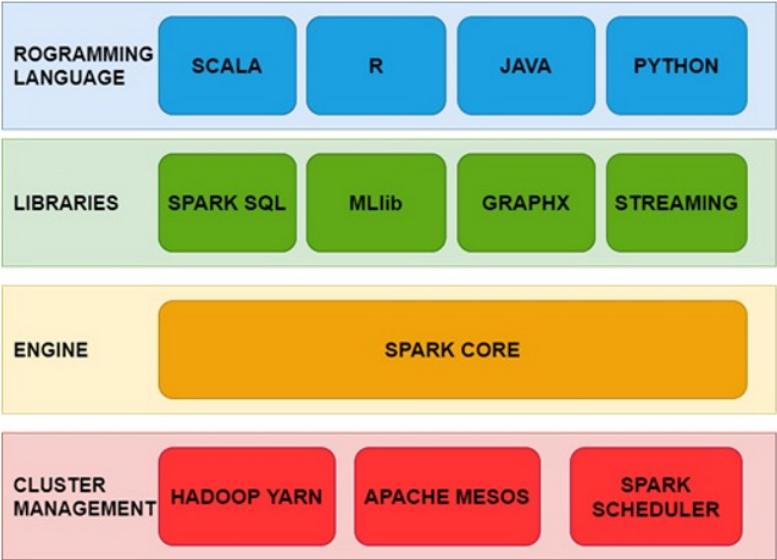
Dalam *analytic big data*, teknologi digunakan untuk membersihkan dan mengubah data menjadi informasi yang dapat digunakan untuk mendorong keputusan bisnis. Langkah selanjutnya setelah proses *data mining* adalah di mana penerapan algoritme, model, dan lainnya menggunakan tools untuk data analitik seperti :

a. Apache Spark



Apache Spark adalah *framework* yang digunakan untuk memproses, menanyakan, dan menganalisis *Big Data*. Apache Spark melakukan pemrosesan data melalui *in-memory*, sehingga waktu pemrosesan lebih cepat daripada *framework* sejenis seperti MapReduce dan lainnya. Perkembangan data dalam tingkat terabyte data diproduksi setiap hari, menjadikan kebutuhan akan solusi yang dapat memberikan *real time analysis* dengan kecepatan tinggi.

Apache Spark memiliki beberapa komponen dan dukungan dari berbagai bahasa pemrograman, ilustrasi mengenai komponen yang ada pada Apache Spark.



b. Kafka



Apache Kafka merupakan alat yang dibangun pada ekosistem Hadoop untuk melakukan penanganan log transaksi serta data dalam waktu nyata atau *real-time*. Platform-nya memiliki aliran dengan sumber terbuka yang dilakukan pengembangan melalui Apache Software Foundation. Dilakukannya penyimpanan dengan menggunakan Kafka yang memproses data secara besar-besaran dan dilakukan perancangan pada log transaksi terdistribusi, hal ini membuat penting sebagai proses pengaliran data. Kafka berperan untuk melakukan penghubungan sistem eksternal melalui Kafka Connect dan memberi Kafka Streams yang merupakan *Java stream library* untuk melakukan pemrosesan data pada Kafka.

Melalui adanya Apache Kafka dapat memudahkan untuk menghadapi adanya tantangan dalam menghadapi pemrosesan *Big Data*. Hal ini disebabkan adanya pemrosesan aliran dan sistem pesan yang toleran terhadap kesalahan karena pemrosesan dilakukan secara cepat dan terukur menyebabkannya kerap kali digunakan dalam pengolahan data dan analisis. Perancang Kafka bertujuan untuk latensi yang rendah dengan throughput yang tinggi. Di dalamnya Kafka membangun adanya jaringan pipa yang bergerak secara real time berhubungan dengan aliran data yang dikelola sehingga dapat diandalkan untuk mendapatkan data antara sistem atau aplikasi. Kemudian, Kafka melakukan penanganan pada ratusan megabyte data dari berbagai kecilnya, hal ini membuat Kafka dapat digunakan sebagai alat untuk menghadapi adanya tantangan pemrosesan *Big Data*.

4. Data Visualization

Teknologi big data dapat digunakan untuk membuat visualisasi dari data. Dalam peran berorientasi data, visualisasi data adalah keterampilan yang bermanfaat untuk menyajikan rekomendasi kepada pemangku kepentingan untuk profitabilitas dan bisnis untuk menceritakan data melalui visualisasi grafik sederhana. Berikut merupakan beberapa *tool* untuk visualisasi data :

a. Tableau



Tableau adalah sebuah *tools* yang dapat menganalisa/menggambarkan suatu kumpulan data untuk disajikan dalam bentuk yang menarik. Tableau ini merupakan *business intelligence* hasil penelitian Gartner Report Februari 2016. Tableau menduduki chart / kuadran teratas dibidang Platform BI (*Business Intelligence*) bersama 2 *platform* sejenis. Memiliki tiga produk utama yaitu tableau desktop, tableau server, dan tableau online. Aplikasi ini memiliki beberapa spesifikasi sebagai berikut :

- Kemampuan koneksi dan pengambilan data dari berbagai sumber data, Kemampuan analisis data yang dinamis dan multidimensi.
- Kemampuan analisis data di peta terintegrasi (tidak perlu *software* tambahan), Kemampuan pembuatan dashboard informasi).
- Menyediakan sarana *collaborative Business Intelligence*.
- Akses informasi di mana saja, kapan saja dan menggunakan perangkat *mobile*.
- Kemampuan menambahkan kalkulasi dari data yang ada dan parameter.
- Memiliki fasilitas data internal yang memiliki unjuk kerja tinggi.
- *Business Intelligence platform* berskala *enterprise* (*High Availability, Multi tenancy*, kinerja dengan *scale up* dan *scale across, data proxy*).
- Mencetak, dan ekspor hasil analisis dan data. Berbagi pakai file template. Implementasi yang cepat

b. Plotly



Plotly adalah *software* berbasis *website* dan bersifat *open source* yang digunakan untuk visualisasi data. Dengan Plotly kita dapat membuat visualisasi data interaktif yang rumit dan langsung pada dashboard atau situs. Plotly menjadi tools visualisasi data yang canggih karena dibangun melalui library visualisasi D3.js. Keunggulan Plotly adalah dapat membuat visualisasi multi grafis untuk membandingkan dataset. Plotly dapat digunakan dalam mode online maupun offline dan di Jupyter Notebook.

Journal

Berikut merupakan hasil Analisa dari beberapa journal terkait implementasi *Big Data* di berbagai sektor.

Tabel 1. Analisa Jurnal Implementasi Big Data

No	Judul Paper	Permasalahan	Strategi/Solusi/ Dampak Big Data
1	<i>Big Data Dreams and Reality in Shenzhen: An Investigation of Smart City Implementation in China</i>	Kota-kota Cina semakin banyak menggunakan teknologi digital untuk mengatasi masalah perkotaan dan mengatur masyarakat. Namun, masih terbatasnya pengetahuan tentang bagaimana transisi digital ini diterapkan. Salah satu nya mengenai implementasi Big Data dalam tata kelola pemerintahan untuk mendukung Smart City di China	The 13th Five-Year Plan (2016–2020) membuka babnya tentang Strategi Big Data Nasional dengan pernyataan bahwa pemerintah 'akan menjadikan Big Data sebagai sumber daya strategis yang mendasar
2	<i>A big Data Analytics Framework for the Integration of Heterogeneous Federated Data Centers</i>	Dengan pertumbuhan data yang cepat, satu lingkungan cluster menjadi tidak memadai untuk mengelola data ini. Pada saat yang sama, ada sumber data yang heterogen di berbagai platform, yang perlu saling terhubung untuk mendapatkan makna analisis. Paradigma perangkat lunak MapReduce telah muncul untuk mengisi kesenjangan, dan telah berhasil beroperasi pada sistem. Tantangannya adalah bagaimana membuat kerangka analitik Big Data yang mendukung integrasi sumber data heterogen pada model komputasi terdistribusi di beberapa pusat data.	Meningkatkan paradigma MapReduce, yang mendukung sumber data heterogeny di beberapa pusat data. Menyusun Arsitektur berdasarkan pada komputasi terdistribusi master/slave model dan Map - Reduce - Merge - GlobalReduce yang disajikan sebagai model pemrograman untuk distribusi integrasi data dalam implementasi Big Data.
3	<i>An Efficient and Scalable Framework for Processing Remotely Sensed Big Data in Cloud Computing Environments</i>	Besarnya data yang dihasilkan oleh satelit yang menjadi tantangan dalam proses <i>remote</i> data yang tersimpan di <i>cloud</i> secara efektif dan efisien	<i>Cloud Computing Framework</i> yang mampu mengintegrasikan yang mengintegrasikan mekanisme pemrosesan terdistribusi dan tugas penjadwalan strategi menjadi prosedur optimasi untuk mengaktifkan pemrosesandata skala besar yang efisien dan terukur melalui <i>cloud</i> .

