

SILABUS SEKOLAH BIG DATA

Copyright :

Sekolah Big Data - www.sekolahbigdata.com

PT. Cilsy Fiolution Indonesia - www.cilsyfiolution.com

Siapapun dilarang untuk menggunakan, mengkopi, menyebarluaskan, menjual sebagian atau seluruh isi dokumen ini tanpa seizin pemilik yaitu PT. Cilsy Fiolution Indonesia. Konten Silabus ini adalah hak cipta dari PT. Cilsy Fiolution Indonesia dan dilindungi oleh Undang-Undang.

Sekolah Big Data - Student Requirements	
Big Data Engineer, Data Engineer, Hadoop Administrator	
Date :	2020
Based on :	Astra Pegi-Pegi IYKRA MNC Traveloka
After graduate from bootcamp, Student should be able to do these jobs :	
Code	Job Description
JD1	Collaborate with Data, Business Intelligence and IT team members to fulfill their requests.
JD2	Build the infrastructure required for optimal extraction, transformation and loading of data from a wide variety of data sources using SQL and other technologies.
JD3	Develop automated ETL Process from various data types & formats
JD4	You will be responsible for the reliability, scalability, and usability of all of our all of our data elements, whether in databases or in raw form.
JD5	Ensure data is queryable for our Data Scientists, Data Analysts and Business team.
JD6	You research new technologies and solutions in the realm of data analytics and data science that can increase the performance, reliability and quality of our data assets.
And student should have these qualifications :	
Code	Qualifications
Q1	Basic understanding of Linux
Q2	Proficient in Programming language: Java/Python
Q3	Working experience developing ETL solutions, data workflows, pipelines and products in Java/Python or equivalent
Q4	Able to write basic to advanced SQL queries
Q5	Relevant experience: <ul style="list-style-type: none"> - AWS Data Infrastructure (Amazon Redshifts) - Hadoop (HDFS/MR/Yarn) - Hadoop File Formats & Compression (Parquet/ORCFile/Snappy/gzip) - SQL on Hadoop (Hive/SparkSQL/Impala) - NoSQL (BigTable/HBase/Cassandra) - RDBMS (MySQL/Postgresql) - Distributed processing engine (Spark/Flink) - Data Ingestion & Message Processing (RabbitMQ/ActiveMQ/ZeroMQ/Kafka/Flume) - Stream processing (Spark Streaming/Kafka)
Q6	Not scared of reading technical documentation or source code.
Q7	Ability to exercise independent, judgement when dealing with competing priorities and

	initiatives, while keeping all stakeholders informed of issues, progress, status and resolution.
--	--

Judul Course 1 : Fundamental Concept & Roadmap Big Data Engineer

Course ini bertujuan memberikan gambaran besar bagaimana pekerjaan Big Data Engineer, apa goal yang dicapai dari pekerjaan ini, bagaimana roadmap belajarnya agar bisa sampai sana.

Week 1

Section 1 : Why Big Data Engineer ? Why now?

- Materi : Masalah yang dipecahkan oleh Big Data Engineer sehingga kenapa butuh Big Data Engineer
- Materi : Big Data Engineer vs Data Scientist
- Materi : Gambaran pekerjaan seorang Big Data Engineer secara garis besar
- Materi : Peluang Karir Big Data Engineer
- Materi : Skill Apa saja yang harus dimiliki untuk menjadi seorang Big Data Engineer yang baik.
- Materi : Roadmap belajar as a Big Data Engineer yang paling efektif.

Section 2 : Lebih dalam memahami Proses Big Data Engineer

- Materi : Mendalami lagi desain arsitektur dan proses dari Big Data
- Materi : Data Warehouse vs Data Lake
- Materi : Kenapa Data Lake lebih make sense di era saat ini
- Materi : Mendalami proses Extract
- Materi : Mendalami proses Transform
- Materi : Mendalami proses Load
- Materi : Mendalami proses Automation
- Materi : Mendalami proses Building scalable and robust infrastructure
- Materi : Tools-tools Big Data Engineer untuk masing-masing fungsi dan proses.
- Materi : Kenapa Hadoop?

