DAFTAR GAMBAR DAN ILUSTRASI

DAFTAR TABEL

DAFTAR ISTILAH

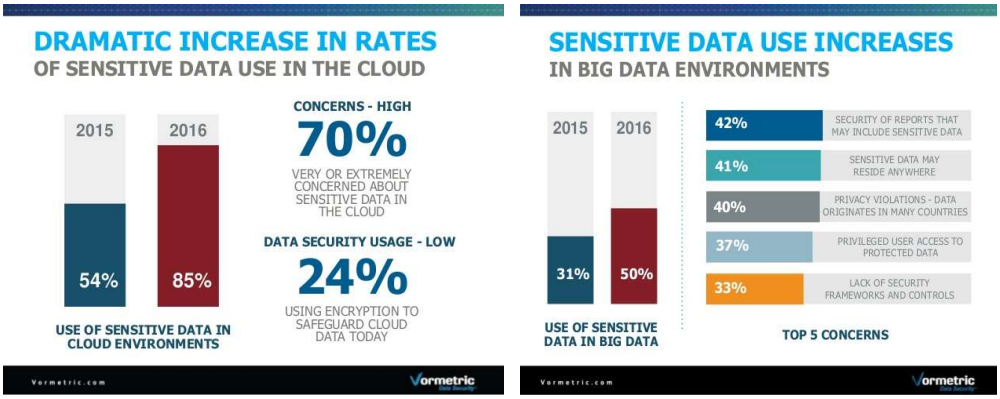
CSA	:	*Cloud Security Aliance*, merupakan organisasi non-for-profit yang mempunyai misi "mempromosikan praktik terbaik untuk memberikan jaminan dalam *cloud computing* dan untuk memberikan edukasi tentang penggunaan *cloud computing*untuk membantu mengamankan semua bentuk komputasi lainnya".
RBAC	:	*Roled Based Access Control,* merupakan pendekatan untuk membatasi pengguna yang mengakses sistem berdasarkan peran pengguna.
ACL	:	*Access Control List,* merupakan daftar akses control pengguna yang dapat mengakses sistem.
TLS	:	*Transport Layer Security,* merupakan sebuah protokol yang bertujuan untuk memberikan komunikasi terenkripsi yang aman antara *device*.
SSO	:	*Single Sign On,* merupakan metode yang memberikan sesi ke suatu akun pengguna.
LDAP	:	*Lightweight Directory Access Protocol,* merupakan sebuah protokol yang dapat memberikan mekanisme untuk mengkoneksikan, mencari, dan modifikasi ke suatu direktori.
HDFS	:	*Hadoop Distributed File System,* merupakan tempat penyimpanan pada Hadoop yang terdistribusi.
DGI	:	*Data Governance Initiative,* merupakan metode untuk mempebaiki kualitas data dengan cara menggunakan tim yang bertanggung jawab untuk *accuracy, accessbilitiy, consistency,* dan *completeness* data.

TDE	:	*Transparent Data Encryption,* merupakan teknologi yang dapat mengenkipsi data pada *database*.
KMS	:	*Key management Server,* merupakan server yang bertanggung jawab untuk mengelola *key* atau enkripsi.
IP	:	*Internet Protocol,* merupakan protokol komunikasi utama dalam rangkaian protokol internet untuk menyampaikan paket keseluruh jaringan.
AD	:	*Active Directory,* merupakan direktori *service* yang berada pada keadaan aktif,
RSA	:	*Rivest-Shamir-Adleman,* meupakan salah satu yang pertama *public key cryptosystem* dan biasa digunakan untuk mengamankan transmisi data
IT	:	*Information Technology,* merupakan pembelajaran atau penggunaan sistem untuk menyimpan, menerima dan mengirimkan informasi berbasis komputer dan telekomunikasi.
PIN	:	*Postal Index Number,* merupakan suatu kode yang biasa digunakan sebagai otentikasi.
API	:	*Application Programming Interface,* merupakan sebuah protokol yang digunakan untuk memberikan perintah antara aplikasi dengan server/service.

BAB I PENDAHULUAN

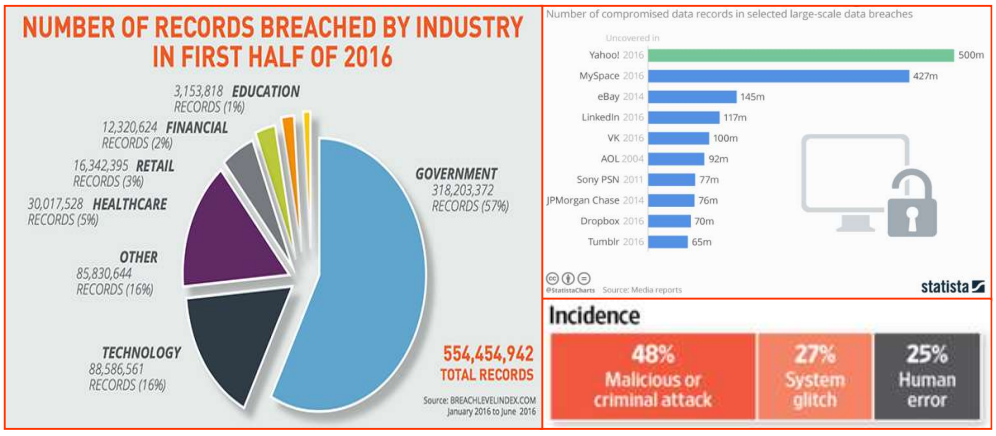
Data berkembang dengan cepat, sehingga sulit untuk menangani data dalam jumlah besar. Setiap tahun data yang dikirmkan melalui internet berkembang secara eksponensial. Pada akhir tahun 2016, hasil riset yang dilakukan Cisco bahwa lalu lintas data pada jaringan mencapai 6,6 *Zettabyte*. Tantangannya bukan hanya untuk mempercepat koneksi internet, tapi juga mengembangkan sistem perangkat lunak yang mampu menangani semua permintaan data yang besar pada waktu yang optimal (TOLE, 2013). *Big data* digunakan dalam teknologi dan bisnis modern untuk menangani jumlah data yang besar (Lamyae HBIBI, 2016). Sulit untuk mengelola informasi yang kompleks dengan menggunkan sistem *database* standar atau pada komputer pribadi, biasanya dibutuhkan sistem perangkat lunak pararel dan infrastruktur tertentu yang mampu menangani informasi yang kompleks. Sementara itu permintaan informasi yang kompleks menigkat setiap tahunnya. Informasi *streaming* secara *real-time* menjadi tantangan besar yang harus diatasi oleh perusahaan – perusahaan yang menyediakan layanan tersebut. Dengan memproses data secara *real-time* perusahaan dapat mengetahui kebutuhan pelanggan dan berguna untuk mendapatkan informasi mengenai pertumbuhan pasar kedepannya. *Big data* merupakan sebuah teknologi yang dapat memproses data dalam jumlah besar yang sangat kompleks dan beragam. *big data* juga dapat digunakan sebagai analisis prediktif, elemen ini digunakan oleh banyak perusahaan sebagai alat bantu untuk mengambil suatu keputusan.

