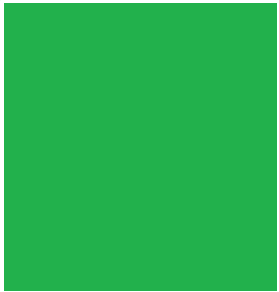
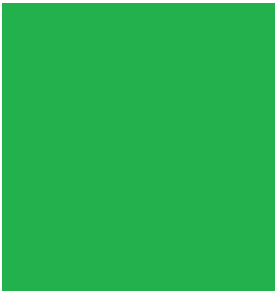
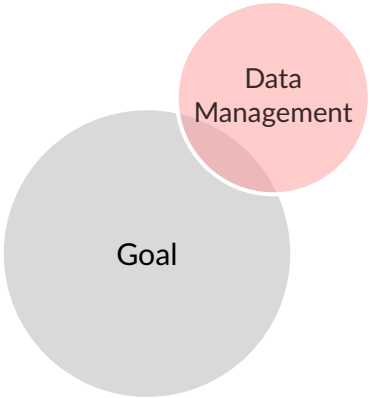
Course Introduction



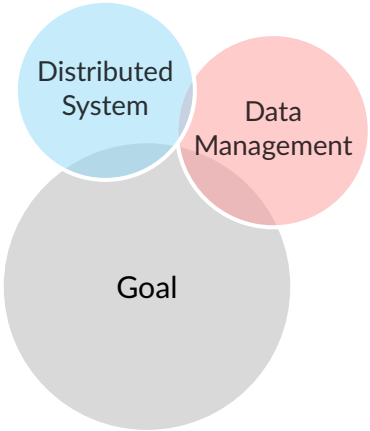
Course Target



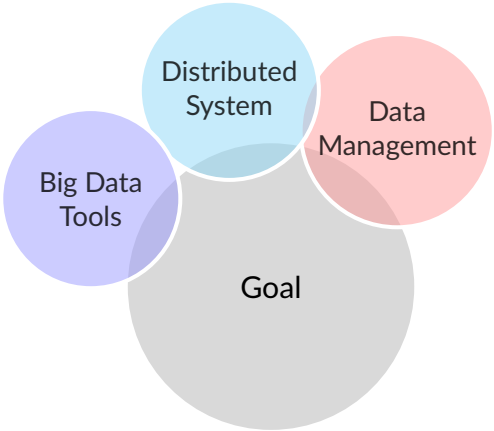
Course Target



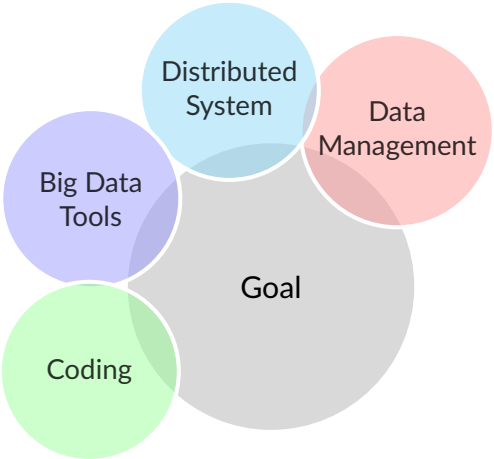
Course Target



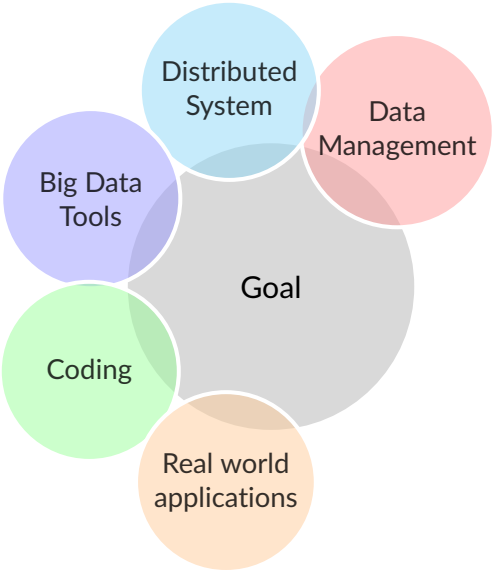
Course Target



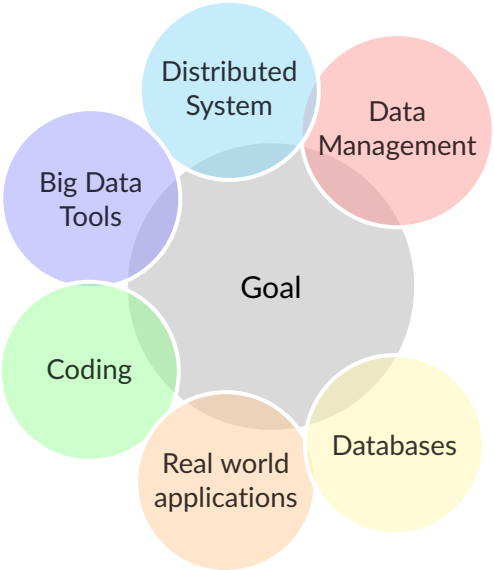
Course Target



Course Target



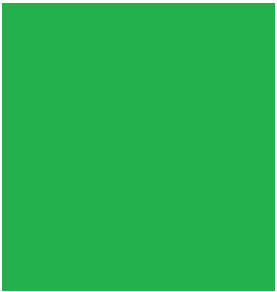
Course Target



Course Target

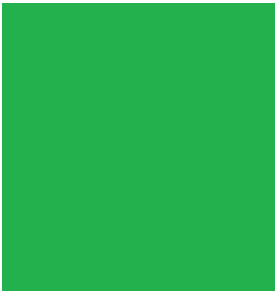
Learning Objectives

- Simplify the concepts in data management.



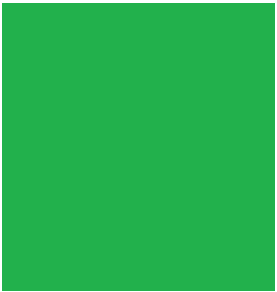
Learning Objectives

- Simplify the concepts in data management.
- Understand the data management life-cycle.



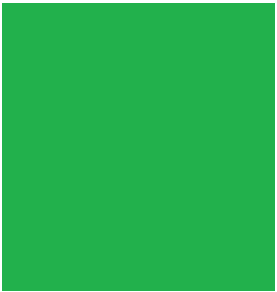
Learning Objectives

- Simplify the concepts in data management.
- Understand the data management life-cycle.
- Illustrate the basics of distributed systems concepts



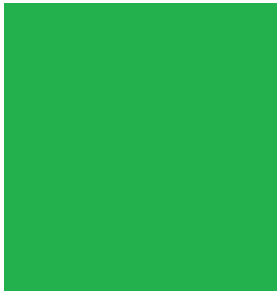
Learning Objectives

- Simplify the concepts in data management.
- Understand the data management life-cycle.
- Illustrate the basics of distributed systems concepts
- Be familiar with ETL for (batch/steaming) data over distributed systems ex: Hadoop & Spark.



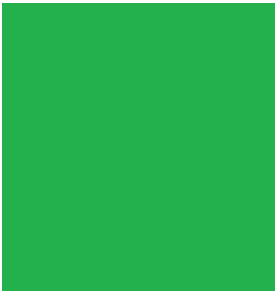
Learning Objectives

- Simplify the concepts in data management.
- Understand the data management life-cycle.
- Illustrate the basics of distributed systems concepts
- Be familiar with ETL for (batch/steaming) data over distributed systems ex: Hadoop & Spark.
- Apply QA and testing for the data pipeline cycle.



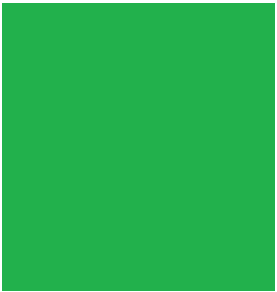
Learning Objectives

- Simplify the concepts in data management.
- Understand the data management life-cycle.
- Illustrate the basics of distributed systems concepts
- Be familiar with ETL for (batch/steaming) data over distributed systems ex: Hadoop & Spark.
- Apply QA and testing for the data pipeline cycle.
- Build and scale your data product.



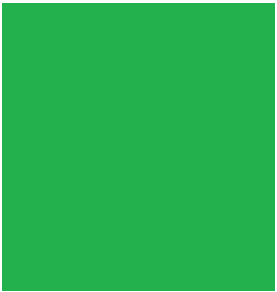
Learning Objectives

- Simplify the concepts in data management.
- Understand the data management life-cycle.
- Illustrate the basics of distributed systems concepts
- Be familiar with ETL for (batch/steaming) data over distributed systems ex: Hadoop & Spark.
- Apply QA and testing for the data pipeline cycle.
- Build and scale your data product.
- Building real-life examples.



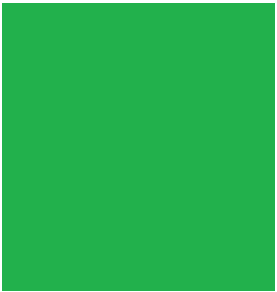
Learning Objectives

- Simplify the concepts in data management.
- Understand the data management life-cycle.
- Illustrate the basics of distributed systems concepts
- Be familiar with ETL for (batch/steaming) data over distributed systems ex: Hadoop & Spark.
- Apply QA and testing for the data pipeline cycle.
- Build and scale your data product.
- Building real-life examples.
- Applying machine learning over big data.



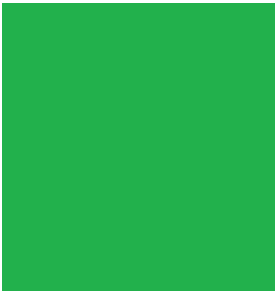
Learning Objectives

- Simplify the concepts in data management.
- Understand the data management life-cycle.
- Illustrate the basics of distributed systems concepts
- Be familiar with ETL for (batch/steaming) data over distributed systems ex: Hadoop & Spark.
- Apply QA and testing for the data pipeline cycle.
- Build and scale your data product.
- Building real-life examples.
- Applying machine learning over big data.
- Automate the data life-cycle process end-to-end (e2e).