Section 3 : Download & Install Software Requirements

- Materi : Kenapa harus Linux Ubuntu dan kenapa harus terbiasa menggunakannya sehari-hari
- Materi : Instalasi Ubuntu 18.04 Dual Boot
- Materi : Perintah dasar Linux untuk instalasi aplikasi
- Materi : Instalasi Hadoop Standalone

Section 4 : Demo memahami proses ETL sederhana

- Materi : Ekstraksi data sampel data pembeli baju di amerika
- Materi : Tranform data sampel agar lebih rapi
- Materi : Load Data yang sudah jadi ke Postgresql
- Materi : Contoh Data Scientist mengolah data yang telah disimpan menjadi informasi grafik berharga untuk perusahaan.
- Materi : Penjelasan bahwa proses ini haruslah cepat, stabil, murah
- Materi : Penjelasan bahwa proses ini nantinya bisa diotomasi.
- Materi : Penjelasan bahwa materi-materi yang perlu dipelajari adalah programming Java & Database dulu.

Section 5 : How to be a Good Engineer

- Materi : Bad Engineer vs Good Engineer
- Materi : Cara bertanya yang baik agar bisa dijawab
- Materi : Cara googling yang baik
- Materi : Cara belajar yang baik
- Materi : Cara bertanya yang baik
- Materi : Attitude yang harus dimiliki oleh seorang good engineer
- Materi : Sebaiknya bisa nulis & baca bahasa inggris
- Materi : Have a dream!
- Tugas : Speaking Drill

Judul Course 2 : Java & Database Fundamentals

Course ini bertujuan agar murid sudah cukup terbiasa dengan coding dan sintaks SQL database sehingga tidak kesulitan saat nanti membuat berbagai script pada proses ETL.

Week 2-3

Section 6 : Basic Programming with Java

- Materi : Mengapa bahasa pemrograman komputer di butuhkan?
- Materi : Kesempatan bekerja menjadi programmer
- Materi : Pengenalan dasar-dasar pemrograman komputer
- Materi : Menginstall JDK atau JRE
- Materi : Install Default Package Manager
- Materi : Pengaturan Java Home Environment
- Materi : Install IntelliJ IDEA (Windows, Mac, Linux)
- Materi : Hands On membuat Hello world
- Tugas : Membuat Hello Nama menggunakan method dan class pada java
- Materi : Membuat projek java
- Materi : Objek, kelas, method, dan tipe data pada java
- Materi : Variabel dan akses modifier pada java
- Materi : Operator pada java

- Materi : Membuat aplikasi kalkulator dengan terminal menggunakan java
- Tugas : Melakukan speaking drill setiap hari sabtu

Section 7 : Lanjutan materi Programming Java - Part 1

- Materi : Kondisional IF, ELSE
- Materi : Hands On membuat aplikasi ganjil genap
- Materi : Kondisional Switch Case
- Materi : Hands On membuat aplikasi Umur (1-2 tahun -> bayi, 3-12 tahun -> anak-anak, 12-17 tahun -> remaja, >18 -> dewasa)
- Tugas : membuat aplikasi Hewan sekreatif mungkin (bersayap -> burung, belalai -> gajah)
- Materi : Array pada java
- Materi : Hands On membuat array list menggunakan java
- Materi : Kondisional menggunakan array list
- Materi : Hands On membuat aplikasi Data Barang menggunakan java
- Tugas : Membuat aplikasi Data Diri menggunakan java
- Tugas : Melakukan speaking drill setiap hari sabtu

Section 8 : Lanjutan materi Programming Java - Part 2

- Tugas : Membuat aplikasi cuaca(< 0 Derajat -> beku, 0 - 25 -> dingin, 25 - 40 -> normal, 40 - 60 -> hangat, > 60 panas) menggunakan java
- Materi : Dictionary (Map) pada java
- Materi : Kondisional menggunakan Dictionary
- Materi : Pengulangan For pada java
- Materi : Hands On membuat pengulangan angka pada pengulangan For
- Tugas : Membuat pengulangan segitiga bintang
- Materi : Pengulangan While pada java
- Materi : Hands On membuat aplikasi yang meminta inputan user secara berulang-ulang
- Tugas Besar : Membuat aplikasi perpustakaan sederhana yang meminta inputan user dan di simpan ke Array atau Map
- Tugas : Melakukan speaking drill setiap hari sabtu

