upGrad

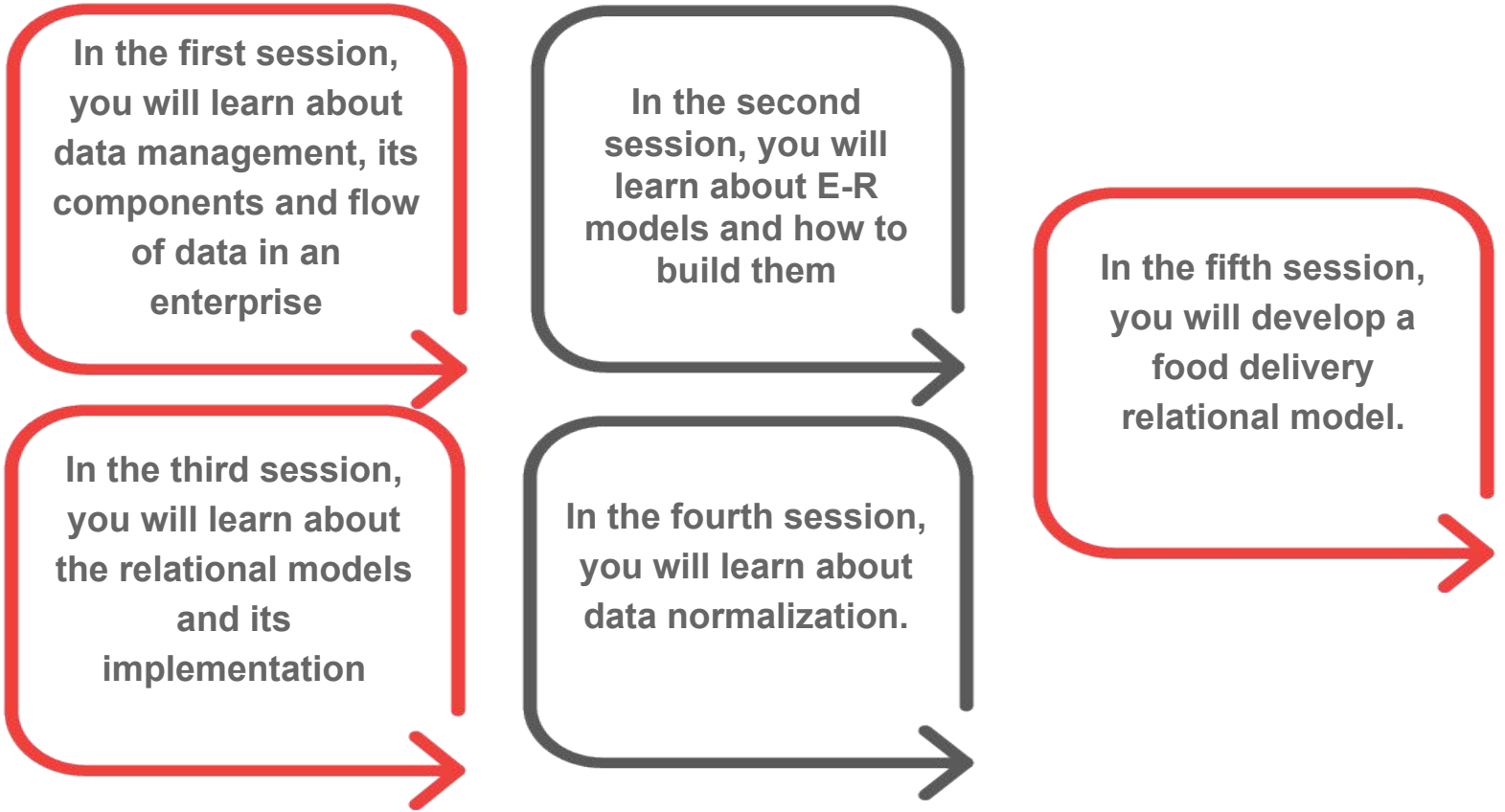
Data Management and Relational Modelling

