IYKRA
Data Fellowship Program

(Deep Dive into Modern Data Engineering Landscape)

by (Bangun Sasongko)

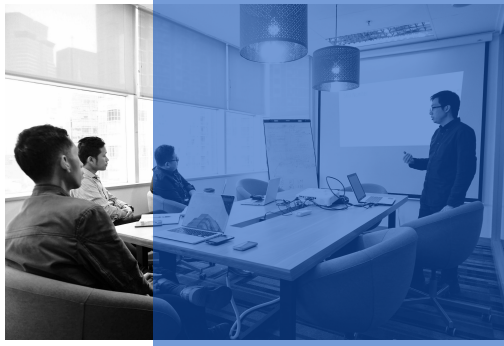
(date of delivery)

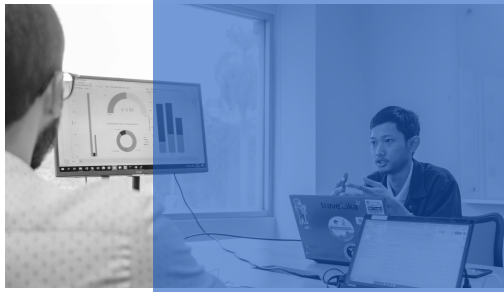
(Section 1)

Proprietary of IYKRA - Data Fellowship

Trainer Profile

Bangun W. Sasongko
Occupation
Data Engineer at BRI
Ex. Devops Engineer
Bachelor of Physics, Electrical and Instrumentations
@sasongkobgn
sasongkobgn@gmail.com





Proprietary of IYKRA - Data Fellowship

Table of Content

About this Course

This course delves into the fundamental responsibilities of data engineers, highlighting distinctions from data scientists, and elucidates the seamless data flow within an organization. Through practical exercises, participants will explore the processes employed artificial banking data from Kaggle the data created from BankSim (is an agent-based simulator of bank payments), to comprehend how their data engineers gather, clean, and organize data. Upon completing the course, participants will gain insight into the roles of their company's data engineers, be equipped to engage in meaningful conversations with them, and possess a solid foundation to embark on their own journey in data engineering.
Data Engineering and Big Data
Data Engineer vs Data Scientist
Data Pipeline

Proprietary of IYKRA - Data Fellowship

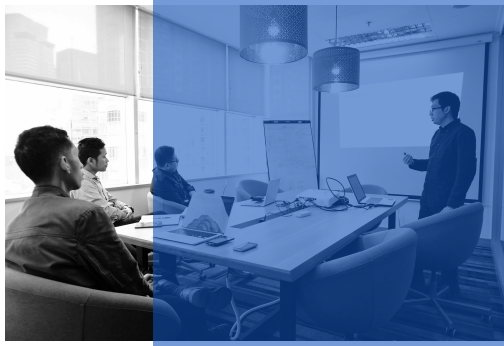
The Objectives

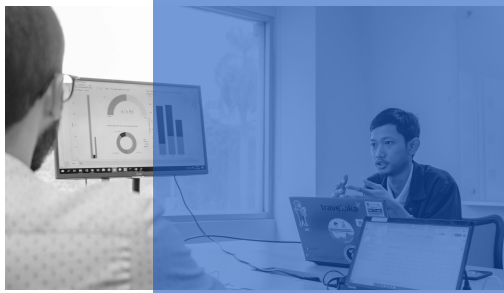
By the end of this course, you will be able to:

1. You will acquire an understanding of data engineering and the reasons behind the growing demand for professionals in this field

2. You will explore the position of data engineering within the data science lifecycle, discern the distinctions between data engineers and data scientists