Section 9 : Lanjutan materi Programming Java - Part 3

- Materi : Java dan data json
- Materi : Library untuk json
- Materi : Membuat data json
- Materi : Json Array dan Json Object
- Materi : Membuat json object didalam json object
- Materi : membuat didalam json object ada json array
- Materi : membuat didalam json array ada json object
- Tugas : Buat lah bentuk data json yang sangat kompleks / json nested

Week 4

Section 10 : Database

- Materi : Kenapa butuh mempelajari SQL Query sebagai seorang Big Data Engineer
- Materi : Pengenalan database, table, dan query
- Materi : Pengenalan Postgresql
- Materi : Pengenalan dasar-dasar Query pada SQL
- Materi : Hands On membuat database dan table kostumer, struk pembayaran, barang, supplier menggunakan PostgreSql
- Materi : Hands On membuat database, table kostumer dengan query PostgreSql
- Tugas : input data minimal 20 data ke setiap table
- Tugas : Menerapkan Create, Update, Delete pada kostumer menggunakan query postgresql
- Materi : Hands On membuat query yang lebih advanced tentang Select, min, max, where, average, count, order by dan lain-lain
- Tugas : membuat query select dengan case-case yang lebih kreatif.
- Materi : Hands on membuat query yang lebih advanced tentang Join untuk kostumer dan struk pembayaran, dan lain-lain
- Tugas : Hands on membuat query yang lebih advanced tentang Join untuk kostumer dan struk pembayaran dan barang dan supply

Small Project

Cilsylify merupakan perusahaan Music Streaming Apps mirip Spotify. Saat ini mereka ingin meng-hire seorang Data Engineer, yaitu Anda. Tapi mereka akan test dulu seberapa bagus skill database, Linux dan ngoding Anda. Anda diberikan database mentah berisi daftar Invoice, Pembelian, dan data User. Silahkan import database ini ke Postgresql di Linux Ubuntu 18.04, lalu buatlah program untuk bisa mengambil data-data berikut :

1. Daftar user yang sudah bayar
2. Daftar user yang sudah cetak invoice tapi belum bayar

Judul Course 4 : Extract, Transform, Load

Course ini bertujuan agar murid benar-benar paham bagaimana mindset utama proses ETL tanpa memikirkan dari sisi infrastruktur sama sekali. Fokus pada terbiasa ngoding di lokal komputer masing-masing berbagai study case proses ETL. Disini tools utama yang digunakan adalah Apache Spark.

Week 5-6

Section 11 : Memahami Pentingnya sinkronisasi dengan kebutuhan perusahaan

- Materi : Bagaimana proses ETL yang benar

- Materi : Memahami sebelum memulai proses ETL apapun, penting bahwa Big Data Engineer harus benar-benar paham apa yang ingin dicapai oleh perusahaan.
- Materi : Bagaimana kolaborasi dengan pemangku kebijakan di perusahaan
- Materi : Demo proses ETL dengan apache Spark

Section 12 : Dasar-dasar penggunaan Apache Spark agar bisa melakukan ETL sendiri

- Materi : Spark Core
- Materi : Penggunaan spark core
- Materi : Apa itu data RDD
- Materi : Apa itu data Tuple2
- Materi : Penggunaan map spark
- Materi : Penggunaan flatmap spark
- Materi : Penggunaan foreach spark
- Materi : Penggunaan reduce spark
- Materi : Penggunaan paralellize spark
- Materi : Penggunaan filter spark
- Materi : Penggunaan distinct spark
- Materi : Read file(txt,free text,csv,json) dengan spark core
- Materi : Fitur save data rdd