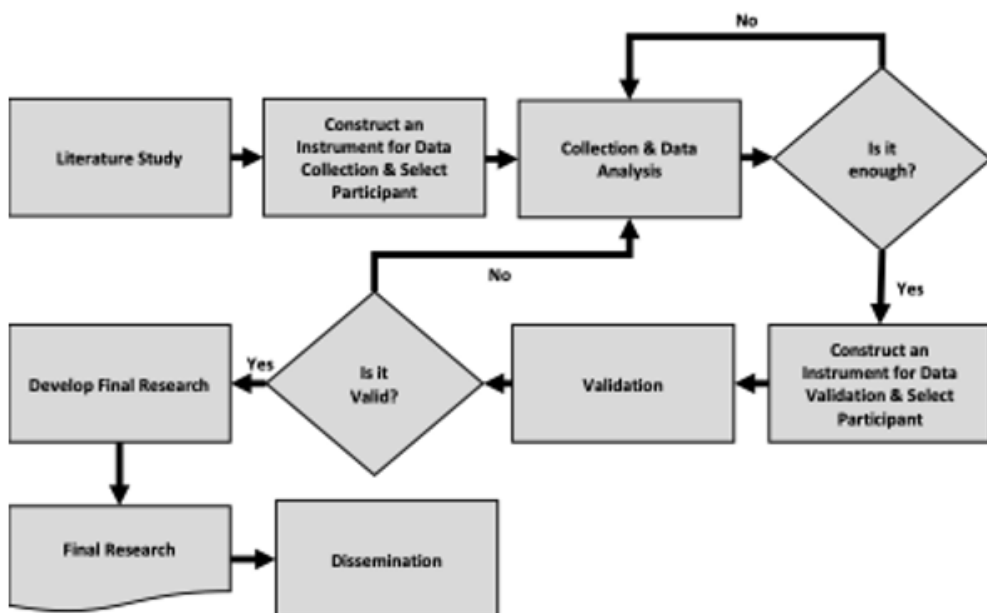
4	<i>An Efficient Data Scheduling Scheme for Cloudbased Big Data Framework for Smart City</i>	Kebutuhan untuk merencanakan dan mengembangkan solusi berbasis Smart City untuk tata kelola kota yang lebih baik menjadi semakin nyata. Sehingga diperlukan <i>Framework</i> teknologi untuk layanan Smart City yang menghubungkan ratusan jaringan perangkat untuk <i>collecting data</i> (misalnya, jaringan sensor) dengan server pusat atau <i>cloud</i> melalui Internet. Penjadwalan data yang sangat besar ini (atau <i>Big Data</i>) baik di jaringan perangkat maupun server pusat/ <i>cloud</i> sangat penting untuk memfasilitasi layanan Smart City berbasis prioritas dan tepat waktu.	<i>Big data framework and Priority-based, Dynamic and Time sensitive data processing and Scheduling (PDTS)</i> berbasis <i>Cloud</i> untuk mengurangi jumlah transmisi data dan waktu pemrosesan data dibandingkan dengan penjadwalan data.
5	<i>An Integration Framework on Cloud for Cyber-Physical-Social Systems Big Data</i>	Pengembangan <i>Cyber-Physical-Social Systems</i> (CPSS) adalah untuk mengintegrasikan <i>volume</i> yang berkembang dan beragam data yang dihasilkan dari berbagai sumber.	<i>Framework</i> yang tertintegrasikan data besar CPSS di <i>cloud</i> yang terdiri dari lima proses yang saling melengkapi secara fungsional, yaitu representasi data, pengurangan dimensi, pembentukan relasi, pemeringkatan data, dan pengambilan data.
6	<i>Big Data and Machine Learning Framework in Healthcare</i>	Tren teknologi di bidang kesehatan dan kompleksitas sumber data dan <i>volume</i> data yang dapat digunakan untuk prediksi medis tingkat lanjut.	Membuat <i>Framework</i> dengan memanfaatkan <i>Big Data</i> untuk perawatan kesehatan dengan mengumpulkan dan memproses data dengan memanfaatkan <i>machine learning</i> , yang memungkinkan untuk memberikan prediksi status kesehatan pasien untuk lebih andal dan tepat dalam perawatan kesehatan.
7	<i>ExNav: An Interactive Big Data Exploration Framework for Big Unstructured Data</i>	Kesenjangan antara pertumbuhan data eksponensial dan kemampuan manusia yang terbatas dalam pengelolaan data tidak terstruktur dalam jumlah besar	<i>Framework</i> untuk eksplorasi data baru, yaitu ExNav (<i>Exploration Navigator</i>), yang memungkinkan pengguna mengelola data tidak terstruktur dengan mudah untuk mendapatkan <i>insight</i> dari data yang diolah

03 METODE PELAKSANAAN

Metode Pelaksanaan

Metodologi penyusunan kajian ini menggunakan metodologi penyusunan laporan koordinasi perencanaan di lingkungan WANTIKNAS seperti pada 2.1 berikut.

Gambar 11. 7 Metode Pelaksanaan



Literature Study

Tahap ini merupakan tahap awal koordinasi perencanaan. Pada tahap ini dilakukan studi terhadap regulasi, *best practices* dan penelitian yang terkait penerapan *Big Data*. Dokumen-dokumen yang dianalisis adalah Peraturan Presiden, Aturan Menteri yang mengikat, whitepaper, jurnal dan lain-lain.

Collection & Data Analysis

Tahap ini adalah proses dilaksanakan pengambilan data serta analisis data. Pengambilan data dilakukan dengan menggunakan metode *Focus Group Discussion* (FGD) dan Wawancara. Metode ini digunakan hingga data dirasakan cukup untuk dilanjutkan pada proses berikutnya. Setelah didapat data dari studi literatur, hasil FGD, data kemudian dianalisis dengan menggunakan teknik analisis *Grounded Theory*.

Construct an Instrument for Data Validation & Select Participant

Apabila data primer yang dihasilkan di dalam proses pengambilan data dirasakan cukup, langkah selanjutnya adalah melakukan validasi data tersebut. Namun, sebelumnya diperlukan instrumen untuk melakukan validasi data serta memilih ahli untuk proses validasi tersebut.

Validation

Tahap ini adalah tahap untuk melakukan validasi data primer yang ada, sehingga data tersebut bisa dijamin keabsahannya. Teknik untuk memvalidasi yang digunakan adalah teknik Delphi, yaitu dengan cara mengumpulkan para ahli dan mencari konsensus untuk mencari kesepakatan.

Develop Final Research

Setelah mendapatkan data yang tervalidasi pada proses sebelumnya, langkah selanjutnya adalah memfinalisasi kajian dengan menuangkan ke dalam dokumen yang sesuai dengan sistematika penulisan yang telah ditentukan.