Course: Data Engineering - I

Lecture On: Data Management and Relational Modelling

Instructor: Vishwa Mohan

Overview of the Module



Session 1 | Introduction to Data Management

Session Overview

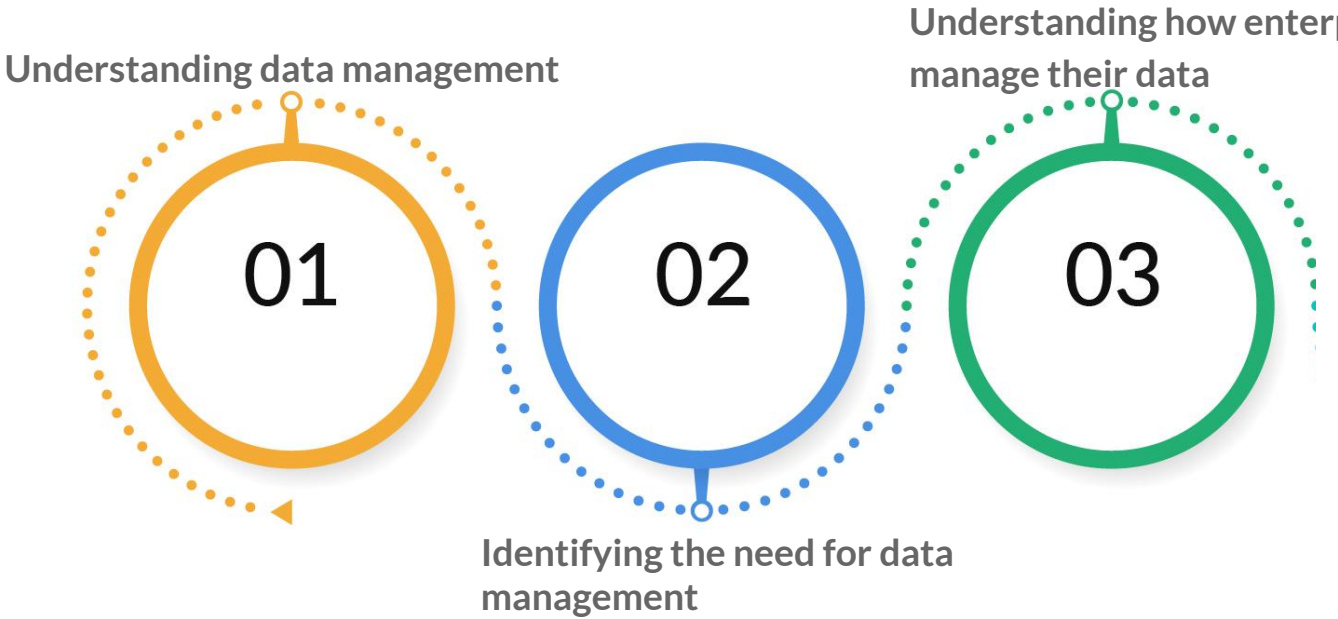
Segment 03 Understanding how and why a company stores data.

Segment 04 Discussion on different components of data management

Segment 05 Understanding the flow of data in an enterprise and why managing data is effective for data-driven decisions.

Segment 3 | Data Management

In This Segment



Where Does Data Come From?

Bus Ticket Booking Company

Customer

User Login
Name
Phone Number
Booking Details
Coupons Used
Date of Joining
Previous Bookings

Source:
Customer App

Bus Agency

Agency Login
Registered Name
Registration Documents
Agency Manager
Phone Number
Date of Joining
Membership Plan

Source:
Agency Portal

Employee

Employee ID
Name
Phone Number
Department
Date of Joining
Attendance
Progress Reports

Source:
Employee Management Software

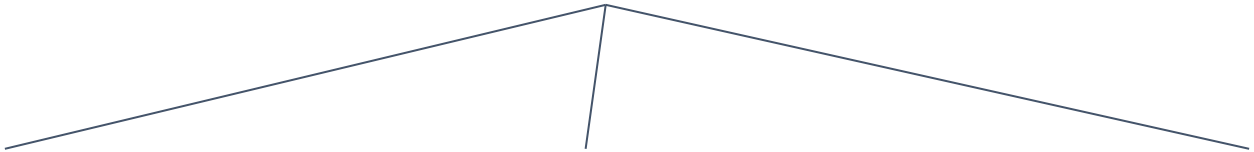
Departments

Sales
Technical
HR
Payroll
Marketing
Daily Activities
Progress Reports

Source:
Department Management Software

Where Does Data Come From?

Bus Ticket Booking Company



Daily Business

New Bookings
Cancelled Bookings
Team Meetings
Daily Tasks
New Customers
New Agencies

Source:
Different Sources

Finance

Operational Costs
Revenue
Profits
Promotional Charges
Financial Reports

Source:
Data Warehouses

Funding

Funding Details
Company Details
Stakeholders
Investors

Source:
Funding Databases

Data Management

1. Collecting, storing and governing the flow of data
2. Integrating data to use it effectively
3. Designing policies and rules to manage data quality and security issues
4. Choosing the right technologies to manage the data at every stage

Why Does an Enterprise Manage Data?