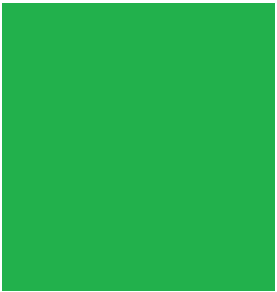
Learning Objectives

- Simplify the concepts in data management.
- Understand the data management life-cycle.
- Illustrate the basics of distributed systems concepts
- Be familiar with ETL for (batch/steaming) data over distributed systems ex: Hadoop & Spark.
- Apply QA and testing for the data pipeline cycle.
- Build and scale your data product.
- Building real-life examples.
- Applying machine learning over big data.
- Automate the data life-cycle process end-to-end (e2e).
- Understanding of the DevOps tools and functions in data life-cycle.



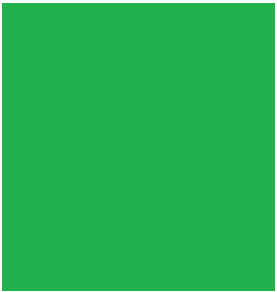
Audience: Who Should Take This Course?

- Data Engineer who needs to get more knowledge in distributed systems and Big Data.



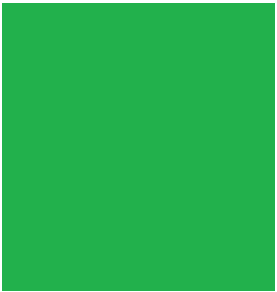
Audience: Who Should Take This Course?

- Data Engineer who needs to get more knowledge in distributed systems and Big Data.
- Data Warehouse Engineer who needs to know more about big data.



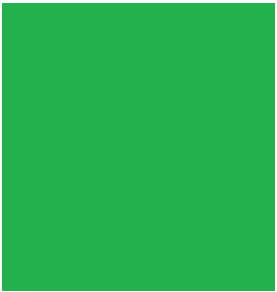
Audience: Who Should Take This Course?

- Data Engineer who needs to get more knowledge in distributed systems and Big Data.
- Data Warehouse Engineer who needs to know more about big data.
- Software developer who needs to change to data engineering track.



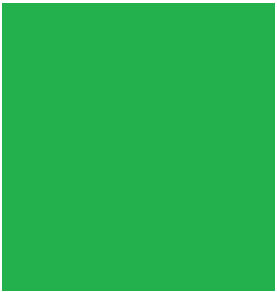
Audience: Who Should Take This Course?

- Data Engineer who needs to get more knowledge in distributed systems and Big Data.
- Data Warehouse Engineer who needs to know more about big data.
- Software developer who needs to change to data engineering track.
- DevOps engineers who needs to understand the concepts of big data.



Audience: Who Should Take This Course?

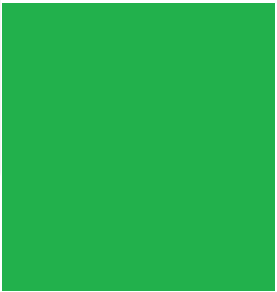
- Data Engineer who needs to get more knowledge in distributed systems and Big Data.
- Data Warehouse Engineer who needs to know more about big data.
- Software developer who needs to change to data engineering track.
- DevOps engineers who needs to understand the concepts of big data.
- Business or entrepreneur who needs to get more information about how to build or manage a data product.



Getting max benefit from this course

Take the course advantage

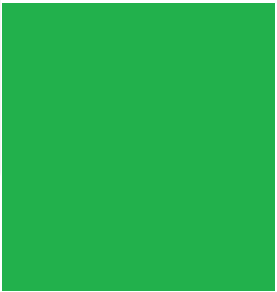
- Follow the videos order as described.



Getting max benefit from this course

Take the course advantage

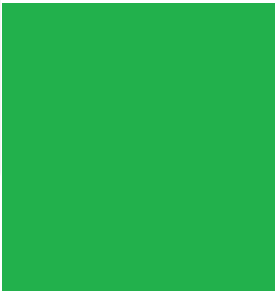
- Follow the videos order as described.
- Read the references for each section (including the implementation of the examples if exists).



Getting max benefit from this course

Take the course advantage

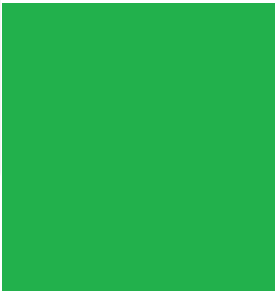
- Follow the videos order as described.
- Read the references for each section (including the implementation of the examples if exists).
- Repeat the lecture code with your own.



Getting max benefit from this course

Take the course advantage

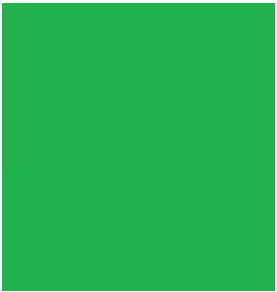
- Follow the videos order as described.
- Read the references for each section (including the implementation of the examples if exists).
- Repeat the lecture code with your own.
- Do the assignments.



Getting max benefit from this course

Take the course advantage

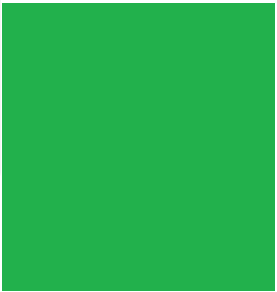
- Follow the videos order as described.
- Read the references for each section (including the implementation of the examples if exists).
- Repeat the lecture code with your own.
- Do the assignments.
- Ask your questions.



Getting max benefit from this course

Take the course advantage

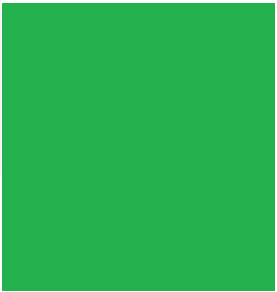
- Follow the videos order as described.
- Read the references for each section (including the implementation of the examples if exists).
- Repeat the lecture code with your own.
- Do the assignments.
- Ask your questions.
- Join the online meeting or discussions.



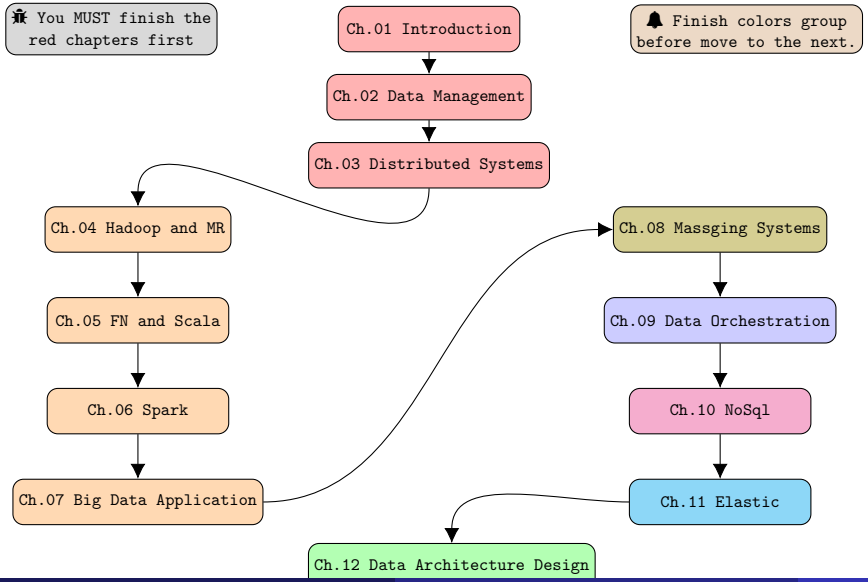
Getting max benefit from this course

Take the course advantage

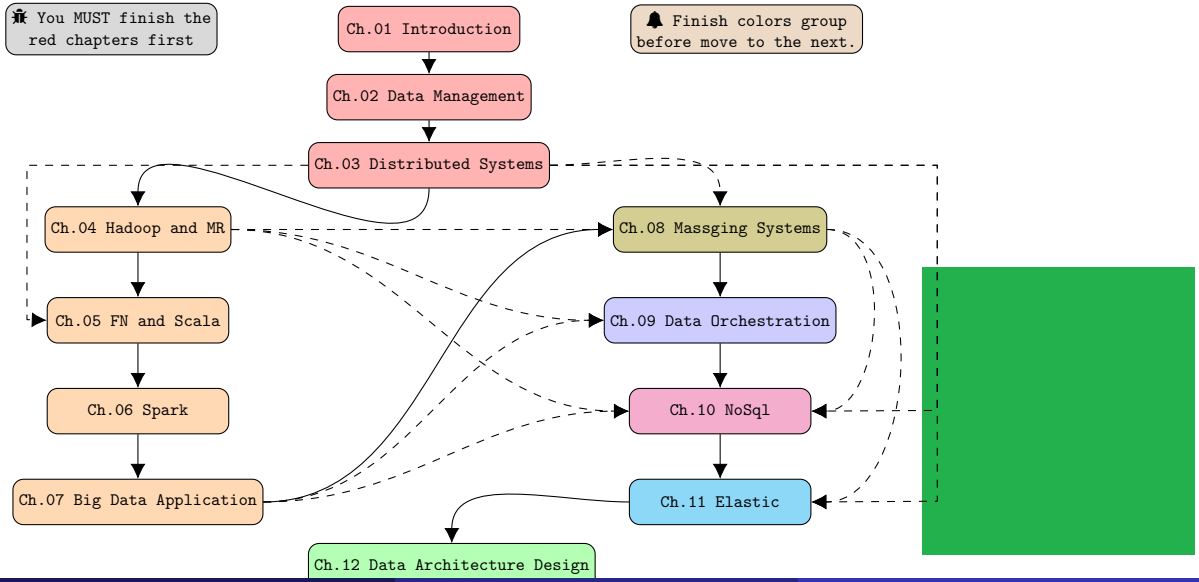
- Follow the videos order as described.
- Read the references for each section (including the implementation of the examples if exists).
- Repeat the lecture code with your own.
- Do the assignments.
- Ask your questions.
- Join the online meeting or discussions.