Dissemination

Tahap ini merupakan tahap akhir, tahap di mana untuk menyebarluaskan hasil kajian agar diketahui oleh masyarakat luas, sehingga manfaat yang diharapkan bisa dirasakan oleh umum.

Pengumpulan dan Analisa Data

Metode analisis yang digunakan dalam mengolah data pada kajian ini adalah analisis tematik. Proses olah data terdiri dari beberapa tahap, yaitu :

- Pengumpulan data

Proses pengumpulan data dalam kajian ini adalah melakukan studi dokumen terhadap regulasi dan dokumen pendukung terkait penerapan *Big Data Analytics*, melakukan observasi terhadap Kementerian/Lembaga/Kabupaten/Kota yang sudah menginisiasi *Big Data Analytics*, melakukan wawancara terhadap narasumber yang kompeten dalam bidangnya serta mengadakan *Focus Group Discussion* dengan topik penerapan *Big Data*.

- Analisis tematik

Data yang terkumpul dari studi dokumen, observasi, wawancara dan FGD dicatat secara sistematis. Selanjutnya data tersebut dibuat *outline* dan dikodifikasi kemudian dibangun tema, pola dan hubungan antar *outline* data agar dapat dievaluasi dan ditarik kesimpulan.

- Validasi

Hasil evaluasi dan penarikan kesimpulan divalidasi kepada narasumber agar mendapatkan kesimpulan yang valid.

Jadwal Pelaksanaan

Pelaksanaan kegiatan dilakukan selama 3 bulan yaitu mulai bulan September 2022 sampai bulan November 2022. Adapun timeline kegiatan kajian ini dapat dilihat pada tabel berikut.

Tabel 2. Jadwal Pelaksanaan

RENCANA KERJA															
National Big Data Architecture: End to End Approach															
No.	Aktifitas	Deliverables	Bulan I				Bulan II				Bulan III				Keterangan
			1	2	3	4	1	2	3	4	1	2	3	4	
1	Start														
2	Weekly Meeting	Progress Pekerjaan													
3	Analisa Inisiatif Penerapan Big Data Pemerintahan	Hasil analisa kondisi saat ini													
	- Analisa Teknologi Big Data saat ini														Terhadap K/L Pilihan
	- Analisa Arsitektur Big Data saat ini														Terhadap K/L Pilihan
	- Analisa Struktur Organisasi Big Data saat ini														Terhadap K/L Pilihan
	- Analisa Kebijakan Big Data saat ini														Terhadap K/L Pilihan
4	Analisis Keselarasan	Hasil analisa kondisi masa mendatang													
	- Analisa Kebijakan terkait SPBE														
	- Analisa Kebijakan terkait Satu Data Indonesia														
	- Analisa Kebijakan terkait Strategi Nasional Artificial Intelligence														
5	Trend Analysis	Rekomendasi trend Big Data													
	- Analisa Best Practice Big Data														
6	Penyusunan Arsitektur Big Data Nasional	Arsitektur Big Data Nasional: End to End Approach													Selaras dengan SPBE dan SDI
7	Penyusunan Strategi Pencapaian	Quick win													
	- Strategi														
	- Roadmap														
	- Catalog Activity														
8	Proses Desain	Buku National Big Data Architecture													
9	Finish														

04

HASIL PELAKSANAAN KEGIATAN KAJIAN IMPLEMENTASI BIG DATA : END TO END APPROACH

Analisa SWOT

Menurut Thompson, Gamble dan Strickland (2006) dalam bukunya berjudul *“Strategy: Winning in the Marketplace. Core Concepts, Analytical Tools, Cases”*, analisis SWOT adalah alat yang sederhana tapi *powerfull* untuk menilai kemampuan dan kelemahan sumberdaya organisasi, peluang yang ada serta ancaman dari luar untuk mencapai kesejahteraan nya (Thompson et al., 2006). Menurut O'brien dan Marakas (2014) dalam bukunya yang berjudul *“Management Information System”*, analisis SWOT digunakan untuk mengevaluasi dampak dari masing-masing kesempatan strategis yang mungkin dapat dimiliki oleh sebuah perusahaan dan penggunaan teknologi informasinya (O'brien & Marakas, 2006). Selanjutnya, menurut Peppard dan Ward (2016), analisis SWOT atau *Strength, Weakness, Opportunities, Threat*, digunakan untuk memetakan kekuatan dan kelemahan proses bisnis di dalam internal organisasi serta peluang dan ancaman yang berasal dari eksternal organisasi yang kemungkinan akan dihadapi di masa yang akan datang (Peppard & Ward, 2016) .

Strength yang telah dipetakan, kemudian dianalisis pemanfaatannya untuk meraih *opportunities* yang ada dan mengatasi *threats* yang akan muncul di kemudian hari. Sedangkan *weakness*, dianalisis agar bisa diatasi dengan memanfaatkan *opportunities* dan agar *threats* yang kemungkinan dihadapi tidak menjadi kenyataan.

Berikut pemetaan SWOT Teknologi Informasi dalam mendukung penerapan *Big Data* Nasional

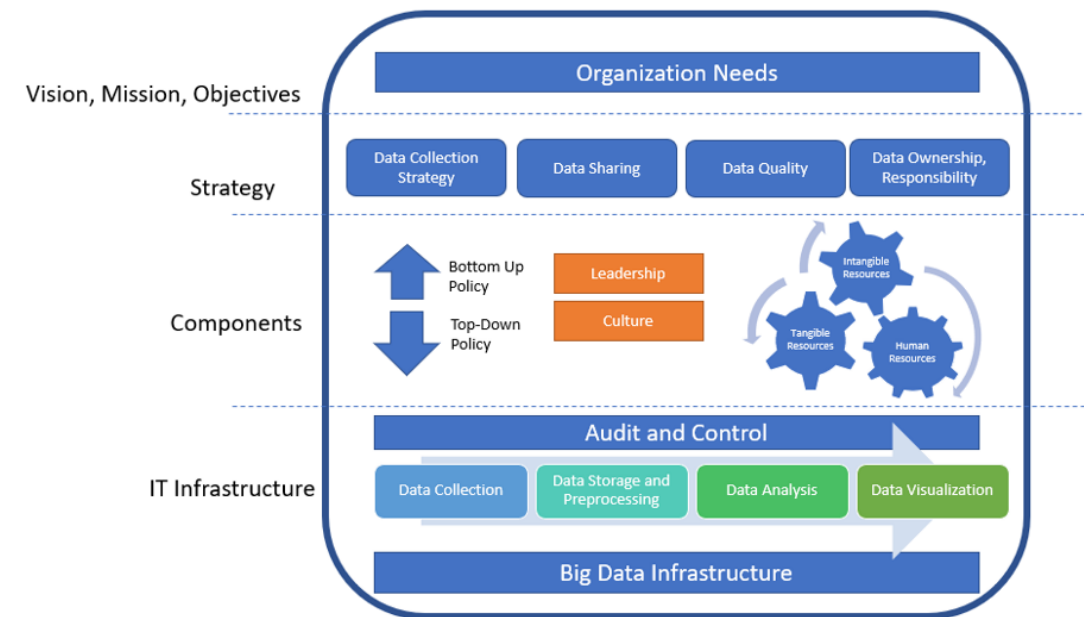
Gambar 12. Analisa SWOT



Tata Kelola

Berikut ini adalah rumusan tata kelola big data.