Akan tetapi, banyaknya penggunaaan data pada *big data* khususnya pada data yang bersifat sensitif, menyebabkan keamanan pada *big data* menjadi rentan. Menurut Vormetric *data security* penggunakan data yang bersifat sensitif mengalami peningkatan yang signifikan sebesar 31% dari tahun 2015 ke 2016. Dari 24% data tersebut kurang mendapat perhatian dari aspek *security*. Sementara itu data sensitif yang ada pada big data mengalami peningkatan sebesar 19% dari tahun 2015 ke 2016.



Gambar I - 1 Presentase Penggunaan Data Sensitif (Vormetric, Inc)

Sedangkan jumlah data yang hilang berdasarkan jenis industri yang didominasi oleh governance dan teknologi pada pertengahan tahun 2016 mencapai 554.454.942 data, Yahoo merupakan perusahaan teratas yang mengalami banyaknya kehilangan data yaitu sebesar 500 juta data. 3 penyebab utama dari data yang hilang tersebut adalah Malicious or Criminal Attack, System Glitch dan Huma Error (breachlevelindex.com).



Gambar I - 2 Data Hilang Berdasarkan Industri (breachlevelindex.com)

Seiring dengan transformasi menjadi Digital Telco Company, PT Telekomunikasi Indonesia menjadikan *big data* sebagai *smart enabler platform* bisnis digitalnya. Ini berarti keberadaan *big data* akan menjadi pendorong utama terciptanya produk-produk digitalnya di Telkom. Hal ini menjadikan posisi *big data* ini rentan terhadap gangguan keamanan yang berasal dari luar, karena seperti dijelaskan sebelumnya

bahwa saat ini data telah menjadi tambang baru bagi ekonomi digital sehingga data tersebut harus dilindungi. Oleh karena itu aspek keamanan *big data* Telkom harus benar-benar diperhatikan.

Hal ini didukung oleh hasil riset dari Vormetric dan Breachlevelindex.com yang mengatakan bahwa pada tahun 2016 jumlah data yang hilang berdasarkan jenis industri didominasi oleh *governance* dan teknologi, 3 penyebab utama dari data yang hilang tersebut adalah *malicious* atau *criminal attack, system glitch* dan *human error.* berdasarkan kondisi ini maka penulis memandang perlu adanya suatu *framework security big data* untuk menjamin keamanan data secara komprehensif.

I.1 Perumusan Masalah

Berdasarkan pada latar belakang yang telah dibuat, maka rumusan permasalahan pada penelitan ini yaitu, seperti apa *framework big data security* untuk PT.Telkomunikasi Indonesia?

I.2 Tujuan Penelitian

Berdasarkan pada rumusan maslah yang telah dijabarkan sebelumnya, maka tujuan dari penelitian ini yaitu untuk memberikan usulan *framework security big data* pada PT. Telekomunikasi Indonesia yang akan menjadi panduan keamanan *big data*.

I.3 Batasan Penelitian

Adapun yang menjadi batasan penelitian tugas akhir ini antara lain:

1. Penelitian ini dibatasi hanya sampai tahap perancangan, tidak sampai pada tahap implementasi atau konfigurasi.
2. Penulis hanya memberikan gambaran terhadap *framework* yang dibuat, tidak menjelaskan secara spesifik terhadap setiap komponen *framework.*
3. Penelitian ini hanya dilakukan pada Divisi Digital Service PT. Telekomunikasi Indonesia, tidak sampai pada divisi lainnya.
4. Penelitian ini tidak sampai memberikan perlindungan terhadap *physical layer.*

I.4 Manfaat Penelitian

Manfaat yang akan didapatkan dari penelitan ini yaitu:

1. *Framework big data security* yang dihasilkan diharapkan menjadi solusi yang mampu melindungi *big data* pada PT. Telekomunikasi Indonesia.
2. Menjadi rekomendasi pada Telkom University untuk menjadikan *big data* sebagai pembelajaran atau mata kuliah baru khususnya pada keamanan *big data*.

I.5 Sistematika Penulisan

Penelitian tugas akhir ini akan diuraikan berdasarkan sistematika laporan sebagai berikut:

BAB I PENDAHULUAN

Bab ini menjelaskan tentang latar belakang masalah, perumusan masalah pada penelitian, batasan penelitian, tujuan penelitian, dan manfaat dari penelitian ini.

BAB II LANDASAN TEORI

Dalam bab ini berisikan landasan teori atau literatur yang terkait sehingga dapat mendukung penulisan penelitian tugas akhir ini. Studi literatur dapat diperoleh dari jurnal atau penelitian terdahulu, buku-buku, dan website yang berkaitan.

BAB III METODOLOGI PENELITIAN

Pada bab metodologi penelitian terdiri dari uraian model konseptual menjelaskan mengenai masukan atau input yang diperlukan untuk melaksanakan penelitian hingga mendapatkan keluaran atau *output* dari penelitian yang digunakan dan sistematika penelitian berupa penjelasan tahapan-tahapan dalam melakukan penelitian mulai dari tahap persiapan hingga tahap pelaporan tugas akhir.

BAB IV PERANCANGAN *FRAMEWORK BIG DATA SECURITY* UNTUK PT. TELEKOMUNIKASI INDONESIA

Pada bab ini berisi mengenai bagaimana peneliti merancang *framework big data security* berdasarkan penggabungan dan perbandingan antara *framework* yang disediakan vendor ternama dan *framework* studi literature.

BAB V EVALUASI *FRAMEWORK BIG DATA SECURITY*

Pada bab ini, peneliti melakukan evaluasi *framework big data security* yang telah dibuat. Evaluasi ini dilakukan untuk memperkut serta memperbaiki *framework big data security*, sehingga *framework* tersebut dapat diterima atau dapat dijadikan rekomendasi *framework* untuk PT. Telekomunikasi Indoensia.

BAB VI KESIMPULAN DAN SARAN

Dalam bab kesimpulan dan saran menjelaskan detail dari kesimpulan pada penelitian tugas akhir dan saran yang diberikan untuk menerapkan *big data security* dari penulis.

BAB II LANDASAN TEORI

II.1 Tinjauan Umum Big Data

Big data merupakan istilah dari kumpulan data-data yang diproses dan memiliki karakteristik 3V. *Volume* (memiliki kapasitas yang besar), *Velocity* (data dengan cepat berubah dan berkembang), *Variety* (data yang masuk berasal dari berbagai sumber dan berbagai format) sehingga menjadikan *big data* sulit untuk ditangani atau diproses jika hanya menggunakan manajemen basis data yang biasa atau tradisional. (Lamyae HBIBI, 2016).

