

## **BAB II**

### **TINJAUAN PUSTAKA**

#### **II.1 Introduksi *Big Data***

Pada bagian ini akan membahas secara rinci mengenai konsep dan definisi *big data* dari berbagai perspektif berdasarkan literatur. Konsep mencakup mengapa muncul penggunaan istilah *big data* yang dipisahkan dengan istilah konvensional dalam pengelolaan data. Konsep juga menjelaskan esensi dari adanya *big data* yang disebut dengan istilah *general notion*. Selanjutnya mengenai definisi akan dijelaskan apa definisi awal *big data* dan perkembangan dari definisi tersebut serta perspektif penggunaan definisi *big data* dari berbagai literatur pada berbagai negara di dunia termasuk di Indonesia.

##### **II.1.1 Konsep *Big Data***

Konsep dari *big data* didasari oleh fenomena digitalisasi. Digitalisasi mendorong integrasi dalam koneksi teknologi dari berbagai layanan maupun proses. Fenomena ini mendorong istilah yang disebut dengan *datafication*. *Datafication* merupakan data yang dihasilkan dari berbagai jenis teknologi seperti sensor, kamera, *smartphone*, dan lain-lain. Istilah *big data* mendeskripsikan “data yang dihasilkan oleh proses *datafication*” (Schonberger, 2013). Melalui teknologi informasi, proses penciptaan informasi setiap harinya dari berbagai sumber mengarah kepada terminologi *big data* (Sirait, 2016). Serupa dengan apa yang dinyatakan oleh (Wong, 2016) bahwa perkembangan teknologi informasi seperti media sosial, *smartphone*, telah memungkinkan munculnya berbagai sumber baru dalam melakukan pengumpulan data. Tentunya, data yang dihasilkan oleh sumber-sumber baru ini memiliki ukuran yang besar dan memicu munculnya era *big data*. Berbeda dengan dekade yang lalu, dimana untuk menciptakan data maka dibutuhkan kesengajaan dan terkadang membutuhkan keterlibatan pihak lain.

Namun, beberapa tahun belakangan, proses penciptaan lebih bersifat *unintentional data*.

Konsep *big data* menimbulkan sedikit pertanyaan dikarenakan praktik untuk mengumpulkan dan menyimpan data dalam jumlah besar pada dekade yang lalu. Namun pembedanya dengan praktik tradisional yaitu bahwa *big data* juga berurusan tidak hanya dengan *structured data* melainkan juga *semi-structured* dan *unstructured data*. Pengembangan *big data* lebih lanjut, salah satunya *big data analytics* juga memungkinkan melakukan analisis yang bersifat prediktif, tidak seperti analisis tradisional yang lebih bersifat deskriptif (Sharda, 2017). Dari perspektif data, maka munculnya istilah *big data* didorong oleh ketidakmampuan dari teknologi tradisional untuk secara efektif menyimpan, mengelola, dan menganalisis data yang ada seiring dengan perkembangan teknologi (Zafira, 2019).

Seorang ekonom asal Austria, Peter Drucker menciptakan gagasan “*you can't manage what you don't measure*” terkait dengan *big data*. Implikasinya adalah ledakan dari *data generation* khususnya data digital dapat diukur oleh manajer perusahaan dan dapat digunakan secara luas melalui bagaimana menafsirkan data menjadi *input* dalam pengambilan keputusan (McAfee, 2012). Didasari oleh *input* bagi proses pengambilan keputusan yang harus dilakukan secara rasional dan bersifat kompleks, maka informasi memegang peranan fundamental dalam proses pengambilan keputusan, khususnya untuk mengidentifikasi berbagai alternatif beserta konsekuensinya (Almeida, 2018). Perusahaan *Phocas Software* menyatakan bahwa *big data* akan menjadi tidak berarti tanpa adanya analisis atau *big data analytics*.

Gagasan umum dari *big data* menjadi semakin jelas melalui publikasi perusahaan *McKinsey Global Institute* bahwa tujuan *big data* adalah menghasilkan *value*. Intinya adalah pengambilan keputusan bersifat *data driven* yang dikaitkan dengan informasi yang tepat sebagai ganti dari intuisi (Hasibuan, 2021) (McAfee, 2012). Tujuan dari *big data* adalah untuk mengungkap pola tersembunyi dalam menghasilkan pengetahuan dari kumpulan data yang besar (Srinavin, 2021). Penelitian dari (Peiffer, 2016) menegaskan bahwa *big data*

sebagai salah satu faktor yang berperan penting dalam menentukan arah perkembangan industri.

### **II.1.2 Definisi *Big Data***

Sebelum tahun 2001, *big data* diartikan sebagai istilah untuk mendeskripsikan fenomena dimana volume data yang terus meningkat dan menyebabkan ukuran data yang besar. Akhirnya pada tahun 2001, (Laney, 2001) mendeskripsikan tiga esensi dari *big data* yaitu *volume*, *velocity*, dan *variety* atau lebih dikenal dengan istilah 3 V's. Definisi ini telah menjadi *milestone* penting bagi perkembangan dari definisi *big data*. Hasil dari definisi Laney tersebut dianggap sebagai definisi pertama yang secara jelas dikemukakan, meskipun istilah *big data* mungkin sudah digunakan beberapa tahun ke belakang (Ylijoki, 2016).

*Volume* berarti jumlah atau ukuran data yang besar, atau seberapa banyak data yang ada. *Volume* dikaitkan dengan ukuran yaitu mulai dari *terabyte* (TB), *petabyte* (PB), dan lain sebagainya (Heripracoyo, 2014). Ukuran tersebut dimulai dari 1 *byte* sama dengan 8 *bit*; 1 *kilobyte* (KB) sama dengan 1024 *byte*; 1 *megabyte* (MB) sama dengan 1024 KB; 1 *gigabyte* (GB) sama dengan 1024 MB; 1 *terabyte* (TB) sama dengan 1024 GB; 1 *petabyte* (PB) sama dengan 1024 TB; 1 *exabyte* (EB) sama dengan 1024 PB; 1 *zettabyte* (ZB) sama dengan 1024 EB; 1 *yottabyte* (YB) sama dengan 1024 ZB (Turner, 2020). Untuk melihat seberapa besar ukuran tersebut, maka terdapat beberapa fakta berikut (Jackson, 2011):

1. 1 TB *hard drive* bisa menyimpan 200.000 foto;
2. 1 ZB membutuhkan pusat data seluas 1/5 kota Manhattan di Amerika atau sekitar 12 km<sup>2</sup>;
3. Jika harga *hard drive* 1 TB standar adalah \$100, maka membutuhkan \$100 triliun untuk harga *hard drive* 1 YB.

Pada tahun 2012 yang lalu, setiap harinya diciptakan sekitar 2,5 EB data. Angka tersebut diperkirakan akan berlipat ganda setiap 40 bulan. Bahkan ukuran

data yang muncul di *internet* setiap detiknya lebih banyak dari yang tersimpan dalam *internet* pada 20 tahun yang lalu. Perusahaan *retailer* terbesar di dunia, Walmart mengumpulkan lebih dari 2,5 PB setiap jam dari transaksi pelanggannya (McAfee, 2012). Melihat perkembangan dari volume data yang diciptakan, kenyataannya pada tahun 2012 dari survei yang dilakukan oleh (Schroeck, 2012) kepada 1144 perusahaan, hanya setengah lebih sedikit dari semua perusahaan tersebut yang memiliki penyimpangan data lebih dari 1 TB. Istilah *volume* mengenai apa yang dianggap sebagai “besar” ini mungkin akan jauh berbeda dengan di masa depan karena melihat historis ke belakang juga kapasitas penyimpanan akan terus meningkat secara signifikan dengan perkembangan teknologi (Gandomi, 2014). Oleh karena itu (Wong, 2016) menyatakan bahwa definisi *volume* yaitu “*sample equals to population*” yang berarti data mencakup seluruh populasi.

*Variety* mengacu kepada heterogenitas dari data yang ada. Heterogenitas mencakup tipe data, sumber data, atau format data (Wong, 2016). Istilah tipe data pada *variety* menjelaskan mengenai tipe *structured*, *semi-structured*, dan *unstructured data*, dimana *structured data* hanya sebesar 5% dari total semua data yang ada (Cukier, 2010). Sebagaimana dinyatakan oleh (Almeida, 2018) bahwa *variety* ini yang membedakan *big data* dengan metode tradisional dimana *big data* mampu mengolah data tidak sebatas kepada *structured data* saja. Berdasarkan (Bernard, 2019) *structured data* merupakan tipe data yang paling mudah diolah dikarenakan sesuai dengan *data model* yang dibuat oleh pihak terkait. Contoh *structured data* adalah transaksi dalam format akuntansi, *star rating* dari pelanggan, nomor kartu kredit, informasi persediaan, dan lain-lain. *Unstructured data* memiliki komposisi yang paling besar dari semua data yang beredar. *Unstructured data* merupakan kebalikan dari *structured data* yang berarti tidak terkait dengan keberadaan *data model* contohnya adalah gambar, video, rekaman audio, teks *email*, konten sosial media, hasil wawancara terbuka, dan lain-lain. *Semi-structured data* merupakan kombinasi dari *structured* dan *unstructured data* artinya sebagian isi data mengikuti *data model* sedangkan sebagian lainnya tidak. Contohnya, jika gambar merupakan *unstructured data* tetapi jika diambil dari *smartphone* pastinya terdapat tanggal dan waktu pengambilan foto dan lokasi.

Contoh lainnya adalah *email* meskipun teks pada *email* merupakan *unstructured data* tetapi pada *email* terdapat format pengirim, penerima, waktu pengiriman, dan lain-lain.

Istilah sumber data yang dimaksud dalam *variety* berarti terdapat beragam sumber yang menghasilkan data. Misalkan gambar yang terdapat dalam media sosial, data yang diperoleh dari sensor yang dipasang, data transaksi dari kartu kredit, lokasi *real time* yang diperoleh dari *smartphone* (McAfee, 2012). Hal ini berarti media sosial, sensor, kartu kredit, dan *smartphone* merupakan sumber data yang beragam. Terakhir, terkait dengan istilah format data dalam *variety* berarti keberagaman dari format konten data. Contohnya adalah DWG (*drawing*), RVT (*revit*), DOCS/XLS/PPT (*microsoft*), JPEG (gambar), MPG (video), dan lain-lain (Bilal, 2016).

Terakhir dari istilah “3V’s” yaitu *velocity* yang berarti bahwa data dihasilkan pada kecepatan yang tinggi (Wong, 2015) dan melampaui kemampuan sistem konvensional. *Velocity* berimplikasi terhadap penambahan *volume* sehingga membutuhkan peningkatan kemampuan sistem dalam menyimpan dan memproses data (Zikopoulos, 2012). Definisi lain mengatakan bahwa *velocity* berarti data harus ditangani tepat waktu (Almeida, 2018) atau *real time* atau *near real time*. Contohnya adalah *posting* pada aplikasi Twitter atau Facebook (Middelburg, 2019). Lengkapnya, *velocity* mengacu kepada tingkat kecepatan data dihasilkan dan itu harus dianalisis tepat waktu (Li Cai, 2015). Perkembangan dari perangkat digital seperti *smartphone*, sensor, dan lain-lain telah menyebabkan tingkat kecepatan penciptaan data yang tidak pernah terjadi sebelumnya sehingga mendorong kebutuhan *real-time analytics* (Gandomi, 2014). Contohnya grup di *Massachusetts Institute of Technology* dapat mengetahui berapa banyak orang yang berada di parkiran toko Macy’s dengan melihat kepada data lokasi *smartphone* (McAfee, 2012).

Seiring dengan berjalananya waktu, definisi *big data* yang awalnya hanya 3 V’s menjadi lebih luas dengan penambahan istilah-istilah baru. *Veracity* berarti seberapa akurat data dalam memprediksi *business value*, dimana terkait pemanfaatan data dalam pengembangan bisnis atau perusahaan (Dimas, 2020).

Istilah *Veracity* diciptakan oleh Perusahaan *International Business Machines* (IBM) dan disebut dengan “*the fourth V*’s” *Veracity* merepresentasikan keandalan yang melekat dari data. Contohnya adalah data sosial media yang secara natur bersifat tidak pasti karena mengandung pandangan subjektif manusia, namun data tersebut tentunya mengandung informasi yang berharga, sehingga bergantung kepada metode analisis data tersebut (Gandomi, 2014). *Veracity* diartikan sebagai data yang berkualitas tinggi (Almeida, 2018). *Veracity* mengacu kepada akurasi dari data yang ada dengan prinsip “*garbage in garbage out*” yang artinya jika data yang menjadi *input* merupakan data yang tidak akurat maka hasilnya tidak akan akurat. Tentunya dalam konteks *big data*, derajat ketidakakuratan tertentu masih diterima karena pertimbangan volume yang sangat besar sehingga memperbesar kemungkinan kesalahan (Wong, 2016). *Veracity* berkaitan dengan kebenaran dalam konteks nilai manfaat dari informasi yang dihasilkan (Fendy, 2019).

Terdapat juga istilah *value* yang dimana masuk ke dalam istilah “5V’s” yang mencakup (*volume*, *variety*, *velocity*, *veracity*, dan *value*) yang dirumuskan oleh (Demchenko, 2014). *Value* berarti bahwa data diolah untuk menghasilkan nilai yang signifikan bagi organisasi (Sergio, 2019). *Value* terkait dengan nilai manfaat dari informasi yang dihasilkan (Emyana, 2016). Sedangkan (Li Cai, 2015) menyatakan bahwa *value* bersifat *inversely proportional* terhadap volume data, yang artinya semakin besar volume data maka nilai yang dihasilkan lebih sedikit relatif dengan volume data yang lebih kecil. *Value* menjadi penting karena untuk memastikan tingkat *return on investment* (ROI) yang dihasilkan didasarkan kepada nilai data yang mengarah kepada peningkatan performa yang terukur (Almeida, 2018). Karakteristik *value* masih didefinisikan secara lemah, mengingat semuanya bergantung kepada kemampuan dalam mengekstraksi nilai dari data tersebut (Cristobal, 2020).