3. You will receive an introduction to constructing your initial end-to-end data pipeline



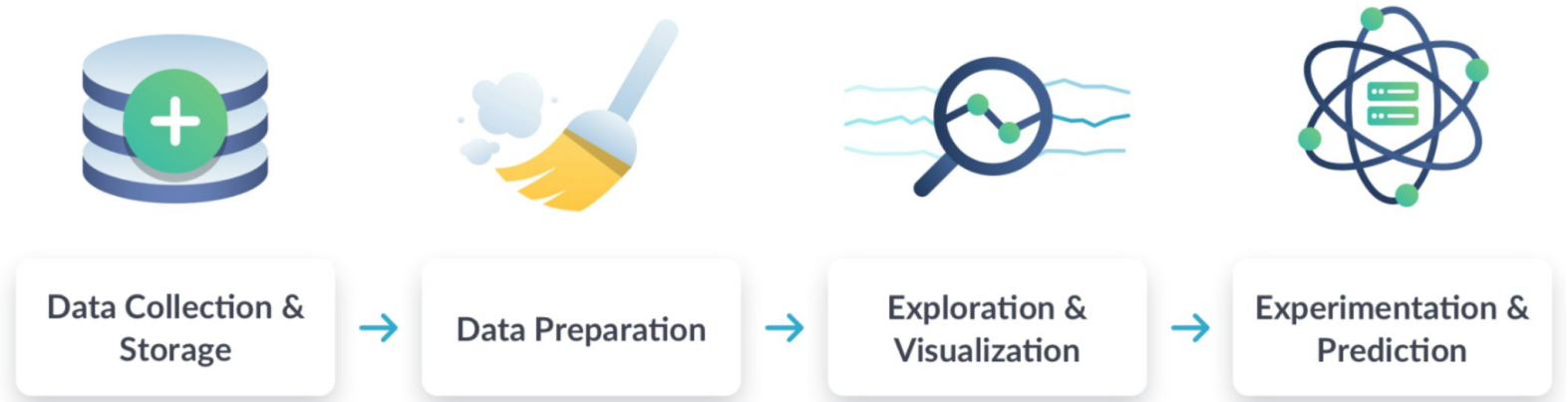


Proprietary of IYKRA - Data Fellowship

Data Fellowship

Data engineering and big data

Data Workflow



Proprietary of IYKRA - Data Fellowship

Data Engineer

Data engineers deliver:

- the correct data
- in the right form
- to the right people
- as efficiently as possible

Proprietary of IYKRA - Data Fellowship

A data engineer’s responsibilities

- Ingest data from different sources
- Optimize databases for analysis
- Remove corrupted data
- Develop, construct, test and maintain data architectures

Proprietary of IYKRA - Data Fellowship

Data engineers and big data

Big data becomes the norm => data engineers are more and more needed

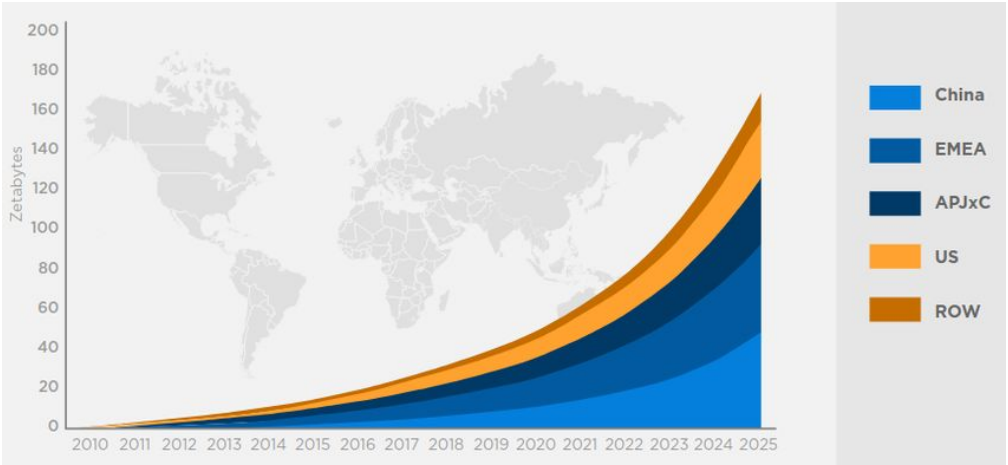
Big data:

- Have to think about how to deal with its size
- So large traditional methods don't work anymore

Proprietary of IYKRA - Data Fellowship

Big data growth

- Sensors and devices
- Social media
- Enterprise data
- VoIP (voice communication, multimedia sessions)



The five Vs

- Volume (how much?)
- Variety (what kind?)
- Velocity (how frequent?)
- Veracity (how accurate?)
- Value (how useful?)