Section 13 : Dasar Apache Spark Session/SQL

- Materi : Spark Session/SQL
- Materi : Penggunaan spark session/sql
- Materi : Pembuatan dataset
- Materi : Transformation data menjadi dataset
- Materi : Integrasi spark dengan datasql (mysql,postgres,etc)
- Materi : Query sql menggunakan sparksql/sparksession
- Materi : Remapping data sparksql
- Materi : dataset to json
- Materi : SPARK UI

Section 14 : ETL in Depth for Relational Database

- Materi : Kenapa butuh Postgresql
- Materi : Instalasi Postgresql
- Materi : Praktek ETL sampel data CSV 1 - dengan beberapa study case
- Materi : Praktek ETL sampel data CSV 2 - dengan beberapa study case
- Materi : Praktek ETL sampel data JSON 1 - dengan beberapa study case
- Materi : Praktek ETL sampel data JSON 2 - dengan beberapa study case
- Materi : Praktek ETL sampel data Postgresql 1 - dengan beberapa study case
- Materi : Praktek ETL sampel data Postgresql 2 - dengan beberapa study case
- Materi : Praktek ETL sampel data API 1 - dengan beberapa study case
- Materi : Praktek ETL sampel data API 2 - dengan beberapa study case
- Materi : Praktek ETL sampel data Excel 1 - dengan beberapa study case

- Materi : Praktek ETL sampel data Excel 2 - dengan beberapa study case

Week 7-8

Section 15 : ETL in Depth for Non-Relational Database

- Materi : Understanding Unstructured vs Structured Data
- Materi : Understanding OLAP VS OLTP
- Materi : Kenapa butuh Apache HBase
- Materi : Instalasi Apache HBase
- Materi : Kapan harus menggunakan HBase dan kapan tidak
- Materi : Praktek ETL sampel data CSV 1 - dengan beberapa study case
- Materi : Praktek ETL sampel data CSV 2 - dengan beberapa study case
- Materi : Praktek ETL sampel data JSON 1 - dengan beberapa study case
- Materi : Praktek ETL sampel data JSON 2 - dengan beberapa study case
- Materi : Praktek ETL sampel data Postgresql 1 - dengan beberapa study case
- Materi : Praktek ETL sampel data Postgresql 2 - dengan beberapa study case
- Materi : Praktek ETL sampel data API 1 - dengan beberapa study case
- Materi : Praktek ETL sampel data API 2 - dengan beberapa study case
- Materi : Praktek ETL sampel data Excel 1 - dengan beberapa study case
- Materi : Praktek ETL sampel data Excel 2 - dengan beberapa study case

Small Project

Cilsylify meng-hire Anda menjadi seorang Data Engineer. Anda diminta untuk bisa mengumpulkan data-data penting yang bisa memberikan insight pada perusahaan terkait lagu-lagu yang paling sering diputar oleh pengguna dari beberapa data CSV, JSON, dan API.

Judul Course 8 : Membangun Data Warehouse & Data Mart

Course ini bertujuan agar murid memahami korelasi antara proses ETL yang sudah dilakukan sebelumnya dengan bagaimana menyimpan data tersebut di Data Warehouse. Disini mulai diberikan pemahaman bahwa proses ETL sebelumnya (yg hanya lokal) juga perlu disimpan di sebuah infrastruktur penyimpanan yang akan digunakan bersama-sama dengan orang lain. Skill ngoding ETL sebelumnya tetap akan banyak digunakan, serta ada sedikit skill pemahaman pembuatan infrastruktur. Disini juga akan diberikan pemahaman perbedaan bagaimana pembuatan Data Warehouse dengan Data Mart.