Gambar 13. Tata Kelola Big Data. Modifikasi dari Kim and Cho (2018)



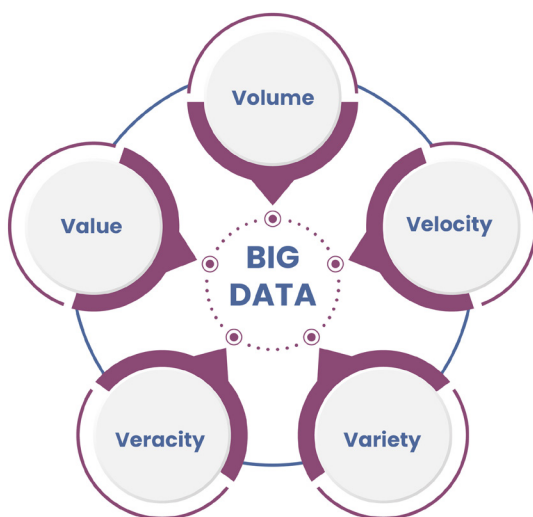
Ada empat lapisan pada kerangka kerja tata kelola *big data* yaitu lapisan visi misi, strategi, komponen dan infrastruktur Teknologi Informasi.

Mariani dan Bagio (2022) menyatakan bahwa tiga komponen utama untuk penerapan *Big Data* adalah sebagai berikut: 1) sumber daya tidak berwujud (mencakup budaya berbasis data dan intensitas pembelajaran organisasi); 2) sumber daya berwujud (termasuk data, teknologi, dan sumber daya dasar [seperti waktu dan investasi]); dan 3) sumber daya manusia (mencakup keterampilan manajerial [seperti ketajaman analitik] dan keterampilan teknis [seperti pendidikan dan pelatihan yang berkaitan dengan keterampilan khusus data])

Sumber Daya Manusia *Big Data*

Data besar bukanlah teknologi tunggal, tetapi kombinasi teknologi lama dan baru yang membantu perusahaan mendapatkan wawasan yang dapat ditindaklanjuti sambil mengelola masalah pemuatan dan penyimpanan data secara efektif. Analisis data besar membutuhkan kemampuan untuk mengumpulkan, mengelola, dan menganalisis *volume* yang berpotensi besar data yang berbeda, pada kecepatan yang tepat, dalam kerangka waktu yang tepat, sambil memberikan waktu yang tepat analisis dan aktivitas ke konsumen akhir.

Gambar 14. Karakteristik 5V pada Big Data

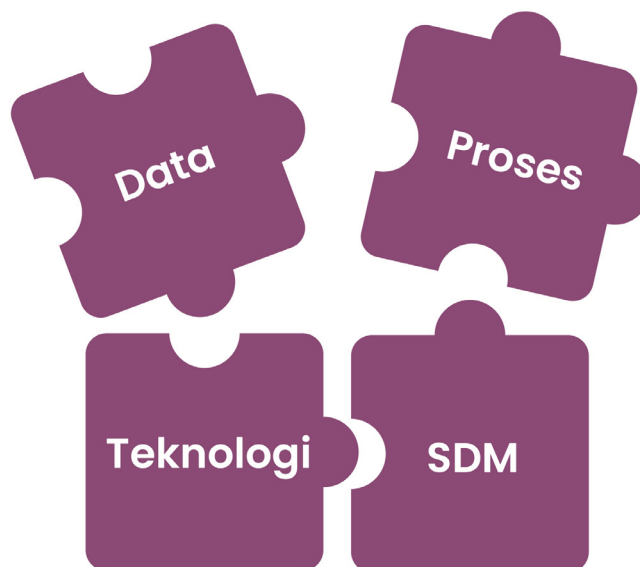


Karakteristik Big Data yang umum digunakan saat ini adalah 5V seperti yang terlihat pada gambar di atas. V yang pertama mengacu pada Volume yang mengacu pada ukuran *Big Data*. Data bisa dikatakan *Big Data* atau tidak berdasarkan volumenya. Namun, besarnya data juga menjadi perdebatan, karena ukuran data yang besar saat ini dapat menjadi kecil di tahun mendatang. Selanjutnya V yang kedua adalah *Velocity* yang mengacu pada kecepatan untuk pergerakan data. Dalam konsep Big Data, data mengalir dari beberapa sumber baik dari mesin, jaringan, media sosial, dan ponsel. Selanjutnya V yang ketiga *Veracity*, yaitu mengacu pada konsistensi dan ketidakpastian dalam data, yang mana data yang tersedia masih berantakan dan kualitas data masih diragukan. Selain itu, konsep *Big Data* juga menyatakan bahwa data

bervariasi karena banyak dimensi data yang dihasilkan dari berbagai tipe dan sumber data yang berbeda. Selanjutnya V keempat adalah *Variety*, yaitu merujuk pada terstruktur, data semi-terstruktur, dan tidak terstruktur, yang mana juga mangacu pada sumber heterogen. Lalu V yang kelima *Value*, yang artinya data yang tidak memiliki nilai tidak berguna bagi perusahaan, kecuali jika dapat mengubah menjadi sesuatu yang berguna. Dengan kata lain, data itu sendiri tidak ada gunanya sampai dengan data tersebut diubah menjadi sesuatu yang berharga dengan melakukan ekstraksi informasi.

Hal yang penting untuk diketahui adalah konsep ‘praktik terbaik *Big Data*’ berkembang karena bidang analitik data itu sendiri berkembang pesat. Menjawab beberapa tantangan big data, yaitu terkait data, proses (salah satunya dukungan dari pemangku kebijakan), teknologi dan kompetensi Sumber Daya Manusia (SDM), berikut adalah analisis untuk praktik di lapangan.

Gambar 15. Tantangan Praktik Big Data



Keempat terkait SDM yang berperan aktif untuk analisis *Big Data*. Berikut ini adalah penjelasan beberapa istilah yang terkait SDM.

- *Data Engineer* terlibat dalam mempersiapkan data. Mereka mengembangkan, membangun, menguji & memelihara arsitektur lengkap.
- *Data Scientist* bertugas menganalisis dan menafsirkan data yang kompleks serta terkadang berhubungan dengan pekerjaan membuat model *machine learning*.
- *Data Analyst* bertugas menganalisis data numerik dan menggunakannya untuk membantu perusahaan membuat keputusan yang lebih baik.
- *Domain Expert* bertugas memberi arahan tentang domain permasalahan.
- *IT People* bertugas menyiapkan infrastruktur IT, terutama terkait *deployment*.
- *Project/ Product Manager* bertugas mengelola proyek atau produk berbasis data serta menyeleraskan tim agar dapat mencapai tujuan bersama.

Berikut ini adalah kondisi ideal terkait SDM dan *skill set* yang seharusnya dimiliki oleh pengelola aplikasi *big data*.