Istilah “V’s” yang baru di luar 5V’s juga mulai dilihat setelah tahun 2013. *Variability* berarti variasi dari tingkat arus data, karena *velocity* dari *big data* itu tidak selalu konsisten, namun memiliki titik puncak dan lembah (Gandomi, 2014). Ada periode dimana *velocity* data berada pada puncak sehingga menimbulkan tantangan dalam mengelola data terlebih untuk tipe *unstructured data* (Almeida,

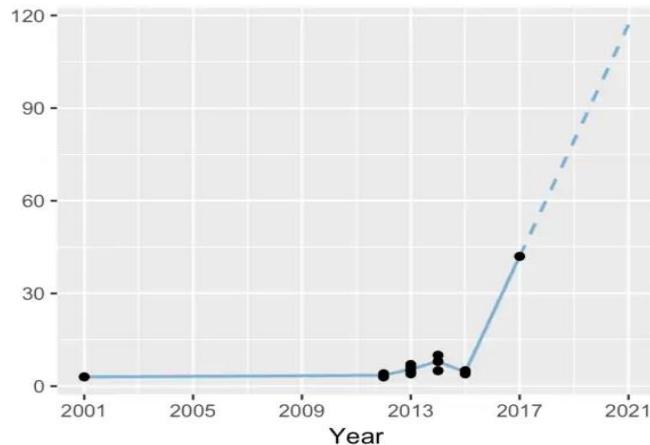
2018). *Variability* berkaitan dengan inkonsistensi dari arus data sehingga beban dari data terutama pada waktu-waktu puncak ketika ada kejadian atau suatu peristiwa menjadi tantangan dalam mengelola, dimana salah satu pemicu munculnya *variability* adalah peningkatan penggunaan media sosial (Katal, 2013) Ada yang mengatakan bahwa *variability* juga lebih terkait kepada *variety* namun lebih menekankan kepada aspek fleksibilitas dalam beradaptasi terhadap format data yang bau melalui proses pengumpulan, penyimpanan, dan pengolahan data (Rogla, 2019).

*Validity* merupakan istilah yang muncul dikarenakan munculnya isu mengenai data pelanggan sehingga perlu adanya penilaian atau kajian lanjut mengenai peraturan yang ada. *Validity* berbeda dengan *veracity*, jika *veracity* menekankan kepada akurasi dan kebenaran data, maka *validity* lebih menekankan kepada peningkatan privasi data dan aspek hukum (Norris, 2018). *Validity* berkaitan dengan aspek integritas data, dimana berkaitan dengan proses validasi bahwa dalam proses pengunggahan maupun pengiriman data, tidak terjadi korup atau kerusakan kepada integritas data (Owais, 2016). *Validity* menekankan kepada proses dalam pemilihan data yang digunakan untuk mengekstrak informasi. Berbeda dengan *veracity* yang berfokus kepada kebenaran data, maka *validity* lebih menekankan kepada penggunaan data yang benar atau tepat (Panimala, 2017).

*Visualization* didasarkan kepada hasil akhir dari *big data* yaitu harus dapat dipahami dan dibaca dengan mudah. Istilah ini menegaskan bahwa meskipun penggunaan analisis dan tepat namun jika tidak dapat dimengerti dengan mudah hasil akhirnya, maka tidak akan berguna. *Visualization* bisa diterjemahkan dalam bentuk grafik atau bagan yang mudah dimengerti dan dibaca (Owais, 2016). Dengan bertambahnya data, akan menjadi semakin penting jika hasil dari data tersebut dapat divisualisasikan, karena jika tidak, akan sulit untuk mengidentifikasi pola dan korelasi antar data (Almeida, 2018). *Visualization* diartikan sebagai proses dalam merepresentasikan data yang awalnya bersifat abstrak menjadi mudah dipahami (Panimala, 2017).

*Volatility* didefinisikan sebagai berapa lama data tersebut masih bersifat valid dan untuk jangka berapa lama data tersebut masih akan disimpan (Mohan, 2018). *Volatility* berbicara mengenai kebijakan *retention* atau penahanan/penyimpanan data dan setelah masa *retention* berakhir, maka data akan dihapus. Contohnya perusahaan *e-commerce* hanya ingin menyimpan data pembelian pelanggan selama 1 tahun, tidak lebih lama daripada itu karena garansi produknya hanya 1 tahun aja (Owais, 2016). Munculnya istilah *volatility* juga didasarkan oleh keterbatasan kapasitas penyimpanan data, sehingga perlu ditentukan berapa lama waktu penyimpanan data, meskipun secara standar pengaturan, data dapat disimpan dalam jangka waktu puluhan tahun (Hurwitz, 2016).

Sebenarnya terdapat beberapa literatur lain yang menyebutkan atau mengagaskan “V’s” lebih banyak dari yang telah disebutkan di atas. (Mohan, 2018) menyatakan “10 V’s” dengan penambahan istilah-istilah lain seperti *venue* (data tersebar di berbagai *platform*), *varmint* (volume data yang semakin besar dapat menyebabkan perangkat lunak mengalami *bug*). (Panimalar, 2017) mengagaskan istilah “17 V’s” dengan penambahan istilah-istilah seperti *virality* (kecepatan dalam penyebaran atau penerimaan data tertentu), *viscosity* (perbedaan waktu atau *lag time* antara kejadian yang terjadi dengan data yang muncul), dan lain-lain. Perkembangan istilah “V’s” terus bertambah hingga muncul istilah “42 V’s” dari *big data* yang dikemukakan oleh (Shafer, 2017) dengan penambahan istilah seperti *vanilla* (bentuk dan pengolahan sederhana dari *big data* tetapi dapat memberikan *value*), *vantage* (*big data* memungkinkan pengguna mendapatkan keunggulan dalam melihat sistem yang kompleks), *vane* (membantu pengambilan keputusan dengan mengarahkan kepada arah yang tepat), dan lain-lain. Adapun perkembangan dari istilah “V’s” dari *big data* dapat dilihat pada gambar berikut:



**Gambar II. 1 Perkembangan Istilah “V’s” *Big Data* (Shafer, 2017)**

Meskipun disebutkan penggunaan istilah-istilah “V’s” di atas, penelitian yang dilakukan oleh (Gandomi, 2014) menyatakan bahwa tidak adanya pemahaman yang koheren dari nomenklatur atau penggunaan istilah *big data*. Kenyataannya, hanya sedikit konsensus mengenai pertanyaan mendasar terkait dengan “seberapa besar data sehingga dapat dikualifikasikan sebagai *big data*?” Jawabannya terdapat dalam pendapat yang tertuang melalui penelitian (Wong, 2016) bahwa meskipun definisi *big data* yang sudah umum digunakan tentang “V’s”, tetapi sampai sekarang masih sulit untuk menentukan apakah suatu data telah memenuhi persyaratan “V’s” tersebut. Premisnya adalah bahwa pada dua dekade yang lalu, cakram liuk hanya memiliki ukuran penyimpanan 1,44 MB, sedangkan pada masa sekarang, *universal serial bus* (USB) yang memiliki ukuran jauh lebih kecil secara fisik namun dapat menyimpan hingga 64 *gigabyte* GB. Oleh karena itu, definisi “big” bersifat relatif dari waktu ke waktu. Sama halnya dengan *variety* dan *velocity* yang bergantung kepada perkembangan teknologi. Tentunya pada suatu waktu yang sama, acuan dari karakteristik *big data* juga berbeda antara satu industri dengan industri lain.

Argumen lain yang dikemukakan oleh (Ylijoki, 2016) dalam penelitiannya dalam melakukan pemetaan terhadap istilah *big data* yang digunakan dalam berbagai literatur terkait *big data* yang ada yaitu bahwa penggunaan istilah “V’s” di luar dari apa yang dikemukakan oleh (Laney, 2001) telah menciptakan inkoherensi. Misalkan *value*, dimana sebenarnya *value* itu bersifat *case-dependent*

sehingga dapat dikatakan bahwa data mentah belum memiliki nilai, dan nilai juga akan muncul bergantung kepada kemampuan dalam mengolah data tersebut. Oleh karena itu, dalam mendefinisikan *big data* perlu dipisahkan antara data dengan kegunaannya. Hal ini sejalan juga dengan (Schonberger, 2013) yang menyatakan bahwa *value* merupakan penggunaan sekunder dari data. Sama halnya untuk *veracity*, dimana diperlukan analisis terlebih dahulu untuk menentukan apakah data tersebut relevan dengan rencana penggunaannya. Definisi *big data* di luar 3 V's tidak mendefinisikan karakteristik dari *big data* melainkan merefleksikan penggunaan dari data.

## II.2 Sejarah *Big Data*

Munculnya *big data* secara tiba-tiba telah meninggalkan berbagai pihak dalam kondisi tidak siap. Sejarah *big data* sulit diketahui dikarenakan jika pada masa lampau, adanya pengembangan atau inovasi teknologi biasanya bermula dari publikasi ataupun riset akademik. Namun, berbeda dengan teknologi *big data* yang muncul bukan dari domain akademik sehingga menyebabkan banyaknya pihak yang belum menyadari atau memahami tentang *big data* (Gandomi, 2014). Kronologis dari Sejarah *big data* disusun dengan menggabungkan penelitian yang dilakukan oleh (Phillips, 2021; Beatrice, 2021; Foote, 2017; Firican, 2022; Rjimenam, 2013; Winegar, 2020; Belyh, 2019) sebagai berikut:

- Sejarah *big data* dimulai pada tahun 1663, meskipun saat itu belum menggunakan istilah *big data*, namun upaya yang dilakukan oleh John Graunt yang dianggap sebagai pendiri demografi asal Inggris adalah melakukan pencatatan statistik terkait angka kematian akibat wabah pes bubo atau *bubonic plague*. Volume data pada saat itu dikatakan sangat besar, mengingat *bubonic plague* menginfeksi hampir seluruh benua Eropa. Tujuan John Graunt adalah untuk meningkatkan kesadaran masyarakat mengenai dampak dari *bubonic plague* saat itu;
- Selanjutnya pada tahun 1865, Richard Millar Denvens mengemukakan istilah “*business intelligence*” yang merupakan proses pengumpulan dan

analisis untuk memberikan informasi yang berguna. Pada bukunya yang berjudul “*Cyclopedia of Commercial and Business Anecdotes*”, dia menjelaskan bagaimana bankir bisa menggunakan informasi untuk menghasilkan keuntungan;

- Pada tahun 1880, *US Census Bureau* mengumumkan bahwa untuk mengolah data sensus pada tahun itu membutuhkan waktu 8 tahun, hingga pada tahun 1881 Herman Hollerith menemukan mesin untuk mengurangi waktu pengolahan data hingga menjadi 3 bulan saja. Beliau pada akhirnya mendirikan perusahaan dan bergabung dengan beberapa perusahaan lain dan membentuk perusahaan *International Business Machines* (IBM) pada tahun 1911;
- Pada tahun 1937, proyek besar pertama yang berkaitan dengan penyimpanan data, digagas oleh Presiden Amerika ke-32, Franklin Delano Roosevelt yang meminta IBM untuk membuat *database* dalam mendukung administrasi dari implementasi *Social Security Act* yang dibuat pada tahun 1935;
- Pada tahun 1965, Pemerintah Amerika Serikat memutuskan untuk membuat *data center* pertama untuk penyimpanan data pajak dan sidik jari yang disimpan di satu lokasi;
- Selanjutnya terjadi sejarah besar pada tahun 1990 dimana dikembangkan *World Wide Web* dan *Hypertext Markup Language* (HTML). Pada saat itu dimulai era internet yang mendorong kepada peningkatan aksesibilitas dari data dan penyimpanan data secara digital.
- Pada tahun 1998 dikembangkan *Not Only Standard Query Language* (NoSQL) oleh Carlo Strozzi yang merupakan *open source database* yang memungkinkan penyimpanan dan pengambilan data yang dimodelkan secara efisien;
- Pada tahun 2000, Francis Diebold, seorang ekonom Amerika menerbitkan makalah berjudul “*Big Data: Dynamic Factor Models for Macroeconomic Measurement and Forecasting*” dan di dalamnya menyebutkan bahwa *big data* mengacu kepada ledakan dari kuantitas data yang tersedia dan

relevan, disebabkan oleh kemajuan teknologi perekaman dan penyimpanan data;

- Pada tahun 2001 (Ylijoki, 2016; Phillips, 2021), Doug Laney menggagas definisi jelas dari *big data* yaitu 3V's yang mencakup *volume*, *variety*, dan *velocity*. Sampai sekarang definisi *big data* menurut Laney merupakan definisi yang secara luas diterima;
- Beberapa literatur (Belyh, 2019; Rjimenam, 2013; Firican, 2022) menyatakan bahwa istilah pertama *big data* digagas oleh Roger Mougalas pada tahun 2005 dan didefinisikan sebagai “*large set of data*” dimana istilah ini muncul karena sudah tidak mungkin pengolahan data dilakukan oleh sistem tradisional;
- Pada tahun 2005 muncul Apache Hadoop yang diciptakan oleh Doug Cutting dan Mike Cafarella. Apache Hadoop merupakan *open source framework* yang digunakan untuk menyimpan dan memproses data dengan volume besar;
- Pada tahun 2006 perusahaan Amazon menyediakan *Amazon Web Services* (AWS) yang merupakan layanan *cloud*;
- Eric Schmidt, *Chief Executive Officer* (CEO) Google mengatakan pada konferensi di California pada tahun 2010 bahwa terdapat 5 EB data yang disimpan dari permulaan hingga 2003 dan mungkin di tahun 2016, data 5 Eb diciptakan setiap dua hari sekali. Kecepatan dari penciptaan data dinilai tidak akan menurun dalam beberapa waktu ke depan;
- Pada tahun 2012, inisiatif investasi dilakukan oleh Mantan Presiden Amerika Serikat, Barack Obama senilai 200 juta dolar yang bertujuan untuk mengekstrak *insight* dari *big data*;
- Riset dari tahun 2016 menyatakan bahwa 90 persen dari data yang diciptakan terdapat pada tahun 2014-2016 (Phillips, 2021);
- Riset terakhir dari *International Data Corporation* (IDC) pada tahun 2017 memprediksi bahwa nilai valuasi pasar *big data analytics* akan mencapai \$203 miliar dibandingkan \$10 miliar pada tahun 2013 lalu.