Proprietary of IYKRA - Data Fellowship

Data engineer and big data: Summary

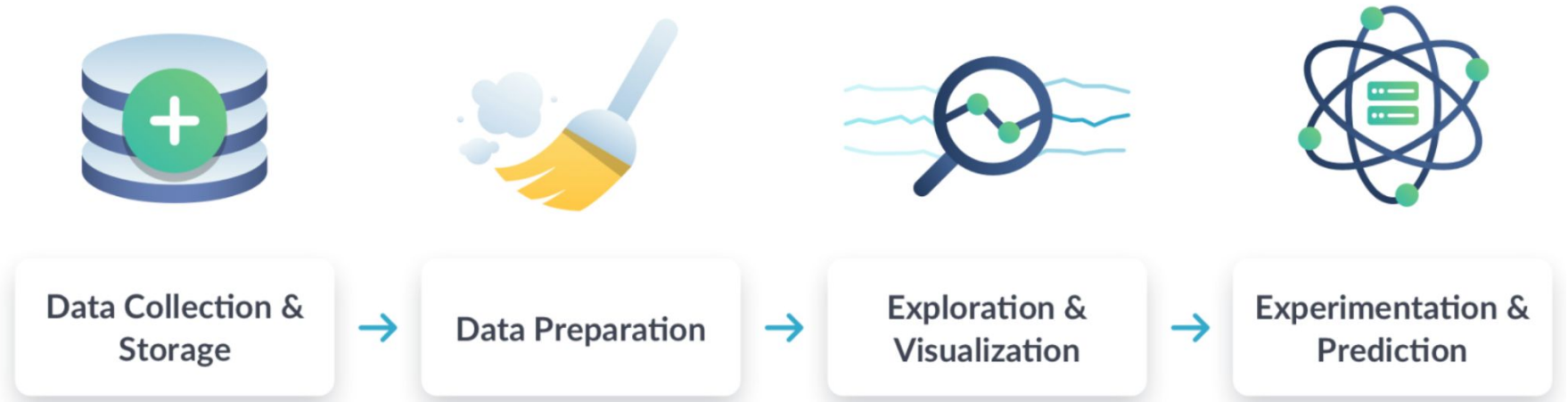
- What's waiting for you
- How data flows through an organization When a data engineer intervenes
- What their responsibilities are
- How data engineering relates to big data