1. **Volume**

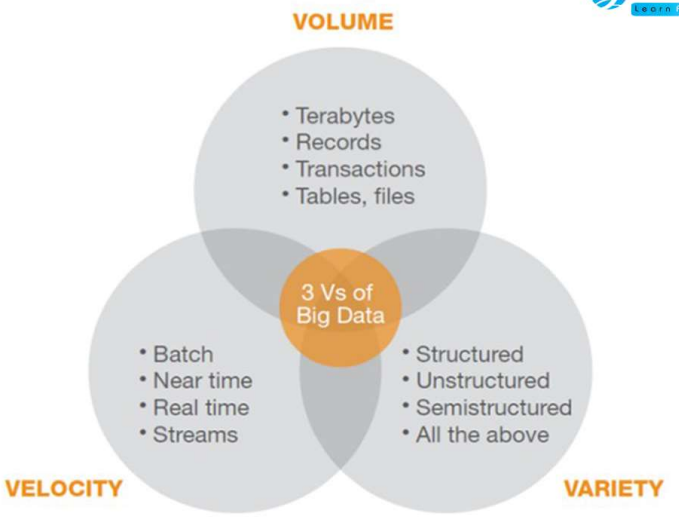
   Volume merupakan atribut utama dari *big data*. *Big data* juga dapat dihitung berdasarkan *records, transaction, tables,* atau *file*.

2. **Velocity**

   Salah satu alasan kenapa disebut *big data* adalah data yang datang dari berbagai banyak sumber seperti *website, logs, clickstream,* dan sosial media. Semakin banyak data yang masuk semakin banyak juga data yang dihasilkan dan diberikan kepada *user*, oleh karena itu data-data yang ada pada *big data* sangat cepat berkembang.

3. **Variety**

   Big data berdasarkan dari berbagai banyak sumber sehingga data yang masuk berbagai jenis juga seperti data terstruktur, tidak terstruktur dan semi struktur (text, sensor data, suara, video, *clickstream data, log file,* dan lain-lain) (Russom, 2011).



Source: Russom, Philip 2011

Gambar II - 1 3V Big Data

II.2 Big Data Framework

Teknologi tradisional *big data* tidak dapat lagi menangani tempat penyimpanan dan analisis perkembangan *volume* dari berbagai jenis data. Para komunitas peneliti bertahan untuk mengembangkan teknologi baru yang *user friendly, dynamic* dan memiliki biaya yang tidak mahal untuk teknologi pada big data yang dapat membantu *data scientists* dalam mengambil keputusan (Lamyae HBIBI, 2016). Berikut framework pada big data yang digunakan vendor terkemuka.

II.2.1 Hortonworks Big Data Framework

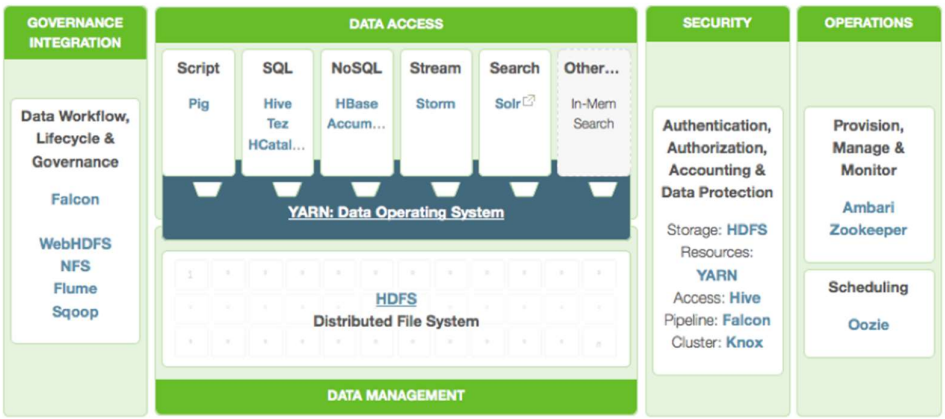
Hortonworks merupakan inovator terkemuka dalam menciptakan, mendistribusikan dan mendukung *platform data* siap pakai pada perusahaan dan aplikasi modern. Hortonwork memliki misi mengelola data dunia, Hortonworks fokus dalam mendorong inovasi dikomunitas *open source* seperti Apache Hadoop, NiFi dan Spark. Hortonworks memiliki mitra lebih dari 1600 yang memberikan keahlian, pelatihan dan layanan yang memungkinkan pelanggan untuk membuka nilai transformasional bagi organisasi mereka disemua bisnis. *Platform* data yang

tehubung diperkuat dengan aplikasi *data* modern yang memiliki kecerdasan terhadap *data-in-motion* dan *data-at-rest* (hortonworks.com).

Menurut hortonworks, *framework big data* terbagi menjadi 5 area inti pada hadoop, yaitu:

1. ***Data Management,*** *Hadoop Distributed File System* (HDFS) menyediakan fondasi untuk menyimpan data dalam format apapun. *Yet Another Resource Negotiator* (YARN) merupakan aplikasi yang bertindak sebagai sistem operasi *data platform* yang menyediakan pengelolaan sumber daya dan arsitektur *pluggable* sehingga Hadoop menjadi fleksibel.
2. ***Data Access,*** MapReduce merupakan aplikasi yang sangat penting digunakan untuk mengatasi proses data yang bersifat *batch-oriented*. Teknologi untuk *scripting*, SQL, NoSQL, *Search* dan *Streaming* sudah terintegrasi dengan Hortonworks *data platform*. Apache Pig dapat menangani *scripting,* Apache Hive merupakan aplikasi yang kemampuan untuk menampilkan *query* berskala petabyte. Apache Hbase dan Apache Accumulo merupakan aplikasi NoSQL yang paling banyak digunakan. Apache Storm dapat mendukung *real-time stream* proses yang biasa digunakan untuk sensor dan *machine data*.
3. ***Data Governance & Integration,*** Apache Falcon mendukung *policy-base* untuk mengatur tata kelola *lifecycle* data pada Hadoop, termasuk menjaga *recovery data* pada saat terjadi bencana. Untuk *data ingest*, Apache Sqoop mempermudah aliran data dari *database* menuju ke Hadoop, dan Apache Flume dapat meyimpan aktivitas apa saja yang terjadi pada *data* pada Hadoop.
4. ***Security,*** otentikasi, otorisasi, akuntansi dan perlindungan data merupakan aspek terpenting yang harus diperhatikan pada keseluruhan komponen Hadoop.
5. ***Operation,*** dengan menggunakan Apache Ambari, operator dapat menyediakan, mengelola dan memantau *cluster* Hadoop serta mengintegrasikannya dengan proses bisnis perusahaan. Untuk mengatur dan manajemen waktu dapat menggunakan Oozie.

Untuk lebih jelas, dapat dilihat pada gambar II-2 dibawah ini.