## II.3 Konstituen *Big Data*

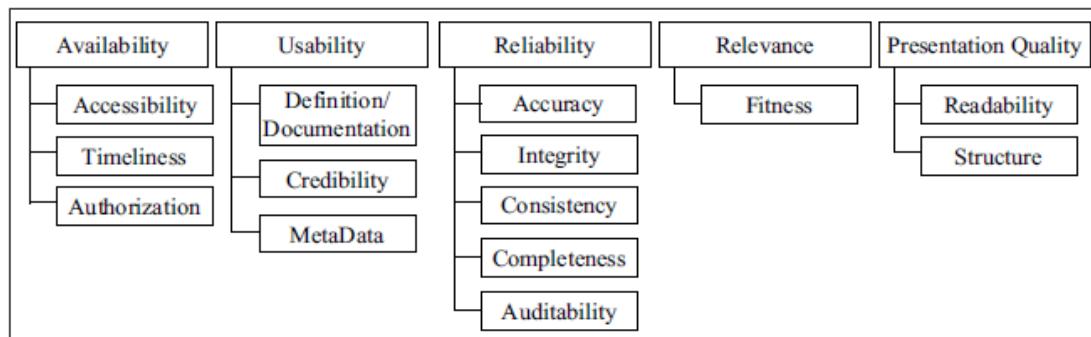
Konstituen menurut KBBI merupakan bagian yang penting atau bagian dari atau pendukung konstruksi. Dalam literatur yang dipublikasi, terdapat beberapa penggunaan istilah lain selain “konstituen” yaitu “*factors*”, “*attributes*”, “*critical success factors*”, “*key factors*”, bahkan “*readiness*”, yang sebenarnya menjelaskan elemen atau aspek yang menentukan keberhasilan dari proyek *big data* yang akan dilakukan. Sebenarnya hasil dari konstituen *big data* yang diperoleh dari studi literatur cukup beragam, namun akan dibahas hanya empat saja konstituen *big data* yang dinilai paling sering muncul dalam hasil penelitian yang ditemukan melalui penelitian (Adrian, 2017; Sai, 2020; Eybers, 2017), dimana ketiga penelitian tersebut telah mengumpulkan banyak penelitian lain yang membahas konstituen *big data* kemudian dipaparkan dalam tabel. Oleh karena itu, empat konstituen yang akan dijelaskan mencakup data, SDM, teknologi, dan organisasi.

### II.3.1 Data

Ketersediaan data menjadi kunci bagi teknologi *big data*. Data dapat diperoleh dari proses bisnis organisasi baik dalam bentuk *structured*, *semi-structured*, ataupun *unstructured*. Selain dari proses bisnis organisasi, data juga dapat diperoleh dengan membeli atau bekerja sama dengan pihak lain (Emyana, 2016). Untuk memastikan kinerja yang optimal, maka harus ditentukan data apa saja yang akan dikumpulkan, kemudian mengidentifikasi sumber-sumber data, dan dalam format apa data akan diterima sebagai *raw data* yang nantinya akan diproses oleh SDM menggunakan teknologi *big data* (Rogla, 2019).

Aspek data tidak hanya terkait ketersediaan data, tetapi juga mengenai kualitas data. Tujuannya adalah untuk memastikan kualitas data mulai dari proses perolehan, transformasi, dan analisis sehingga bisa menghasilkan *output* yang valid (Rogla, 2019). Data dengan kualitas tinggi merupakan prasyarat untuk menganalisis dan menggunakan *big data* sehingga dapat menjamin *value* yang dihasilkan berdasarkan prinsip *garbage in garbage out*. Kualitas data tentunya

juga dipengaruhi oleh sumber data karena menentukan integritas dan keandalan data (Eybers, 2017). Kualitas data mencakup *timeliness*, kredibilitas, aksesibilitas, *readability*, dan lain-lain (Li Cai, 2015). Secara lebih lengkap, aspek kualitas data dapat dilihat pada gambar berikut:



**Gambar II. 2 Aspek Kualitas Data (Li Cai, 2015)**

Di sisi lain, kualitas data saja tidak menggambarkan secara utuh mengenai aspek data. Data dalam *big data* harus terjamin keamanan dan kerahasiaannya, terlebih banyaknya organisasi yang takut untuk melakukan *data sharing* dengan organisasi lain karena alasan keamanan dan kerahasiaan data (Almeida, 2018). Hal ini terjadi karena *big data* berurusan dengan data yang bersifat sensitif (Rogla. 2019). Oleh karena itu perlu adanya *standard operating procedure* (SOP) misalnya berapa lama data akan disimpan atau apa prasyarat untuk mengakses data (Wong, 2016).

### II.3.2 Sumber Daya Manusia

Aspek SDM menjadi sangat penting karena meskipun keberadaan data itu merupakan prasyarat bagi *big data*, tetapi SDM yang mengolah dan menggunakan teknologi *big data* tersebut. Semua teknologi termasuk *big data* bersifat netral, semua akan bergantung kepada SDM yang menggunakannya dan memberikan dampak baik positif maupun negatif. *Big data* dapat digunakan untuk meningkatkan pengalaman dan kepuasan pelanggan, tetapi juga dapat digunakan untuk melanggar kerahasiaan data pelanggan (Almeida, 2018). Oleh karena itu, SDM dalam organisasi harus memiliki kemampuan seperti mengelola data

(McAfee, 2012), keahlian analitik, kerativitas dalam menentukan metode baru, menginterpretasi data, pemrograman, dan lain-lain (Emyana, 2016). Ketersediaan SDM yang berkualitas memiliki dampak yang besar bagi keberhasilan implementasi *big data*, dimana SDM harus memiliki kemampuan analitik dan keterrampilan bekerja dalam tim (Al Sai, 2020).

Profesi yang dikenal memiliki hubungan erat dengan *big data* seperti *data scientists* (McAfee, 2012). Mereka memiliki kemampuan dan pengetahuan mengenai teknologi seperti Hadoop dan instrumen *big data* lainnya (Pedro, 2019). Tentunya tidak hanya *data scientists* yang merupakan satu-satunya SDM penting bagi implementasi *big data*, melainkan peran manajer perusahaan yang harus memiliki pengetahuan dan kemampuan dalam bisnis dan manajemen. Aspek bisnis berarti mengetahui cara kerja dari *big data* bukan dalam arti *coding*, tetapi lebih kepada pemahaman akan data, bagaimana data tersusun, dan apa hubungan antara data-data yang ada. Aspek manajemen berbicara mengenai bagaimana mengalokasikan sumber daya untuk setiap pekerjaan *big data* (Pedro, 2019). Oleh karena itu, kapabilitas manajer yang dibutuhkan yaitu memiliki kemampuan berkomunikasi dan berkoordinasi, serta memiliki pemahaman dan mampu mengevaluasi hasil atau *output* dari *big data* (Adrian, 2017).

Berdasarkan penelitian (Ajah, 2019) dibedakan SDM yang menunjang *big data* berdasarkan fungsi sebagai berikut:

1. Fungsi bisnis: Terdiri dari pimpinan perusahaan dan manajer perusahaan. Pimpinan perusahaan menetapkan tujuan dan strategi, mencari peluang, dan memimpin melalui permasalahan yang muncul. Sedangkan manajer perusahaan melakukan identifikasi risiko, memelihara, dan meningkatkan pemahaman dan *value* dari *big data* dalam organisasi;
2. Fungsi analitis: Terdiri dari *data scientist* yaitu SDM yang memiliki keahlian *big data* mulai dari proses pengumpulan, analisis, dan interpretasi baik *structured*, *semi-structured*, dan *unstructured data*;
3. Fungsi arsitektur data: Terdiri dari *platform engineer* atau pemrogram yang memiliki keahlian dalam bahasa pemrograman untuk membuat *data architecture* yaitu pengaturan mencakup data apa yang dikumpulkan,

- bagaimana data diolah, diintegrasikan, hingga digunakan dalam sistem organisasi;
4. Fungsi operator: Terdiri dari *hadoop engineer* untuk melaksanakan *debugging* dan *troubleshooting* jika terjadi masalah. Fungsi ini merupakan fungsi *maintenance* dari teknologi *big data*.

SDM yang mengoperasikan *big data* dapat diperoleh secara internal maupun eksternal (*outsourcing*). Untuk SDM internal berarti SDM dalam organisasi harus memiliki kompetensi terkait *big data*. Upaya yang dapat dilakukan yaitu mengadakan *continuous training* sehingga memungkinkan SDM untuk memperoleh dan meningkatkan kemampuannya untuk melakukan pekerjaannya (Ghaleb, 2021).

### **II.3.3 Organisasi**

Aspek organisasi merupakan aspek yang paling banyak ditemukan dalam literatur mengenai *critical success factors* dari *big data* (Al Sai, 2020). Organisasi secara utuh harus memahami *perceived benefits* dari teknologi *big data* dalam organisasi sehingga meningkatkan semangat dalam implementasi *big data* (Adrian, 2017). Dengan kata lain, organisasi harus memiliki *big data understanding* karena misalnya jika pimpinan organisasi belum memahami *big data*, maka pimpinan akan berpikir bahwa *big data* merupakan investasi yang mahal dan sulit sehingga sejak awal tidak akan muncul kesediaan dan semangat dalam mengeksplorasi *big data*. *Big data understanding* penting dalam memberantas potensi penolakan atau hambatan yang terjadi selama proses implementasi *big data* (Maaz, 2019).

Tantangannya adalah *big data* memiliki sifat *disruptive* sehingga organisasi secara keseluruhan harus memiliki pemahaman tentang *big data* dan bagaimana *big data* dapat memberikan peluang proses bisnis yang baru. Organisasi harus meninjau kembali proses bisnisnya untuk dapat mengidentifikasi ancaman dan peluang, sehingga organisasi dapat menciptakan strategi implementasi *big data* (Ylijoki, 2016). Ukuran organisasi memiliki perilaku yang

berbeda terhadap *big data*, dimana organisasi besar lebih unggul karena memiliki sumber daya yang lebih banyak sehingga bisa mengambil risiko yang lebih besar dibandingkan dengan organisasi kecil (Ghaleb, 2021).

Dalam menjalankan proses implementasi *big data*, dibutuhkan perubahan budaya organisasi seperti yang awalnya dalam pengambilan keputusan hanya berdasarkan intuisi menjadi *data-driven* yang artinya mengambil keputusan berdasarkan data yang akurat dan relevan (Emyana, 2016). Istilah lainnya disebut dengan *analytics culture* atau budaya analitis yang mengacu kepada kebijakan organisasi untuk selalu mengikutsertakan proses analitis dalam pengambilan keputusan. Budaya tersebut telah terbukti mendorong penerimaan terhadap implementasi *big data* dalam organisasi (Adrian, 2017). Praktik yang sering terjadi adalah eksekutif perusahaan mengambil keputusan menggunakan pendekatan *highest paid person's opinion* (HiPPO) kemudian meminta bawahannya untuk mencari data yang mendukung pendapatnya tersebut. Budaya organisasi lain yang harus dihindari adalah *not invented here syndrome* yang menolak penggunaan apapun dari pihak eksternal (McAfee, 2012). Dengan demikian, optimisme dan budaya kreatif dalam organisasi merupakan faktor yang mendukung adopsi teknologi *big data*, sebaliknya ketidaknyamanan dan kegelisahan merupakan faktor penghambat adopsi teknologi *big data* (Ghaleb, 2021).

Organisasi juga mencakup aspek kepemimpinan, dimana keberhasilan dalam implementasi *big data* tidak sekadar karena memiliki data yang lebih banyak atau lebih berkualitas, tetapi juga karena memiliki pimpinan organisasi yang mendefinisikan visi yang jelas. Pimpinan organisasi harus mampu melihat kesempatan serta menyampaikan hingga mempengaruhi seluruh organisasi untuk berpegang kepada visi tersebut (McAfee, 2012). Dukungan dari manajemen puncak atau pimpinan organisasi sangat penting karena secara langsung mempengaruhi alokasi sumber daya dalam organisasi untuk melaksanakan implementasi *big data* (Ghaleb, 2021). Oleh karena itu, organisasi harus mempertimbangkan visi, misi, dan strategi *big data* serta menyelaraskan dengan

tujuan organisasi. Organisasi harus sadar mengenai apa yang mereka punya dan apa yang ingin dicapai (Al Sai, 2020).

#### **II.3.4 Teknologi**

Teknologi mengacu kepada kemampuan dari infrastruktur teknologi informasi untuk mentransformasikan *big data* dari *raw data* menjadi data yang memiliki *value* dan memberikan pengetahuan yang bisa dijadikan dasar pengambilan keputusan (Adrian, 2017). Teknologi yang dimaksud merupakan infrastruktur ataupun *tools* dalam menangani *volume*, *velocity*, dan *variety* dari *big data* misalnya *cloud* atau *data storage* untuk menangani penyimpanan *big data* (McAfee, 2012). Kategori teknologi terkait dengan proses mulai dari pengumpulan, penyimpanan, analisis, hingga aplikasi data. Lebih lanjut, teknologi harus bisa menyediakan kemampuan *real-time analytics*, kualitas data (keamanan, kerahasiaan, dan validitas) selama proses pengolahan, serta integrasi dengan sistem yang ada (Al Sai, 2020).

Infrastruktur *big data* secara lebih spesifik harus memiliki fungsi yaitu manajemen data, integrasi sistem, dan visualisasi hasil. Manajemen data menjadi penting karena data bersumber dari banyak unit bisnis secara internal maupun juga eksternal, dimana manajemen data berkaitan dengan *flow* dari data tersebut. Integrasi sistem menjadi tantangan terutama untuk sumber data dari eksternal karena berurusan dengan *compatibility* antar sistem. Sedangkan visualisasi hasil berarti infrastruktur *big data* harus mampu menghasilkan *output* dalam bentuk dasbor, laporan, grafik, dan lain sebagainya yang mudah dipahami untuk mendukung pengambilan keputusan yang lebih tepat (Pedro, 2019).