Proprietary of IYKRA - Data Fellowship

Data Fellowship

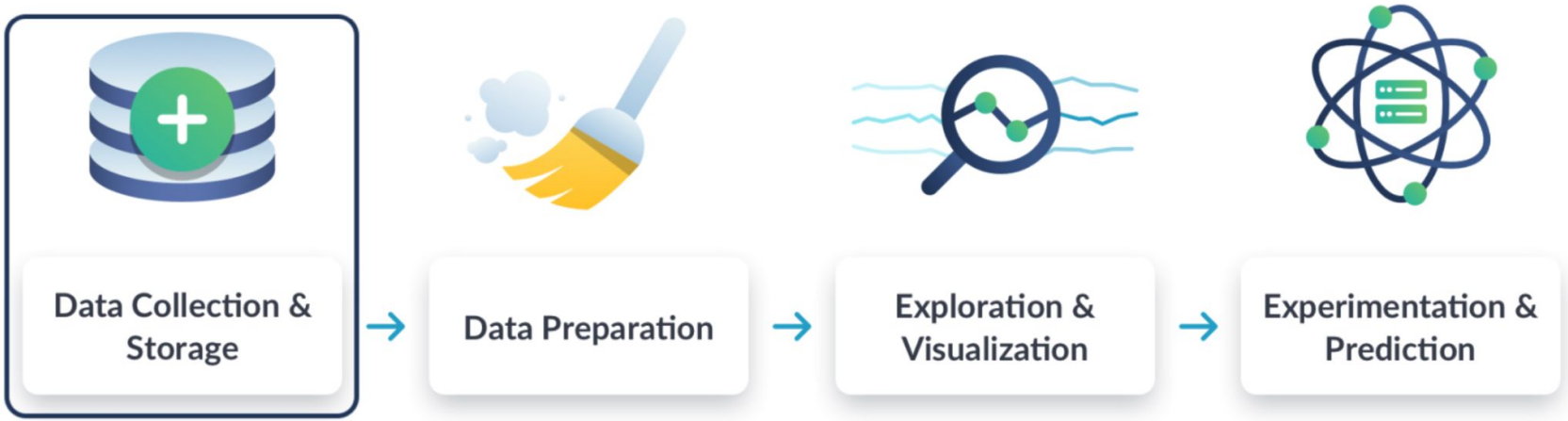
Data Engineer vs Data Scientist

Data workflow



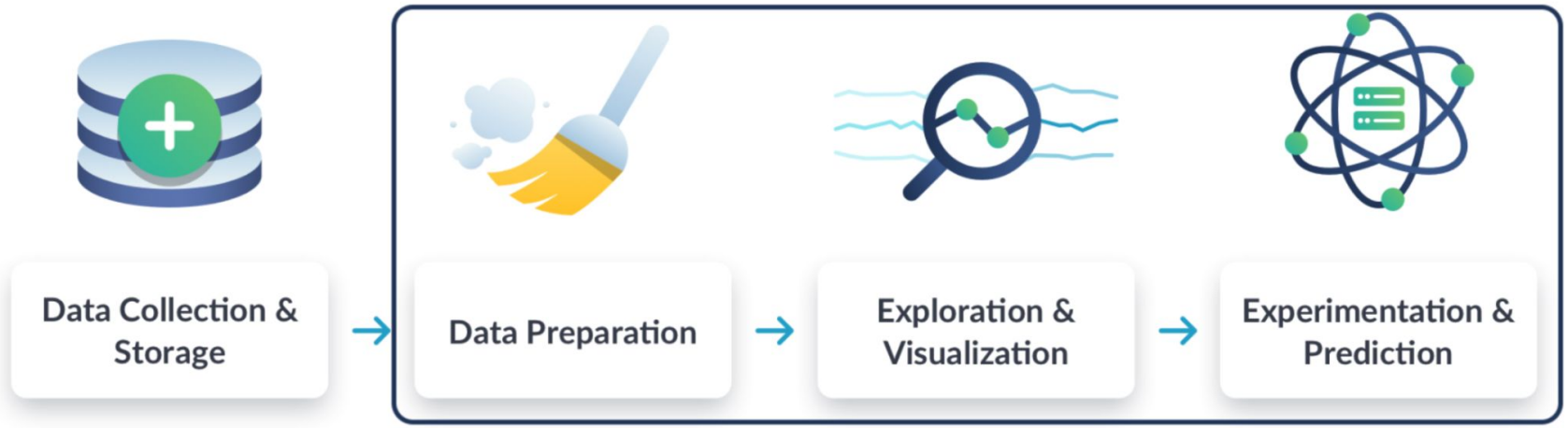
Proprietary of IYKRA - Data Fellowship

Data engineer



Proprietary of IYKRA - Data Fellowship

Data scientist

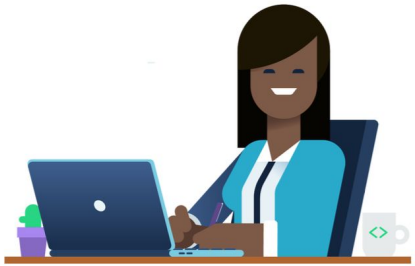


Proprietary of IYKRA - Data Fellowship

Data engineers enable data scientists

Data engineer

- Ingest and store data
- Set up databases
- Build data pipelines Strong software skills



Data scientist

- Exploit data
- Access databases
- Use pipeline outputs
- Strong analytical skills

Proprietary of IYKRA - Data Fellowship

Data engineer vs data scientist: Summary

- At which stages data engineers and data scientists intervene
- How data engineers enable data scientists