Source: Hortonworks Data Platorm 2018

Gambar II - 2 Hortonwork Big Data Framework

II.2.2 Cloudera Big Data Framework

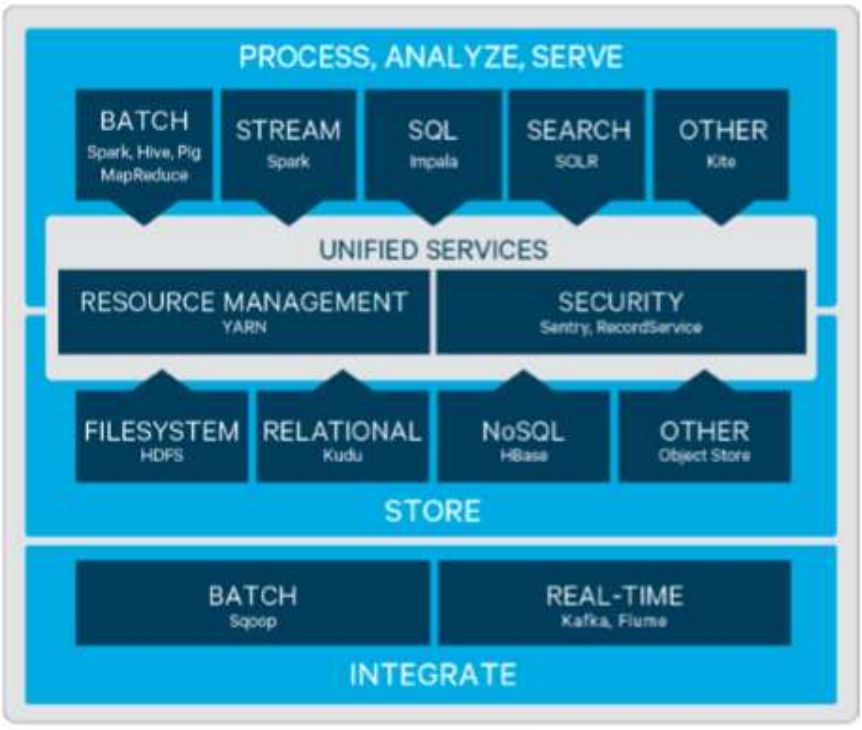
Cloudera merevolusi pengelolaan *data* perusahaan dengan menawarkan platform untuk *big data.* Cloudera menawarkan perusahaan sauatu tempat untuk menyimpan, memproses dan menganalisis semua data mereka untuk memperluas nilai investasi yang ada dan membuat cara baru yang mendasar untuk mendapatkan nilai dari data mereka. Cloudera didirikan pada tahun 2008 oleh beberapa orang paling cerdas diperusahaan terkemuka Silicon Valley, termasuk Google (Christophe Bisciglia), Yahoo! (Amr Awadallah), Oracle (Mike Olson) dan Facebook (Jeff Hammerbacher). Pendiri Cloudera memegang inti kepercayaan mereka bahwa *open source*, standar terbuka dan pasar terbuka adalah yang terbaik. Keyakinan itu tetap penting bagi nilai-nilai mereka. Doug Cutting (Co-creator) dari Hadoop, bergabung dengan Cloudera pada tahun 2009 sebagai *Chief Architect.* Saat ini Cloudera memiliki lebih dari 1.600 karyawan dan memiliki kantor di 24 negara seluruh dunia dengan kantor berpusat di Palo Alto, California (Cloudera, 2018).

Menurut Cloudera *framework big data* terbagi menjadi 5 area yaitu:

1. ***Integrate*,** terbagi menjadi dua yaitu *batch* dan *real-time*. Cloudera menggunakan Apache Sqoop untuk memproses *data batch* dan menggunakan Apache Kafka dan Apache Flume untuk menangani *data real-time*.

2. ***Store,*** terbagi menjadi lima yaitu *Filesystem, Relational* dan NoSQL, pada *Filesystem* Cloudera menggunakan HDFS, *Relational* Cloudera menggunakan Kudu, NoSQL Cloudera menggunakan Apache HBase.
3. ***Unified Service,*** terbagi menjadi dua yaitu *Resource Management* dan *Security. Resoucer Management* Cloudera menggunakan YARN dan pada *Security* Cloudera menggunakan Apache Sentry. Apache Sentry merupakan sistem yang dapat mengatur otorisasi berbasi *role* dan metadata pada Hadoop Cluster.
4. ***Process, Analyze, Serve*****,** terbagi menjadi empat yaitu *Batch, Stream, SQL, Search*. Untuk *batch* Cloudera menggunakan Apache Spark, Hive, Pig dan MapReduce, pada *stream* Cloudera menggunakan Apache Spark, pada SQL Cloudera menggunakan Apache Impala, dan pada *search* Cloudera menggunakan SOLR.

Untuk lebih jelas, dapat diliat pada gambar II-3 dibawah ini.



Source: Cloudera Introduction

Gambar II - 3 Cloudera Big Data Framework

II.3 *Framework Big Data Security Vendor* dan Studi Literatur

Setelah mengetahui *framework big data* yang digunakan *vendor,* kemudian pada sub bab ini menjelaskan *framework big data* berdasarkan vendor. Menurut (Oxford, 2018)*, vendor* merupakan seseorang atau perusahaan yang menawarkan sesuatu atau jasa untuk dijual. Dengan menggunakan vendor akan mempermudah dan meningkatkan kinerja perusahaan. Berikut merupakan *framework security* pada big data berdasarkan *vendor* dan studi literatur.

II.3.1 *Framework Big Data Security* Menurut Hortonworks

Hortonworks menyediakan *framework* yang dapat digunakan untuk melindungi *big data*. Untuk memberikan manajemen keamanan yang konsisten, Hortonworks menggunakan Apache Ranger yang dapat digunakan untuk mendefinisikan, mengelola kebijakan keamanan yang konsisten dan secara terpusat disemua komponen Hadoop. Berikut framework *big data security* menurut Hortonworks.

II.3.1.1 *Authentication*

Menetapkan identitas pengguna dengan otentikasi yang kuat adalah dasar untuk akses aman di Hadoop. Pengguna perlu mengidentifikasi diri mereka sendiri dan kemudian memiliki identitas yang disebarkan diseluruh *cluster* hadoop untuk mengakses sumber daya *cluster.* Hortonworks menggunakan Kerberos untuk otentikasi, Kerberos adalah standar industri yang digunakan untuk otentikasi pengguna dan sumber daya dalam kelompok Hadoop.

Apache Knox digunakan untuk membantu memastikan keamanan perimeter bagi pelanggan Hortonworks. Dengan Knox, perusahaan dengan percaya dapat memperluas Hadoop API ke pengguna baru tanpa adanya kerumitan, Knox menyediakan *gateway* utama untuk Hadoop API yang memiliki kemampuan otorisasi, otentikasi, *Transport Layer Security* (TLS), *Single Sign On* (SSO) yang bervariasi untuk memungkinkan satu jalur akses ke Hadoop.

II.3.1.2 *Authorizarion*

Apache Ranger mengelola kontrol akses melalui antarmuka pengguna yang memastikan konsistenitas administrasi kebijakan diseluruh komponen akses *data*

Hadoop. Administrator dapat menentukan kebijakan kemanan ditingkat *database,* table, kolom dan *file* serta dapat mengelola izin untuk grup yang berbasis LDAP (*lightweight Directory Access Protocol)* atau pengguna individual tertentu. Aturan berdasarkan kondisi dinamis seperti waktu atau geolokasi juga dapat ditambahkan kedalam kebijakan yang ada. Model otorisasi Ranger adalah *pluggable* dan dapat dengan mudah disebarluaskan ke sumber *data* manapun.