Teknologi juga termasuk hal yang sederhana misalnya jaringan internet (Srinavin, 2021). Secara umum, teknologi yang dibutuhkan tidak terlalu mahal karena kebanyakan merupakan *open source* (McAfee, 2012), sehingga organisasi tidak memiliki kendala yang signifikan karena teknologi bisa diperoleh dengan membeli ataupun bekerja sama dengan pihak lain (Emyana, 2016). Pemilihan teknologi untuk mendukung *big data* pelu memperhatikan faktor kompleksitas

dan kesesuaian. Teknologi dengan kompleksitas yang tinggi akan memiliki efek yang buruk bagi upaya implementasi *big data* karena berarti membutuhkan upaya yang tinggi. Sedangkan untuk kesesuaian berarti melihat sejauh apa tingkat relevansinya dengan sistem *existing* organisasi (Ghaleb, 2021).

## **II.4 Potensi dan Implementasi *Big Data***

Pada bagian ini akan membahas potensi dan implementasi dari *big data* baik di berbagai sektor hingga secara lebih spesifik pada sektor konstruksi. Pada bagian II.4.1 dan II.4.2 akan dijelaskan mulai dari potensi dan implementasi di luar negeri hingga spesifik di Indonesia. Arti kata potensi menurut KBBI merupakan kemampuan yang mempunyai kemungkinan untuk dikembangkan. Sedangkan arti kata implementasi menurut KBBI merupakan pelaksanaan atau penerapan. Potensi dan implementasi akan dibedakan sehingga tidak menimbulkan ambiguitas terkait seberapa jauh penelitian terkait potensi maupun implementasi *big data* pada berbagai sektor termasuk sektor konstruksi baik di luar negeri maupun di Indonesia.

### **II.4.1 Potensi dan Implementasi *Big Data* pada Berbagai Sektor**

Identifikasi potensi dan implementasi *big data* pada berbagai sektor dibatasi hanya untuk literatur yang terpublikasi baik untuk lingkup luar negeri maupun untuk lingkup dalam negeri atau Indonesia.

#### **II.4.1.1 Potensi *Big Data* pada Berbagai Sektor**

Pada sektor keuangan, *big data analytics* dapat digunakan untuk menurunkan *churn rate* dengan mengantisipasi potensi pelanggan yang akan keluar. Selain itu, juga dapat digunakan untuk mengetahui perilaku pelanggan menggunakan kartu kredit, pinjaman apa yang dibutuhkan, serta membantu pihak

perbankan menciptakan produk baru yang sesuai dengan kebutuhan pelanggan (Almeida, 2018).

Pada sektor retail, penggabungan CCTV dengan *big data analytics* mungkin dapat digunakan untuk mengumpulkan informasi yang bersifat demografis tentang pelanggan dalam toko seperti usia, jenis kelamin, dan etnis. Kemudian menghitung jumlah pelanggan dalam toko, berapa lama mereka berada dalam toko, hingga pola pergerakan mereka. Tujuannya adalah untuk penempatan produk yang lebih tepat, penentuan harga, promosi, *cross-selling*, dan penempatan karyawan toko (Gandomi, 2014). *Big data* mampu melakukan segmentasi terhadap pelanggan berdasarkan perilaku dan kebutuhannya sehingga penawaran dan promosi produk dapat menjadi lebih bersifat *personalized* (Almeida, 2018). Pengembangan lebih lanjut dari teknologi *big data* memungkinkan sektor retail menentukan harga maupun tata letak toko (Aktas, 2017), dimana untuk harga salah satunya dapat ditentukan dari *sell-through rate* maupun dari harga kompetitor, sedangkan untuk tata letak toko dilakukan melalui pola pergerakan CCTV dan barang yang dibeli oleh pelanggan.

Pada sektor kesehatan, *big data* dapat mewujudkan *digital health*. *Digital health* membutuhkan sistem informasi untuk mengumpulkan, menggabungkan, dan menganalisis data untuk memastikan pemberian obat yang tepat hingga mencegah epidemi dengan memantau data kesehatan populasi. Dengan demikian, institusi kesehatan memiliki waktu yang lebih dalam menyesuaikan diri serta penetapan kebijakan (Almeida, 2018). Penelitian lain oleh (Wong, 2016) menyatakan potensi *big data* yaitu untuk kesiapan terhadap bencana, contohnya dapat memprediksi kebutuhan mobil ambulans setiap harinya menggunakan data variabel independen seperti cuaca maupun data pasien yang di-update secara *real-time*.

Pada sektor publik, pemerintah dapat memanfaatkan *social media analytics* untuk mendeteksi komunitas yang dilihat dari *network* atau interaksi dalam akun media sosial. Pengembangan lebih lanjut bisa melakukan *social influence analysis* yang bertujuan untuk melihat perilaku tokoh tertentu di media sosial dan bagaimana pengaruhnya terhadap masyarakat (Gandomi, 2014).

Pemerintah dapat memanfaatkan *big data* dalam menetapkan program prioritas, anggaran setiap program, hingga pencegahan bencana alam (Almeida, 2018).

Pada sektor transportasi, diprediksi pada tahun 2025 bahwa kendaraan *autonomous* masa depan mampu berkomunikasi dengan kendaraan lain melalui informasi secara *online* menggunakan data dari CCTV, radar, dan lain-lain untuk memberikan informasi terkait jalan sehingga dapat meningkatkan efisiensi penggunaan bahan bakar (Kanniyappan, 2014).

Pada sektor agrikultur, dikenal dengan istilah *smart farming* yaitu menggunakan teknologi *big data* dalam mengoptimalkan produksi dengan menentukan jenis benih yang akan ditanam, bagaimana cara menanamnya, bagaimana memupuknya, kapan waktu panen, bagaimana penyimpanan hingga transportasi hasil tersebut (Wolfert, 2017).

Pada sektor manufaktur, teknologi *big data* dapat mengurangi *waste* dan *variability* dengan tujuan untuk meningkatkan efisiensi dan efektivitas proses produksi. Ke depannya, *big data* dapat digunakan untuk mengintegrasikan *supply chain* atau rantai pasok sehingga bisa mengimplementasikan *pull manufacturing* yaitu membuat barang atau memesan *raw material* ketika terdapat pesanan saja. *Pull manufacturing* dapat menghemat biaya mulai dari *storage cost* hingga biaya sumber daya (Almeida, 2018).

Pada sektor energi, *big data* dapat mengidentifikasi pola pemakaian listrik dari pelanggan sehingga dapat meningkatkan efisiensi distribusi enegeri dari pusat untuk menyimbangkan data *supply* dan *demand* listrik (Almeida, 2018). Generator listrik dapat dioptimasi secara *real-time* menggunakan baik *real-time data* maupun *demand forecast*. Ke depannya, bangunan-bangunan yang menggunakan *smart meter* dapat terintegrasi dengan generator listrik sehingga menghasilkan data yang lebih akurat (Porter, 2017). Konsep *smart building* yang bertujuan untuk meningkatkan efisiensi energi berdasarkan *input* yaitu temperatur dan paparan sinar matahari juga dapat diimplementasikan menggunakan teknologi *big data* (Koseleva, 2017).

Pada sektor media, yang masih merupakan *early adopter* dari *big data* (Lippell, 2016), *big data* dapat digunakan untuk mengidentifikasi pola perilaku penonton pada jam-jam tertentu mulai dari program apa yang ditonton melalui saluran digital maupun tradisional sehingga dapat dijadikan dasar pengambilan keputusan seperti promosi dan konten seperti apa yang disukai (Almeida, 2018).

Di Indonesia, penelitian oleh (Sudarsono, 2020) membahas mengenai peran *big data* dalam *e-commerce* yaitu *content marketing*, dimana dengan mengidentifikasi perilaku pelanggan maka dapat menyediakan promosi sesuai kebutuhan pelanggan dan memberikan kepuasan serta kemudahan. Penelitian oleh (Mubaroq, 2020) membahas mengenai bagaimana *big data* seharusnya dimanfaatkan oleh pemerintah Indonesia untuk menangani pandemi COVID-19 misalnya menginformasikan melalui aplikasi ponsel jika di dekat mereka ada orang yang terinfeksi COVID-19, serta dengan mengintegrasikan data kependudukan dan data COVID-19 maka seharusnya dapat menjadi *input* dalam penetapan kebijakan COVID-19 di Indonesia.

#### **II.4.1.2 Implementasi *Big Data* pada Berbagai Sektor**

Pada sektor kesehatan, *audio analytics* yang termasuk dalam *big data analytics* digunakan untuk menganalisis *input* berupa suara tangisan bayi untuk mengetahui kondisi kesehatan dan emosional dari bayi (Patil, 2010). Tidak hanya untuk bayi, melainkan juga bisa mengidentifikasi penyakit seperti depresi, kanker, dan lain-lain dengan *input* yaitu pola komunikasi dari pasien (Hirschberg, 2010). Implementasi lain dapat dilihat di Malaysia yaitu dengan dibentuk *The Malaysian Health Data Warehouse* (MyHDW) pada tahun 2017 oleh Kementerian Kesehatan Malaysia. MyHDW merupakan *platform* tunggal yang bertujuan untuk mensinkronisasi data medis pasien baik dari rumah sakit milik negara, rumah sakit swasta, dan lembaga-lembaga kesehatan lainnya, sehingga menyediakan layanan kesehatan dapat melakukan pengambilan keputusan dalam memberikan *treatment* secara tepat (Ghaleb, 2021). Penelitian lain oleh (Mubaroq, 2020) membahas mengenai negara Korea memanfaatkan *big data analytics* yang terintegrasi

dengan aplikasi ponsel dalam menangani pandemi COVID-19, yaitu ketika seseorang dinyatakan positif, akan langsung diumumkan melalui aplikasi ponsel untuk orang-orang yang sebelumnya berpergian ke tempat yang sama agar segera melakukan tes COVID-19.

Pada sektor yang melibatkan kerja di pabrik atau lapangan, maupun di sektor publik, dapat menggunakan CCTV yang dilengkapi dengan *big data analytics* yang memungkinkan pengawasan terhadap *restricted area*, objek yang dilepas, aktivitas yang mencurigakan, hingga terintegrasi secara *real-time* dengan alarm, kunci pintu, menyalakan atau mematikan lampu, dan lain-lain (Hakeem, 2012).

Pada sektor retail, perusahaan Sears Holdings menggunakan teknologi Hadoop untuk mengolah *big data* menggunakan data dari seluruh *brand* yang bekerja sama serta data dari gudang penyimpanan. Tujuannya adalah mengurangi waktu promosi dan sudah terbukti dari yang awalnya membutuhkan delapan minggu menjadi satu minggu saja (McAfee, 2012). Implementasi pada sektor retail dapat dibedakan yaitu *availability* (menggunakan penjualan masa lampau serta perilaku konsumen misalnya skema *loyalty* untuk memprediksi kebutuhan masa depan atau menentukan kapan jenis produk tertentu akan dibeli oleh pelanggan) dan *assortment* (sama seperti *availability* namun lebih berfokus kepada produk apa yang perlu di-*stock* maupun ditampilkan pada waktu tertentu) (Aktas, 2017).

Pada sektor logistik, produk *parcelcopter* yang diproduksi oleh perusahaan *Dalsey, Hillblom, and Lynn* (DHL) merupakan suatu alat yang bergerak di udara dan digunakan untuk mengirimkan barang kepada pelanggan (Hern, 2014). Terkait dengan *big data*, *parcelcopter* membutuhkan data dalam *volume* yang besar, *variety* atau jenis data yang berbeda dari berbagai sumber, serta *velocity* atau kecepatan data yang tinggi dikarenakan membutuhkan data yang *real time*. Jenis data yang dibutuhkan bisa mencakup analisis video, data *geolocation*, dan lain-lain. Semua ini digunakan agar *parcelcopter* dapat terbang dan menghindari tabrakan dengan objek lain di udara hingga mengirimkan tepat kepada lokasi pelanggan yang dituju (Ylijoki, 2016).

Pada sektor transportasi khususnya bandara, terdapat perusahaan PASSUR yang menggunakan *big data* dalam meningkatkan ketepatan *estimated time of arrival* (ETA) dari pesawat. Ketepatan ETA penting karena jika pesawat mendarat sebelum *ground staff* siap, maka penumpang akan menunggu, sebaliknya maka *ground staff* akan *idle* dan meningkatkan biaya. Data yang dibutuhkan PASSUR seperti data cuaca *real-time*, jadwal penerbangan, data pesawat lain di langit, dan lain-lain. Data diperoleh dan *di-update* setiap 4,6 detik (McAfee, 2012). Penelitian lain dari (Kochhar, 2016) menyatakan bahwa *big data* telah digunakan untuk transportasi rel dalam fase *planning* dan *operation and maintenance*, dimana *big data* mampu melakukan pemodelan terhadap *demand*, pemilihan rute, hingga *predictive maintenance*.

Pada sektor agrikultur, *Integrated Field Systems* (IFS) telah diterapkan di Amerika Utara, dimana IFS mengumpulkan data mengenai kondisi tanah, cuaca, dan varietas rumput liar. Tujuannya adalah membantu petani untuk mengidentifikasi kebutuhan dari tanaman seperti senyawa kimia, bahkan memungkinkan kebutuhan unsur kimia baru sehingga dapat meminimalkan risiko dan mendukung inovasi (Bronson, 2016).

Pada sektor manufaktur, *big data* dapat meningkatkan kualitas produk dan keselamatan dengan memanfaatkan sensor-sensor pada *assembly line* atau lini produksi sehingga jika terjadi variabilitas produksi, dapat terdeteksi serta juga jika terdapat perilaku membahayakan dari pekerja (Delgado, 2017). *Big data* dapat memberikan pandangan menyeluruh atau *bird's eye view* dari proses produksi dengan menghasilkan informasi mulai dari efisiensi mesin dan tingkat pengembalian pelanggan sehingga dapat meningkatkan kualitas produk maupun penurunan biaya (Caldwell, 2015).

Pada sektor publik, *The Home Office Child Abuse Image Database* di Inggris telah memanfaatkan *big data* dalam proses investigasi pelaku penyalahgunaan atau kriminalitas pada anak dengan memanfaatkan data CCTV, dimana yang awalnya membutuhkan waktu 3 hari untuk melakukan peninjauan terhadap hasil CCTV, sekarang hanya menjadi 1 jam dengan menggunakan *big data*, sehingga lebih cepat dan murah (Almeida, 2018).