Week 9

Section 16 : Mengenal Hive sebagai Data Warehouse

- Materi : Penjelasan tujuan dari Course ini, serta kenapa proses ini sudah disebut membangun Data Warehouse.
- Materi : Mengenal Hive & Arsitekturnya
- Materi : Bagaimana arsitektur Hive, Spark, dan Hadoop bisa bersama.
- Materi : Perbedaan Data Warehouse dan Database
- Materi : Kapan menggunakan Hive kapan menggunakan Spark
- Materi : Instalasi Hive
- Materi : Basic command Hive

Section 17 : Praktek membangun Hive

- Materi : Lebih detail memahami cara kerja Hive
- Materi : Deep dive command Hive
- Materi : Praktek ETL sampel data CSV 1 - dengan beberapa study case
- Materi : Praktek ETL sampel data JSON 1 - dengan beberapa study case
- Materi : Praktek ETL sampel data Postgresql 1 - dengan beberapa study case
- Materi : Praktek ETL sampel data API 1 - dengan beberapa study case
- Materi : Praktek ETL sampel data Excel 1 - dengan beberapa study case

Section 18 : Spark dengan Hive

- Materi : Kenapa harus Spark dengan Hive?
- Materi : Mengkoneksikan Spark dengan Hive
- Materi : Basic cara kerja Spark jika bersama-sama dengan Hive
- Materi : Praktek ETL sampel data CSV 1 - dengan beberapa study case
- Materi : Praktek ETL sampel data JSON 1 - dengan beberapa study case
- Materi : Praktek ETL sampel data Postgresql 1 - dengan beberapa study case
- Materi : Praktek ETL sampel data API 1 - dengan beberapa study case
- Materi : Praktek ETL sampel data Excel 1 - dengan beberapa study case

Section 19 : Membuat Data Mart

- Materi : Perbedaan data mart dan data warehouse
- Materi : Fungsi data mart
- Materi : Arsitektur data mart
- Materi : Tahapan membuat data mart
- Materi : Praktek ETL Data Warehouse ke Data Mart 1
- Materi : Praktek ETL Data Warehouse ke Data Mart 2
- Materi : Praktek ETL Data Warehouse ke Data Mart 3

Judul Course 6 : Crawling Data Mentah

Pada course-course sebelumnya murid selalu disediakan sumber file data JSON, CSV dll yang sudah jadi. Sedangkan pada kenyataannya banyak startup-startup baru yang sama sekali tidak punya data internal. Benar-benar nol tidak punya data apapun. Course ini

bertujuan agar murid paham bagaimana mindset dan cara mengumpulkan data dari sumber mana saja. Sehingga startup yang sama sekali tidak punya data, menjadi bisa punya data untuk di olah.

Week 10-11

Section 20 : Praktek Dasar-dasar berbagai teknik Crawling

- Materi : Apa itu Crawler/Crawling
- Materi : Apa itu Scraper/Scraping
- Materi : Apa itu Selenium
- Materi : Apa itu Rest-API
- Materi : Praktek crawler
- Materi : Praktek scraper
- Materi : Praktek selenium
- Materi : Praktek get by rest-API

Section 21 : Study Case berbagai teknik Crawling

- Materi : Study Case buat engine crawler dan masukkan ke Data Warehouse 1
- Materi : Study Case buat engine crawler dan masukkan ke Data Warehouse 2
- Materi : Study Case buat engine crawler dan masukkan ke Data Warehouse 3
- Materi : Study Case buat engine crawler dan masukkan ke Data Warehouse 4

Judul Course 7 : Membangun Data Lake

Course ini bertujuan agar murid mulai memahami bahwa ekosistem Big Data itu luas. Tidak hanya menyimpan data-data yang sifatnya cenderung statik, tapi juga menyimpan data-data yang realtime dan super besar. Skill ngoding ETL tetap akan banyak digunakan disini, dan murid harus paham perbedaan besar bagaimana mengolah data statik dengan data realtime sehingga jika disatukan akan terwujud sebuah Data Lake yang bagus.

