Data Modeling By SkyEss

Let Us Start With The Concepts of Logical Data Model

SkyEss Techno Solutions Pvt. Ltd.

What is A Logical Data Model?

SkyEss Techno Solutions Pvt. Ltd.

- A Logical Data Model in Short is Called As LDM And is A Systems Engineering Process.
- A Logical Data Model is A Representation of An Organization's Data With Standardized Vocabulary And Organization.
- Logical Data Model Organizes The System in Terms of Entities And Relationships.
- Logical Data Model is Designed Independent of Any Particular Data Management Technology And Data Storage Device.
- Logical Data Model Represents The Abstract Structure of A Domain of Information Belonging To The Business Environment.
- Logical Data Model is A Diagrammatical Representation of The Business System in Terms of Entities, Attributes And Relationships.
- Logical Data Models Are Most Typically Used in Business Processes That Seek To Capture Things of Importance To An Organization And How They Relate To One Another.
- Logical Data Model Should Be Based on The Structures Identified in The Conceptual Data Model Phase.
- The Logical Data Model Need Not Anticipate Implementation on A Specific Computing System, But The Content of The Logical Data Model Should Be Adjusted To Achieve Certain Efficiencies.
- The Logical Data Model is Built Using An Entity Relationship Diagram (ERD) Which is A Standard Modeling Technique.

Few Questions Before We Begin

SkyEss Techno Solutions Pvt. Ltd.

Who Uses The Logical Data Model?

- In The Process of System Development The Business Area Experts Own The Logical Data Model.
- The Business Area Experts Use The Logical Models For Impact Analysis of Changes To Business Requirements.

What A Data Modeler Should Do To Build A Logical Data Model?

- The Data Modeler Should Conduct Facilitated Sessions With Business Area Experts To Gather The Data Requirements And Then Build The Logical Data Model Using And Analyzing The Data Requirements of The Business System.

With Whom The Data Modeler Should Coordinate?

- The Data Modeler Should Coordinate With "Process Analyst" To Link The Data With Various Business Processes That Are Playing Crucial Role in System Existence.

Who Should Approve The Logical Data Model?

- The Data Modeler After Completing The Process of Development of Logical Data Model is Responsible For Getting Approval of The Logical Data Model From The Business Area Experts.

What is The Role of Data Modeler After Approval?

- The Data Modeler Should Work in Coordination With Database Administrator(DBA) To Transition The Logical Model To Physical Model Considering The System Architecture And Technology.

What Does The Database Administrator Do?

- The Database Administrator (DBA) Builds The Physical Data Model From The Logical Data Model.
- The Database Administrator Reviews The Logical Model To Select Technology Appropriate Keys, Create Indexes, Detail Data Types, And Builds The Referential Integrity To Protect The Consistency of The Data Values Across The System.

What is The Goal of Database Administrator?

- To Create A Good Quality Database Design Following The Standards of The Physical Model.

Can The Database Administrator Make Any Changes To The Model?

- Yes, The Database Administrator May De-Normalize The Database For increasing The Efficiency And Performance of The Database System.

What Are The Core Responsibilities of A Database Administrator?

- Database Administrators Are Responsible For
  - Creating All The Required Physical Structures of The Business System.
  - Establishing The Data Integrity Concepts Using Constraints.
  - Designing And Deriving The Stored Procedures, Functions And Database Triggers.
  - Planning The Composition of Packages For High Level Reusability.
  - Monitoring The Database Performance.

SkyEss Techno Solutions Pvt. Ltd.

Why Build a Logical Data Model?

SkyEss Techno Solutions Pvt. Ltd.

- The Reason To Build A Logical Data Model is To Confirm The Users And Analysts Understandability of The Business Requirements That Have Been Derived.
- Logical Data Model Assures The System Being Developed, Satisfies The Business Need As Existing in The Physical Business Process.
- Logical Data Modeling Provides The System And Database Analyst With A Tool And Technique To Conduct Analysis More Easily At Enterprise Level.
- Logical Data Model Helps Most Business Area Experts To Articulate The Business Problems And Recognize The Business Isolations That May Be Existing.
- Logical Data Model Makes The Business Experts Assume A Perfect Technology Forcing Users And Analysts To Look Beyond The Current System Limitations.
- Logical Data Model Allows The Business To Drive The System And Also Stimulates Discussion And Thoughts Among The Business Users.
- The Logical Data Model is The Basic Foundation For Designing A Database That Supports The Business Requirements.
- Database Designers Can Start Their Database Design With A Complete Picture of The Business Requirements And Then Proceed With The Best Implementation Approach Using LDM.