Bus Ticket Booking Company

To run the daily business

Tracking available seats for each bus from different agencies and displaying the updated data to customers

Booking and managing tickets

Receiving payment from customers and paying agencies

Tracking employee performance

To make data-driven decisions

Enhancing customer experience on the application based on customer activity

Providing the best and useful coupons to every customer

Choosing the right promotional and marketing channels to reach every customer

Two Methods of Data Management

File Systems

01 The data is stored in files. Different files are created for each activity. Customers, daily transactions, employees, sales, marketing and other such business activities have their own files.

02 Programs are written to update any information or query these files.

03 One program is for adding daily transactions,
One program is for querying a customer file, and
One program is for data manipulation

DBMS

01 The data is stored in databases. Databases are a collection of tables. There is a different database for each activity. Customers, daily transactions, employees, sales, marketing and other business activities have their own databases.

02 SQL commands are written to update any information or query these files.

Summary | Data Management

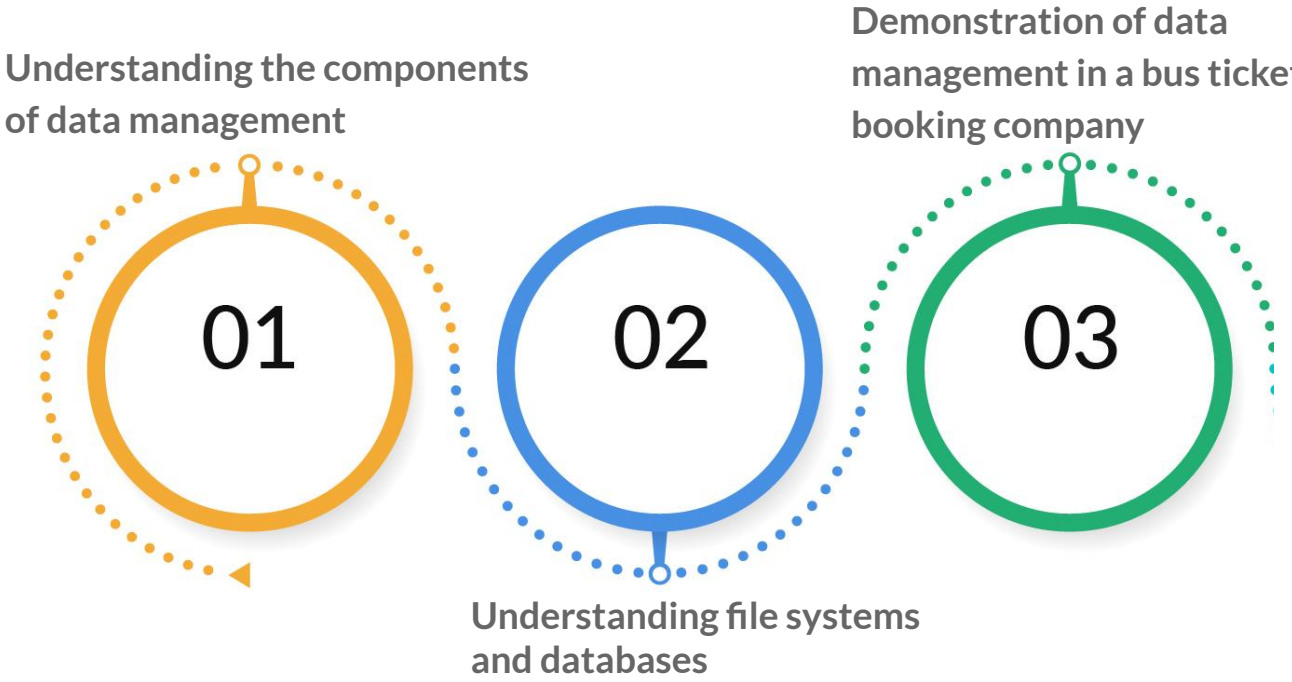
A company collects data from various sources.

Data management is the ability of a company to store and manage its data efficiently.