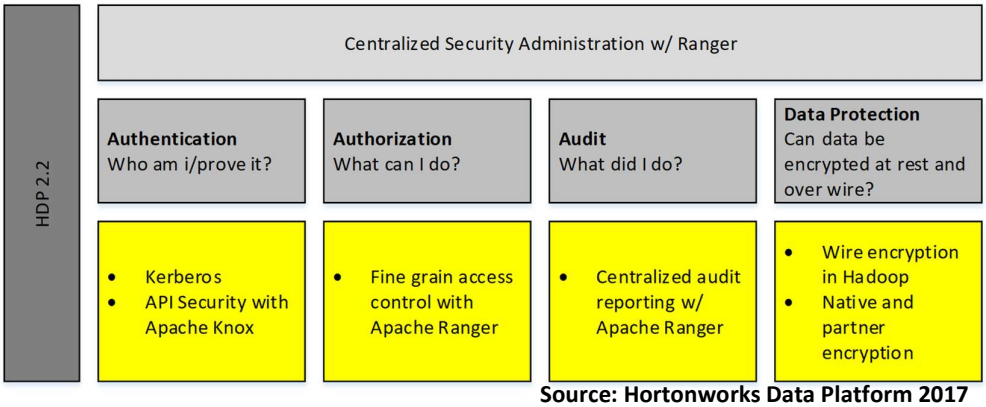
II.3.1.3 Audit

Metadata dan tata kelola harus menjadi bagian penting dari data perusahaan yang siap untuk memiliki *big data*. Hortonworks mendirikan *Data Governance Initiative* dengan Aetna, Merck, Target dan SAS untuk memperkenalkan pendekatan umum terhadap tata kelola data Hadoop ke dalam komunitas *open source*. Inisiatif ini telah berkembang menjadi proyek *open source* baru bernama Apache Atlas. Apache Atlas adalah seperangkat layanan tata kelola inti yang memungkinkan perusahaan untuk melakukan tata kelola *data*. Layanan ini meliputi:

- Pencarian *dataset,*
- *Metadata-driven data access control,*
- Index dan pencarian terpusat audit,
- *Data lifecycle* manajemen dari data masuk sampai data digunakan.

II.3.1.4 *Data Protection*

Fitur perlindungan data membuat data tidak dapat dibaca baik saat transit melalui jaringan dan saat *at-rest* pada disk. Hadoop menggunakan *Transparent Data Encryption* (TDE) untuk mengenkripsi data yang masuk ke HDFS. Dengan menggunakan KMS (*key Management Server)* memungkinkan Ranger administrator untuk mengelola *key* dan kebijakan otorisasi lebih mudah. Hortonworks juga bekerja sama dengan mitra enkripsi untuk mengintegrasikan enkripsi HDFS dengan manajemen *key* kelas *enterprise* (Hortonworks, 2017).



Source: Hortonworks Data Platform 2017

Gambar II - 4 Big Data Security Framework menurut Hortonworks

II.3.2 *Framework Big Data Security* Menurut Claudera

Tujuan untuk sistem manajemen data seperti *big data* merupakan *confidentiality, integrity* dan *availability* mengharuskan sistem itu diamankan dalam berbagai dimensi, dimensi ini dapat dihasilkan dari segi tujuan operasional dan konsep teknis. Tidak halnya Hortonworks, Cloudera juga menyediakan *framework* untuk mengamankan sistem manajemen data seperti *big data,* Berikut *framework security big data* menurut cloudera:

II.3.2.1 Perimeter

Akeses ke *cluster* harus dilindungi dari berbagai jenis ancaman yang datang dari jaringan eksternal maupun internal serta dari berbagai pengguna. Isolasi jaringan dapat disediakan dengan mengkonfigurasi Firewall, Router, Subnet dan penggunaan IP *public dan private* dengan baik. Mekanisme otentikasi memastikan bahwa *people, process* dan aplikasi benar-benar mengidentifikasikan bahwa pengguna teridentifikasi untuk dapat mengakses *cluster.*

II.3.2.2 *Data*

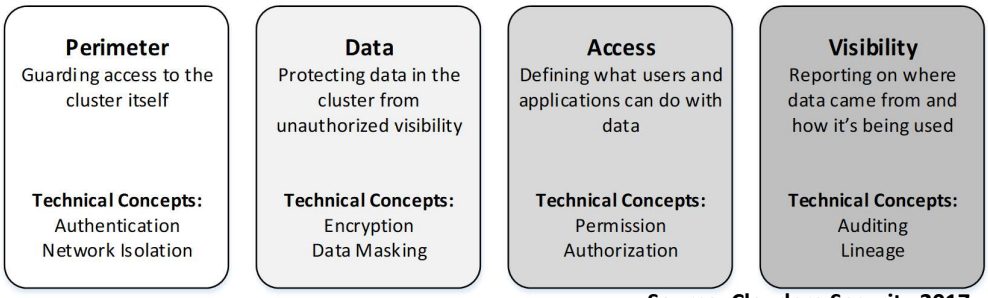
Data pada *cluster* harus selalu dilindungi dari pengguna yang tidak sah. Demikian juga dengan komunikasi antara *node* pada *cluster* harus dilindungi. Mekanisme enkripsi memastikan bahwa paket *network* diambil atau fisikal *hard-disk* bebas dari *bad sector* dan konten yang tidak digunakan.

II.3.2.3 *Access*

Mengakses ke beberapa layanan spesifik atau *data* pada *cluster* harus diberikan kepada pengguna khusus. Mekanisme otorisasi memastikan bahwa pengguna harus terotorisasi untuk mengakses *cluster*, pengguna hanya bisa melihat *data* dan menggunakan *data* tersebut berdasarkan hak akses mereka.

II.3.2.4 *Visibilty*

Visibilitas berarti *history* dari perpindahan atau perubahan *data* berdasarkan kebijakan yang ada. Mekanisme audit memastikan bahwa seluruh aktivitas pada *big data* terdokumentasi (Cloudera, Inc, 2017).



Source: Cloudera Security 2017

Gambar II - 5 Big Data Security Framework menurut Cloudera

II.3.3 Tinjauan Big Data Security Framework Menurut Ajit Gaddam

Hadoop merupakan elemen terpenting pada *big data*, hadoop merupakan *software* yang berbasis *open source*. Oleh karena itu, keamanan pada hadoop memiliki banyak celah, berikut kelemahan pada hadoop:

1. **Infrastruktur *Security & Integrity***
   - **a.** *File* konfigurasi keamanan pada hadoop tidak mengandung kebijakan, sehingga menghasilkan integritas data dan masalah ketersediaan.
   - **b.** Pada versi hadoop sebelumnya tidak terdapat otentikasi, sehingga siapa saja dapat mengakses hadoop.
2. **Identitas & *Access management***
   - **a.** *Role Based Access Control* (RBAC) dan *Access Control List* (ACL) untuk komponen seperti MapReduce dan Hbase biasanya dikonfigurasi

melalui *clear-text file*, sehingga *file-file* ini dapat diubah oleh akun istimewa seperti akun *root* dan akun lainnya.