Chapter Dependencies



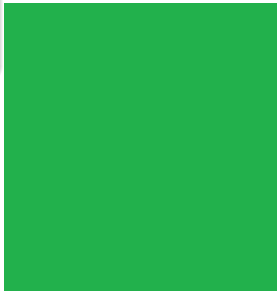
Chapter Dependencies (Jump Out Path)



Assignments and Labs

Remark

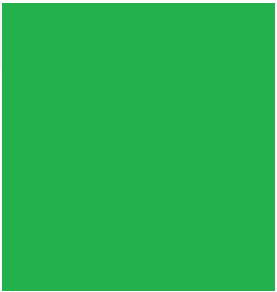
- Full project code.



Assignments and Labs

Remark

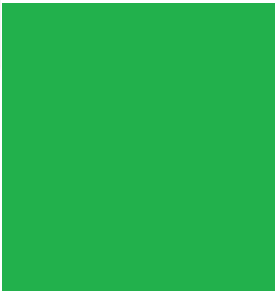
- Full project code.
- Notebooks (Jupyter or Zeppelin).



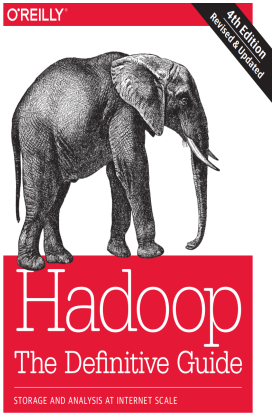
Assignments and Labs

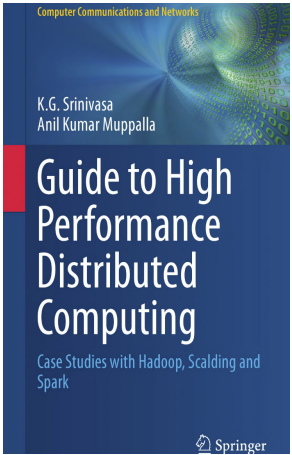
Remark

- Full project code.
- Notebooks (Jupyter or Zeppelin).
- Read the reference.



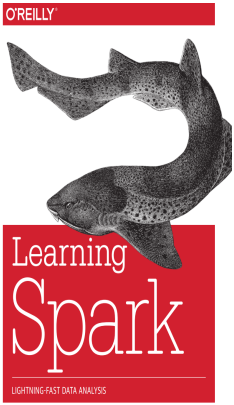
Textbooks-1


O'REILLY
4th Edition
Revised & Updated
Hadoop
The Definitive Guide
STORAGE AND ANALYSIS AT INTERNET SCALE
Tom White


Computer Communications and Networks
K.G. Srinivasa
Anil Kumar Muppalla
Guide to High Performance Distributed Computing
Case Studies with Hadoop, Scalding and Spark
Springer

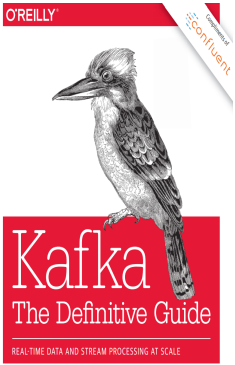
Textbooks-2

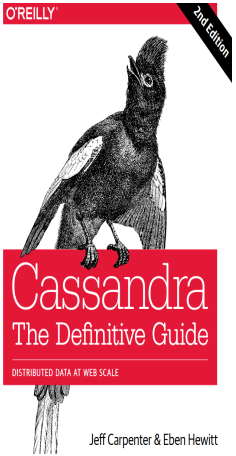
Holden Karau &
Rachel Warren

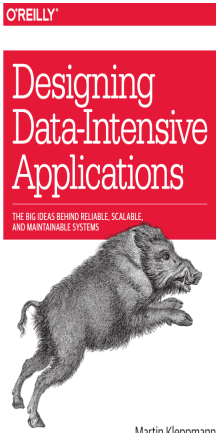


Holden Karau, Andy Konwinski,
Patrick Wendell & Matei Zaharia

Textbooks-3


O'REILLY
Compliments of
Confluent
Kafka
The Definitive Guide
REAL-TIME DATA AND STREAM PROCESSING AT SCALE
Neha Narkhede,
Gwen Shapira & Todd Palino


O'REILLY
2nd Edition
Cassandra
The Definitive Guide
DISTRIBUTED DATA AT WEB SCALE
Jeff Carpenter & Eben Hewitt


O'REILLY
Designing
Data-Intensive
Applications
THE BIG IDEAS BEHIND RELIABLE, SCALABLE,
AND MAINTAINABLE SYSTEMS
Martin Kleppmann

Ugly but important

- User stories or technical discussions are not related to any of my current work or my previous companies.
- I am working at EPAM Systems. My company approved me for doing this online course public but the materials are not reviewed or assessed by my company. It is on my own responsibilities.

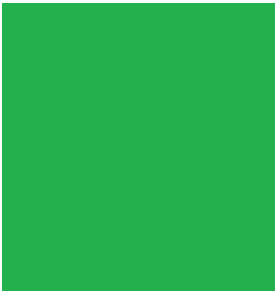


Table of Contents I

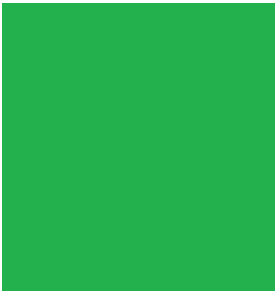


Table of Contents II

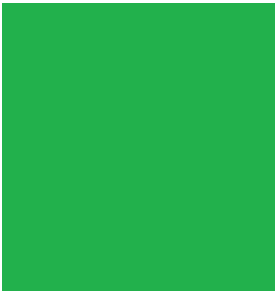


Table of Contents III

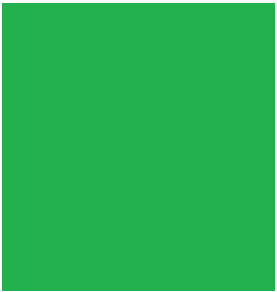


Table of Contents IV

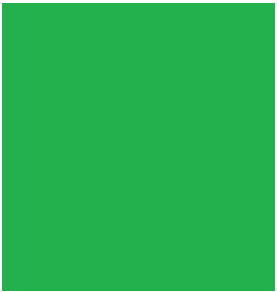


Table of Contents V

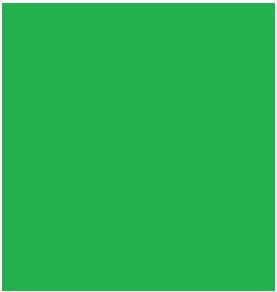


Table of Contents VI

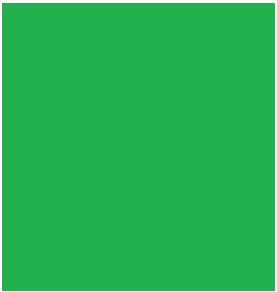


Table of Contents VII

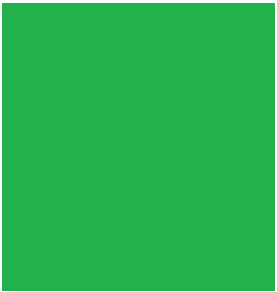
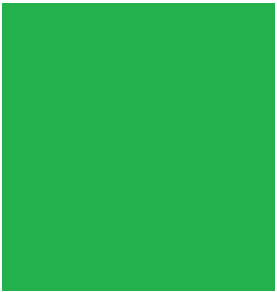
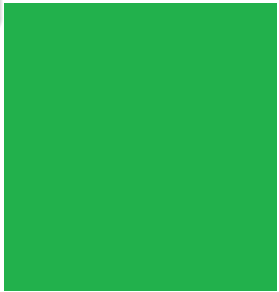


Table of Contents VIII

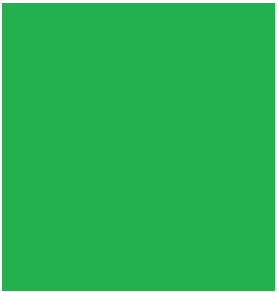


Introduction To Data Management and Data Warehouse



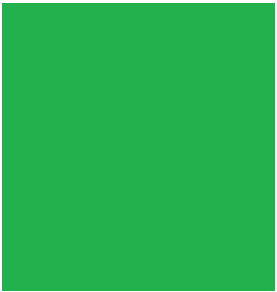
Chapter Objectives

- Be familiar with data management life-cycle.



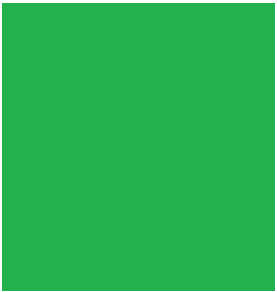
Chapter Objectives

- Be familiar with data management life-cycle.



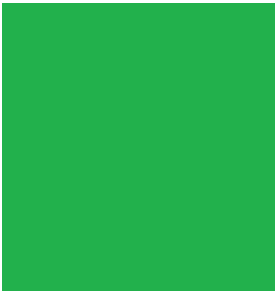
Chapter Objectives

- Be familiar with data management life-cycle.
- Introduction to data warehouse and its usage.