Pada sektor pertambangan, *big data* telah diimplementasi dengan mengumpulkan data dari proses operasional dengan tujuan untuk deteksi masalah dan pengendalian (Matthew, 2016). *Big data analytics* juga digunakan untuk melakukan *predictive maintenance* dan telah terbukti menurunkan *maintanance cost* hingga 13% (Choudhry, 2016).

Di Indonesia, penelitian yang membahas mengenai implementasi *big data* yaitu (Emyana, 2016) pada sektor publik terhadap empat entitas pemerintah yaitu LKPP, Pemerintah Kota Bandung, Dirjen Pajak, dan BIG. Untuk LKPP dan BIG akan dibahas dalam II.4.2.2 karena termasuk *stakeholders* industri konstruksi Indonesia dan untuk penjelasan Pemerintah Kota Bandung dan Dirjen Pajak sebagai berikut:

1. Pemerintah Kota Bandung (Pemkot Bandung): Memiliki *digital command center* yang mulai dibangun pada tahun 2015 bernama *Jabar Command Center*. Di dalamnya terdapat informasi dan laporan kinerja Satuan Kerja Perangkat Daerah (SKPD). Terdapat *monitoring CCTV* untuk memantau dan mengatasi kejadian secara cepat misalnya kemacetan atau kebakaran. Selain itu juga memiliki fitur *real-time global positioning system* (GPS) dari kendaraan dinas seperti mobil pemadam kebakaran dan ambulans. Terakhir, terdapat *panic button* untuk pertolongan dalam kondisi darurat dari *end users* serta *social media analytics* berdasarkan *twitter* untuk melihat topik terhangat yang dibahas masyarakat Bandung;
2. Direktorat Jenderal Pajak (Kementerian Keuangan): Memiliki teknologi informasi untuk memerangi kecurangan pajak menggunakan data rekening tabungan. Dirjen Pajak memanfaatkan teknologi Hadoop sehingga memungkinkan integrasi data dari berbagai *personal computer* (PC). Implementasinya, Dirjen Pajak mampu melihat berapa banyak kasus penerbitan faktur tidak berdasarkan transaksi sebenarnya maupun penerbitan faktur ganda.

Penelitian lain yang dilakukan oleh (Suratnoaji, 2018) untuk melihat pengaruh atau kekuatan dari calon presiden Republik Indonesia di tahun 2019 berdasarkan opini publik dengan menggunakan Twitter melalui “*replies*”,

“*hashtag*”, “*retweet*” yang membentuk jaringan atau *network*. Penelitian oleh (Nafah, 2021) membahas mengenai penggunaan *big data* yaitu *Google Trends* dalam mengetahui negara apa saja yang tertarik melakukan pencarian kata kunci pariwisata serta destinasi apa saja yang paling banyak dicari oleh wisatawan mancanegara sehingga hasilnya dapat digunakan untuk pengambilan keputusan oleh pemerintah dalam kegiatan promosi dan pengembangan. Penelitian oleh (Suratnoaji, 2020) menjelaskan mengenai penggunaan *social media analytics* dalam mengidentifikasi opini publik terhadap kebijakan Pembatasan Sosial Berskala Besar (PSBB), caranya menggunakan Twitter serta menentukan *keyword* untuk mengkategorikan opini publik. Penelitian oleh (Nugroho, 2019) menjelaskan implementasi *big data* dalam sistem penerimaan calon pegawai negeri sipil (CPNS) yang mengambil data dari sistem kependudukan untuk memverifikasi NIK dari peserta yang mengikuti tes tersebut. Penelitian oleh (Nursiyono, 2021) membahas mengenai penentuan harga tanah di Indonesia menggunakan *big data* dengan menggunakan variabel independen yaitu luas tanah dan lokasi tanah. Penelitian oleh (Hambalah, 2019) membahas bahwa klub bola di Indonesia sudah mulai mengumpulkan data penggemar seperti nama lengkap, tanggal lahir, alamat, dan lain-lain, dengan tujuan misalnya mengirimkan surat ucapan ulang tahun secara personal, sehingga dapat menambah minat beli penggemar terhadap produk yang ditawarkan.

#### **II.4.2 Potensi dan Implementasi *Big Data* pada Sektor Konstruksi**

Identifikasi potensi dan implementasi *big data* pada sektor konstruksi dibatasi hanya untuk literatur penelitian yang terpublikasi untuk lingkup luar negeri. Sedangkan untuk implementasi *big data* dengan lingkup dalam negeri atau Indonesia, karena keterbatasan literatur maka juga mengacu kepada sumber lain seperti *website* resmi ataupun dokumen resmi entitas terkait.

#### II.4.2.1 Potensi *Big Data* pada Sektor Konstruksi

*Big data* dapat dimanfaatkan untuk fungsi manajemen proyek (*monitoring* dan pengambilan keputusan), keselamatan (perilaku pekerja dan tempat kerja), manajemen energi (pengendalian konsumsi energi pada tahap *operation*), dan manajemen sumber daya (efisiensi sumber daya) (Ismail, 2018). Penelitian oleh (Hatoum, 2020) membahas potensi implementasi dari *big data* berdasarkan *project life cycle* sebagai berikut:

1. Tahap perencanaan: Data yang dibutuhkan adalah data geospasial sehingga dapat memberikan informasi menyeluruh tentang lokasi proyek, infrastruktur sekitar, ruang publik, dan sumber daya (Sorensen, 2016). Jika terintegrasi secara lebih lanjut maka dapat mengetahui proyek lainnya yang sedang berjalan di sekitar wilayah tersebut. Pada tahap ini *big data* memungkinkan melakukan simulasi dan melihat konsekuensinya dari berbagai perencanaan sehingga menjadi *input* yang tepat untuk tahap desain atau perancangan (Loyola, 2018). Potensi lain seperti evaluasi harga tender berdasarkan *input* data proyek terdahulu mencakup informasi proyek umum, metode konstruksi, dan lain-lain sehingga dapat memberikan kisaran harga proyek secara akurat berdasarkan hasil perencanaan dan perancangan. Tujuannya dapat untuk mengestimasi keuntungan yang akan diperoleh kontraktor (Zhang, 2015). Selain itu, manfaat *big data* juga dapat digunakan oleh *owner* dalam memilih kontraktor, dimana jika terwujudnya transparansi informasi, maka perusahaan kontraktor dapat dinilai menggunakan *big data* berdasarkan kondisi finansial, kemampuan teknis, kemampuan manajemen, kesehatan dan keselamatan kerja (K3), dan reputasi (Taylan, 2017);
2. Tahap perancangan: *Big data analytics* yang diintegrasikan dengan *building information modeling* (BIM) dapat memungkinkan saran desain yang optimal serta menentukan risiko yang mungkin terjadi dari hasil desain bangunan (Tejjiy, 2022). *Input* lebih luas mencakup masukkan dari *stakeholder*, data lingkungan, catatan komunikasi digital, dan lain-lain

- dapat menghasilkan desain bangunan yang sesuai kebutuhan maupun keinginan *stakeholder* (Proforma, 2019);
3. Tahap konstruksi: *Big data* dapat membantu melihat kemajuan proyek secara *real-time* atau *near real-time* menggunakan *input* seperti gambar, video, dan disesuaikan dengan *progress* yang diakui (Hatoum, 2020). Untuk kebutuhan *monitoring* juga dapat diakomodasi dengan *input* data seperti status peralatan sehingga dapat dinilai efisiensi penggunaan alat (Akhavian, 2015). Informasi yang diperoleh dari hasil dokumentasi atau *output* misalnya peralatan, mesin, perangkat komunikasi, dan lain-lain dapat dijadikan acuan dalam menilai apakah pekerja konstruksi menggunakan waktu dan bekerja secara efektif atau tidak (Proforma, 2019). Produktivitas peralatan dapat diukur menggunakan *photogrammetry* dan *video analysis*, kemudian juga dapat ditentukan faktor performa dari lokasi proyek tersebut. *Photogrammetry* dapat digunakan untuk menentukan volume dari galian tanah, sedangkan *video analysis* dapat digunakan untuk menentukan waktu gali maupun waktu *idle* (Bugler, 2014);
  4. Tahap *operation and maintenance*: Dengan menggunakan sensor yang ditempatkan pada elemen-elemen bangunan tertentu, maka komponen bangunan tersebut dapat pantau dan dipelihara dengan mengetahui kondisinya. Sensor tersebut harus diintegrasikan dengan perangkat *internet of things* (IOT) sehingga dapat memberikan *alarm* jika terdapat elemen bangunan yang harus dilakukan *maintenance* (Chen, 2015);
  5. Tahap demolisi: Melakukan simulasi rencana demolisi terhadap polusi udara dan suara yang dihasilkan (Chen, 2018). Selain itu juga bisa mengoptimalkan *waste management* seperti pembuangan jenis material, rute transportasi, hingga pendapatan yang diperoleh dari penjualan *waste* (Yuan, 2013).

Terdapat penelitian di Indonesia oleh (Dewandaru, 2020) yang masih dalam tahap perancangan, yaitu *big data* jalan dan jembatan di Kementerian PUPR yang bernama *Indonesian Road Data Center Operation* (IRODCO). IRODCO mewujudkan integrasi data di bidang jalan yang awalnya dimiliki oleh

instansi-instansi terkait. Penggerjaan ini merupakan kolaborasi dengan Negara Korea melalui *Korea International Cooperation Agency. Progress* dari IRODCO jika mengacu kepada penelitian tersebut baru sampai pengembangan arsitektur dan alur *big data* dalam IRODCO.

#### **II.4.2.2 Implementasi *Big Data* pada Sektor Konstruksi**

Pada proyek *metro construction* di China, *big data* digunakan untuk menganalisis perilaku pekerja khususnya mengidentifikasi perilaku tidak aman. *Monitoring* dilakukan menggunakan CCTV dan diintegrasikan dengan aplikasi *smartphone*. Selain itu juga dapat digunakan untuk analisis faktor-faktor yang mempengaruhi perilaku tersebut kemudian disesuaikan dengan *job hazard analysis* (Guo, 2015).

Aplikasi seperti Sefaira maupun Autodesk mampu melakukan *interative design* khususnya untuk efisiensi energi. Contoh lain, perusahaan CallisonRTKL telah mengembangkan *software* yang mampu menyimpan data proyek seperti total luas bangunan, rasio luas lantai, persen ruang hijau di lokasi, persen atap hijau, luas muka bangunan, dan lain-lain untuk menganalisis kebutuhan bukaan bangunan, dan performa pencahayaan alami (Barista, 2014).

Perusahaan *Skidmore, Owings, and Merrill* telah membuat *database* terkait informasi proyek terdahulu untuk nantinya dapat menjadi *input* penggerjaan proyek selanjutnya. Misalnya jika desain tertentu berhasil pada suatu proyek, maka jika spesifikasi proyek barunya kurang lebih sama, dapat memanfaatkan informasi tersebut untuk digunakan pada desain yang baru. Tujuannya adalah untuk mengurangi kemungkinan *rework* serta menghemat waktu dan biaya (Barista, 2014).

Di Hong Kong, *Hong Kong Environmental Protection Departement* (HKEPD) mengumpulkan data untuk setiap *construction waste* yang diterima di fasilitas. Data tersebut digunakan untuk melihat performa proyek berdasarkan *waste generation rate* yaitu perbandingan antara kuantitas *waste* dalam ton

terhadap nilai kontrak. *Output* yang diperoleh yaitu bahwa performa *waste management* pada proyek swasta lebih baik dibandingkan dengan proyek pemerintah. Hal ini tentunya bisa menjadi *insight* dalam penetapan kebijakan serta fokus perhatian dari pemerintah (Lu, 2015).

Di Indonesia, hanya satu penelitian yang ditemukan membahas mengenai implementasi *big data* pada sektor konstruksi yaitu (Emyana, 2016) yang meninjau pada sektor publik terhadap empat entitas pemerintah yang dua diantaranya termasuk *stakeholders* industri konstruksi Indonesia yaitu LKPP dan BIG sebagai berikut:

1. LKPP: Memiliki Layanan Pengadaan Secara Elektronik (LPSE) dan sudah 90% terintegrasi dengan *cloud* untuk memastikan *database* dapat di-*backup*. Antara tahun 2014-2016 LKPP bekerja sama dengan pihak ketiga untuk membuat *software* SPLUNK yang bertujuan sebagai *alert system* untuk *warning*, *security*, dan *capacity overlead*, dan menyediakan data informasi perkembangan pengadaan setiap LPSE. *Progress*-nya masih sampai tahap investasi kepada infrastruktur;
2. BIG: Memiliki kebijakan satu peta atau *one map policy* yang didasarkan oleh Undang-Undang Nomor 4 Tahun 2011 tentang Jaringan Informasi Geospasial Nasional (JIGN). Dengan demikian, data yang dihasilkan akan berukuran sangat besar karena mencakup data mengenai posisi suatu objek atau kejadian yang berada di bawah, pada, atau di atas permukaan bumi.

Implementasi *big data* pada sektor konstruksi di Indonesia dapat dilihat melalui sumber *website* entitas resmi maupun publikasi resmi entitas terkait. Penjabaran dibatasi hanya untuk kedua entitas yaitu BPS dan DJBK dengan literatur yang dapat diakses publik secara terbatas.