3. ***Data Privacy & Security***
   a. Semua masalah yang terkait dengan jenis serangan *SQL Injection* tidak hilang. Serangan ini berkerja pada komponen hadoop seperti Hive dan Impala.
   b. Kurangnya control kriptografi untuk melindungi data sensitive, dan kontrol seperti ini sebaiknya dilakukan diluar tumpukan *data* dan aplikasi . (Gaddam, 2015)

Berdasarkan penyataan diatas perlu adanya *framework* khusus yang menangani *security* pada *big data*, berikut beberapa *framework big data security*:

II.3.3.1 Data Management

Komponen *data management* didekomposisi menjadi tiga sub komponen inti, yaitu *Data Classification, Data Discovery* dan *Data Tagging.*

II.3.3.1.1 Data Classification

Klasifikasi *data* yang efektif merupakan salah satu kegiatan terpenting yang dapat menyebabkan implementasi kontrol *security* pada *big data* menjadi efektif. Berikut adalah cara yang dapat menghasilkan matriks klasifikasi *data*:

1. Bekerja sama dengan *privacy office, legal, intellectual property, finance* dan *information secrutiy* untuk menentukan bidang data yang berbeda.
2. Lakukan *security control assessment.*
   a. Tentukan lokasi data,
   b. Lakukan penomoran pengguna dan sistem dengan akses,
   c. Tentukan *security control*.
3. Tentukan nilai data dari penyerang.
   a. Apakah data mudah untuk dijual pada pasar gelap?
   b. Apakah data memiliki informasi yang sangat berharga?
4. Tentukan penyesuaian dan dampak dari pendapatan.
   a. Tentukan persyaratan pelaporan pelanggaran untuk semua bidang yang berbeda.

b. Apakah kehilangan data tertentu dapat menyebabkan bisnis terhambat.
c. Estimasi biaya untuk memperbaharui sistem.
d. Biaya lainnya seperti audit, mengekspansi kebijakan tertentu.

5. Tentukan penyebab hilangnya data pribadi.
    a. Apakah dikarenakan phissing.

II.3.3.1.2 Data Discovery

Kurangnya kesadaran situasional sehubungan dengannya data sensitif dapat membuat organisasi terkena risiko yang signifikan. Mengidentifikasikan apakah data sensitif ada di Hadoop, menentukan keberadaan *data* dan kemudian memicu tindakan perlindungan data yang sesuai, seperti penyembunyian data, redaksi data, tokenisasi atau enkripsi data. Berikut cara yang sangat penting dilakukan untuk menemukan data pada *big data:*

1. Pastikan dan validasi struktur *data* dan skema *data*.
2. Kumpulkan matriks data seperti *volume, counts, unique count* dan lain-lain.
3. Bagikan informasi yang telah didapat dengan tim *data science* agar mereka dapat membangun model ancaman.
4. Buat rutinitas pencarian seperti laporan.

II.3.3.1.3 Data Tagging

Mengerti *end-to-end* aliran *data* pada *big data*, khususnya pada *data* masuk dan *data* dikases.

1. Identifikasi semua data yang baru masuk pada *cluster*.
2. Mengetahui apakah data berasal dari sumber terpercaya.
3. Ikuti *data* keluar dan identifikasi data apa saja yang keluar dari *big data*.
4. Pernyataan diatas juga dapat membantu aktivitas *data discovery* dan *data acess management* lainnya.

II.3.3.2 *Identity & Access Management*

POSIX-style permission pada HDFS merupakan dasar dari akses kontrol yang digunakan untuk menentukan siapa dan apa saja yang dapat dilihat atau mengakses

Hadoop. *Identity & Access Management* terbagi menjadi lima komponen inti, yaitu *Authorization, Authentication, RBAC Authorization, Data Metering+user Entitlement, server, DB, Table, View based AuthorizationI.*

II.3.3.2.1 Authentication

Otentikasi merupakan tindakan untuk memverivikasi identitas pengguna untuk mengakses suatu objek. Otentikasi juga bisa digunakan sebagai pesan yang dapat digunakan untuk membantu memvalidasi pengguna pada mekanisme otorisasi (B. Aboba, 2003). Otentikasi pada *big data* menggunakan AD (*Active Directory*), LDAP dan Kerberos (Gaddam, 2015).

II.3.3.2.2 Authorization

Otorisasi merupakan tindakan untuk menentukan hak akses pada pengguna, seperti objek apa saja yang dapat diakses oleh pengguna (B. Aboba, 2003). Otentikasi dapat membatasi pengguna untuk mengakses suatu objek. Otorisasi sebaiknya diimplementasikan keseluruh komponen hadoop termasuk antara *data node –to-data node, name node-to-name node, name node-to-name node* dan antara komponen lainnya (Gaddam, 2015).

II.3.3.2.3 User Entitlement + Data Metering

Menyediakan pengelolaan kebijakan akses secara terpusat. Berikut langkah-langkah untuk membuat perizinan kebijakan akses secara terpusat:

b. Penting untuk menambahkan kebijakan pada data dan bukan hanya pada akses data.
c. Mengumpulkan atribut berbasis akses kontrol dan proteksi data yang berdasarkan label pada data.
d. Lakukan *metering data* dengan membatasi akses ke data setelah data melewati aplikasi (Gaddam, 2015).

II.3.3.2.4 RBAC Authorization

Memberikan otorisasi harus melalui RBAC, berikut langkah-langkah untuk memberikan otorisasi berdasarkan RBAC:

- Kelola *data access* berdasarkan peran pengguna.
- Tentukan relasi antara pengguna dan peran melalui grup, seperti menggunakan AD/LDAP dan menentukan peraturan *data access* yang lewat (Gaddam, 2015).

II.3.3.2.5 Server, Database, Table, View Based Authorization

Mekanisme otorisasi dapat membatasi pengguna untuk mengakses suatu objek, pada *framework* ini *server, database, table* harus memiliki otorisasi pengguna untuk mengakses komponen tersebut, sehingga user yang tidak memiliki wewenang tidak dapat mengakses secara keseluruhan komponen tersebut (Gaddam, 2015).

II.3.3.3 *Data Protection & Privacy*

Sebagian besar distribusi Hadoop dan *add-ons* vendor memasang enkripsi pada *data-at-rest* atau pada keseluruhan *file.* Perlindungan kriptografi tingkat aplikasi seperti *colomn level encryption, tokenization* dan *data masking* perlu dilakukan (Gaddam, 2015).

II.3.3.3.1 *Transparent Encryption (disk/HDFS layer)*

Full Disk Encryotion dapat mencegah akses pada media penyimpanan. Enkripsi *file* juga dapat mencegah terhadap akses pada tingkat sistem operasi node.