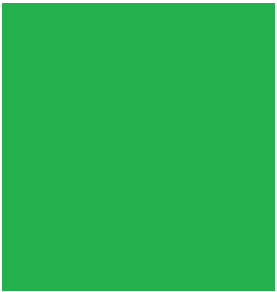
Chapter Objectives

- Be familiar with data management life-cycle.
- Introduction to data warehouse and its usage.



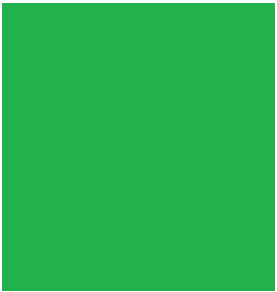
Chapter Objectives

- Be familiar with data management life-cycle.
- Introduction to data warehouse and its usage.
- Motivation to DWH.



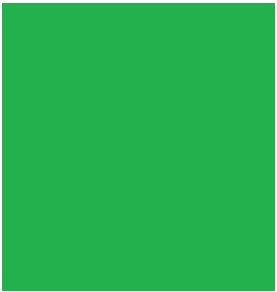
Chapter Objectives

- Be familiar with data management life-cycle.
- Introduction to data warehouse and its usage.
- Motivation to DWH.



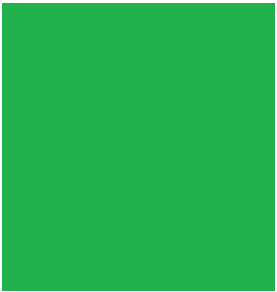
Chapter Objectives

- Be familiar with data management life-cycle.
- Introduction to data warehouse and its usage.
- Motivation to DWH.
- What is the different types of DWH?



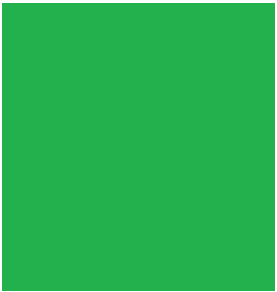
Chapter Objectives

- Be familiar with data management life-cycle.
- Introduction to data warehouse and its usage.
- Motivation to DWH.
- What is the different types of DWH?



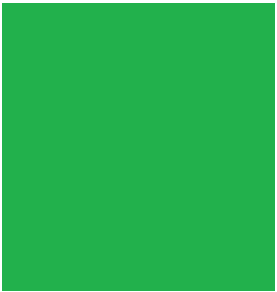
Chapter Objectives

- Be familiar with data management life-cycle.
- Introduction to data warehouse and its usage.
- Motivation to DWH.
- What is the different types of DWH?
- Use cases



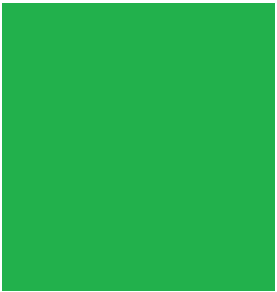
Chapter Objectives

- Be familiar with data management life-cycle.
- Introduction to data warehouse and its usage.
- Motivation to DWH.
- What is the different types of DWH?
- Use cases



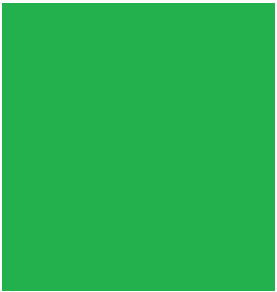
Chapter Objectives

- Be familiar with data management life-cycle.
- Introduction to data warehouse and its usage.
- Motivation to DWH.
- What is the different types of DWH?
- Use cases
- Data Encoding and Formats



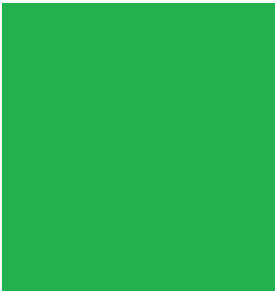
Chapter Objectives

- Be familiar with data management life-cycle.
- Introduction to data warehouse and its usage.
- Motivation to DWH.
- What is the different types of DWH?
- Use cases
- Data Encoding and Formats



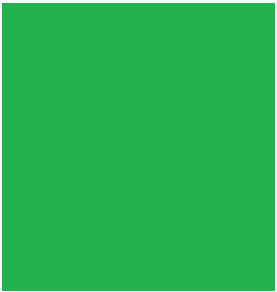
Chapter Objectives

- Be familiar with data management life-cycle.
- Introduction to data warehouse and its usage.
- Motivation to DWH.
- What is the different types of DWH?
- Use cases
- Data Encoding and Formats
- Challenges to build a DWH.



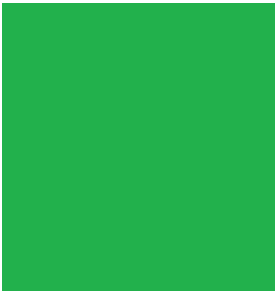
Chapter Objectives

- Be familiar with data management life-cycle.
- Introduction to data warehouse and its usage.
- Motivation to DWH.
- What is the different types of DWH?
- Use cases
- Data Encoding and Formats
- Challenges to build a DWH.



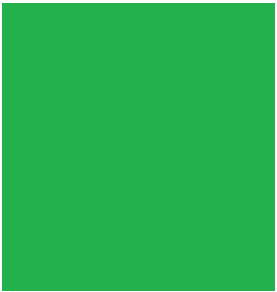
Chapter Objectives

- Be familiar with data management life-cycle.
- Introduction to data warehouse and its usage.
- Motivation to DWH.
- What is the different types of DWH?
- Use cases
- Data Encoding and Formats
- Challenges to build a DWH.
- Data modeling design.



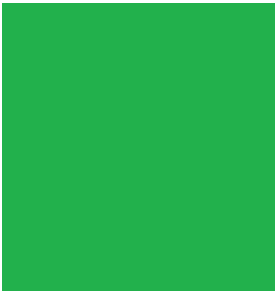
Chapter Objectives

- Be familiar with data management life-cycle.
- Introduction to data warehouse and its usage.
- Motivation to DWH.
- What is the different types of DWH?
- Use cases
- Data Encoding and Formats
- Challenges to build a DWH.
- Data modeling design.



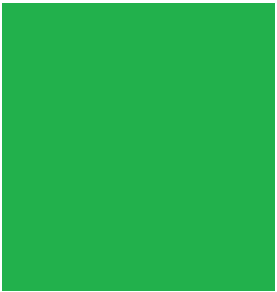
Chapter Objectives

- Be familiar with data management life-cycle.
- Introduction to data warehouse and its usage.
- Motivation to DWH.
- What is the different types of DWH?
- Use cases
- Data Encoding and Formats
- Challenges to build a DWH.
- Data modeling design.



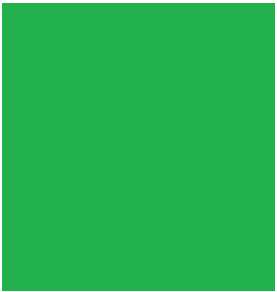
Data Management

- Data are a product.



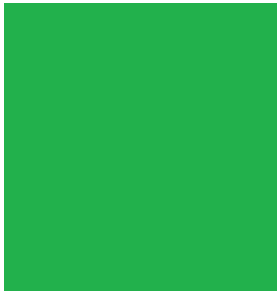
Data Management

- Data are a product.
- Data product has a life-cycle as following (simplified):



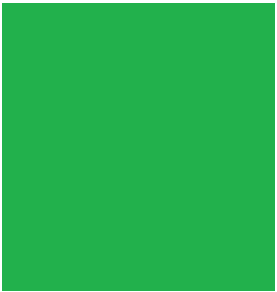
Data Management

- Data are a product.
- Data product has a life-cycle as following (simplified):
  - **Question**, Idea, or service.



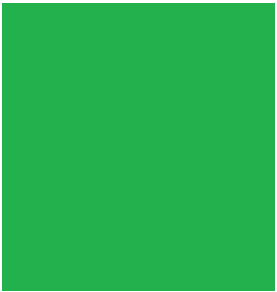
Data Management

- Data are a product.
- Data product has a life-cycle as following (simplified):
  - **Question**, Idea, or service.
  - **Identifying** the source of information and the data type ex: (text, images, videos, audio, or sensors).



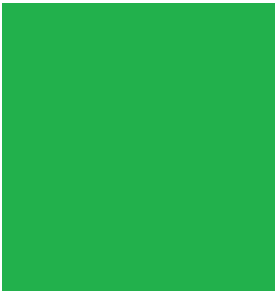
Data Management

- Data are a product.
- Data product has a life-cycle as following (simplified):
  - **Question**, Idea, or service.
  - **Identifying** the source of information and the data type ex: (text, images, videos, audio, or sensors).
  - **Document** all details regarding the data including quality, security, efficiency, and access (consideration during the cycle).



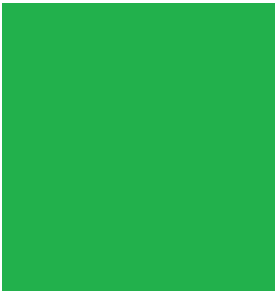
Data Management

- Data are a product.
- Data product has a life-cycle as following (simplified):
  - **Question**, Idea, or service.
  - **Identifying** the source of information and the data type ex: (text, images, videos, audio, or sensors).
  - **Document** all details regarding the data including quality, security, efficiency, and access (consideration during the cycle).
  - Delivery automation (Tools and Process) AKA **DevOps** cycle.



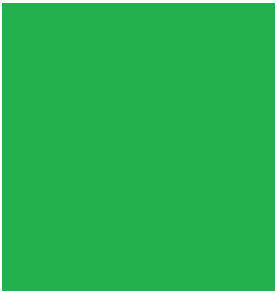
Data Management

- Data are a product.
- Data product has a life-cycle as following (simplified):
  - **Question**, Idea, or service.
  - **Identifying** the source of information and the data type ex: (text, images, videos, audio, or sensors).
  - **Document** all details regarding the data including quality, security, efficiency, and access (consideration during the cycle).
  - Delivery automation (Tools and Process) AKA **DevOps** cycle.
  - **Extraction** Process (collection).



