

First Module

Modern Data Engineering:

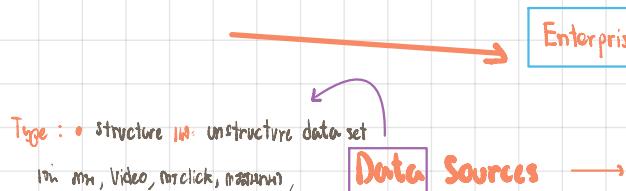
Data និងបច្ចេកទេស សម្រាប់ប្រើប្រាស់ tools និងការអនុវត្តន៍ទៅការ
និងការប្រើប្រាស់ សម្រាប់ប្រើប្រាស់ (Consumption), និងការរួមចូលរួមជាមួយ
នឹងការប្រើប្រាស់ទិន្នន័យ និងការប្រើប្រាស់ទិន្នន័យដែលបានបង្កើតឡើង

ផ្តល់ព័ត៌មានទិន្នន័យ

Accuracy និង data

និងការប្រើប្រាស់នូវការប្រើប្រាស់ Data ដើម្បីបង្កើតនិងប្រើប្រាស់

Reading step



Enterprise Data Environment

- Data engineer (integrated) និងអាជីវកម្ម
- ពន្លាអនុវត្តន៍ ឬ skills នាយកធម្មត សំខាន់សំខាន់ (Insight) (in consumer)
- និត្យភាពនិងរាយការនៃ stakeholder នូវទិន្នន័យ និងការប្រើប្រាស់
- Tools, Applications ឬ Infrastructures (infrastructure) និងធនធាន, បច្ចនា
- ឬ (disseminate) data នូវការប្រើប្រាស់

User



source

→ Data Repository

Skill Data Engineer ត្រូវ:

- DE: ឲាយទាន់ទិន្នន័យ ឬ data និងការប្រើប្រាស់ទិន្នន័យ
- DA: ឲាយ data និងការប្រើប្រាស់ទិន្នន័យ (insight)
- DS: ឲាយ data នៃ DE ឬ DA ដើម្បី predict នូវការងារ data ទូទៅ
- BA ឬ BI ឲាយ insight ឬ prediction នូវការងារ នៅក្នុងការប្រើប្រាស់

The field of DE involves:

Data Collecting → Processing Data
Marking Data ← Storing Data

available to user

securely (administer)

- Collecting source data
 - ឈរិញ្ញា (Extracting), ស្វែង (integrating), និងរិន (organize) និងប្រើប្រាស់នាមីនាមី
 - ទិន្នន័យ (acquisition Data) និងរិនិក
 - Architecture Data និងឈរិញ្ញា

• Storing Data

- រកបានការងារក្នុងការប្រើប្រាស់នូវការប្រើប្រាស់ទិន្នន័យ
- រួម Data នៃ storage នូវ Processed Data
- Scalable system
- និងការងារ data នៃគុណភាព, ឱ្យការ, compliance (compliance), monitoring (consistency), back up and recovering

• Processing Data

- Cleansing, Transformation ឬ ការចែកចាយទិន្នន័យទិន្នន័យ
 - Distributed System នូវ Data នៃ data
 - Pipeline ឬ Extracting, Transforming ឬ load data
 - គឺជាការរំលែកការងាររបស់ក្រុមហ៊ុន, នាយកដែលត្រូវការងារ និងការប្រើប្រាស់ទិន្នន័យ
 - metrics: និងការងារ
 - និងការងារ (Adherence) និងការងារ (Compliance) ក្នុង line

• Making Data available to user securely

- API, services, ឬ programs និងការងារទិន្នន័យ
- User និងការងារ interface នៃ dashboard
- និងការងារ balance និងការងារ data និងការងារ

Conclusion

The field of Data Engineering:

- Provides a robust and scalable structure to make quality data available for decision-making
- Includes the tools and technologies involved in data manipulation
- Involves understanding the complexities of data and how it is leveraged for fact-finding and decision-making

7pm, 7/06

• Goal of DE

ត្រូវ data នៃការងារទិន្នន័យ និងការងារទិន្នន័យ និងការងារទិន្នន័យ

និងការងារទិន្នន័យ

In this lesson, you have learned:

Modern data ecosystem includes a network of interconnected and continually evolving entities that include:

- Data, that is available in a host of different formats, structures, and sources.
- Enterprise Data Environment, in which raw data is staged so it can be organized, cleaned, and optimized for use by end-users.
- End-users, such as business stakeholders, analysts, and programmers who consume data for various purposes.

Emerging technologies such as Cloud Computing, Machine Learning, and Big Data, are continually reshaping the data ecosystem and the possibilities it offers.

Data Engineers, Data Analysts, Data Scientists, Business Analysts, and Business Intelligence Analysts, all play a vital role in the ecosystem for deriving insights and business results from data.

The goal of Data Engineering is to make quality data available for analytics and decision-making. And it does this by collecting raw source data, processing data so it becomes usable, storing data, and making quality data available to users securely.

Functional Skills of a Data Engineer:



Convert business requirements into technical specifications



Work with the complete software development lifecycle
Ideation -> Architecture -> Design -> Prototyping -> Testing ->
Deployment -> Monitoring



Understand data's potential application in business



Understand risks of poor data management
Data Quality | Data Privacy | Security | Compliance

Conclusion

Data engineering requires a broad set of skillsets.

You need to select one or more specialization areas, but have a good understanding of all areas, so you can make more informed decisions.

Your skills will grow over time with experience, the areas you choose to focus on, and the time you invest in upskilling yourself.