Proprietary of IYKRA - Data Fellowship

Data Fellowship

The data pipeline

If data is new oil

The Economist, 2017-05-06, by David Parkins

Proprietary of IYKRA - Data Fellowship

Proprietary of IYKRA - Data Fellowship

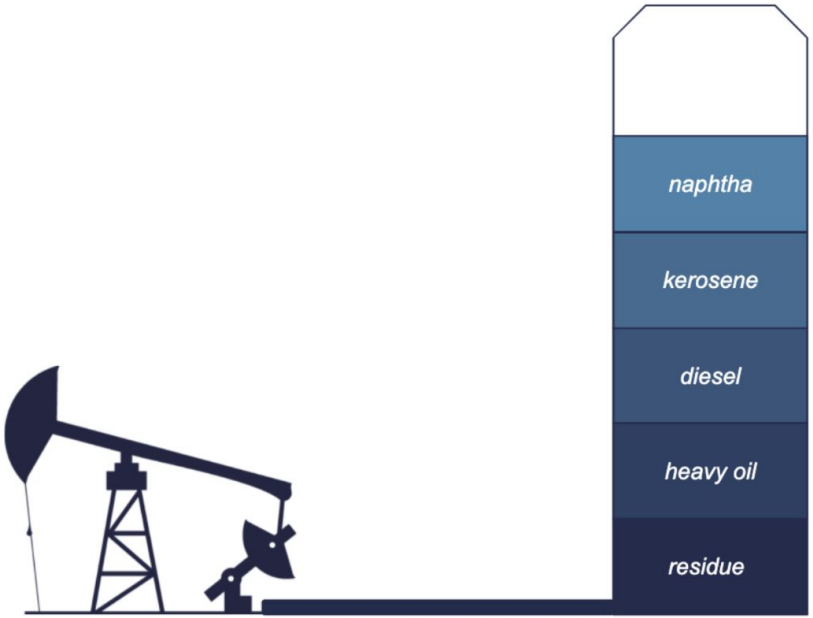
Proprietary of IYKRA - Data Fellowship

Proprietary of IYKRA - Data Fellowship

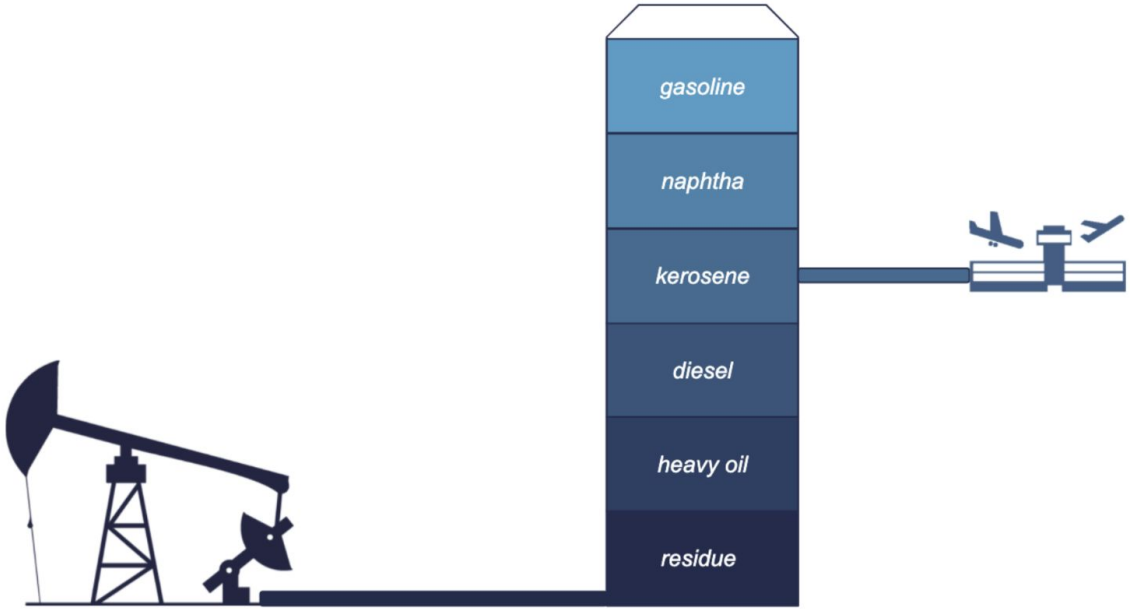
Proprietary of IYKRA - Data Fellowship

Proprietary of IYKRA - Data Fellowship

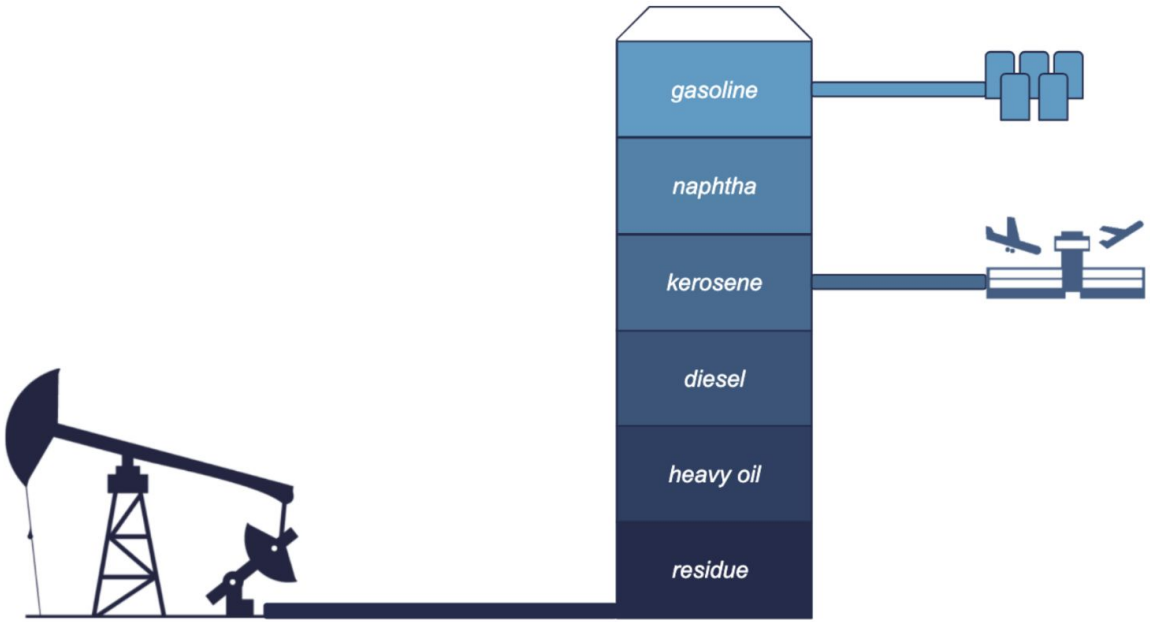
Proprietary of IYKRA - Data Fellowship



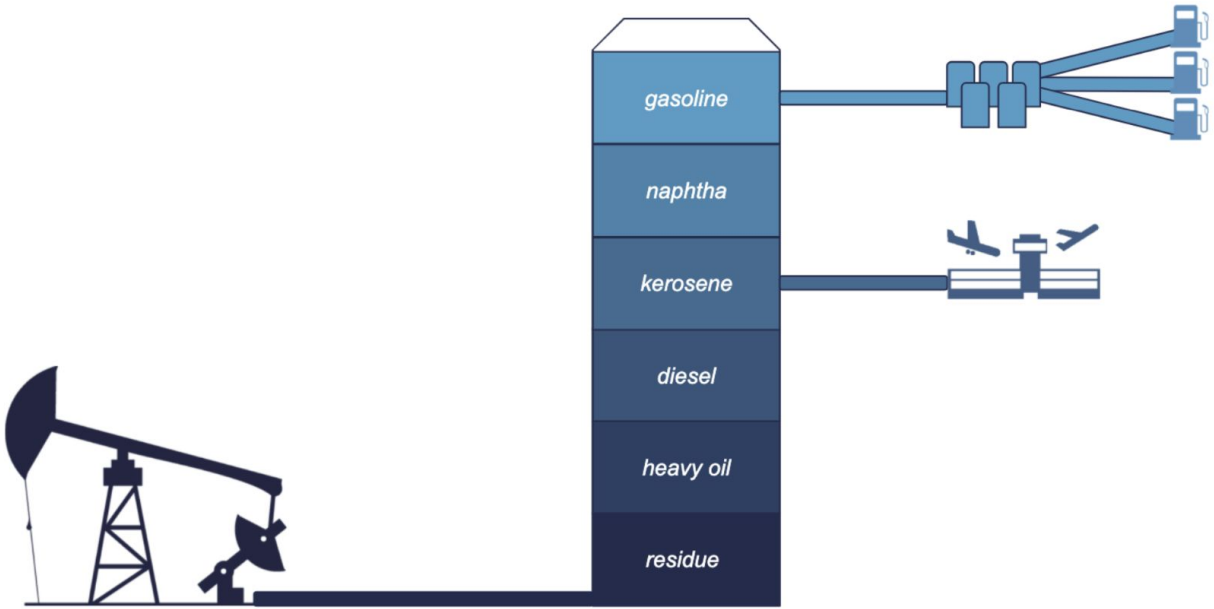