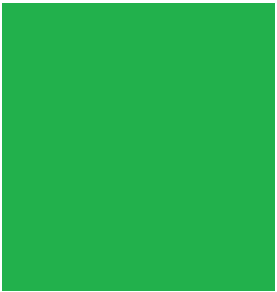
Data Management

- Data are a product.
- Data product has a life-cycle as following (simplified):
  - **Question**, Idea, or service.
  - **Identifying** the source of information and the data type ex: (text, images, videos, audio, or sensors).
  - **Document** all details regarding the data including quality, security, efficiency, and access (consideration during the cycle).
  - Delivery automation (Tools and Process) AKA **DevOps** cycle.
  - **Extraction** Process (collection).
  - **Transformation** ex: (cleansing, Apply business logic, Organize).



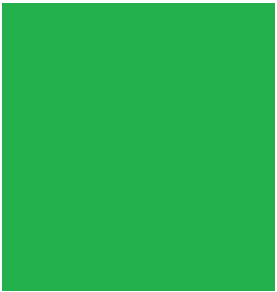
Data Management

- Data are a product.
- Data product has a life-cycle as following (simplified):
  - **Question**, Idea, or service.
  - **Identifying** the source of information and the data type ex: (text, images, videos, audio, or sensors).
  - **Document** all details regarding the data including quality, security, efficiency, and access (consideration during the cycle).
  - Delivery automation (Tools and Process) AKA **DevOps** cycle.
  - **Extraction** Process (collection).
  - **Transformation** ex: (cleansing, Apply business logic, Organize).
  - **Loading** or store the transformed data based on our usage or use case.



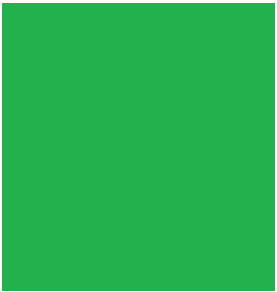
Data Management

- Data are a product.
- Data product has a life-cycle as following (simplified):
  - **Question**, Idea, or service.
  - **Identifying** the source of information and the data type ex: (text, images, videos, audio, or sensors).
  - **Document** all details regarding the data including quality, security, efficiency, and access (consideration during the cycle).
  - Delivery automation (Tools and Process) AKA **DevOps** cycle.
  - **Extraction** Process (collection).
  - **Transformation** ex: (cleansing, Apply business logic, Organize).
  - **Loading** or store the transformed data based on our usage or use case.
  - Business Intelligence (**BI**) or data discovery (continues process).



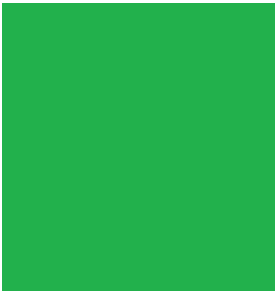
Data Management

- Data are a product.
- Data product has a life-cycle as following (simplified):
  - **Question**, Idea, or service.
  - **Identifying** the source of information and the data type ex: (text, images, videos, audio, or sensors).
  - **Document** all details regarding the data including quality, security, efficiency, and access (consideration during the cycle).
  - Delivery automation (Tools and Process) AKA **DevOps** cycle.
  - **Extraction** Process (collection).
  - **Transformation** ex: (cleansing, Apply business logic, Organize).
  - **Loading** or store the transformed data based on our usage or use case.
  - Business Intelligence (**BI**) or data discovery (continues process).
  - **Integration** and publishing.

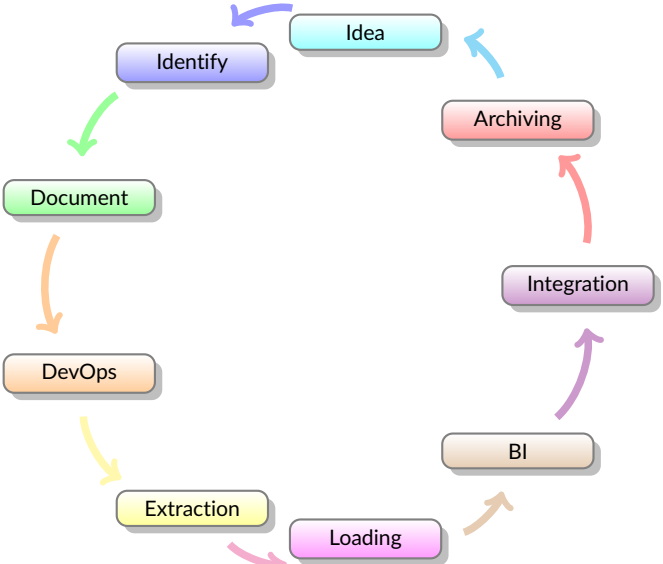


Data Management

- Data are a product.
- Data product has a life-cycle as following (simplified):
  - **Question**, Idea, or service.
  - **Identifying** the source of information and the data type ex: (text, images, videos, audio, or sensors).
  - **Document** all details regarding the data including quality, security, efficiency, and access (consideration during the cycle).
  - Delivery automation (Tools and Process) AKA **DevOps** cycle.
  - **Extraction** Process (collection).
  - **Transformation** ex: (cleansing, Apply business logic, Organize).
  - **Loading** or store the transformed data based on our usage or use case.
  - Business Intelligence (**BI**) or data discovery (continues process).
  - **Integration** and publishing.
  - Data retention or **archiving** process ex: (Hot or Cold storage).

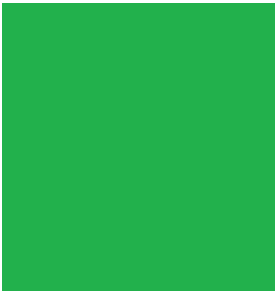


Data Management Life-Cycle



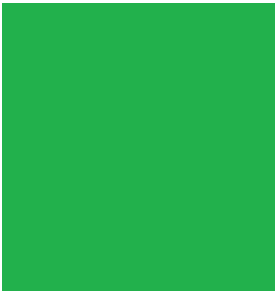
Motivation to Data Warehouse

- Data could be a product for some companies.



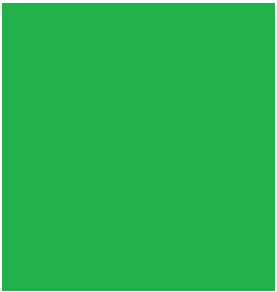
Motivation to Data Warehouse

- Data could be a product for some companies.
- It could be decision support for other products or businesses.



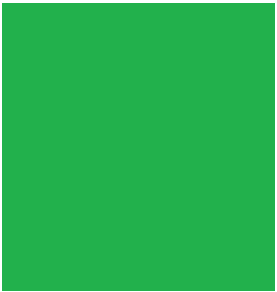
Motivation to Data Warehouse

- Data could be a product for some companies.
- It could be decision support for other products or businesses.
- Reporting the results after pass the data life-cycle will be from storage (Database).



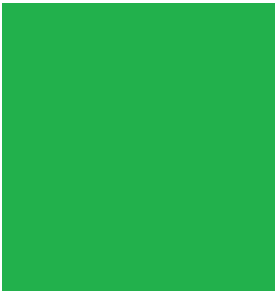
Motivation to Data Warehouse

- Data could be a product for some companies.
- It could be decision support for other products or businesses.
- Reporting the results after pass the data life-cycle will be from storage (Database).
- There are some challenges facing the people who work on data management backend:



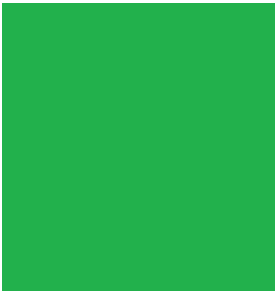
Motivation to Data Warehouse

- Data could be a product for some companies.
- It could be decision support for other products or businesses.
- Reporting the results after pass the data life-cycle will be from storage (Database).
- There are some challenges facing the people who work on data management backend:
  - Performance.



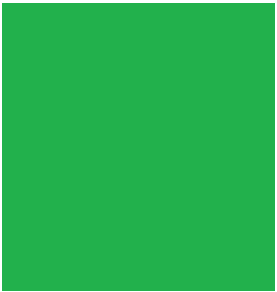
Motivation to Data Warehouse

- Data could be a product for some companies.
- It could be decision support for other products or businesses.
- Reporting the results after pass the data life-cycle will be from storage (Database).
- There are some challenges facing the people who work on data management backend:
  - Performance.
  - Integration.



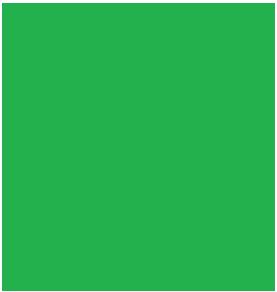
Motivation to Data Warehouse

- Data could be a product for some companies.
- It could be decision support for other products or businesses.
- Reporting the results after pass the data life-cycle will be from storage (Database).
- There are some challenges facing the people who work on data management backend:
  - Performance.
  - Integration.
  - Applying analytical functions.



Motivation to Data Warehouse

- Data could be a product for some companies.
- It could be decision support for other products or businesses.
- Reporting the results after pass the data life-cycle will be from storage (Database).
- There are some challenges facing the people who work on data management backend:
  - Performance.
  - Integration.
  - Applying analytical functions.
- Vendors who are working to solve the above challenges creating their own product of DWH and their ultimate work is to optimize the above points.

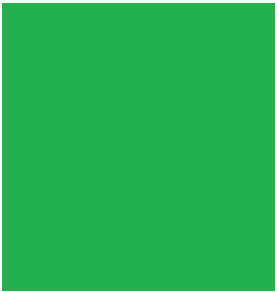


Motivation to Data Warehouse (DWH)

Definition (What is Data Warehousing?)