#### **II.4.2.2.1 Badan Pusat Statistik (BPS)**

Implementasi *big data* dilakukan oleh BPS dalam bentuk sistem informasi dikenal dengan istilah DATAin dan juga *dashboard* yang dapat diakses publik

pada laman [bigdata.bps.go.id](http://bigdata.bps.go.id). Semua penjelasan mengenai fitur dari implementasi *big data* BPS diperoleh dari laman tersebut. Untuk mengakses *dashboard big data* harus menggunakan *virtual private network* atau *VPN* untuk sebagian tampilan *dashboard*. Tampilannya sebagai berikut:

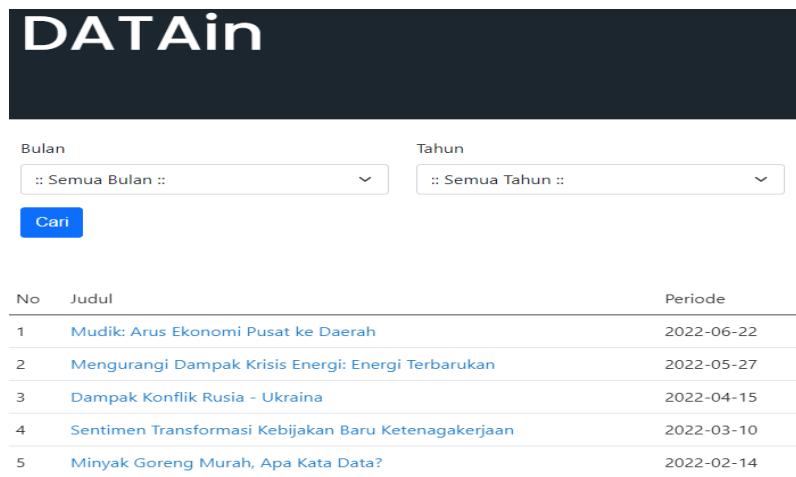


**Gambar II. 3 Dashboard Big Data BPS (Sumber: [bigdata.bps.go.id](http://bigdata.bps.go.id))**

Untuk informasi lain terkait *big data* BPS dapat dikategorikan menjadi dua, yaitu yang membutuhkan *VPN* dan yang tidak membutuhkan *VPN* sebagai berikut:

1. Membutuhkan *VPN*: *big data* perdagangan elektronik, *big data* perhotelan, Google mobilitas, *big data* penerbangan, dan *big data* lowongan pekerjaan;
2. Tidak membutuhkan *VPN*: *big data* penerbangan *near real-time*, *big data* harga *near real-time*, dan *big data* perhotelan *near real-time*.

Selain itu juga terdapat pulikasi yang disebut dengan istilah DATAin, dimana di dalam publikasi tersebut terdapat hasil pengolahan data dari *big data*. Publikasi dari DATAin mencakup berbagai sektor secara luas. Tampilan pada DATAin sebagai berikut:



The screenshot shows a search interface with two dropdown menus: 'Bulan' (Month) and 'Tahun' (Year), both set to ':: Semua Bulan ::' and ':: Semua Tahun ::' respectively. Below the dropdowns is a blue 'Cari' (Search) button. Underneath the search bar is a table with five rows of publication data. The table has three columns: 'No' (Number), 'Judul' (Title), and 'Periode' (Period). The titles are: 'Mudik: Arus Ekonomi Pusat ke Daerah' (Period: 2022-06-22), 'Mengurangi Dampak Krisis Energi: Energi Terbarukan' (Period: 2022-05-27), 'Dampak Konflik Rusia - Ukraina' (Period: 2022-04-15), 'Sentimen Transformasi Kebijakan Baru Ketenagakerjaan' (Period: 2022-03-10), and 'Minyak Goreng Murah, Apa Kata Data?' (Period: 2022-02-14).

No	Judul	Periode
1	<a href="#">Mudik: Arus Ekonomi Pusat ke Daerah</a>	2022-06-22
2	<a href="#">Mengurangi Dampak Krisis Energi: Energi Terbarukan</a>	2022-05-27
3	<a href="#">Dampak Konflik Rusia - Ukraina</a>	2022-04-15
4	<a href="#">Sentimen Transformasi Kebijakan Baru Ketenagakerjaan</a>	2022-03-10
5	<a href="#">Minyak Goreng Murah, Apa Kata Data?</a>	2022-02-14

**Gambar II. 4 DATAin BPS (Sumber: [bigdata.bps.go.id](#))**

Misalnya jika membuka nomor pertama dari gambar II.4, maka terdapat publikasi yang di dalamnya disajikan grafik hasil dari *big data* misalnya “jumlah keberangkatan penerbangan dari lima Bandara Internasional di Indonesia”, “mobilitas harian di pusat transportasi umum dan tempat perdagangan retail rekreasi”, “tingkat penghunian kamar berdasarkan platform pemesanan akomodasi *online*”, dan lain-lain. Sejauh ini, baru terdapat lima publikasi saja yang dimulai dari bulan Februari tahun 2022 hingga bulan Juni tahun 2022 dengan frekuensi publikasi adalah satu bulan sekali.

#### **II.4.2.2.2 Direktorat Jenderal Bina Konstruksi (DJBK)**

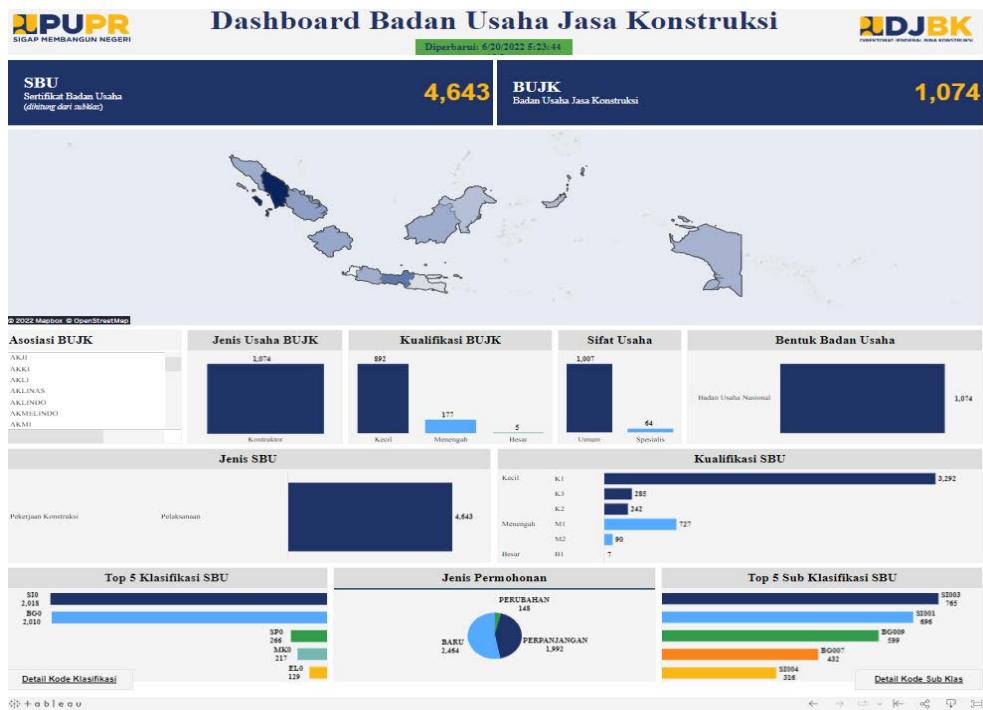
Implementasi *big data* yang dilakukan oleh DJBK dalam bentuk sistem informasi yang dikenal dengan istilah SIJK Terintegrasi. SIJK telah ditetapkan dalam Rancangan Peraturan Perundang-Undangan Cipta Kerja Sektor Jasa Konstruksi tanggal 5 Oktober tahun 2020 dimana dalam Pasal 6 tertulis “pencatatan pengalaman badan usaha melalui sistem informasi jasa konstruksi terintegrasi dan pencatatan pengalaman profesional tenaga kerja melalui sistem informasi jasa konstruksi terintegrasi.” Berdasarkan (PUPR, 2020), uji coba secara terbatas telah dilakukan untuk SIJK terkait badan usaha dan tenaga ahli jasa konstruksi pada tanggal 1 Oktober 2020.

Informasi yang bisa diakses oleh publik merupakan data agregat atau gabungan, bukan data individu. Untuk bagian tenaga kerja konstruksi, data tersebut mencakup jumlah tenaga kerja konstruksi mencakup tenaga ahli, tenaga asing, dan tenaga terampil beserta jumlah Sertifikat Keahlian (SKA) dan Sertifikat Keterampilan (SKT) dan detail mengenai sub bidang SKA seperti ahli teknik bangunan, ahli teknik jalan, dan lain-lain, serta detail mengenai sub bidang SKT seperti tukang pasang batu, pelaksana bangunan gedung, juru gambar, dan lain-lain. Tampilan yang dapat diakses oleh publik berupa *dashboard* tenaga kerja konstruksi yang terdapat dalam laman Direktorat Jenderal Bina Konstruksi sebagai berikut:



Gambar II. 5 *Dashboard* Tenaga Kerja Konstruksi (Sumber: [binakonstruksi.pu.go.id](http://binakonstruksi.pu.go.id))

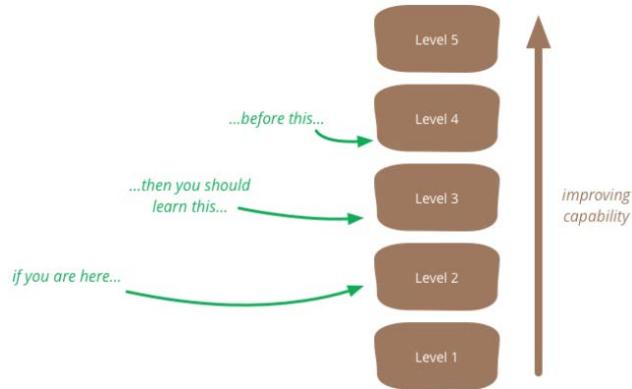
Untuk informasi terkait badan usaha jasa konstruksi mencakup jumlah Badan Usaha Jasa Konstruksi (BUJK), jumlah Sertifikat Badan Usaha (SBU), jenis usaha BUJK, kualifikasi BUJK, bentuk badan usaha, jenis SBU, kualifikasi SBU, dan lain-lain yang termasuk data agregat atau gabungan. Tampilan yang dapat diakses oleh publik berupa *dashboard* BUJK yang terdapat dalam laman Direktorat Jenderal Bina Konstruksi sebagai berikut:



Gambar II. 6 Dashboard BUJK (Sumber: binakonstruksi.pu.go.id)

## II.5 Big Data Maturity Model

Definisi *maturity model* menurut (Indeed, 2021) merupakan alat yang digunakan oleh bisnis untuk mengukur seberapa baik bisnis atau proyek yang sedang dikerjakan dan menunjukkan seberapa mampu mereka melakukan perbaikan yang berkelanjutan. Umumnya *maturity model* digunakan untuk mengukur parameter kualitatif dan bertujuan sebagai bahan evaluasi bagi perusahaan dalam melihat apakah perusahaannya semakin matang atau tidak. *Maturity model* dapat membantu SDM dalam organisasi untuk mengetahui kapabilitas apa yang dibutuhkan untuk meningkatkan performanya (Fowler, 2014). Ilustrasinya sebagai berikut:

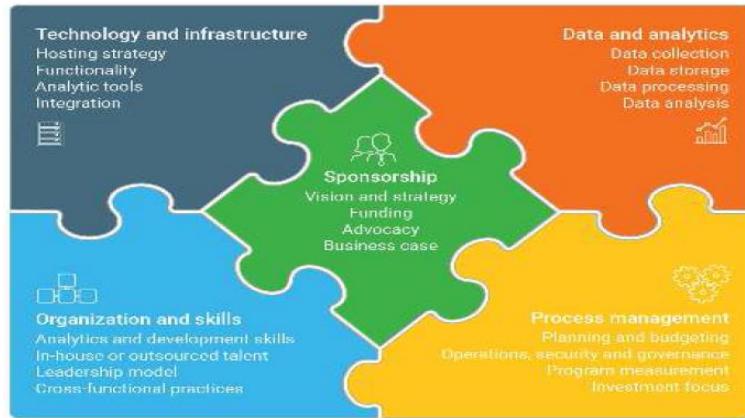


**Gambar II. 7 Konsep *Maturity Model* (Fowler, 2014)**

Konsep pentingnya adalah hasil sesungguhnya yang diperoleh dari menggunakan *maturity model* bukan pada level atau tingkat berapa melainkan apa yang harus dikerjakan untuk terus berkembang (Fowler, 2014). Dengan melakukan implementasi *maturity model*, organisasi dapat mengetahui tolok ukur yang harus dicapai serta untuk mengukur peningkatan (Good, 2020).

### **II.5.1 Hortonworks Big Data Maturity Model**

*Maturity model* ini dikembangkan oleh perusahaan Hortonworks yang merupakan anak perusahaan Cloudera. Melansir dari website Cloudera, Hortonworks merupakan perusahaan yang bergerak dalam menyelesaikan permasalahan *big data* sebagai penyedia jasa *open enterprise hadoop*. Keseluruhan informasi mengenai *Hortonwoks Big Data Maturity Model* diperoleh dari *Hortonworks White Paper* yang dipublikasi pada tahun 2016 dan digagas oleh (Dhanuka, 2016) sebagai *managing director* dari Hortonworks. *Maturity model* ini meninjau lima aspek yang bersifat holistik untuk suatu organisasi yaitu *sponsorship, data and analytics, technology and infrastructure, organization and skills, dan process management*. Masing-masing aspek terdiri dari empat tingkatan mulai dari *aware, exploring, optimizing, dan transforming*. Penjelasan dari setiap aspek dan tingkatan sebagai berikut:



**Gambar II. 8 Hortonworks Big Data Maturity Model (Dhanuka, 2016)**

### ***Sponsorship***

Aspek ini didasarkan oleh seberapa strategis visi *big data* dibuat dan seberapa dalam organisasi telah memvalidasi dan bergerak menuju visi tersebut. Aspek ini memiliki empat komponen yang harus diperhatikan sebagai berikut:

1. Visi dan strategi: memiliki satu visi dan strategi yang padu dan berasal dari tingkatan tertinggi dalam perusahaan atau organisasi. Pada tingkat kematangan yang rendah, visi dan strategi hanya berada pada divisi yang bersangkutan dengan *big data*, tetapi untuk mencapai tingkat yang lebih tinggi diperlukan keselarasan dari semua *C-level*.
2. Pendanaan: Pada tingkat kematangan yang rendah, program *big data* dilakukan secara tidak formal sehingga tidak memiliki anggaran yang jelas dan bersifat *one-off project*. Sedangkan untuk mencapai tingkat yang lebih tinggi, organisasi atau perusahaan perlu mendanai program *big data* secara berkala.
3. Dukungan organisasi: proses transformasi *big data* membutuhkan perubahan budaya. Pada tingkat kematangan yang rendah, hanya divisi *information and technology* (IT) yang memiliki budaya *big data* misalnya perencanaan berbasis data. Hal yang harus diketahui bahwa dengan menggunakan *big data* memungkinkan munculnya ide-ide baru melalui pemanfaatan data secara maksimum.