Gambar 16. Sumber Daya Manusia Implementasi Big Data



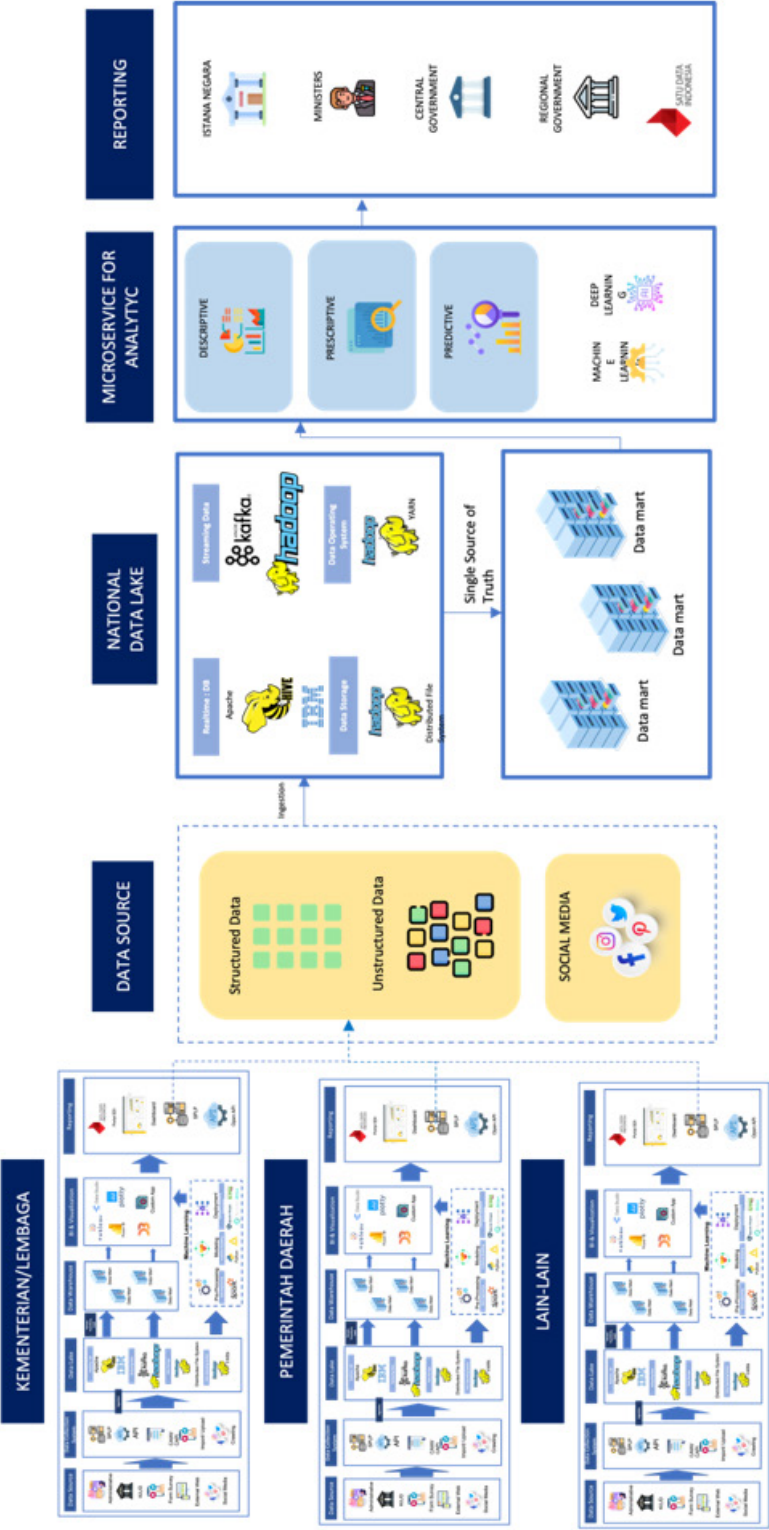
Sebuah tim untuk implementasi *big data* secara ideal memiliki enam peran seperti yang disebutkan di atas. Masing-masing peran akan berfungsi untuk tahapan-tahapan implementasi *big data*. Seperti yang digambarkan di atas bahwa pada saat perumusan masalah diperlukan seorang atau lebih dari satu orang yang mendefinisikan permasalahan atau kebutuhan organisasi. Selanjutnya, *data engineer* yang berperan untuk mengumpulkan data dari berbagai macam sumber dan membersihkan data sehingga dapat dihasilkan *dataset* yang siap untuk digunakan oleh *data scientist*. *Data scientist* akan berperan dalam membangun model dan melakukan evaluasi, selanjutnya *data analyst* akan melakukan analisis dan melakukan visualisasi dan juga membangun ke dalam sebuah sistem yang dapat digunakan oleh pihak pemangku kepentingan. *Data scientist* akan dibantu juga oleh orang *Information Technology* (IT) dalam proses *deployment*. *Project manager* yang akan mengawasi, mengkoordinasikan di tiap tahapan.

Penting untuk menjadi catatan, bahwa pada praktiknya setiap peran yang digambarkan dapat dirangkap oleh satu orang, misal nya saja *data scientist* dan *data analyst* dibebankan kepada satu orang. Begitu pula sebaliknya, satu peran dapat diisi oleh beberapa orang terutama untuk organisasi yang memiliki data yang cukup banyak.

Arsitektur Big Data Nasional

Berdasarkan hasil kajian yang telah dilaksanakan, Dewan TIK Nasional telah menyusun Arsitektur Big Data Nasional dengan pendekatan end to end untuk penerapan *Big Data Analytics* dan langkah-langkah strategis sebagai panduan dalam penerapan satu data Indonesia. Berikut merupakan rancangan Arsitektur Big Data Nasional.

Gambar 17. Arsitektur Big Data Nasional



Teknologi Big Data

Teknologi big data dirancang untuk menganalisis, memproses, dan mengekstrak informasi dari kumpulan data besar dengan struktur yang sangat kompleks yang tidak dapat ditangani oleh teknologi pemrosesan data secara manual.

Sumber Data

Data source adalah sumber data untuk *big data*. Data umumnya dipompa masuk *Big Data* dengan menggunakan API ataupun dengan operasional *file system* seperti transfer *file*. Ada dua jenis *data source* yaitu *streaming data source* dan *bulk data source*.

Data Collection

Data collection adalah proses pengumpulan, pengukuran, dan analisis berbagai tipe informasi menggunakan teknik berstandar. Tujuan utama *data collection* adalah untuk mengumpulkan informasi dan data terpercaya sebanyak-banyaknya, yang kemudian dianalisis untuk membuat sebuah keputusan bisnis yang krusial. Ketika sudah berhasil dikumpulkan, data ini kemudian melalui sejumlah proses meliputi pembersihan dan pemrosesan data agar dapat digunakan oleh perusahaan. Berikut merupakan beberapa teknik pengumpulan data :

- *Application Programming Interface (API)*
Application Programming Interface (API) adalah Sebuah bahasa dan pesan format yang digunakan oleh program aplikasi untuk berkomunikasi dengan sistem operasi atau program pengendalian lainnya seperti sistem manajemen *database* (DBMS) atau protokol komunikasi . API diimplementasikan dengan menulis panggilan fungsi dalam program, yang menyediakan hubungan ke subrutin yang diperlukan untuk eksekusi. Dengan demikian, API menyiratkan bahwa *driver* atau program modul tersedia dalam komputer untuk melakukan operasi atau perangkat lunak yang harus dihubungkan ke dalam program yang ada untuk melakukan tugas-tugas. API dapat menjelaskan cara sebuah tugas (*task*) tertentu dilakukan. Dalam pemrograman *procedural* seperti bahasa C, aksi biasanya dilakukan dengan media pemanggilan fungsi. Karena itu, API biasanya menyertakan penjelasan dari fungsi/rutin yang disediakannya.
API (*Application Programming Interface*) merupakan *interface* yang digunakan untuk mengakses aplikasi atau layanan dari sebuah program. API memungkinkan pengembangan untuk memakai fungsi yang sudah ada dari aplikasi lain sehingga tidak perlu membuat ulang dari awal
- *CAWI/CAPi*
- *Import/Upload*
- *Crawling*
Pada umumnya *crawling* diterapkan pada *web* yang banyak disebut *Web Crawling*. *Web Crawling* pada umumnya digunakan pada *search engine* yang dilakukan oleh sekelompok komputer yang dikluster dimana setiap komputer menjalankan beberapa *thread* (Hatzi, 2014).

Prosedur dari *Web Crawling* dimulai dari memilih satu set URL yang akan dilakukan proses *crawling* dimana URL tersebut menyediakan banyak *link* ke halaman yang penting untuk proses *Crawling*. *The crawler* akan men-download konten dari halaman tersebut ke dalam penyimpanan *local* seperti *hardisk* dll. Secara bersamaan *thread* yang ada pada *Crawler* mencari link baru dari halaman yang terhubung dengan halaman yang

telah diunduh, *link* baru membentuk yang disebut dengan *crawler's frontier*. Tujuan utama dari *crawler's frontier* adalah mendapatkan halaman baru sebanyak mungkin dan mengupdate tingkat kebaruan halaman yang telah diunduh.