A DWH is defined as a technique for collecting and managing data from varied sources to **provide meaningful business insights**. It is a blend of technologies and components which aids the strategic use of data.

- The DWH is not a product but an environment.

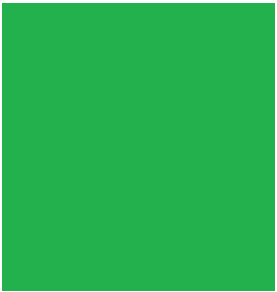


Motivation to Data Warehouse (DWH)

Definition (What is Data Warehousing?)

A DWH is defined as a technique for collecting and managing data from varied sources to **provide meaningful business insights**. It is a blend of technologies and components which aids the strategic use of data.

- The DWH is not a product but an environment.
- It is a process of transforming data into information and make it available to users in a **timely manner** to make a difference.

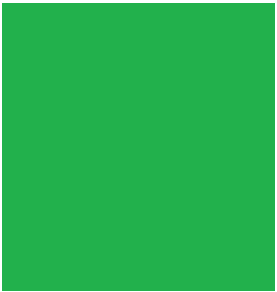


Motivation to Data Warehouse (DWH)

Definition (What is Data Warehousing?)

A DWH is defined as a technique for collecting and managing data from varied sources to **provide meaningful business insights**. It is a blend of technologies and components which aids the strategic use of data.

- The DWH is not a product but an environment.
- It is a process of transforming data into information and make it available to users in a **timely manner** to make a difference.
- It is an architectural construct of an information system which provides users with current and historical decision support information which is difficult to access or present in the traditional operational data store.

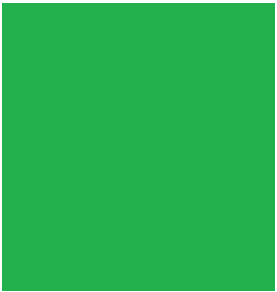


Motivation to Data Warehouse (DWH)

Definition (What is Data Warehousing?)

A DWH is defined as a technique for collecting and managing data from varied sources to **provide meaningful business insights**. It is a blend of technologies and components which aids the strategic use of data.

- The DWH is not a product but an environment.
- It is a process of transforming data into information and make it available to users in a **timely manner** to make a difference.
- It is an architectural construct of an information system which provides users with current and historical decision support information which is difficult to access or present in the traditional operational data store.
- The DWH is the core of the BI system which is built for data analysis and reporting.

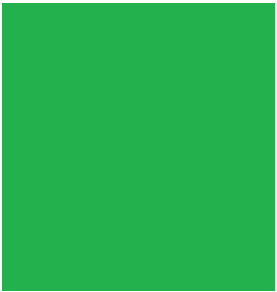


Motivation to Data Warehouse

Data warehouse system is also known by the following names:

- Decision Support System (DSS).
- Business Intelligence Solution.
- Executive Information System.
- Management Information System.
- Analytic Application.
- Data Warehouse.

The real concept was given by Inmon Bill. He was considered as a father of the DWH. He had written about a variety of topics for building, usage, and maintenance of the warehouse & the Corporate Information Factory



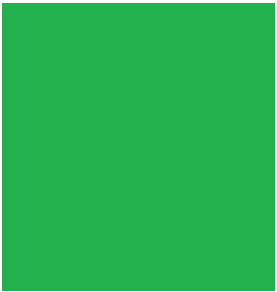
Motivation to Data Warehouse

Types of Data Warehouse

Enterprise Data Warehouse (EDWH) It provides decision support service across the enterprise. It offers a unified approach for organizing and representing data (DWH Model). It offers data classifications according to the subject with privileges policy.

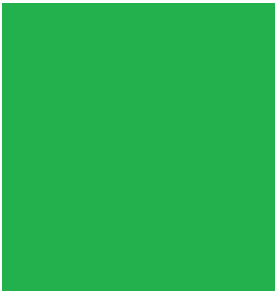
Operational Data Store (ODS): is a central database that provides an up-to-date (real-time) data from multiple transnational systems for operational reporting into a single DWH.

Data Mart: A data mart is a subset of the data warehouse. It specially designed for a particular line of business, such as sales, finance, sales or finance. In an independent data mart, data can collect directly from sources.



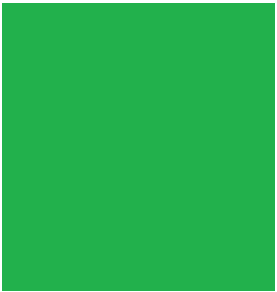
DWH vs ODS vs Data Mart

Metric	DWH	ODS	Data Mart
Latency	Day -1	Real-time	Day -1
Data level	Transnational	Transnational	Summary
Historical	Long-term	Snapshot	Aggregated Long-Term
Size	TB/PB	GB	GB/TB
Orientation	Multi sources	Multi sources	Product
Business Units	Multi organizational units	Product team	Business team



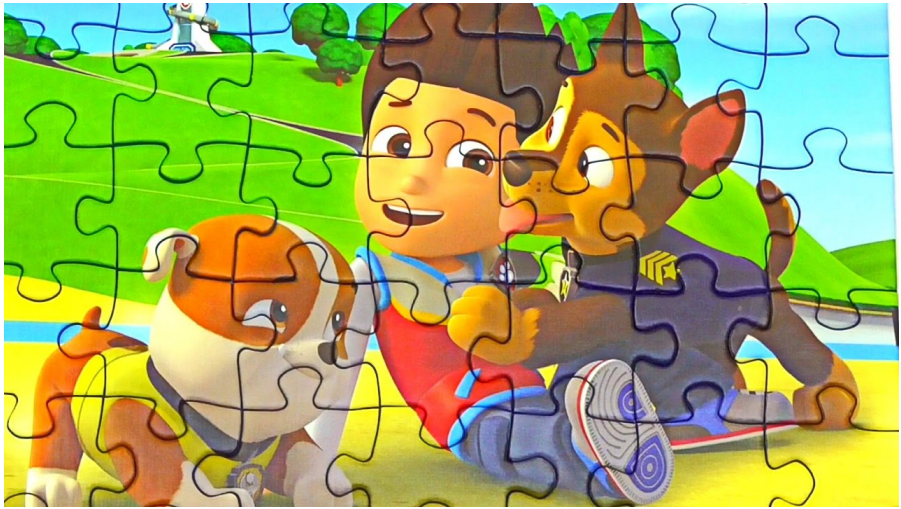
DWH vs Operational databases

Metric	Transactions DB	DWH
Volume	GB/TB	TB/PB
Historical	Short-term	Long-Term
rows	¡1000M	1000M¿
Orientation	Product	Subject or multi products
Business Units	Product team	Multi organizational units
Normalization	Normalized	Not required (De-normalized in many use cases)
Data Model	Relational	Star Schema or Multi-dim
Intelligence	Reporting	Advanced reporting and Machine Learning
Use cases	Online transactions & operations	Centeralized storage (360°)



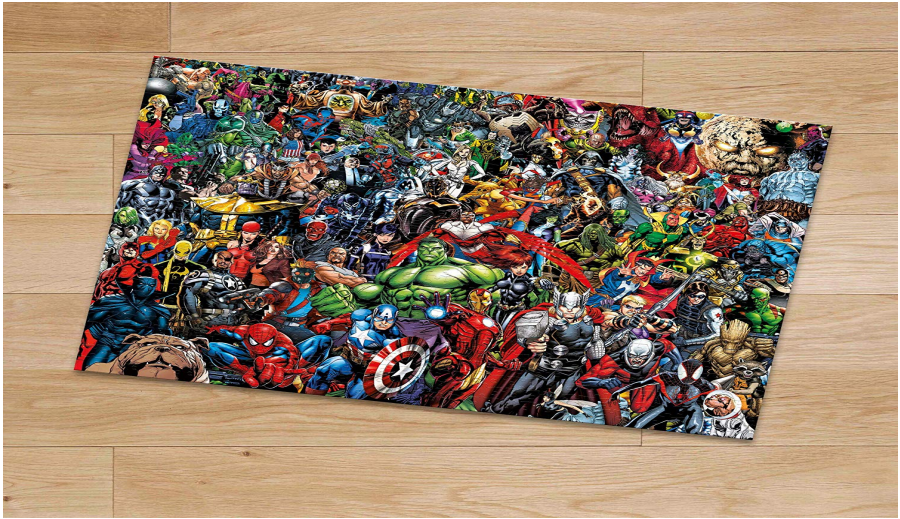
Transnational DB Use cases

Transnational DB Use cases



DWH Use cases

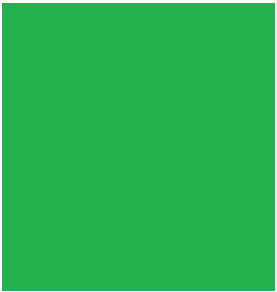
DWH Use cases



DWH Use cases

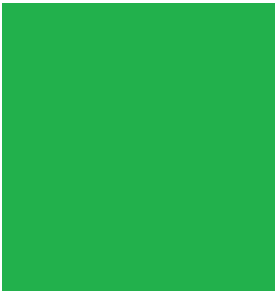
Use case (Operational DB)

- A telecommunication company named **XTec**.



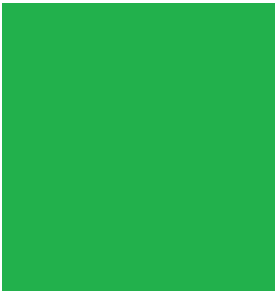
Use case (Operational DB)

- A telecommunication company named **XTec**.
- They have lots of systems. One of this systems is a CRM system as example of operational DB.