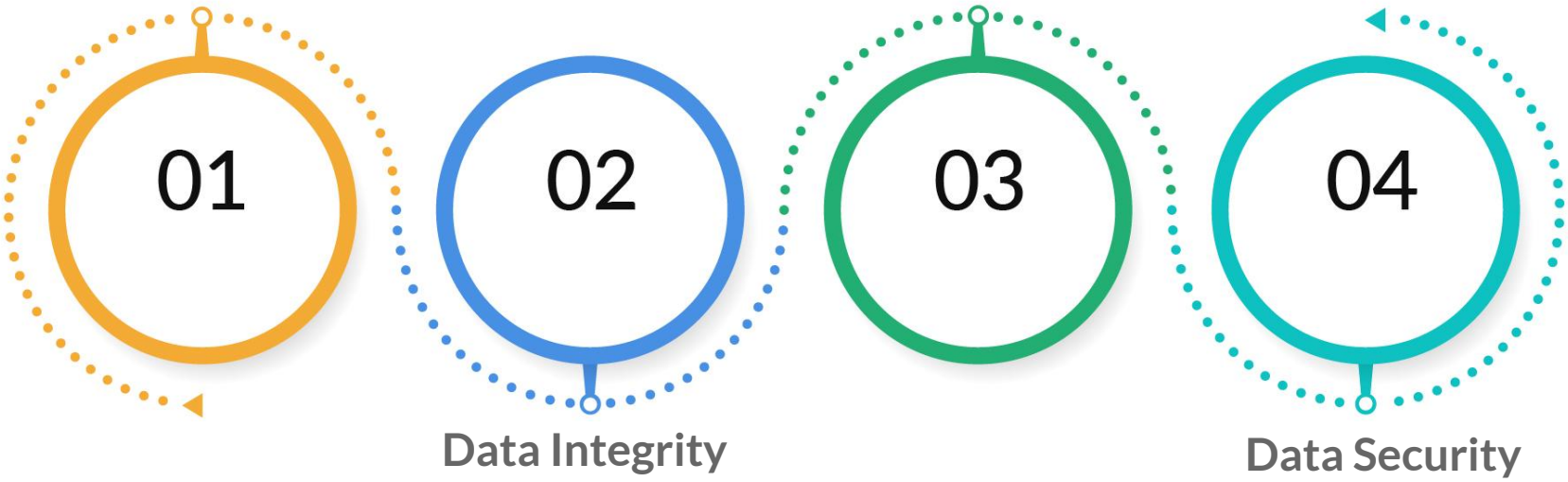
A company manages data to run the daily business and make data-driven decisions.

Segment 4 | Components of Data Management

In This Segment



Components of Data Management



Data Governance

Data Governance is a set of rules and policies that govern the flow of data in a company.

Every bit of data is collected, stored and managed as per the rules set up by a company.

How to handle incorrect data

- In case certain fields in a table are null, whether to use default values according to the data type or add a descriptive detail such as 'Not Specified'
- In case there are many null fields in a row, whether to reject it or make changes to the row and add it to the data processing stream

- Where to store the activity log data of a customer on the app
- Which applications to use to store various kinds of data
- Who can access the data stored at various places

- Which connectors to use to extract data from various sources
- Which applications to use to analyse the data
- Whether to use a warehouse or data lakes as the central repository

Data Governance

Data governance decides the tools that need to be used in order to implement data security, data integration and data quality.

It answers the following questions:

- What is the need for setting certain rules for data usage?
- What are the rules?
- Where can specific company data be found?
- Who has access to certain data?
- Which technologies can be used for data management and governance?
- How can certain rules be implemented for efficient usage of data?

It sets clear and standardised rules throughout the organisation.

Data Governance

Bus Ticket Booking Company

Let's consider how data governance works in file systems and a data management system (DBMS).

- Specific programs are written for every updation or manipulation of the data stored in files. Hence, it is difficult to keep track of and govern every such application in terms of who is accessing and manipulating the data.
- Rules and policies can be designed, but their implementation can be difficult.
- The data is unstructured, and the metadata of every field is not stored.

- In a DBMS, data is stored in the form of tables. The data in these tables can be updated or used through different SQL commands.
- The data is structured, and the data type and metadata of every field is stored.

Data Integration

Data Integration is the process of compiling data in one central location.

Data inconsistency occurs when the same data copied at two different places displays different values for a particular field.

Every piece of data should have only one copy that can be accessed by business users.

Multiple copies of the same data should not be available at various locations.

If data integration is not done, there can be multiple copies of the same data, making the data inconsistent.

Data Integrity

Bus Ticket Booking Company

Let's consider how data integrity works in file systems and DBMS. The company offers special membership plans to customers.

- Consider the two different files: customers and members.
- Customers who are members have their entire information stored at two places.
- If the phone number of a customer is changed, then the number may change in the customers file but not in the members file. This can lead to data inconsistency.

- In a DBMS, data is stored in the form of tables.
- The members table does not store all the information about every customer who is a member; it only stores the link to such information.
- Hence, the data is stored at one place.

Data Quality

Data Quality is the process of cleaning and standardising the collected data.

Data quality ensures that the data can be used for transactional and analytical purposes.

A complete description of the stored data is available at every stage of data management.

The data must describe every detail about the entity that it represents, and the stored information must not be incorrect.