- **Scraping**

Scraping atau yang dikenal sebagai ekstraksi *web* adalah teknik untuk mengekstraksi data dari *World Wide Web* (WWW) dan menyimpannya ke file sistem atau basis data untuk dijadikan analisis data. *Web Scraping* dapat dilakukan baik secara manual oleh seorang pengguna atau secara otomatis oleh bot atau *crawler web*. Proses *Web Scraping* dari Internet dapat dibagi menjadi dua langkah berurutan, yaitu mengakuisisi sumber daya *web* dan kemudian mengekstraksi informasi yang diinginkan dari data yang diperoleh.

Secara khusus, program *Web Scraping* dimulai dengan meminta HTTP untuk memperoleh sumber daya dari yang ditargetkan oleh situs *web*. Permintaan ini dapat diformat kedalam URL yang berisi permintaan GET atau HTTP yang berisi POST. Setelah permintaan berhasil diterima dan diproses oleh situs *web* yang ditargetkan, sumber daya yang diminta akan diambil dari situs web dan kemudian dikirim kembali ke program *Web Scraping*. Sumber daya ini bisa dalam berbagai format, seperti halaman *web* yang dibangun dengan HTML, XML atau JSON, atau data multimedia seperti gambar, *audio*, atau video. Terdapat dua modul penting dari *Web Scraping* - modul untuk menulis permintaan HTTP, seperti *Urllib2* atau *selenium* dan satu lagi untuk parsing dan mengekstraksi informasi dari kode HTML mentah, seperti *BeautifulSoup* atau *Pyquery*.

Data Lake

Data Lake adalah gudang penyimpanan yang dapat menyimpan data terstruktur, semi-terstruktur, dan tidak terstruktur dalam jumlah besar. Ini adalah tempat untuk menyimpan semua jenis data dalam format aslinya tanpa batas tetap pada ukuran atau file akun. Ini menawarkan kuantitas data yang tinggi untuk meningkatkan kinerja analitik dan integrasi asli. *Data Lake* merupakan cara yang hemat biaya karena menyimpan semua data dari suatu organisasi yang akan diproses nantinya. Seorang Riset Analis dapat berfokus pada penemuan pola-pola makna dalam data dan bukan data itu sendiri. Lain halnya dengan *Data warehouse hierarkis* di mana data disimpan dalam File dan *Folder*, *Data Lake* memiliki arsitektur datar. Setiap elemen data di *Data Lake* diberi pengidentifikasi unik dan diberi *tag* dengan satu set informasi *metadata*.

Tujuan utama dari membangun *Data Lake* adalah untuk menawarkan pandangan data yang belum diolah ke para ilmuwan atau peneliti data. Berikut ini alasan menggunakan *Data Lake* adalah:

- Dengan munculnya mesin penyimpanan seperti Hadoop, menyimpan informasi yang berbeda menjadi lebih Tidak perlu memodelkan data ke skema perusahaan secara meluas apabila dengan menggunakan *Data Lake*.
- Dengan peningkatan *volume* data, kualitas data, dan *metadata*, kualitas analisis juga meningkat.
- *Data Lake* juga menawarkan kecepatan proses dalam bisnis
- Pembelajaran mesin dan Kecerdasan Buatan dapat digunakan untuk membuat prediksi yang menguntungkan dan menawarkan keunggulan kompetitif bagi organisasi pelaksana.

Berikut ini adalah konsep-konsep kunci dari *Data Lake* yang perlu dipahami sepenuhnya untuk memahami *Data Lake Architecture* :

Gambar 18. *Data Lake*



Data Warehouse

Data Warehouse merupakan sebuah gudang data yang berisi data dalam jumlah besar dan digunakan untuk proses analisa dan pembuatan laporan yang dibutuhkan perusahaan.

Hal ini diperkuat oleh teori dari Han Jiawei (2011:p143) yang mengatakan bahwa *data warehouse* adalah tempat penyimpanan informasi yang dikumpulkan dari berbagai sumber, disimpan dalam skema yang terpadu dan biasanya berada dalam suatu lokasi / situs.

Menurut W.H. Inmon dan Richard D.H., *data warehouse* adalah koleksi data yang mempunyai karakteristik berorientasi subjek, terintegrasi, time-variant, dan bersifat tetap dari koleksi data dalam mendukung proses pengambilan keputusan management.

Berdasarkan definisi yang dikemukakan Han Jiawei (2011:p144) tentang *data warehouse*, maka *data warehouse* mempunyai empat buah karakteristik yaitu :

- ***Subject Oriented***

Data warehouse diorganisasikan ke dalam banyak *subject* yang utama seperti *customer*, *product*, dan *sales*. *Data warehouse* tidak berkonsentrasi pada kegiatan operasional sehari-hari dan proses transaksi pada suatu organisasi, *data warehouse* fokus terhadap *modeling* dan analisis data untuk pengambilan keputusan, oleh karena itu, *data warehouse* menghasilkan pandangan sederhana dan ringkasan terhadap *subject* tertentu dan mengabaikan data yang tidak *relevant* terhadap proses pendukung keputusan.

- *Integrated*

Data Warehouse biasanya dibangun dengan mengintegrasikan sumber data yang berbeda-beda, seperti *relational database*, *flat files*, dan *online transactional*. Teknik *data cleaning* dan *data integration* digunakan untuk memastikan konsistensi dalam konversi penamaan, struktur pengkodean, ukuran atribut, dan sebagainya.

- *Non-volatile*

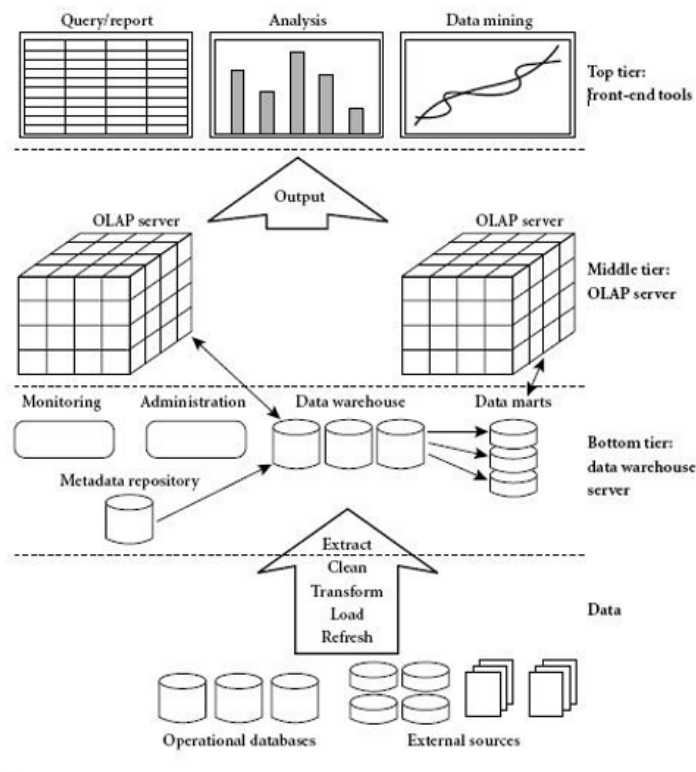
Data warehouse secara fisik memisahkan pengumpulan data dari aplikasi data yang ditemukan dalam *operational environment*. Di dalam pemisahan *data warehouse* tidak memerlukan proses transaksi, *recovery*, dan *concurrency control mechanism*. Biasanya hanya membutuhkan dua operasi dalam mengakses data, *initial loading data* dan *access of data*.

- *Time variant*

Data disimpan untuk memberikan informasi dari perspektif *history* (misalnya, 5-10 tahun terakhir). Setiap struktur kunci dalam *data warehouse* berisi baik secara implisit maupun eksplisit elemen waktu.

Data warehouse mengandung tiga tingkatan dalam arsitektural, seperti digambarkan pada gambar.

Gambar 19. Data warehousing Architecture



12 A three-tier data warehousing architecture.

Data Mart

Data mart adalah suatu bagian pada *data warehouse* yang mendukung pembuatan laporan dan analisa data pada suatu unit, bagian atau operasi pada suatu perusahaan. Dalam beberapa implementasi *data warehouse*, *data mart* adalah *miniature data warehouse*. *Data mart* sering digunakan untuk memberikan informasi kepada segmen fungsional organisasi.