Differences Between Logical And Physical Data Model

SkyEss Techno Solutions Pvt. Ltd.

Logical Data Model	Physical Data Model
• Includes All Entities, Relationships, And Attributes And Their Information Types Whether Supported By A Technology OR Not.	• Includes Tables, Columns, Keys, Datatypes, Validation Rules. DB Triggers, Stored Procedures, Domains, And Access Constraints For Security.
• Uses Business Names.	• Names May Be Limited By The DBMS.
• Captures And Records Information Necessary For The Business.	• Includes Technology-Specific Data Elements Such As Flags, Switches, And Timestamps.
• Includes Only Unique Identifiers.	• Includes Primary Keys, Foreign Keys, And Indices For Fast Data Access.
• It is Normalized Upto At Least 3rd Normal Form	• May Be De-Normalized To Meet Performance Requirements.
• Does Not Include Any Redundant Data.	• May Include Redundant Data Elements.
• Does Not Include Any Derived Data.	• May Include Results of Complex OR Difficult To Recreate Calculations.
• Business Experts Drive The Model.	• DBA's Drive The Model.

Additional Benefits Provided By The Logical Data Model

SkyEss Techno Solutions Pvt. Ltd.

- A Logical Data Model Facilitates Data Re-Use And Sharing, As Additional Project Teams Scope Out Their Areas, They Can Re-Use Model Components That Are Shared By The Business.
- Logical Data Model Leads To Physical Data Sharing And Less Storage of Redundant Data.
- Logical Data Model Helps The Organization Recognize That, Information is An Organization-Wide Resource, Not The Property of One Department OR Another.
- Data Sharing Property Provided By The Logical Data Model Makes The Organization More Cohesive And Increases The Quality of Service To Outside Customers And Suppliers.
- Building And Maintaining A Logical Data Model Decreases The System Development And Maintenance Time And Cost of Development.
- Logical Data Model Identifies All Business Requirements At The Beginning of A Project Hence Makes The Design, Coding, Testing, And Implementation Phases Go Much Smoother And Faster.
- A Model is Easier And Cheaper To Modify During The Development Life Cycle Hence Deriving A Logical Data Model Makes The Organization Achieve Cost Effective Standards.
- Logical Data Model Makes Mistakes, Missed Data, And Misinterpretations Less Costly For Correction Than A Correction in An Implemented System.

- Logical Data Model Decreases The User Requests For Changes.
- If Changes Are Considered Necessary Then The Logical Model Can Be Used For Impact Analysis.
- A Logical Data Model is Documented in The Form of An Entity Relationship Diagram Which is A Pictorial Representation of The Business Area.
- Every Entity, Attribute, And Relationship Created And Defined in The Logical Data Model is A Piece of Documentation For The Business System.
- The Objects in The Logical Data Model Contains Textual Definitions That Describe Their Characteristics in Business Language.
- The Logical Data Modeling Once Completed Provides The System Documentation As An Automatic Output.
- A Logical Data Model Confirms A Logical Process Model of The Business System And Provides Documentation of The Information Requirements of The Business Area For Ongoing Impact Analysis.
- Each Business Process is Tied To The Logical Data Model To Assure That All Data And Process Model Components Have Been Discovered.
- Logical Data Model Acts As A Pilot Process For System Implementation And Raises Any Signals of Failure That May Arise in The System Implementation.

Let Us The Concepts of Logical Data Model Using Entity Relationship Model(ER Model)

SkyEss Techno Solutions Pvt. Ltd.

- An Entity – Relationship Model (ER Model) is A Software Engineering Process Which is An Abstract Way To Describe A Database.
- An Entity-Relationship Model (E-R Model) is A Detailed, Logical Representation of The Data For An Organization OR For A Business Area.
- An E-R Model is Expressed in Terms of
  - Entities in The Business Environment
  - The Relationships OR Associations Among Those Entities
  - The Attributes OR Properties of Both The Entities And Their Relationships
- An E-R model is Normally Expressed As An Entity-Relationship Diagram in Short Called As E-R Diagram, OR ERD, Which is A Graphical Representation of An E-R Model.
- Entities Which Are The Objects of An Organization Are Represented By The Rectangle Symbol.
- Relationships Between The Entities Are Represented By Lines Connecting To The Related Entities.
- An E-R Model Clearly Defines All The Objects of The Business System As Metadata For Each Entity Identified in The Business System.
- The Symbols At The End of Each Line on An ERD Specify Relationship Cardinalities, Which Represent How Many Entities of One Kind Relate To How Many Entities of Another Kind.