A description of how the stored data is related must be available.

Data Quality

Bus Ticket Booking Company

Let's consider how data quality works in file systems and DBMSes.

- To update the data in files, specific programs are written.
- For example, to apply the condition that no booking price should go below ₹100, a program is written to ensure the value is greater than 100 during updation.
- However, it becomes difficult to do this when many fields in several files have to be checked.

- In a DBMS, data is stored in the form of tables.
- Only permitted values are checked and stored in every column of a table.

Data Security refers to the process of securing the data within an organisation as well as outside it.

Data security enables only authorised users to access the data.

Data governance policies decide the rules regarding data usage and access within the organisation.

Data Security

Bus Ticket Booking Company

Let's consider how data security works in file systems and DBMSes.

- The technical team should not have access to the financial details of a company.
- The payroll team should have access to employee data only.
- Specific programs are written in file systems to manage the data.
- It is difficult to enforce data security when there are a large number of programs.

- In a DBMS, data is stored in the form of tables.
- It can only be accessed through a DBMS.
- A DBMS can enforce data security to the stored data.

Summary | Components of Data Management

Data can be managed in two different ways: File systems and DBMSes.

Data governance refers to designing policies for managing data. Data integrity is the process of compiling data at one place and ensuring consistent data.

Data quality ensures that the data is correct and metadata about the data is known. Data security ensures that only authorised users can access the data.

Segment 5 | Uses of Managed Data

In This Segment

How the data flows within an enterprise?

01

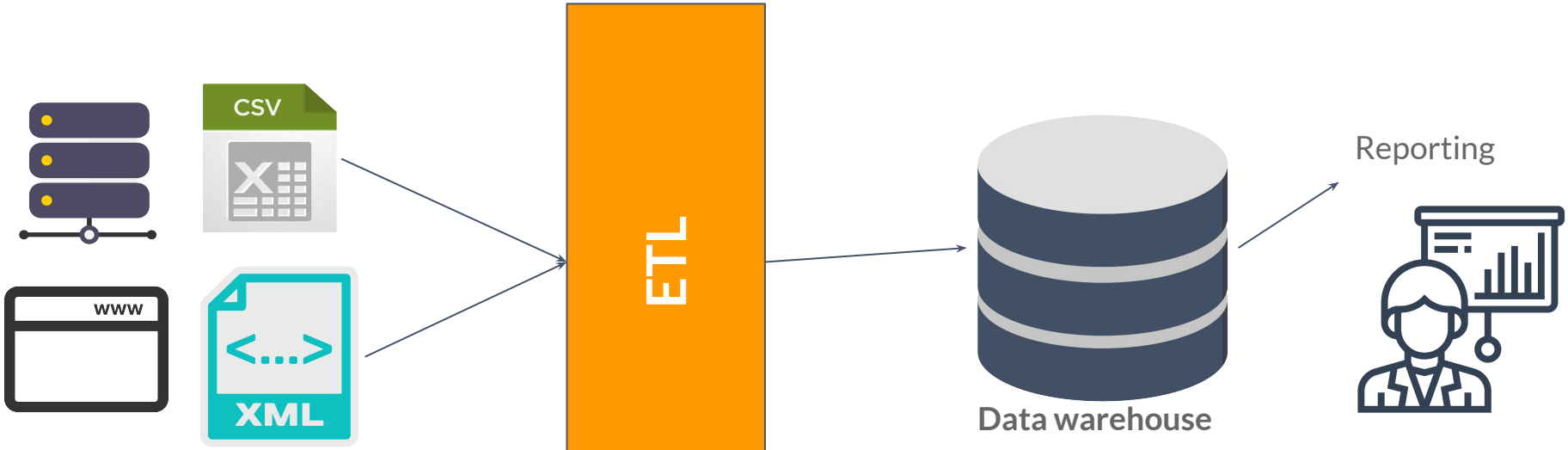
02

A glance at ETL process

03

Reporting various query answers after analysis.

Flow of Data from Data Sources to Reporting Queries



Data sources

File formats and relational modelled data etc.

Data warehouse

Dimensional modelled data

Session Summary

01 **Data Management** is the ability of an enterprise to collect, store, manage and analyse the data.

02 **Data Management** is important to run daily business and to make data-driven decisions

03 **Data Management** has various components that must be followed during the designing process of a data management system.

04 **Data Governance, Data Integrity, Data Quality and Data security** are various components that define the effectiveness of a data management system

05 A company extracts data from its various sources, transforms that data and then load it to a central repository.

06 Analytical Queries can be performed on this data stored in data warehouse.

Thank you