Data Mart adalah fasilitas penyimpanan data yang berorientasi pada *Subject* tertentu atau berorientasi pada Departemen tertentu dari suatu organisasi, fokus pada kebutuhan Departemen tertentu seperti *Sales*, *Marketing*, *Operation* atau *Collection*. Sehingga suatu Organisasi bisa mempunyai lebih dari satu *Data Mart*. *Data Mart* pada umumnya di organisasikan sebagai suatu *Dimensional Model*, seperti *Star-Schema* (OLAP Cube) yang tersusun dari sebuah tabel *Fact* dan beberapa tabel *Dimension*.

Contoh umum *data mart* adalah untuk departemen penjualan, departemen persediaan dan pengiriman, departemen keuangan, manajemen tingkat atas, dan seterusnya. *Data mart* juga dapat digunakan untuk gudang data segmen data untuk mencerminkan bisnis secara geografis terletak di mana masing-masing daerah relatif otonom. Sebagai contoh, sebuah organisasi layanan yang besar mungkin memperlakukan pusat operasi regional sebagai unit usaha perorangan, masing-masing dengan *data mart* sendiri yang memberikan kontribusi untuk gudang *data master*.

Microservice for Analytic

Microservices adalah arsitektur perangkat lunak untuk mengembangkan suatu aplikasi yang dibentuk dari susunan berbagai *service* yang kecil, berjalan dengan prosesnya sendiri, dan berkomunikasi menggunakan mekanisme yang sederhana, seringkali menggunakan HTTP sebagai media komunikasi API.

Komponen adalah unit perangkat lunak yang berdiri sendiri (*independent*), *replaceable*, dan *upgradeable*. *Microservices* dapat dipandang sebagai komponen perangkat lunak.

Untuk melakukan perubahan dari arsitektur monolitik ke *microservices* ada tiga tahap untuk melakukanya yaitu *stop digging*, *split fronted and backend* dan *extrack services* :

- **Stop Digging**, berhenti mengali aplikasi monolitik dan harus berhenti membuat monolitik lebih besar lagi. Ini berarti bahwa ketika Anda menerapkan fungsi baru Anda tidak boleh menambahkan lebih banyak kode ke monolitik. Sebagai gantinya, gagasan besar dengan strategi ini adalah memasukkan kode baru itu ke dalam *microservices*.
- **Split Fronted and Backed**, Sebuah strategi yang mengecilkan aplikasi monolitik adalah dengan membagi lapisan presentasi dari logika bisnis dan lapisan akses data. Aplikasi perusahaan yang khas terdiri dari setidaknya tiga jenis komponen yang berbeda:
 - » Lapisan presentasi – Komponen yang menangani permintaan HTTP dan menerapkan API (REST) atau UI web berbasis HTML. Dalam sebuah aplikasi yang memiliki antarmuka pengguna yang canggih, tingkat presentasi seringkali merupakan kumpulan kode yang besar.
 - » *Business logiclayer* – *Components that are the core*he aplikasi dan menerapkan peraturan bisnis
 - » Lapisan akses data – Komponen yang mengakses komponen infrastruktur, seperti database dan pialang pesan

- **Extract Services**, ketiga adalah mengubah modul yang ada dalam monolit menjadi *microservices* mandiri. Setiap kali Anda mengekstrak modul dan mengubahnya menjadi layanan, monolit menyusut. Setelah Anda cukup mengubah modul, monolit akan berhenti menjadi masalah. Entah itu akan hilang sama sekali. Atau menjadi cukup kecil sehingga hanya layanan lain. Memprioritaskan Modul mana untuk Dikonversi ke Layanan aplikasi monolitik besar dan kompleks terdiri dari puluhan atau ratusan modul, yang kesemuanya merupakan kandidat untuk ekstraksi.

Visualisation

Data visualization merupakan hal yang sangat krusial bagi perkembangan suatu bisnis atau karir. Dibandingkan menggunakan membaca saja, Wyzowl berkata 80% orang lebih gampang mengerti sesuatu secara visual.

Menurut Tableau, visualisasi data merupakan representasi grafis menurut keterangan & data. Dengan memakai elemen visual misalnya bagan, grafik, dan peta, indera visualisasi data menyediakan cara yg bisa diakses buat melihat dan mengetahui tren, *outlier*, dan pola pada.

Gambar 20. Contoh Visualisasi Data



05 STRATEGI PENCAPAIAN

Strategi Pengembangan

a. Strategi Tata Kelola *Big Data*

Rencana penyesuaian pada organisasi dan manajemen SI/TI (Pusat Data Dan Sistem Informasi) yang akan dilakukan adalah sebagai berikut:

1. Penyusunan dan Penetapan Kebijakan Pengelolaan *Big Data*
2. Penyusunan dan Penetapan SOP penerapan *Big Data*
3. Penyusunan dan Penetapan SK Tim Data Analitik K/L/D
4. Penataan SDM TI melalui pelatihan dan sertifikasi kompetensi *Big Data* dilingkungan Kementerian/Lembaga/Pemerintah Daerah
5. Pendampingan dan transfer knowledge penerapan Analitik Data dalam beberapa use case critical oleh pihak ketiga
6. Sosialisasi pemanfaatan *Big Data* di lingkungan K/L/D
7. Sosialisasi keamanan data di lingkungan K/L/D
8. Pilot project pengembangan use case pada beberapa satuan kerja di lingkungan K/L/D
9. Membangun koordinasi yang kuat dengan seluruh Satuan Kerja di Lingkungan K/L/D
10. Membangun kerjasama dengan lintas Kementerian, Perguruan Tinggi dan/atau Industri dalam hal pengembangan dan pemanfaatan *big data*

b. Strategi Teknologi *Big Data* dan Sumber Daya Manusia

Beberapa usulan pembangunan dan pengembangan Arsitektur Teknologi *Big Data* untuk mendukung proses kegiatan dalam pencapaian implementasi Arsitektur *Big Data* Nasional. Rencana pembangunan dan pengembangan terkait teknologi big data yang akan dilakukan adalah sebagai berikut.

1. Perluasan dan pengembangan teknologi penarikan data melalui berbagai kanal seperti *Web Scraping*, *CAWI/CAPI* dan *Import/ Upload*.
2. Pengembangan *Data Mart* melalui perluasan akuisisi data
3. Adopsi penerapan *tools* visualisasi modern (Power BI, Plotly, Data Studio dan Tableau)
4. Penerapan *Artificial Intelligence* (AI) melalui pengembangan *tools Machine Learning*
5. Penerapan Ekosistem Hadoop dalam pengembangan *Data Lake*

c. Strategi Infrastruktur dan Keamanan *Big Data*

Dalam mendukung usulan pembangunan dan pengembangan Teknologi *Big Data*, maka diperlukan penyesuaian infrastruktur dan keamanan *big data*. Seluruh server menggunakan komputasi awan (*cloud*) yang akan ditempatkan pada Pusat Data Nasional (*National Data Center*) sesuai amanah Peraturan Presiden Nomor 95 Tahun 2018 tentang sistem Pemerintahan Berbasis Elektronik (SPBE). Selain itu, perlu dilakukan *backup* atau *mirroring* terhadap pusat data di lingkungan K/L/D sehingga dapat terjamin ketersediaan dan keamanannya. Rencana pengembangan Infrastruktur dan Keamanan dalam mendukung teknologi *big data* di lingkungan K/L/D yang akan dilakukan adalah sebagai berikut.

1. Integrasi Portal Satu Data Indonesia
2. Membuat *staging development* pada setiap server untuk QA dan QC sistem aplikasi *Big Data*
3. Pengembangan Sistem Penghubung Layanan Pemerintah (SPLP)
4. Membuat *environment Data Logs* untuk analisis aktifitas aplikasi *Big Data*
5. Migrasi menggunakan arsitektur *microservices* dalam mempercepat prosesing data
6. Penerapan *Internet of Things (IoT)*.