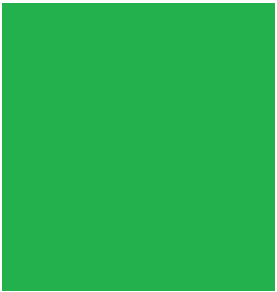
Use case (Operational DB)

- A telecommunication company named **XTec**.
- They have lots of systems. One of this systems is a CRM system as example of operational DB.
  - The CRM system handles the customer activities with the company including (sales, change in customer plans, and other activities).



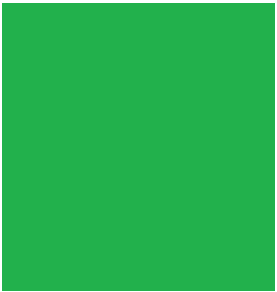
Use case (Operational DB)

- A telecommunication company named **XTec**.
- They have lots of systems. One of this systems is a CRM system as example of operational DB.
  - The CRM system handles the customer activities with the company including (sales, change in customer plans, and other activities).
  - This system has a backend database (MySQL).



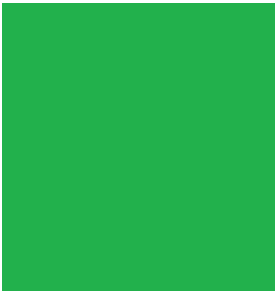
Use case (Operational DB)

- A telecommunication company named **XTec**.
- They have lots of systems. One of this systems is a CRM system as example of operational DB.
  - The CRM system handles the customer activities with the company including (sales, change in customer plans, and other activities).
  - This system has a backend database (MySQL).
  - CRM team can report their sales and customer activities from their database.



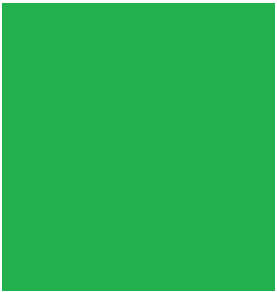
Use case (Operational DB)

- A telecommunication company named **XTec**.
- They have lots of systems. One of this systems is a CRM system as example of operational DB.
  - The CRM system handles the customer activities with the company including (sales, change in customer plans, and other activities).
  - This system has a backend database (MySQL).
  - CRM team can report their sales and customer activities from their database.
  - Product owner can take a decision based on their system backend reports.



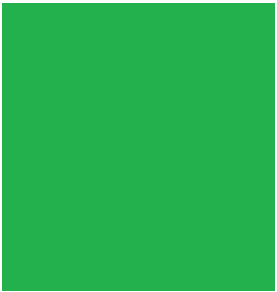
Use case (DWH)

- What is the need for DWH?



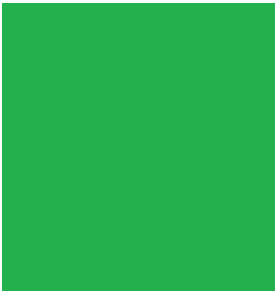
Use case (DWH)

- What is the need for DWH?
  - This company has other systems for example: billing, charging, signaling.



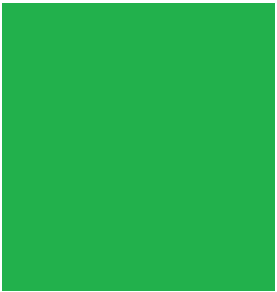
Use case (DWH)

- What is the need for DWH?
  - This company has other systems for example: billing, charging, signaling.
  - They need to report information related to the CRM, billing, and signaling source systems in one report.



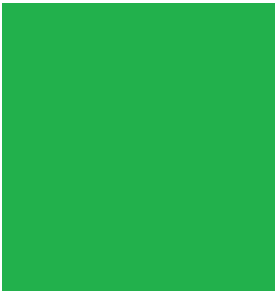
Use case (DWH)

- What is the need for DWH?
  - This company has other systems for example: billing, charging, signaling.
  - They need to report information related to the CRM, billing, and signaling source systems in one report.
  - So, they need to ingest (transfer) the data from the source systems to one single database.



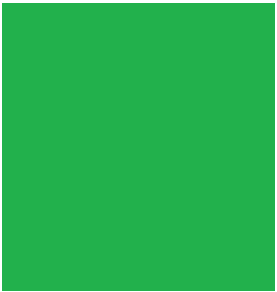
Use case (DWH)

- What is the need for DWH?
  - This company has other systems for example: billing, charging, signaling.
  - They need to report information related to the CRM, billing, and signaling source systems in one report.
  - So, they need to ingest (transfer) the data from the source systems to one single database.
  - The decision from the DHW is a **global and strategical decision.**



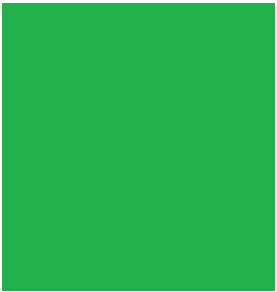
Use case (DWH)

- What is the need for DWH?
  - This company has other systems for example: billing, charging, signaling.
  - They need to report information related to the CRM, billing, and signaling source systems in one report.
  - So, they need to ingest (transfer) the data from the source systems to one single database.
  - The decision from the DHW is a **global and strategical decision.**
  - If the company needs to build a machine learning model which needs data from different sources. They need to load the data from a centralized database rather than read each source alone.



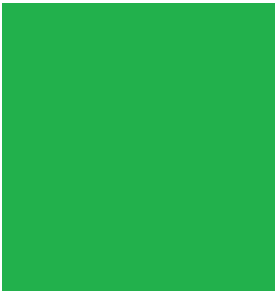
Use case (DWH)

The Full picture required a DWH. However, we still need the other operational databases for product development perspective.



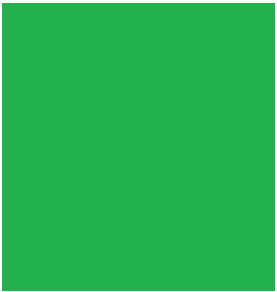
Use case (ODS)

- Why do we need the ODS?



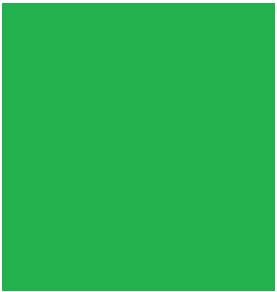
Use case (ODS)

- Why do we need the ODS?
- How does it fit in our system?



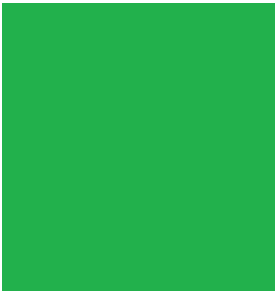
Use case (ODS)

XTec has a call center system which handles the customer inquiries. This system requires the some data related to usage, customer information, billing details to be calculated and accumulated in **real-time** to be able to give the customer the right answer for his inquires.



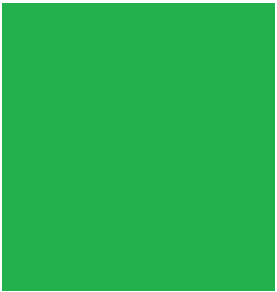
Use case (ODS)

- So, What is the challenge for this system?



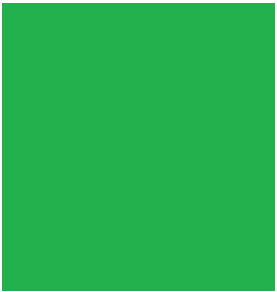
Use case (ODS)

- So, What is the challenge for this system?
  - It needs specific information from different source systems.



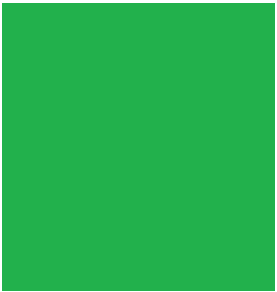
Use case (ODS)

- So, What is the challenge for this system?
  - It needs specific information from different source systems.
  - It requires to track the source system database changes or update in real-time.



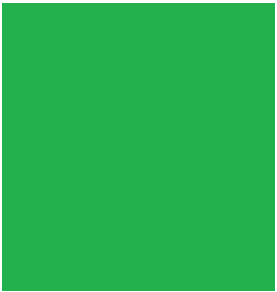
Use case (ODS)

- So, What is the challenge for this system?
  - It needs specific information from different source systems.
  - It requires to track the source system database changes or update in real-time.
  - It's functionality is based on the aggregate data not the transactions for example (It needs the total outgoing calls till time or it needs the total charging amounts from prepaid or the available limits from billing if it is postpaid).



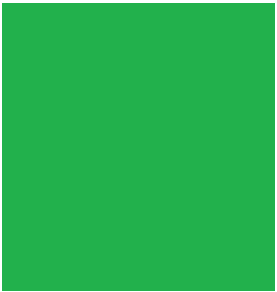
Use case (ODS)

- ODS is based on change data capture (CDC). This approach used to determine the data change and apply action based on this change.



Use case (ODS)

- ODS is based on change data capture (CDC). This approach used to determine the data change and apply action based on this change.
- ODS uses the real-time aggregations to support the online systems from different source systems.



DWH Architecture Overview

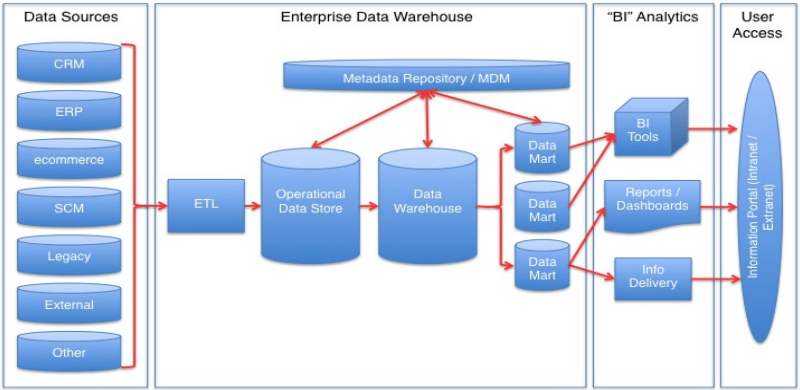
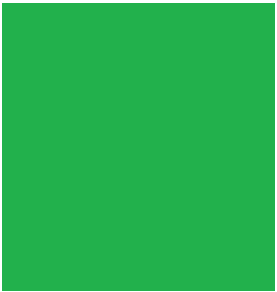


Figure: taken from XXXX

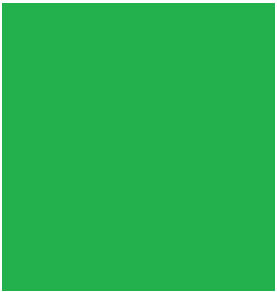
Thinking about data as a product

- How can we think about a data solution?