Week 12-13

Section 22 : Memahami ekosistem Big Data yang lebih luas

- Materi : Data Lake vs Data Warehouse vs Data Mart
- Materi : Stream Processing vs Volume Processing
- Materi : Kenapa Data Lake lebih relevant di era saat ini
- Materi : Alur Data Lake yang biasa ada di industri
- Materi : Sedikit penjelasan komponen-komponen yang ada

Section 23 : Apache Kafka sebagai Real Time Data Processing

- Materi : Apa itu Apache Kafka? Kapan butuh menggunakannya?
- Materi : Instalasi Apache Kafka & Start Kafka

- Materi : Topics, Partitions and Offsets
- Materi : Brokers and Topics
- Materi : Topic Replication
- Materi : Producers and Message Keys
- Materi : Consumers & Consumer Groups
- Materi : Consumer Offsets & Delivery Semantics
- Materi : Kafka Broker Discovery
- Materi : Zookeeper
- Materi : Kafka Guarantees
- Materi : Kafka CLI

Section 24 : Apache Spark Streaming & Kafka bersama

- Materi : Apa itu Spark Streaming? Kapan butuh menggunakannya?
- Materi : Bagaimana Spark Streaming & Kafka digunakan bersama
- Materi : Kapan menggunakan Spark Streaming bersama, kapan tidak.
- Materi : Instalasi dan basic cara penggunaannya
- Materi : Spark Streaming project : Twitter - Study case 1
- Materi : Spark Streaming project : Twitter - Study case 2
- Materi : Spark Streaming project : Facebook - Study case 1
- Materi : Spark Streaming project : Facebook - Study case 2

Small Project

Cilsylify merupakan perusahaan Music Streaming Apps mirip Spotify. Cilsylify adalah startup yang baru saja dibentuk, sehingga data-data yang mereka punya masih sangatlah kurang. Mereka perlu untuk mendapatkan sumber-sumber data mentah dari manapun utk memperkaya data yang mereka miliki. Selain itu karena perkembangan jaman yang sangat pesat, tidak mungkin Cilsylify hanya mengandalkan data-data statik yang cenderung lama utk mengumpulkannya. Saat ini kebutuhan mereka adalah untuk bisa mendapatkan semua data apapun saat ini juga, secara realtime (baik yang penting dan tidak penting) terkait tren masa kini anak muda yang mendengarkan musik. Anda diminta untuk mengumpulkannya.

Judul Course 10 : Otomasi Proses Data Pipeline

Course ini bertujuan agar murid paham bahwa proses sebelumnya yang serba manual akan sangat repot jika dikerjakan terus menerus. Sehingga disini akan diberikan cara bagaimana cara mengotomasi proses-proses manual yang telah dipelajari sebelumnya.

Week 14-15

Section 25 : Mengenal Apache Airflow

- Materi : Kenapa butuh mengotomasi proses data pipeline?
- Materi : Kenapa menggunakan Airflow?

- Materi : Lalu apa itu Airflow?
- Materi : Bagaimana cara kerja Airflow?
- Materi : Instalasi Airflow
- Materi : Penjelasan Airflow UI
- Materi : Penjelasan Airflow CLI

Section 26 : Membuat Data Pipeline

- Materi : Apa itu DAG?
- Materi : Coding DAG pertama kita
- Materi : Apa itu Operator?
- Materi : Coding Operator
- Materi : Operator Relationships dan Bitshift Composition
- Materi : Menambahkan dependencies
- Materi : Bagaimana cara kerja Scheduler?
- Materi : Bermain dengan Backfill dan Catchup
- Materi : Apa itu Workflow?

Section 27 : Database & Executors

- Materi : Local Executor dengan PostgreSQL
- Materi : Konfigurasi DAG dengan PostgreSQL
- Materi : Celery Executor dengan PostgreSQL dan Kafka
- Materi : Konfigurasi DAG dengan Celery Executor, PostgreSQL and RabbitMQ

Section 28 : Advanced Concept in Airflow

- Materi : Meminimalisir Repetitive Patterns dengan SubDAGs
- Materi : Praktek Meminimalisir Repetitive Patterns dengan SubDAGs
- Materi : Bagaimana berinteraksi dengan External Sources menggunakan Hooks
- Materi : Praktek mendapatkan data dari PostgreSQL menggunakan Hooks
- Materi : Bagaimana cara share data antara Tasks dengan XCOM
- Materi : Praktek Bagaimana cara share data antara Tasks dengan XCOM
- Materi : Cara eksekusi Tasks berdasarkan kriteria menggunakan Branching
- Materi : Praktek membuat Conditional Task menggunakan Branching
- Materi : Mengontrol Tasks dengan SLAs
- Materi : Praktek mendefinisikan SLA dalam DAG

Small Project

Cilsylify merupakan perusahaan Music Streaming Apps mirip Spotify. Cilsylify melihat kinerja Anda sangat lambat karena terlalu banyak pekerjaan yang manual. Jadi anda harus mengotomasi seluruh data pipeline yang ada yang sudah anda buat sebelumnya, agar lebih efektif.