4. *Business case*: Pada tingkat kematangan yang rendah, fokus dari menggunakan *big data* yaitu menghemat biaya dalam penyimpanan data karena teknologi *big data* menggunakan Apache Hadoop. Namun pada tingkat kematangan yang lebih tinggi, penggunaan *big data* terkait dengan analitis dalam menghasilkan *big data insights*.

### ***Data and Analytics/ Data Management***

Aspek ini mencakup empat tahap alur atau *flow* dari data sebagai berikut:

1. Pengumpulan data: Pada tingkat kematangan yang rendah, target dan cara pengumpulan data yaitu hanya untuk data terstruktur dan mencakup banyak upaya yang bersifat manual. Pada tingkat kematangan yang lebih tinggi, data tidak pernah dibuang dan pengumpulan data dilakukan secara otomatis.
2. Penyimpanan data: Pada tingkat kematangan yang lebih tinggi, organisasi menyadari pentingnya menyimpan seluruh data meskipun nilai data tersebut tidak dapat ditentukan dalam jangka waktu dekat.
3. Pemrosesan data: Pada tingkat kematangan yang rendah, organisasi memproses hanya *structured data* secara manual yang sebenarnya berakibat kepada biaya yang mahal secara volume. Pada tingkat kematangan yang lebih tinggi, adanya *metadata* dalam melakukan pemrosesan data dan proses ini dilakukan tanpa henti dalam menghasilkan *real-time data*.
4. Analisis data: Pada tingkat kematangan yang lebih tinggi, analisis data dilakukan untuk menghasilkan skenario “*what if*” maupun juga pemodelan prediktif dan analisis *real-time*.

### ***Technology and Infrastructure***

Aspek ini didasarkan dari penggunaan Apache Hadoop dan tingkat integrasi dari *tools big data* dalam organisasi. Aspek ini memiliki empat komponen yang harus diperhatikan sebagai berikut:

1. Strategi *hosting*: Komponen ini berfokus kepada upaya dalam menjaga keamanan data. Pada tahap awal umumnya organisasi menggunakan *hosting* lokal untuk tujuan keamanan namun seiring perkembangan volume data maka harus beralih kepada *hybrid cloud*.
2. Fungsionalitas: Komponen ini mengacu kepada kemampuan infrastruktur organisasi dalam mendukung pekerjaan yang dibutuhkan. Pada tingkat kematangan yang tinggi, infrastruktur organisasi harus dapat mendukung pemrosesan *real-time data*.
3. *Analytic tools*: Pada tingkat kematangan yang rendah, investasi pada *analytics tools* berisfat *project-specific*. Pada tingkat kematangan yang lebih tinggi, tujuan penggunaan *analytics tools* akan menyatu dalam organisasi atau perusahaan misalnya untuk setiap pengambilan keputusan penting yang disebut dengan *centralized analytics*.
4. Integrasi: Pada tingkat kematangan yang rendah, organisasi akan menghabiskan banyak waktu untuk menyesuaikan teknologi dan infrastruktur untuk kebutuhan bisnis yang berubah. Pada tingkat kematangan yang lebih tinggi sudah menciptakan *centralized data service*.

### ***Organization and Skills***

Aspek ini berfokus kepada pendekatan perusahaan baik dalam melakukan *outsourcing*, manajemen, kepemimpinan, dan kerja sama antar divisi perusahaan. Aspek ini memiliki empat komponen yang harus diperhatikan sebagai berikut:

1. Kemampuan SDM: Mengacu kepada kemampuan analitik. Pada tingkat kematangan yang rendah, hanya terdapat segelintir *data analysts* yang memiliki kemampuan atau keahlian terkait *big data*. Pada tingkat kematangan yang lebih tinggi, telah dilakukan pelatihan formal di seluruh perusahaan untuk memberikan keahlian analitik di seluruh *business role*.
2. Sumber SDM: Komponen ini didasarkan bahwa ketersediaan SDM sangat berpengaruh terhadap peningkatan kematangan *big data*. Tingkat kematangan organisasi tidak memiliki hubungan dengan apakah sumber SDM diperoleh dari *in-house* atau *outsourced*.

3. Model kepemimpinan: Transformasi *big data* tidak sekadar membutuhkan *centralized data*, melainkan yang penting juga adalah *centralized leadership*. Model kepemimpinan ini memungkinkan penurunan budaya dan perubahan prosedur sehingga mengarah kepada transformasi organisasi.
4. Kolaborasi: Pada tingkat kematangan yang rendah, kolaborasi hanya dilakukan untuk optimasi pengumpulan data dan standardisasi *metadata*. Pada tingkat kematangan yang lebih tinggi, kolaborasi akan lebih luas hingga mencakup keberadaan program *big data* masing-masing divisi melalui *cross-functional business teams*.

### ***Process Management***

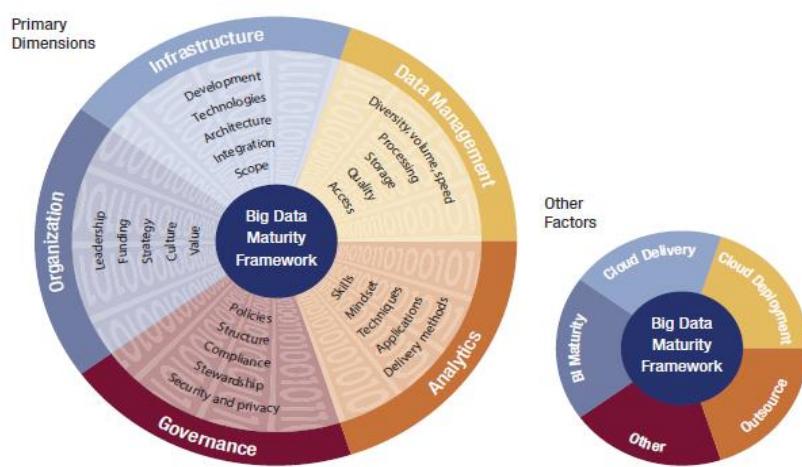
Aspek ini berfokus kepada bagaimana organisasi mengelola dan mengevaluasi program *big data*. Aspek ini memiliki empat komponen yang harus diperhatikan sebagai berikut:

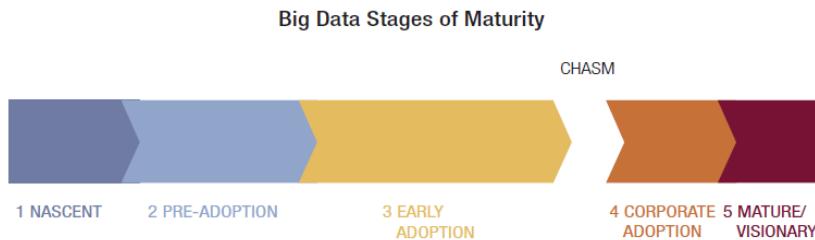
1. Perencanaan dan penganggaran: Pada tingkat kematangan yang rendah, perencanaan dan penganggaran dilakukan secara *ad hoc*, yaitu ketika muncul ide (tanpa perencanaan awal) atau terjadi permasalahan. Pada tingkat kematangan yang lebih tinggi, proses ini terintegrasi dengan *roadmap* perusahaan.
2. Operasional, keamanan, dan tata kelola: Pada tingkat kematangan kedua, fokus dari komponen ini adalah keamanan. Selanjutnya pada tingkat kematangan ketiga terkait dengan operasional dan tata kelola. Pada tingkat kematangan keempat, organisasi secara ketat memberlakukan *enterprise grade security* dan rencana pemulihan dengan *backup* data.
3. Pengukuran program: Tujuannya adalah optimasi program *big data*. Jika terdapat program yang gagal, harus terdapat budaya “*test and learn*” dalam organisasi. Kegagalan harus dilihat sebagai pembelajaran dan keberhasilan menjadi aturan bisnis baru yang menguntungkan. Pengukuran program juga termasuk perhitungan *benefit/cost ratio* untuk mengevaluasi ROI dari program *big data*.

4. Fokus investasi: Pada tingkat kematangan yang rendah, investasi masih berada pada lingkup Apache Hadoop dan infrastruktur *big data*. Pada tingkat kematangan yang lebih tinggi, organisasi mampu memperluas investasi seperti untuk pengembangan *existing* dan penciptaan model bisnis baru.

### II.5.2 *Transforming Data with Intelligence* (TDWI) *Big Data Maturity Model*

Berdasarkan TDWI.org, TDWI merupakan perusahaan yang didirikan pada tahun 1995 dan menyediakan layanan *big data analytics* untuk mewujudkan pengambilan keputusan berdasarkan data dan meningkatkan performa perusahaan. TDWI juga mengembangkan *TDWI Big Data Maturity Model* yang digagas oleh (Halper, 2014). Oleh karena itu, seluruh informasi mengenai model ini diperoleh dari publikasi tersebut. *Maturity model* ini meninjau lima aspek yang bersifat holistik untuk suatu organisasi yaitu *organization*, *infrastructure*, *data management*, *analytics*, dan *governance*. Masing-masing aspek terdiri dari lima tingkatan mulai dari *nascent*, *pre-adoption*, *early adoption*, *corporate adoption*, dan *mature*. Penjelasan dari setiap aspek dan tingkatan sebagai berikut:





Gambar II. 9 TDWI *Big Data Maturity Model* (Halper, 2014)

### ***Nascent***

Pada tahap ini, organisasi memiliki kesadaran yang rendah mengenai nilai dari *big data* dan masih belum mendapatkan dukungan dari pimpinan organisasi, sehingga hanya dirasakan oleh segelintir orang di posisi bawah hierarki organisasi. Tahap *nascent* masih belum memulai pengembangan *big data*-nya. Adapun kematangan dari aspek-aspeknya sebagai berikut:

1. Manajemen data: Data yang dikumpulkan tidak memiliki standar penamaan dan belum memiliki strategi manajemen data. Praktiknya, data bisa berguna mungkin dapat dibuang karena perusahaan belum mampu melihat nilai dari data tersebut. Fokusnya hanya kepada data yang dapat memberikan nilai secara langsung;
2. Infrastruktur: Organisasi pada tahap ini masih belum memahami kebutuhan untuk memiliki infrastruktur untuk mendukung karakteristik *big data* yaitu *volume*, *variety*, keamanan, dan lain-lain;
3. Analitis: Organisasi pada tahap ini mungkin atau masih belum memiliki tim analitis, namun biasanya jika ada, tim analitis tersebut ada di dalam divisi lain seperti *marketing*. Selain itu juga tipe data yang diolah hanya terbatas pada *structured data*.

### ***Pre-Adoption***

Pada tahap ini, SDM dalam organisasi sudah mulai diminta untuk menghadiri seminar atau konferensi serta membaca mengenai topik *big data*. Organisasi mungkin sudah berinvestasi di teknologi seperti Hadoop. Perusahaan tahu bahwa dalam beberapa waktu ke depan akan mulai mengimplementasikan *big data*. Adapun kematangan dari aspek-aspeknya sebagai berikut:

1. Manajemen data: Organisasi umumnya sudah memiliki *metadata* dan mulai mengidentifikasi dan mengumpulkan *big data* dari berbagai sumber;
2. Infrastruktur: Organisasi mungkin sudah mencoba teknologi Hadoop atau teknologi *big data* lainnya. Infrastrukturnya hanya dapat menangani penggunaan dasar dan secara spesifik untuk satu departemen perusahaan saja;
3. Analitis: Organisasi mulai mengarah kepada *big data analytics* yang bersifat *advanced*. Namun, kemampuan terkait *big data* hanya terdapat pada beberapa orang saja yang dapat melihat nilai dari implementasi *big data*;
4. Organisasi: Organisasi baru memulai inisiatif *big data* ditandai dengan dukungan *chief information officer* atau mungkin beberapa eksekutif lainnya. Kebutuhan mengenai implementasi *big data* berdasarkan permasalahan bisnis sudah mulai dicari. Pada tahap ini, organisasi sudah menggunakan *big data* tetapi budaya *data driven* masih belum tertanam dalam organisasi;
5. Tata kelola: Organisasi mungkin sudah memiliki komite *big data* yang mengawasi program *big data* dengan adanya perwakilan dari departemen terkait untuk memberikan laporan kemajuan.

### *Early Adoption*

Pada tahap ini organisasi sudah memiliki *proof of concept* yang menjadikan program *big data* sudah dapat dijalankan. Tahap ini merupakan tahapan paling lama karena sulit untuk melewati tahapan *chasm*. Adapun kematangan dari aspek-aspeknya sebagai berikut:

1. Manajemen data: Data yang dikumpulkan sudah memiliki aspek *variety* serta memiliki standar penamaan dan manajemen penyimpanan data. Organisasi tidak secara sembarangan membuang data kecuali memang memiliki tujuan tertentu. Pada tahapan ini organisasi mulai menyadari bahwa permasalahannya adalah keamanan dan kualitas data;
2. Infrastruktur: Teknologi *big data* mungkin sudah diimplementasikan seperti Hadoop atau NoSQL namun masih belum terdapat arsitektur

- terpadu. Infrastruktur yang ada mencakup pusat data perusahaan atau penyimpanan di *cloud*;
3. Analitis: Organisasi mungkin menggunakan analisis deskriptif atau bahkan prediktif dalam program *big data*. Organisasi secara umum masih menggunakan satu jenis data namun dengan *volume* yang besar, lebih dari 10 TB *structured data*. Penggunaan analitis masih belum terintegrasi misalnya pada suatu departemen organisasi menggunakan *structured data* tetapi departemen menggunakan *unstructured data*;
  4. Organisasi: Pada tahap ini, paling tidak ada satu *C-level* yang mendukung implementasi *big data*. Organisasi secara keseluruhan sudah bersemangat untuk bergabung kepada program *big data*. Tim *big data* sudah dibentuk dan menyusun strategi dalam cakupan yang lebih luas.