Roadmap

Beberapa program disusun dalam waktu 3 tahun kedepan, untuk implementasi *Big Data* seperti pada tabel berikut.

Tabel 3. Roadmap Implementasi Big Data

No	Aktivitas	2023				2024				2025				
		Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	
A	Tata Kelola													
1	Penyusunan dan Penetapan Kebijakan Pengelolaan Big Data yang selaras dengan Satu Data Indonesia													
2	Penyusunan dan Penetapan SOP penerapan Big Data													
3	Penataan SDM kompetensi Big Data melalui pelatihan dan sertifikasi													
4	Sosialisasi pemanfaatan Big Data secara nasional													
6	Pilot project pengembangan use case Big Data pada beberapa Instansi Pemerintah													
7	Membangun kerjasama multi-stakeholder baik Kementerian, Perguruan Tinggi dan/atau Industri dalam hal pengembangan dan pemanfaatan big data													
B	Teknologi													
1	Mendorong penerapan single platform Big Data terintegrasi													
2	Sosialisasi pengembangan teknologi penarikan data melalui berbagai kanal seperti Web Scraping, Crawling, CAWI/CAPI dan Import/ Upload.													
5	Mendorong adopsi penerapan tools visualisasi modern (Power BI, Plotly, Data Studio dan Tableau)													
6	Penerapan Artificial Intelligence (AI) melalui pengembangan tools Machine Learning													
7	Penerapan Ekosistem Hadoop dalam pengembangan Data Lake													
C	Infrastruktur dan Keamanan													
1	Mendorong penerapan Sistem Penghubung Layanan Pemerintah (SPLP)													
2	Penerapan Internet of Things (IoT) dalam perluasan akuisisi data streamline													
3	Mendorong Keamanan Data melalui penerapan Komputasi Awan Pusat Data Nasional													

Tata Kelola

Roadmap Tata Kelola *Big Data* disusun berdasarkan penilaian prioritas terhadap perencanaan organisasi dan SDM Teknologi Informasi. Beberapa program tersebut diatas akan disusun dalam waktu 3 tahun kedepan, seperti pada tabel berikut.

Tabel 4. Roadmap Tata Kelola Big Data

No	Aktivitas	2023				2024				2025			
		Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4
A	Tata Kelola												
1	Penyusunan dan Penetapan Kebijakan Pengelolaan Big Data yang selaras dengan Satu Data Indonesia												
2	Penyusunan dan Penetapan SOP penerapan Big Data												
3	Penataan SDM kompetensi Big Data melalui pelatihan dan sertifikasi												
4	Sosialisasi pemanfaatan Big Data secara nasional												

6	Pilot project pengembangan use case Big Data pada beberapa Instansi Pemerintah													
7	Membangun kerjasama multi-stakeholder baik Kementerian, Perguruan Tinggi dan/atau Industri dalam hal pengembangan dan pemanfaatan big data													

Teknologi Big Data

Roadmap Teknologi Big Data disusun berdasarkan penilaian terhadap Arsitektur Teknologi Big Data. Beberapa program tersebut akan dibangun dan dikembangkan dalam waktu 3 tahun kedepan, seperti pada tabel berikut.

Tabel 5. Roadmap Teknologi Big Data

No	Aktivitas	2023				2024				2025			
		Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4
	Teknologi												
1	Mendorong penerapan single platform Big Data terintegrasi												
2	Sosialisasi pengembangan teknologi penarikan data melalui berbagai kanal seperti Web Scraping, Crawling, CAWI/CAPI dan Import/ Upload.												
5	Mendorong adopsi penerapan tools visualisasi modern (Power BI, Plotly, Data Studio dan Tableau)												

6	Penerapan Artificial Intelligence (AI) melalui pengembangan tools Machine Learning												
7	Penerapan Ekosistem Hadoop dalam pengembangan Data Lake												

Infrastruktur dan Keamanan

Roadmap Infrastruktur dan Keamanan *Big Data* disusun berdasarkan penilaian prioritas terhadap penyediaan infrastruktur yang selaras dengan Peraturan Presiden Nomor 95 Tahun 2018 tentang Sistem Pemerintahan Berbasis Elektronik (SPBE). Beberapa program tersebut akan dibangun dan dikembangkan dalam waktu 3 tahun kedepan, seperti pada tabel berikut.

Tabel 6. Roadmap Infrastruktur dan Keamanan Big Data

No	Aktivitas	2023				2024				2025			
		Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4
	Infrastruktur dan Keamanan												
1	Mendorong penerapan Sistem Penghubung Layanan Pemerintah (SPLP)												

2	Penerapan Internet of Things (IoT) dalam perluasan akuisisi data streamline												
3	Mendorong Keamanan Data melalui penerapan Komputasi Awan Pusat Data Nasional												

08 PENUTUP

Laporan kajian Implementasi *Big Data : End to End Approach* disusun guna memberikan pedoman dalam pemanfaatan *Big Data* baik di Pemerintah Pusat maupun Pemerintah Daerah. Strategi dalam roadmap ini akan dijadikan landasan kegiatan dalam pemanfaatan *Big Data* di Pemerintah Pusat maupun Pemerintah Daerah untuk beberapa tahun kedepan dalam rangka pencapaian target yang selaras dengan Sistem Pemerintahan Berbasis Elektronik dan Satu Data Indonesia.

REFERENSI

- [1] The Gartners IT Glossary : What Is Big Data
<https://www.gartner.com>, diakses 2 September 2022

- [2] Global Datasphere Seagate. "The Digitization of the World: From Edge to Core"
<https://www.seagate.com/files/www-content/our-story/trends/files/idc-seagate-dataage-whitepaper.pdf> , diakses 13 September 2022

- [3] MongoDB Big Data Explain
<https://www.mongodb.com/basics/big-data-explained>

- [4] Rizqifaluthi, H., & Yaqin, M. A. (2019). Process Mining Akademik Sekolah menggunakan RapidMiner. *Matics*, 10(2),47-49.
<https://doi.org/10.18860/mat.v10i2.5158>

- [5] Shree, R., Choudhury, T., Gupta, S. C., & Kumar, P. (2018). KAFKA: The modern platform for data management and analysis in big data domain. 2nd International Conference on Telecommunication and Networks, TEL-NET 2017, 2018-January, 1-5. <https://doi.org/10.1109/TEL-NET.2017.8343593>

- [6] Apache Spark Explanation
<https://spark.apache.org/>, diakses 23 September 2022

- [7] About Tableau
<https://www.tableau.com/>, diakses 24 September 2022

- [8] Hatzi, Vassiliki, dkk. 2014. Web Page Download Scheduling Policies for Green Web Crawling. 22nd International Conference on Software, Telecommunications and Computer Networks (SoftCOM).

- [9] Jing Wang, Yuchun Guo. 2012. Scrapy-based Crawling and User-behavior Characteristics Analysis on Taobao. 2012 International Conference on Cyber-Enabled Distributed Computing and Knowledge Discover.

- [10] B. Zhao, "Web Scraping," in Springer International Publishing AG, USA, 2017.

- [11] Inmon, B. (2016). Data Lake Architecture. New Jersey USA: Technics Publications.

- [12] Ibrahim, Nugroho Setyabudhi, Takariyana Heni A., 2004,Perancangan Data Warehouse Pada Pusat Data dan Informasi

- [13] G. Grunman and A. Marrison, "Microservices :The Resurgence of SOA Principles and an Alternative to the Monolithic," 2014. [Online].Available: <http://www.pwc.com/us/en/technology-forecast/2014/cloud-computing/features/microservices.html>, diakses 24 September 2022
- [14] C. Richardson and F. Smith, Microsevicees From Design to Deployment, NGINX, 2016.