In this lesson, you have learned:

The role of a Data Engineer:

- Gathering data from disparate sources.
- Integrating data into a unified view for data consumers.
- Preparing data for analytics and reporting.
- Managing data pipelines for a continuous flow of data from source to destination systems.
- Managing the complete infrastructure for the collection, processing, and storage of data.

To be successful in their role, Data Engineers need a mix of technical, functional, and soft skills.

Oj

- Technical Skills include working with different operating systems and infrastructure components such as virtual machines, networks, and application services. It also includes working with databases and data warehouses, data pipelines, ETL tools, big data processing tools, and languages for querying, manipulating, and processing data.
- An understanding of the potential application of data in business is an important skill for a data engineer. Other functional skills include the ability to convert business requirements into technical specifications, an understanding of the software development lifecycle, and the areas of data quality, privacy, security, and governance.
- Soft Skills include interpersonal skills, the ability to work collaboratively, teamwork, and effective communication.

1. Which one of these functional skills is essential to the role of a Data Engineer?

- Proficiency in Mathematics
 The ability to work with the software development lifecycle
 Inspect analytics-ready data for deriving insights
 Proficiency in working with ETL Tools

Correct

As a Data Engineer, you will be required to work through different phases of the software development lifecycle, which includes, ideation, architecture, design, prototyping, testing, deployment, and monitoring.

3. To ensure business stakeholders can see real-time data each time they log into the dashboard, Sarah decided to build _____ to extract, transform, and load data on an ongoing basis.

- A Python program
 A sentiment analysis algorithm
 APIs
 A Data Pipeline

Correct

Data pipelines cover the journey of data from source to destination systems which include extracting, transforming, and loading data.

2. Oracle Exadata, IBM Db2 Warehouse on Cloud, IBM Netezza Performance Server, and Amazon RedShift are some of the popular _____ in use today.

- Big Data Platforms
 ETL Tools
 NoSQL Databases
 Data Warehouses

Correct

These are some of the popularly used data warehouses.

Second Module

Ecosystem:

Ecosystem នៃ Data Engineer ត្រូវរាយការងារ Infrastructure (Toolset), tool, framework និង resources ទាំងអស់

- Extracting Data ពី source ទីមួយ
- Architecting នៃ data និង pipeline នៅ transformation, integration, និង storage (storage of data)
- Architecting នៃ infrastructure
- Automating ឬ Optimizing workflow នៃ infrastructure នៃ system
- running Application នៃការងារបច្ចុប្បន្ននៃ DE

Data Repository :

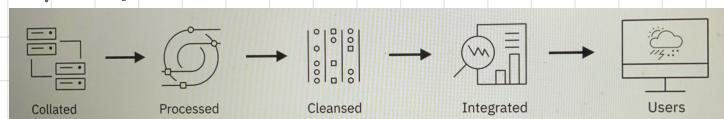
1. Transactional នៃ Online Transaction Processing (OLTP) systems

- បញ្ជាក់នូវទិន្នន័យ Data និងតម្លៃ
 - ឬ ATM Transaction, Online Banking Transaction
 - ឬ Air-line booking
- ផ្តល់ព័ត៌មាន Data ដោយ Rational សំគាល់ (primary key)

2. Analytical នៃ Online Analytical Processing (OLAP) systems

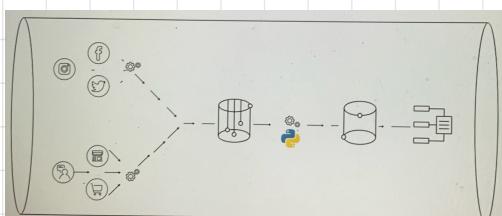
- បញ្ជាក់នូវទិន្នន័យបានបង្ហាញឡើងទៅ data និងនាយករដ្ឋមន្ត្រី
- បន្ថែមទិន្នន័យ Relational ឬ Non-Relational data bases
 - នាម Data Warehouses, Data Marts, Data lakes ឬ Big Data stores

Data Integration:



- នូវការ collect data នៃរាយការណ៍ នៅក្នុងពីរភេទ ឬជាអាមេរិក, សំគាល់ (query)
 - ឬការរាយការណ៍

Data Pipeline:



Types of Data :

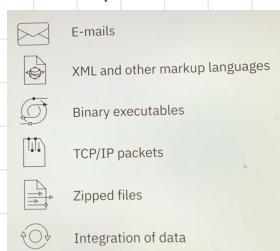
Structured

- Data នឹងមកក្នុងការណែនាំ (rigid) និងក្នុង row ឬ column
- ផ្តល់ព័ត៌មាន Facts ឬ Number ស្ថិតិយោគ និងតម្លៃ collected, Exported, stored
- Organized SOURCE នៃ structure



Semi- Structured

- នឹង mixនូវទិន្នន័យទូទៅនិងទិន្នន័យតាមតារាង
 - ទិន្នន័យក្នុង row ឬ column
 - ផ្តល់ព័ត៌មាន tags, events, events, និង metadata និងផ្តល់ព័ត៌មាន group data និងពីរបាលាដុំ (hierarchy)
 - Email → sender ឬ receiver នៃ information
 - និងទិន្នន័យ structure
 - ឬ content និង text និង XML, pdf នៃព័ត៌មាន unstructured



Unstructured

- ឬ Data នឹងមកក្នុងការណែនាំទូទៅនិងការណែនាំ (Qualitative Information) និងតម្លៃទិន្នន័យ row ឬ column និង
 - នូវ text file, doc, Video, pdf, social media content