Judul Course 6 : Big Picture of Big Data in Cloud

Sejauh ini murid lebih banyak fokus pada sisi ngoding. Pada Course ini murid akan lebih melihat secara big picture dari sisi skill arsitektur infrastrukturnya itu sendiri. Minimal paham dan punya gambaran seperti apa infrastruktur data yang baik untuk skala industri, dan memiliki mindset paham “kenapa” menggunakan tools ini, “kenapa” menggunakan tools itu. Misalnya paham kenapa tidak pakai Cloudera saja dibanding harus membuat infrastruktur sendiri? Kenapa tidak pakai S3 dan amazon redshift saja dibanding simpan data di harddisk server sendiri?

Week 16

Section 29 : Mengenal arsitektur Big Data di Cloud

- Materi : Problem yang ada pada Hadoop & On-Premise Infrastruktur
- Materi : Kenapa harus di Cloud?
- Materi : Beberapa contoh alternatif komponen Hadoop yang lebih efisien & murah di Cloud
- Materi : Contoh beberapa service Cloud besar yang ada (Cloudera, dll)
- Materi : Mindset yang harus dimiliki dalam memilih suatu service
- Materi : Contoh gambaran best practice arsitektur Big Data di Cloud - study case 1
- Materi : Contoh gambaran best practice arsitektur Big Data di Cloud - study case 2
- Materi : Contoh gambaran best practice arsitektur Big Data di Cloud - study case 3

Section 30 : Simple Practice AWS S3 & Amazon Redshifts

- Materi : Problem data warehouse, data mart & data lake pada Hadoop
- Materi : Kenapa AWS S3 & Amazon Redshifts bisa menjadi solusi
- Materi : Praktek membuat akun AWS
- Materi : Praktek membuat EC2
- Materi : Praktek membuat S3
- Materi : Praktek membuat Amazon Redshifts
- Materi : Praktek mengintegrasikan salah satu study case yang sudah dilakukan sebelumnya jika menggunakan EC2, S3, dan Amazon Redshifts.

Section 31 : Extras : Memanfaatkan Elasticsearch & Kibana Tools

- Materi : Apa itu Elasticsearch kenapa kita bisa memanfaatkannya
- Materi : Apa itu Kibana tools kenapa kita bisa memanfaatkannya
- Materi : Pembuatan index di elasticsearch
- Materi : Insert data to index elasticsearch
- Materi : Spark integrasi dengan elasticsearch
- Materi : Penggunaan kibana tools
- Materi : Study case pemanfaatan Elasticsearch & Kibana tools 1
- Materi : Study case pemanfaatan Elasticsearch & Kibana tools 2

Final Project

Cilsylify merupakan perusahaan Music Streaming Apps mirip Spotify. Anda diminta untuk benar-benar membuat seluruh data infrastruktur dari awal sesuai yang sudah dipelajari dan running dengan baik. Ada nilai tambah jika Anda bisa memberikan saran + alternatif infrastruktur di Cloud kepada Cilsylify.

Disini diharapkan semua komponen yang sudah dipelajari harus bisa running dengan baik, seperti :

1. Sudah Bisa membuat Engine Crawling
2. Berhasil menyimpan data mentah ke HDFS
3. Bisa melakukan batch data, data modelling, dan mapping data menggunakan Spark
4. Realtime processing dengan Kafka dan Spark Streaming
5. Pembuatan Data Lake
6. Pembuatan Data Warehouse dan Data Mart dengan analisis data lebih dalam dari Data Lake yang ada.
7. Otomasi Workflow dengan Apache Airflow
8. Bonus jika bisa memanfaatkan Cloud dan Elasticsearch dan Kibana.