### ***Chasm***

Tahapan ini menyebabkan fase *early adoption* menjadi tahapan paling lama, karena berbagai tantangan nyata mulai terlihat. Kemampuan *big data* terkait Hadoop atau analitis mungkin masih belum dimiliki SDM dalam organisasi. Konflik politik mulai muncul ketika berbagai departemen berebut wewenang atas siapa yang memiliki data dan secara khusus memiliki visinya sendiri. Organisasi mulai menyadari bahwa isu-isu lain seperti keamanan data harus menjadi pertimbangan penting. Beberapa tantangan yang dihadapi sebagai berikut:

1. Pendanaan: Kenyataannya beberapa perusahaan tidak bisa melewati tahap *prototyping* karena tidak memiliki cukup pendanaan. Umumnya membutuhkan biaya yang besar untuk bagian IT;
2. Manajemen data: Data harus dapat tersebar di seluruh organisasi yang artinya perlu adanya *information architecture*. Tujuannya, data harus koheren untuk setiap departemen yang dibutuhkan untuk melakukan *data analytics*;
3. Arsitektur: Perusahaan cenderung mencoba merancang arsitektur yang terbaik tetapi sering kali perusahaan tidak mampu mengimplementasikan arsitektur tersebut;

4. SDM: Tantangannya berada pada *skill set* terkait *big data* yaitu menggunakan teknologi *big data* seperti Hadoop atau NoSQL maupun *big data analytics*. Perusahaan memiliki opsi untuk melakukan *outsourcing* SDM maupun mengembangkan kompetensi dari SDM dalam perusahaan;
5. Budaya dan politik: Tantangan dalam budaya organisasi adalah mengubah fokus dari yang awalnya penghematan biaya menjadi keunggulan kompetitif. Tantangan politik terkait dengan kepentingan pengambilan keputusan;
6. Tata kelola: Perusahaan membutuhkan tim *big data* yang mengatur program *big data* dan bagaimana menyelaraskan dengan program besar perusahaan. Tim tersebut juga mengatur perencanaan dan penganggaran untuk masing-masing kegiatan.

### ***Corporate Adoption***

Pada tahap ini, pengguna akhir ikut terlibat dan memperoleh wawasan. Untuk internal perusahaan sudah terlihat budaya pengambilan keputusan yang berdasarkan data. Adapun kematangan dari aspek-aspeknya sebagai berikut:

1. Manajemen data: Organisasi sudah mampu mengolah *variety* dari data. *Data sharing* juga sudah dilakukan antar departemen dalam organisasi melalui kebijakan data yang telah ditetapkan. Organisasi selalu mencari sumber data baru untuk meningkatkan kemampuan analitisnya. *Data lifecycle management* juga sudah didefinisikan dan ada pihak yang mampu memonitor;
2. Infrastruktur: Arsitektur teknologi *big data* sudah dibentuk mencakup NoSQL, Hadoop, *data warehouse* atau *cloud*, dan juga termasuk jika terjadi bencana pada data sehingga perlu dilakukan rencana pemulihan data. Telah ditetapkan juga standar perusahaan bagi instalasi, konfigurasi, dan pemeliharaan infrastruktur *big data*.
3. Analitis: Pada tahap ini, organisasi mampu secara cepat mengolah data yang masuk dan dianalisis dengan cepat sesuai dengan kebutuhan organisasi. Proses *analytics* mampu secara efisien mengetahui jika kebutuhannya juga mungkin dapat digunakan kembali dalam departemen

lain perusahaan. Dengan kata lain, *analytics* sudah terintegrasi dengan proses bisnis organisasi;

4. Organisasi: Seluruh organisasi sudah menyadari bahwa *big data analytics* memberikan keunggulan kompetitif dan inovasi. Penetapan strategi bisnis dilakukan baik secara *top-down* maupun *bottom-up*;
5. Tata kelola: Perusahaan sudah berfokus untuk menjawab pertanyaan seperti “data siapa itu?” “kemana data itu akan digunakan?” “berapa lama data itu akan disimpan?” dengan kebijakan yang ditetapkan oleh tim *big data*. Program *big data* dilaksanakan sesuai penganggaran yang telah ditetapkan.

### ***Mature***

Pada tahap ini, organisasi menjalankan program *big data* menggunakan infrastruktur dan telah mengimplementasikan tata kelola data yang matang. Organisasi sudah memiliki pandangan yang visioner sehingga muncul antusias terhadap *big data*. Adapun kematangan dari aspek-aspeknya sebagai berikut:

1. Manajemen data: Organisasi mampu mengambil sumber data baru baik yang bersifat internal maupun eksternal terhadap organisasi dan kemudian melakukan analitis dengan setiap departemen mampu mengakses data secara penuh sesuai kebutuhannya masing-masing;
2. Infrastruktur: Organisasi sudah dengan memanfaatkan secara penuh Hadoop, NoSQL, serta menyelesaikan isu terkait keamanan data, rencana pemulihan data terhadap bencana, serta *monitoring* infrastruktur;
3. Analitis: Organisasi terus mengembangkan kapabilitas analitik secara kontinu. Organisasi mampu mengolah semua jenis data dan menghasilkan *real-time data* dalam membantu mengambil keputusan dan terintegrasi dengan proses bisnis. Hasil pengambilan keputusan digunakan untuk meningkatkan kualitas produk atau jasa.
4. Organisasi: Seluruh organisasi memahami bahwa peran *big data* bersifat kritis dan memberikan keunggulan kompetitif serta mendorong pola pikir kreatif. Implementasi *big data* tidak hanya sebatas memberikan wawasan

melainkan juga untuk melihat peluang bagaimana menggunakan *analytics* untuk proses bisnis yang baru.

### II.5.3 *United Nation (UN) Big Data Maturity Matrix*

*Maturity model* ini dikembangkan oleh *United Nation Global Working Group* (UNGWG) pada tahun 2020, sehingga seluruh informasi mengenai UN *big data maturity matrix* diperoleh dari (UNGWG, 2020). *Maturity model* ini terdiri dari 4 tinjauan aspek dan 4 tingkat kematangan. Untuk deskripsi singkat mengenai 4 tinjauan aspek sebagai berikut:

1. Hukum dan kebijakan: Seberapa baik hukum dan kebijakan untuk mengakses data maupun dan berbagi data. Secara spesifik terkait dengan memelihara keamanan dan kerahasiaan dari *big data*;
2. Infrastruktur: Keberadaan infrastruktur yang menunjang *big data analytics* dalam lingkungan yang aman;
3. SDM: Jumlah dari tenaga *big data* seperti *data scientist*, kemampuan, serta proses rekrutmen;
4. Aplikasi: Bagaimana *big data analytics* digunakan dalam organisasi

Untuk penjelasan mengenai setiap aspek dalam 4 tingkat kematangan sebagai berikut:

	Pre-Foundation	Foundation	Practitioner	Expert	Target
Legal & Policy Framework	Description	Description	Description	Description	Practitioner
IT Infrastructure	Description	Description	Description	Description	Practitioner
Human Resources	Description	Description	Description	Description	Practitioner
Application	Description	Description	Description	Description	Practitioner

Gambar II. 10 *United Nation Big Data Maturity Matrix* (UNGWG, 2020)

## ***Pre-Foundation***

Pada tahapan ini, organisasi masih mendiskusikan dan mempertimbangkan bagaimana memulai dalam upaya program *big data* serta strategi-strategi yang dibutuhkan untuk mewujudkan implementasi *big data*. Penjelasan setiap aspeknya sebagai berikut:

1. Hukum dan kebijakan: Organisasi mulai mempertimbangkan kebutuhan prasyarat hukum untuk pengaksesan dan pembagian data dalam organisasi, serta sanksi-sanksi dan *copyright*;
2. Infrastruktur: Organisasi mulai mempertimbangkan kebutuhan infrastruktur penunjang *big data*. Jika terdapat infrastruktur, maka masih sangat terbatas dari segi penyimpanan data atau secara *ad hoc* terhadap proyek. Pada tahap ini masih belum terdapat standar keamanan untuk *import* dan *export big data*;
3. SDM: Baru timbul sedikit kesadaran mengenai *big data analytics* dalam organisasi. Pimpinan perusahaan mulai berdiskusi dan mempertimbangkan *big data*, hingga rencana merekrut atau mengembangkan keterampilan *big data* dalam organisasi;
4. Aplikasi: *Big data* dilaksanakan oleh segerintir SDM dalam organisasi. Prosedur mengenai bagaimana mendapatkan data, proses mengolah data, dan bagaimana *sharing big data* masih belum ditetapkan tapi sedang dalam tahap penggerjaan.

## ***Foundation***

Pada tahapan ini, kepemimpinan dan strategi sedang dikembangkan dan sudah mulai terlihat sejumlah kecil proyek *big data* terkait kemungkinan penggunaannya. Penjelasan setiap aspeknya sebagai berikut:

1. Hukum dan kebijakan: Organisasi sudah mulai mengerjakan dokumen-dokumen yang diperlukan meskipun hasilnya belum selesai. Organisasi sudah memiliki pernyataan yang jelas mengenai *copyright* maupun juga jika melakukan kerja sama dengan pihak luar;

2. Infrastruktur: Penggunaan infrastruktur *big data* belum terintegrasi untuk kegiatan operasional sehari-hari melainkan hanya kepada proyek tertentu. Keamanan untuk *import* dan *export* data sudah mulai dilakukan meskipun tidak konsisten. Selain itu juga organisasi masih berdiskusi mengenai kebutuhan infrastruktur yang tersedia di luar organisasi;
3. SDM: Kesadaran *big data* dalam organisasi sudah berkembang. Mekanisme rekrutmen untuk SDM *big data* sudah ditetapkan serta telah mengidentifikasi kemampuan yang dibutuhkan. Tingkat kematangan ini mencakup sampai tahapan dimana sudah terdapat beberapa analis *big data* namun masih bereksperimen dengan penggunaan *big data*;
4. Aplikasi: *Big data* dilaksanakan dalam tim yang dibentuk untuk menjalankan program *big data*. Prosedur mengenai bagaimana mendapatkan data, proses mengolah data, dan bagaimana *sharing big data* telah dibuat tapi hanya untuk tim tersebut. Lingkup pertukaran informasi yang dihasilkan oleh *big data* masih digunakan secara spesifik oleh tim tersebut dan belum

### ***Practitioner***

Pada tahapan ini, organisasi sudah memiliki *framework big data*, SDM untuk *big data* sudah lebih terstruktur dan proyek *big data* sedang dikerjakan dan dikelola secara strategis. Penjelasan setiap aspeknya sebagai berikut:

1. Hukum dan kebijakan: Organisasi memiliki kebijakan ataupun pedoman untuk internal meskipun hukum pemerintah secara eksternal belum mewajibkan organisasi tersebut membuat peraturan. Perjanjian dengan pihak luar sudah jelas ketentuannya serta memiliki kejelasan tentang *copyright*;
2. Infrastruktur: Kapabilitas dari infrastruktur dalam organisasi sudah mampu mengolah data dengan *volume* atau *variety* yang besar. Dikatakan bahwa *volume* dan *variety* bukan menjadi hambatan bagi infrastruktur yang ada, meskipun terkadang beberapa intervensi secara manual masih terjadi khususnya untuk penyebaran data. Infrastruktur juga sudah mendukung pemulihan data jika terjadi bencana. Infrastruktur yang sama juga

- digunakan untuk seluruh organisasi. Kebutuhan untuk *import* dan *export* data masih melalui proses permintaan terpusat;
3. SDM: Mayoritas dalam organisasi sudah menyadari manfaat atau *value* yang diberikan oleh *big data analytics*. Mekanisme rekrutmen SDM untuk *big data* sudah ditetapkan serta SDM internal bertanggungjawab dengan *big data* memiliki tim. Ketika dibutuhkan, organisasi mungkin dapat memanfaatkan *outsourcing* untuk membantu pengembangan *big data* maupun dari sisi SDM-nya. Pada tahapan ini juga organisasi sudah beberapa kali melakukan pelatihan internal terkait *big data*;
  4. Aplikasi: Program *big data* dikerjakan oleh tim yang secara tetap ada dalam organisasi dan selain dari *data analyst*, juga terdapat SDM dari berbagai departemen lain dalam organisasi yang tergabung dalam tim tersebut. Solusi yang dihasilkan dari *big data* benar-benar diimplementasikan oleh organisasi dan disebarluaskan ke seluruh organisasi.

### *Expert*

Tahapan ini merupakan tingkatan yang paling matang dalam implementasi *big data*, dimana *big data* sudah menjadi bagian integral dari organisasi dan diterapkan secara luas dalam organisasi. SDM yang ada juga sudah memiliki pengetahuan, kemampuan, dan pengalaman untuk memimpin dan melaksanakan proyek *big data*. Program pelatihan, pembinaan, dan pendampingan sudah tersedia secara internal dan bahkan organisasi mampu menawarkan program tersebut kepada eksternal. Penjelasan setiap aspeknya sebagai berikut:

1. Hukum dan kebijakan: Organisasi diwajibkan oleh pemerintah untuk menetapkan hukum dan kebijakan internal organisasi. Organisasi sudah secara matang memiliki hukum dan kebijakan misalnya terkait dengan *data disclosure*;
2. Infrastruktur: Kapabilitas infrastruktur *big data* mampu melakukan proyek yang kompleks. SDM yang membutuhkan *big data* dapat secara mudah mengakses data tersebut sebagai kegiatan sehari-hari. Keamanan dan kerahasiaan data sudah tidak lagi menjadi masalah. *Data pipeline*

*framework* sudah ditetapkan sehingga tidak perlu adanya intervensi manual terkait penyebaran data dalam organisasi;

3. SDM: Budaya *big data analytics* sudah mengakar dalam organisasi dan sudah menjadi pertimbangan umum saat pengambilan keputusan dalam proyek. Organisasi menjadi semakin bergantung kepada SDM *big data* internal organisasi. Seluruh departemen dalam organisasi bekerja sama untuk meningkatkan serta mengadopsi *big data*. Organisasi mampu menawarkan keahliannya kepada organisasi lain melalui bentuk pelatihan, konsultasi, bimbingan, dan lain-lain;
4. Aplikasi: Tim *big data* sudah memiliki tempat yang permanen dalam struktur organisasi, sehingga arahan dapat dilakukan secara struktural dan sistematis serta melibatkan departemen lain dalam organisasi maupun secara eksternal. Solusi yang dihasilkan dari *big data* sudah diimplementasikan dalam organisasi. Organisasi memiliki kapabilitas dalam mengolah secara luas permasalahan *big data*. Program *big data* juga dipantau secara menerus dalam organisasi.