Proprietary of IYKRA - Data Fellowship



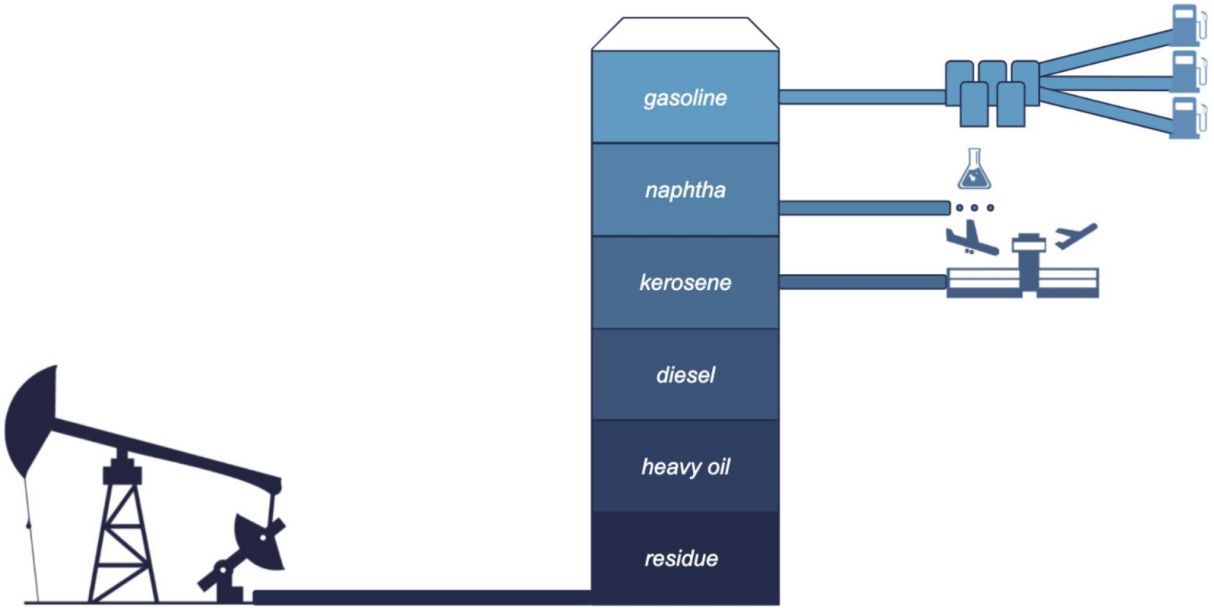
Proprietary of IYKRA - Data Fellowship



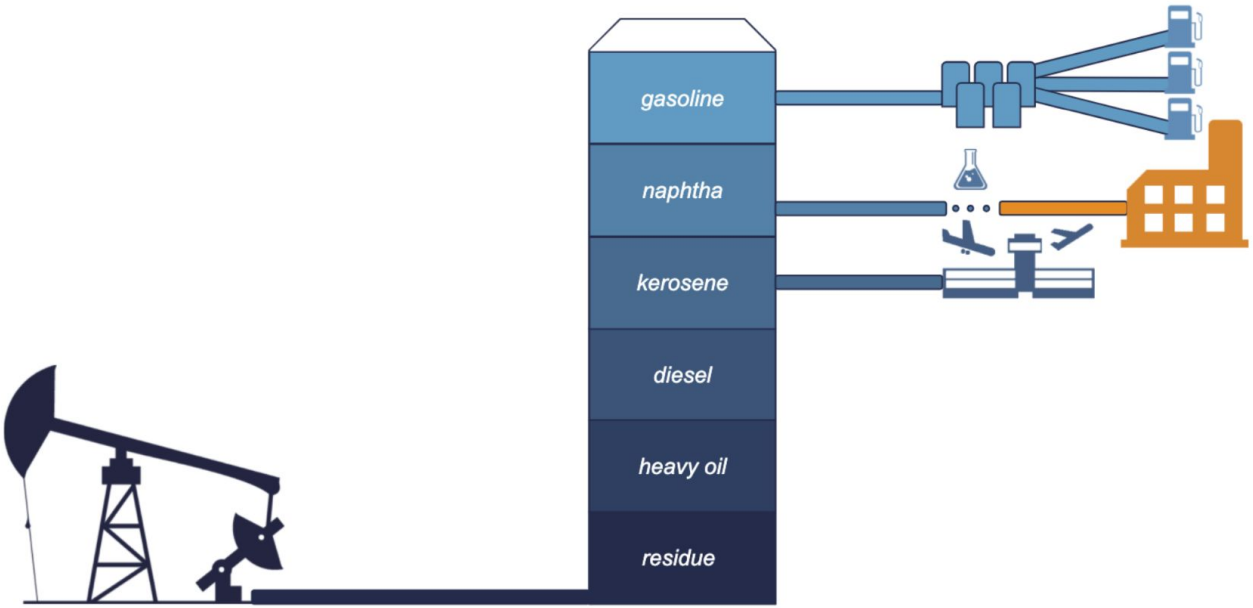
Proprietary of IYKRA - Data Fellowship



Proprietary of IYKRA - Data Fellowship



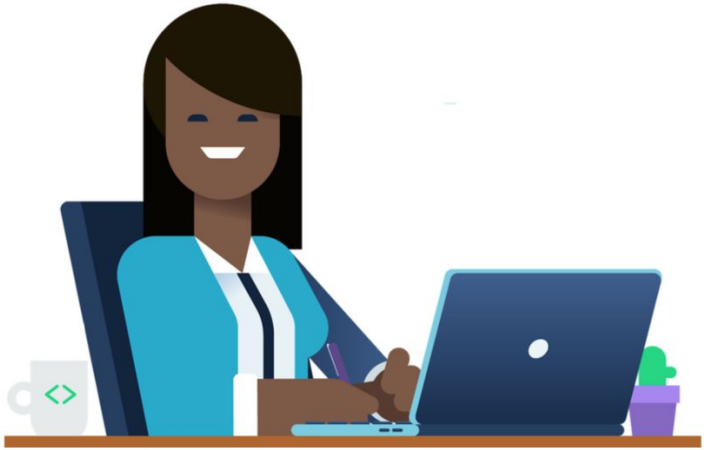
Proprietary of IYKRA - Data Fellowship



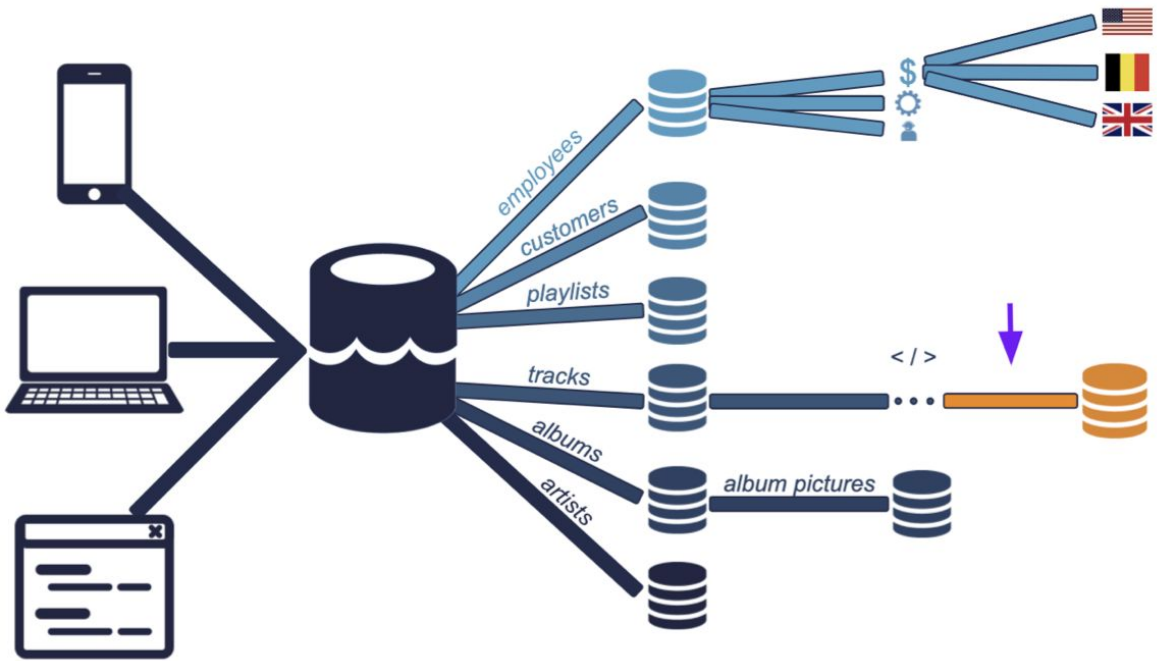
Proprietary of IYKRA - Data Fellowship

Back to data engineering

- Ingest
- Process
- Store
- Need pipelines
- Automate flow from one station to the next
- Provide up-to-date, accurate, relevant data