- Pada kasus tertentu, data sensitive perlu disimpan dan diproses pada Hadoop, enkripsi pada *data-at-rest* akan melindungi data sensitive organisasi dan membantu untuk menentukan batasan audit.
- Dalam *cluster* Hadoop yang besar, *disk* seringkali perlu dikeluarkan dari *cluster* dan diganti. *Disk level encryption* memastikan bahwa tidak ada data sisa yang dapat dibaca (*no human-readable*) pada saat data dipindahkan atau dihapus.
- *Full disk encryption* dapat juga sebagai *os-native disk encryption* seperti *dm-crypt.*

II.3.3.3.2 Application Level Cryptography (Tokenization, field-level Encryption)

Ketika enkripsi pada tingkat *field level encryption* dapat memberikan kemampuan keamanan dan *audit tracking* (Gaddam, 2015). Enkripsi asimetris seperti RSA (Rivest-Shamir-Adleman) merupakan algoritma asimetris yang paling banyak digunakan. Tokenisasi biasanya digunakan pada transaksi kartu kredit atau debit, dengan menggunakan.

II.3.3.3.3 Data Loss Prevention

Data Loss Prevention (DLP) merupakan strategi untuk memastikan bahwa user tidak membawa data yang bersifat sensitif keluar perusahaan. DLP menggunakan *business rules* untuk mengklaisfikasikan dan melindungi data. DLP berada pada *data-at-use* dan *data-in-transit*.

- *Data-at-use* berarti disaat data digunakan seperti pada laptop, *flashdisk*, MP3 *player* dan lain-lain.
- *Data-in-transit* berarti disaat data bergerak melalui *network* via email, *peer-to-peer* dan mekanisme komunikasi lainnya (Simon Liu, 2010).

II.3.3.4 *Network Security*

Lapisan keamanan jaringan didekomposisi menjadi empat sub komponen, yaitu *data protection in-transit* dan *network zoning (authorization component).*

II.3.3.4.1 Data Protection in-transit

Mengamankan komunikasi dibutuhkan HDFS untuk melindungi *data-in-transit.* Dengan menggunakan *Trasnport Layer Security* (TLS) untuk mengotentikasi dan memastikan komunikasi antara *node,* nama server, dan aplikasi dengan aman. Berikut beberapa skenario penyerangan pada *data-in-transit:*

- Penyerang bisa mendapatkan akses tidak sah ke data dengan cara mencegat komunikasi ke konsol Hadoop.
- Komunikasi antara *name node* dan *data node* dapat dilacak oleh penyerang.
- *Token* yang memberikan pengguna mengotentikasi diri ke Kerberos dapat juga dilacak dan digunakan sebagai penyamaran untuk mengakses *name node.*

Berdasarkan skenario diatas, berikut langkah-langkah yang dapat diimplementasikan pada *big data cluster:*

1. *Packet Level Encryption* menggunakan TLS dari *client* ke Hadoop.
2. *Packet Level Encryption* menggunakan TLS antara *cluster* meliputi *http* antara *name node –to- Job Tracker –to- Data Node*.
3. *Packet Level Encrytion* menggunakan TLS pada *cluster* seperti *mapper* ke *reducer*.
4. Menggunakan LDAP melalui SSL ketika berkomunikasi dengan direktori perusahaan untuk mengetahui pelacakan penyerang.
5. Biarkan admin untuk mengkonfigurasi dan menghidupkan enkripsi acak dan TLS untuk HDFS, MapReduce, YARN, Hbase dan lain-lain (Gaddam, 2015).

II.3.3.4.2 Network Security Zoning

Network security zoning merupakan wilayah dalam jaringan khususnya pada *big data* yang berdasarkan komponen-komponen yang melindungi informasi didalamnya. Berikut cara melindungi perimeter *big data:*

- User hanya boleh terkoneksi dengan name nodes, tidak boleh terkoneksi dengan data nodes.
- API harus berkemampuan untuk mengontrol protokol yang keluar masuk pada hadoop.
- Firewall harus bisa memberikan akses hanya pada name nodes dan API yang masuk ke hadoop.
- Harus adanya IDS/IPS.

II.3.3.5 *Infrastructure Security & Integrity*

Infrastructure security & integrity layer terbagi menjadi 4 komponen inti, yaitu *logging/audit, Secure Enhanced Linux (SELinux), File Integrity (Data Tamper)* dan *Privileged user and activity monitoring.*

II.3.3.5.1 Logging Audit

Seluruh perubahan sistem atau ekosistem yang unik pada *cluster* Hadoop perlu diaudit dengan *log* audit yang dilindungi, sebagai contoh:

a. Penambahan dan pengurangan data dan manajemen *node.*
b. Perubahan manajemen *node* termasuk *job tracker node* dan *name node.*

Bila *data* tidak terlpepas pada salah satu komponen inti Hadoop, maka tinggi juga fragmentasinya, akibatnya banyak tangkapan *metadata* dan *log* audit di seluruh fragmen. Berikut rekomendasi teknologi untuk mengatasi *data fragmentation:*

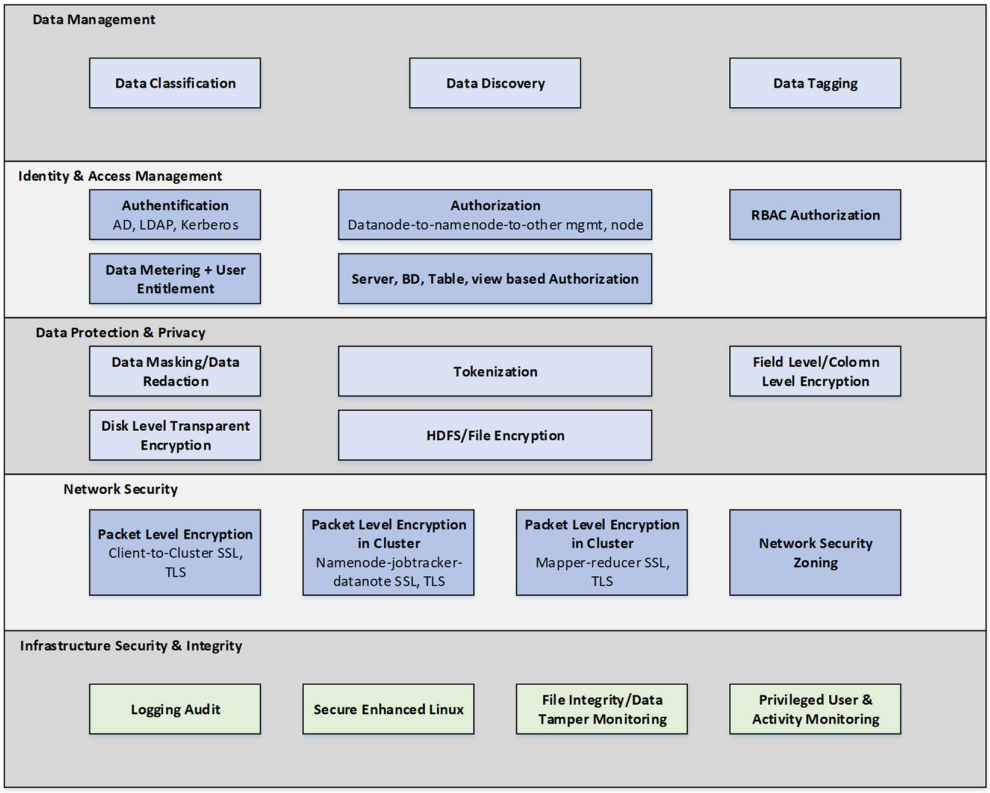
1. Apache Falcon, merupakan ingkubasi dari proyek Apache OSS yang berfokus pada manajemen *data*. Apache Falcon mampu memberikan grafik turunan *data* dan kontrol yang aktif pada *data lifecycle*.
2. Cloudera Navigator, merupakan bagian dari distribusi Cloudera dan Apache Hadoop. Cloudera Navigator merupakan alat untuk mengatasi *logging metadata.*
3. Zettaset Orchestrator, merupakan produk yang memanfaatkan keseluruhan fragmentasi keamanan Hadoop dengan menggabungkan *workflow*. Zettaset memiliki repositori metadata tersendiri dimana metadata dari semua kompenen Hadoop dikumpulkan dan disimpan.