- There Will Be Actually Two Business Rules For Each Relationship, One For Each Direction From One Entity To The Other.
- Each of The Business Rules Roughly Follows Certain Grammar That Should Be Followed As Per The Given Syntax.

Syntax

- <entity><minimum cardinality><relationship><maximum cardinality><entity>

Illustrative Example

- <CUSTOMER><may><Submit><any number><ORDER>

E-R Model Notation

Point of Concentration Before We Begin

- There is No Industry-Standard Notation For E-R Modeling.
- The Notation Combines Most of The Desirable Features of The Different Notations That Are Commonly Used in E-R Drawing Tools Existing in The Markets Today.
- E-R Model Allows Us To Model Accurately Most Situations That Are Encountered in Practice.
- E-R Models Are Associated With Additional Notation For Enhanced Entity-Relationship Modeling Which Includes Class-Subclass Relationships.
- An E-R Model Should Speak About The Business Requirements in Detail, With All The Aspects of The Entities And Relationships in Practice.
- An E-R Model Should Detail All The Attributes That Describe The Entity in That Business Environment.

Data Modeling By SkyEss

Let Us Make E-R Modeling Notations Practical

SkyEss Techno Solutions Pvt. Ltd.

Entity

- An Entity is A Person, Place, Object, Event, OR Concept in The User Environment About Which The Organization Wishes To Maintain Data.
- An Entity is Identified With A Noun Name.

Classified Categories To Identify Entities

Person Wise

- Every Person Based Instance Identified in The Business Environment Can Be Considered As An Entity.

Illustrative Examples → EMPLOYEE, STUDENT, PATIENT

Place Wise

- Every Place That is An Operational Area For The Business Environment Can Be Considered As An Entity.

Illustrative Examples → STORE, WAREHOUSE, STATE

Object Wise

- Every Object That is An Identified in The Business Environment Can Be Considered As An Entity.

Illustrative Examples → MACHINE, BUILDING, AUTOMOBILE

Event Wise

- Every Event That is Identified in The Business Environment Can Be Considered As An Entity.

Illustrative Examples → SALE, REGISTRATION, RENEWAL

Concept Wise

- Every Concept That is Identified in The Business Environment Can Be Considered As An Entity.

Illustrative Examples → ACCOUNT, COURSE, WORK CENTER

Entity Type

- An Entity Type is A Collection of Entities That Share Common Properties OR Characteristics.
- Each Entity Type in An E-R Model is Given A Name, As The Name Represents A Collection OR Set of Items, it is Always Singular.
- We Should Use Capital Letters For Names of Entity Type(s).
- In An E-R Diagram, The Entity Name is Placed Inside The Box Representing The Entity Type.

Entity Instance

- An Entity Instance is A Single Occurrence of An Entity Type.
- An Entity Type is Described Just Once Using Metadata in A Database, Whereas Many Instances of That Entity Type May Be Represented By Data Stored in The Database.

EMPLOYEE	
Employee Number	NUMBER(4)
Name	VARCHAR2(15)
Address	VARCHAR2(100)
Date Hired	DATE

Entity Type

Entity Instance

Employee Number	Name	Address	Date Hired
1000	Sathish Yellanki	7-1-215/B/1/C/6	12-SEP-2009
1001	Chiranjeevi Anisetty	10-2-345/6/4	10-JUN-2010

Strong Entity

- A Strong Entity Type is One That Exists Independently of Other Entity Types, It is Also Called As Independent Entity.
- Instances of A Strong Entity Type Always Have A Unique Characteristic Called An Identifier.
- An Identifier is An Attribute OR A Combination of Attributes That Uniquely Distinguish Each Occurrence of That Entity.