Back to data engineering



Proprietary of IYKRA - Data Fellowship

Data Pipeline

Proprietary of IYKRA - Data Fellowship

Data pipelines ensure an efficient flow of the data

Automate

- Extracting
- Transforming
- Combining
- Validating
- Loading

Reduce

- Human intervention
- Errors
- Time it takes data to flow

Proprietary of IYKRA - Data Fellowship

ETL and data pipelines ETL

ETL

Popular framework for designing data pipelines

- Extract data
- Transform extracted data
- Load transformed data to another database

Data pipelines

- Move data from one system to another
- May follow ETL or ELT
- Data may not be transformed
- Data may be directly loaded in applications

Proprietary of IYKRA - Data Fellowship

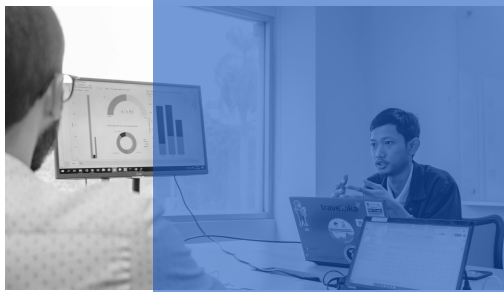
Data pipeline: Summary

- What a data pipeline is
- What it does
- Why it's important
- How data pipelines are implemented at Spotflix What ETL is and its nuances

Proprietary of IYKRA - Data Fellowship

References

Datacamp : understanding data engineering



Proprietary of IYKRA - Data Fellowship

iykra Data Fellowship

Thank you!

nothin’ is impossible until its done