II.3.3.5.2 Secure Enhanced Linux (SELinux)

SELinux diciptakan oleh United State National Security Agency (NSA) sebagai seperangkat *patch* ke Linux Kernel menggunakan Linux Security Modules (LSM). SELinux adalah contoh dari *Mandatory Access Control* (MAC) untuk Linux. Secara historis Hadoop dan *platform big data* lainnya yang dibangun atas sistem Linux dan UNIX memiliki kontrol akses tanpa pamrih atau secara bebas. Berikut keuntungan jika menggunakan SELinux;

a. Dengan menetapkan dan menkonfigurasi SELinux di lingkungan *big data,* maka kebijakan dapat diatur dan ditetapkan secara administrative.
b. Jika pengguna mengubah pengaturan apa pun pada direktori mereka, kebijakan tersebut dapat mencegah pengguna lain atau proses untuk mengaksesnya.

c. Jika pengguna berbahaya dapat memperoleh akses ke *root* menggunakan SSH, namun pengguna tersebut tidak dapat mengeksekusi apapun.



Source: Ajit Gaddam 2015

Gambar II - 6 Framework Big Data Security menurut Ajit Gaddam

II.3.4 Kelebihan dan Kekurangan *Framework Big Data Security*

Framework yang telah disajikan sebelumnya pasti memiliki kekurangan dan kelebihan, yang mana kekurangan ini akan ditutupin oleh *framework* yang akan dibuat penulis dan kelebihannya dijadikan pertimbangan bagi penulis. Berikut tabel II-1 menjelaskan kelebihan dan kekurangan *framework big data security*.

	Kelebihan	Kekurangan
Hortonworks dan Cloudera	• Memberikan jaminan atas kemanan pada big data. • Memberikan pelatihan kemanan pada big data.	• Tidak memiliki pemananan terhadap aplikasi. • Tidak memberikan keamanan terhadap

		perimeter terhadap environment big data. • Tidak menyediakan bagaimana cara mengelola data.
Ajit Gaddam	• Framework ini dibuat berdasarkan permasalahan yang sering terjadi pada keamanan big data.	• Tidak memiliki pengamanan terhadap aplikasi.

Tabel I - 1 Kelebihan dan Kekurangan Framework

II.4 Arsitektur *Platform* Big Data pada PT.Telekomunikasi Indonesia

Sebagaimana halnya data *warehouse*, *web stores* maupun platform IT lainnya, infrastruktur *big data* juga memiliki kebutuhan khusus. Dengan mempertimbangkan semua komponen dari *platform big data*, penting untuk diingat bahwa tujuan dari *big data* adalah untuk mempermudah konsolidasi berbagai jenis data dalam perusahaan sehingga memungkinkan user menggunakan data. Berikut *platform big data* pada PT. Telekomunikasi Indonesia.

platform big data Telkom memiliki 4 bagian besar, yaitu *acquired, accessed, analytic* dan *application.*

1. ***Acquired***

   Acquired merupakan bagian pertama dimana pada bagian ini data masuk ke *big data*, data yang masuk berupa data yang terstruktur, tidak terstruktur dan semi terstruktur.

2. ***Accessed***

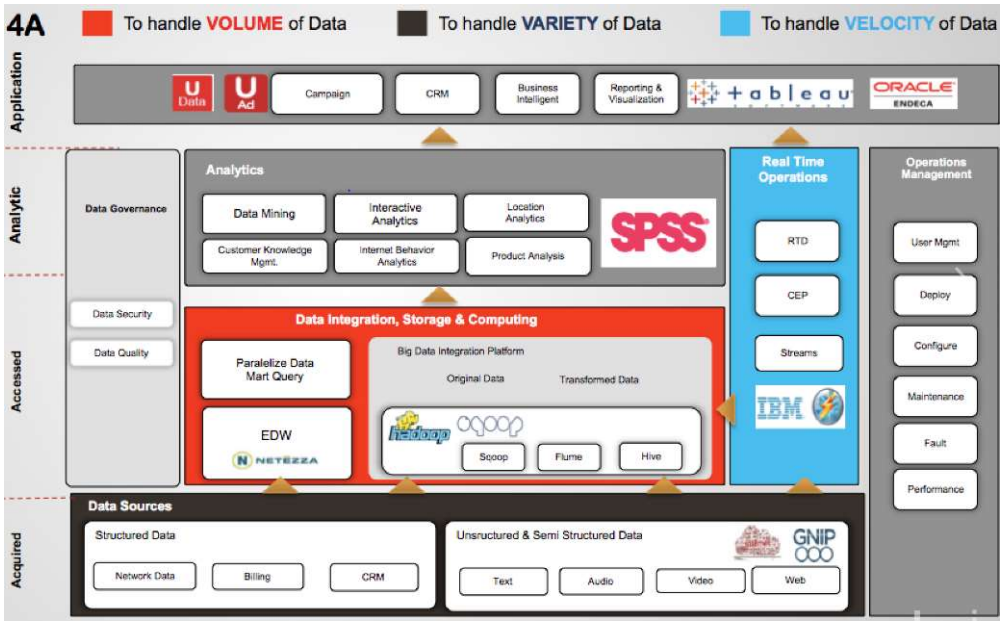
   Accessed merupakan bagian dimana data diproses dan disimpan. Pada bagian ini terdiri dari *data governance* dan *real time operation*. *Data governance* bertugas untuk menjaga keamanan dan kualitas data, sedangkan *real time operation* bertugas memonitor aktifitas komponen pada *accessed* dan *analytics.*

3. ***Analytics***

   Analytics merupakan bagian dimana data pada *big data* dianalisa. Bagian ini dapat membantu menjada kualitas data pada *big data*. Teknik analis pada bagian ini berupa *data mining, interactive analytics, location analytics, customer knowledge management* dan lain-lain.

4. ***Application***

   Application merupakan bagian terakhir pada *big data*. Bagian ini bertujuan untuk menampilkan data pada *big data*. Aplikasi pada bagian ini meliputi UData dan UAD.



Gambar II - 7 Platform Big Data Telkom (Unit Big Data Telkom DDS)