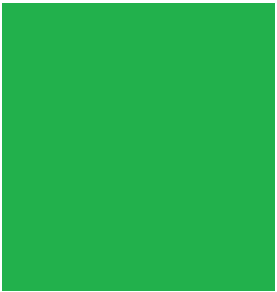
Thinking about data as a product

- How can we think about a data solution?
  - Requirements analysis.



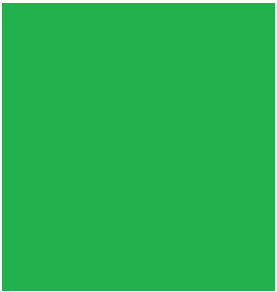
Thinking about data as a product

- How can we think about a data solution?
  - Requirements analysis.
  - Identify the problem (challenges)



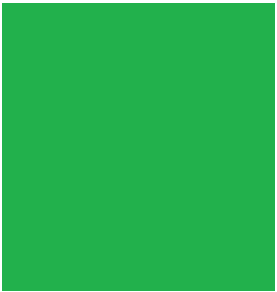
Thinking about data as a product

- How can we think about a data solution?
  - Requirements analysis.
  - Identify the problem (challenges)
  - Think about how to overcome the challenges.



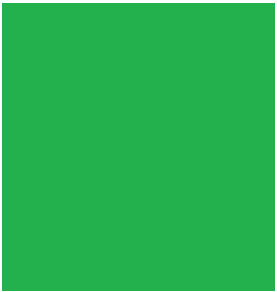
Thinking about data as a product

- How can we think about a data solution?
  - Requirements analysis.
  - Identify the problem (challenges)
  - Think about how to overcome the challenges.
  - Ask your self the following questions:



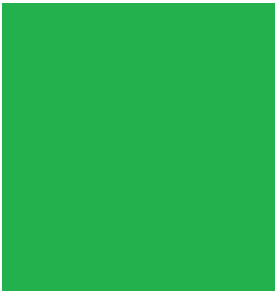
Thinking about data as a product

- How can we think about a data solution?
  - Requirements analysis.
  - Identify the problem (challenges)
  - Think about how to overcome the challenges.
  - Ask your self the following questions:
    - Can we solve the problem using the current data structure with adding new features?



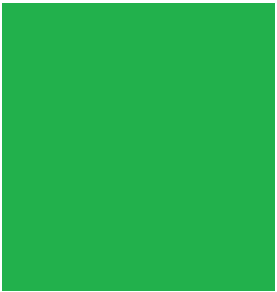
Thinking about data as a product

- How can we think about a data solution?
  - Requirements analysis.
  - Identify the problem (challenges)
  - Think about how to overcome the challenges.
  - Ask your self the following questions:
    - Can we solve the problem using the current data structure with adding new features?
    - What if we enhance/change the data structure or modeling?



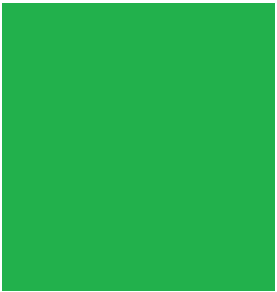
Thinking about data as a product

- How can we think about a data solution?
  - Requirements analysis.
  - Identify the problem (challenges)
  - Think about how to overcome the challenges.
  - Ask your self the following questions:
    - Can we solve the problem using the current data structure with adding new features?
    - What if we enhance/change the data structure or modeling?
    - Could it help if we change the backend engine (ex: database)?



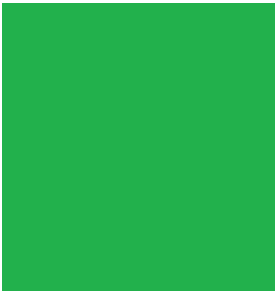
Thinking about data as a product

- How can we think about a data solution?
  - Requirements analysis.
  - Identify the problem (challenges)
  - Think about how to overcome the challenges.
  - Ask your self the following questions:
    - Can we solve the problem using the current data structure with adding new features?
    - What if we enhance/change the data structure or modeling?
    - Could it help if we change the backend engine (ex: database)?
- To answer these questions you need to understand the **data layers**.



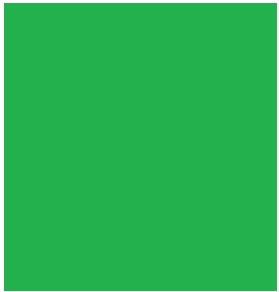
Data Layers (Abstraction)

- Any data product (database) contains multi-layers.



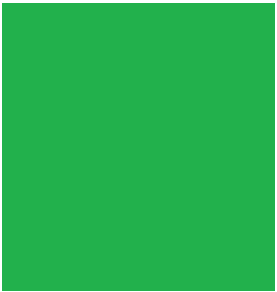
Data Layers (Abstraction)

- Any data product (database) contains multi-layers.
- Each layer responsible for different tasks and operations.



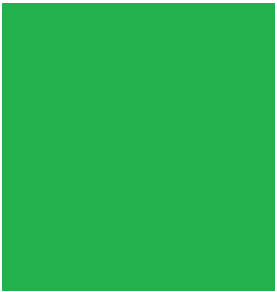
Data Layers (Abstraction)

- Any data product (database) contains multi-layers.
- Each layer responsible for different tasks and operations.
- Each layers interacts with (hardware or software or mixed).



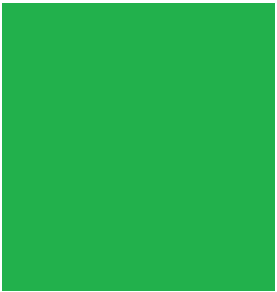
Data Layers (Abstraction)

- Any data product (database) contains multi-layers.
- Each layer responsible for different tasks and operations.
- Each layers interacts with (hardware or software or mixed).
- To eliminate the complexity for data interactions not all internal processes are shared or available for the user.



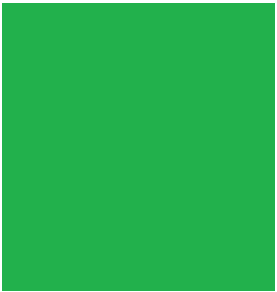
Data Layers (Abstraction)

- Any data product (database) contains multi-layers.
- Each layer responsible for different tasks and operations.
- Each layers interacts with (hardware or software or mixed).
- To eliminate the complexity for data interactions not all internal processes are shared or available for the user.
- The developer for each layer hide internal irrelevant details from developer (users).



Data Layers (Abstraction)

- Any data product (database) contains multi-layers.
- Each layer responsible for different tasks and operations.
- Each layers interacts with (hardware or software or mixed).
- To eliminate the complexity for data interactions not all internal processes are shared or available for the user.
- The developer for each layer hide internal irrelevant details from developer (users).
- The process of **hiding** irrelevant details from developer (user) is called data **abstraction**.

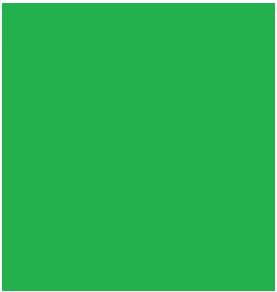


Data Layers (Abstraction)

Definition

Data Abstraction and Data Independence: Database systems comprise of complex data-structures. In order to make the system efficient in terms of retrieval of data, and reduce complexity in terms of usability of users, developers use abstraction i.e. hide irrelevant details from the users. This approach simplifies database design.

- There are 3 levels of data abstraction.

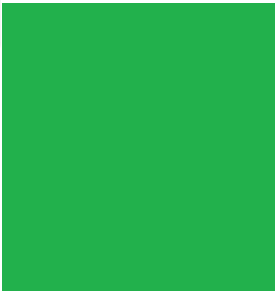


Data Layers (Abstraction)

Definition

Data Abstraction and Data Independence: Database systems comprise of complex data-structures. In order to make the system efficient in terms of retrieval of data, and reduce complexity in terms of usability of users, developers use abstraction i.e. hide irrelevant details from the users. This approach simplifies database design.

- There are 3 levels of data abstraction.
  - Physical level

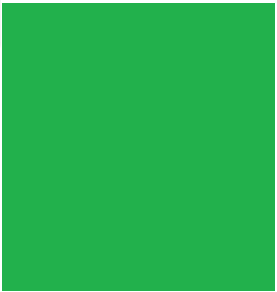


Data Layers (Abstraction)

Definition

Data Abstraction and Data Independence: Database systems comprise of complex data-structures. In order to make the system efficient in terms of retrieval of data, and reduce complexity in terms of usability of users, developers use abstraction i.e. hide irrelevant details from the users. This approach simplifies database design.

- There are 3 levels of data abstraction.
  - Physical level
  - Logical/ Conceptual level.



Data Layers (Abstraction)

Definition

Data Abstraction and Data Independence: Database systems comprise of complex data-structures. In order to make the system efficient in terms of retrieval of data, and reduce complexity in terms of usability of users, developers use abstraction i.e. hide irrelevant details from the users. This approach simplifies database design.

- There are 3 levels of data abstraction.
  - Physical level
  - Logical/ Conceptual level.
  - View level.