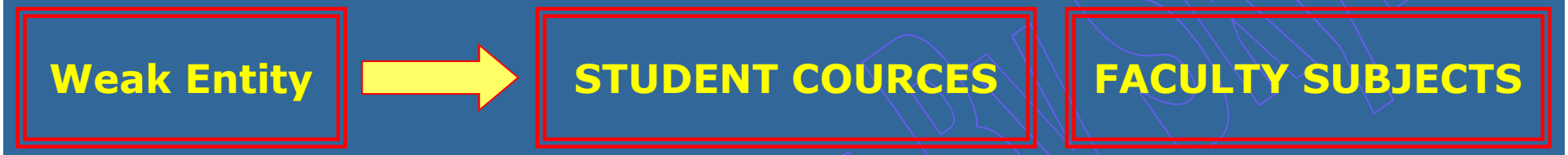
Notation

Weak Entity

- Weak Entity Type is An Entity Type Whose Existence Depends on Some Other Entity Type, It is Also Called As Dependent Entity.
- A Weak Entity Type Has No Business Meaning in An E-R Diagram Without The Entity on Which it Depends.
- The Entity Type On Which The Weak Entity Type Depends is Called The Identifying Owner OR Simply Owner.
- A Weak Entity Type Need Not Have its Own Identifier Definitely.
- In An E-R Diagram, A Weak Entity Type Has An Attribute That Serves As A Partial Identifier.

- During A Later Design Stage A Full Identifier Will Be Formed For The Weak Entity By Combining The Partial Identifier With The Identifier of its Owner OR By Creating A Surrogate Identifier Attribute.

Notation



Identifying Relationship

- The Relationship Between A Weak Entity Type And its Owner.

Associative Entity

- An Associative Entity is An Entity Type That Associates The Instances of One OR More Entity Types And Contains Attributes That Are Peculiar To The Relationship Between Those Entity Instances.
- An Associative Entity is Represented With A Rounded Corner Rectangle.
- Associative Entities Are Also Referred To As Gerunds, As The Relationship Name Which is A Verb is Usually Converted To An Entity Name That is A Noun.
- An Associative Entity Always Represents The Relationship Between Two Entities, Which is A Collection of Represented Data in Business Operations.

When To Convert A Relationship To An Associative Entity Type?

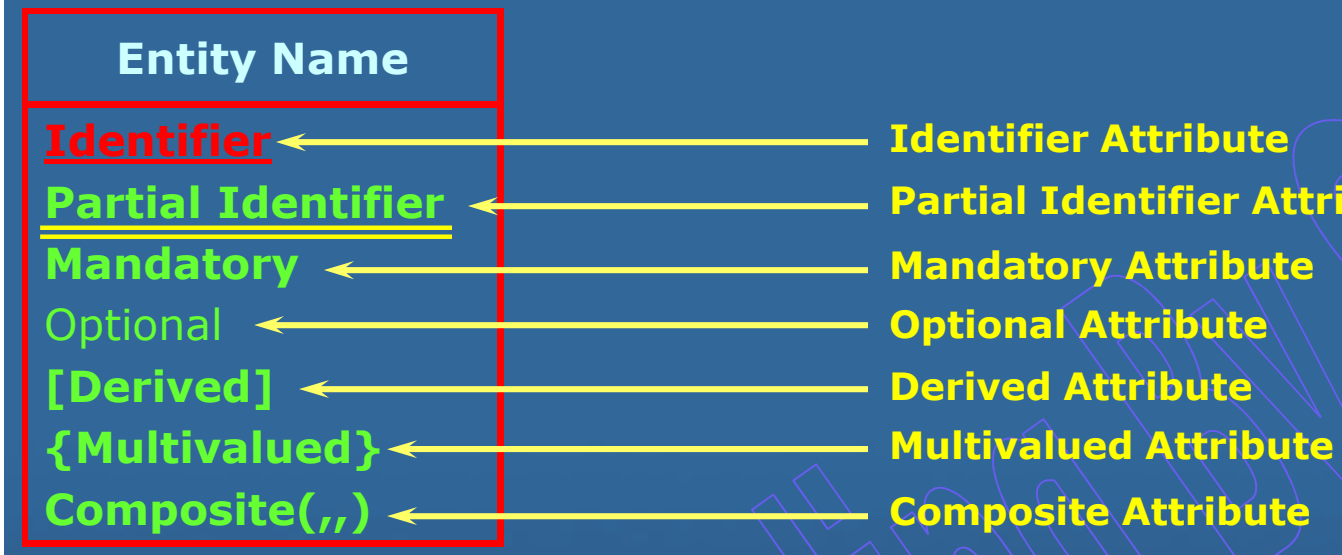
- There Are Four Conditions That Should Be Existing To Convert A Relation To An Associative Entity
  - All The Relationships For The Participating Entity Types Are "Many" Relationships.
  - The Resulting Associative Entity Type Has Independent Meaning To End Users And, Preferably, Can Be Identified With A Single-Attribute Identifier.
  - The Associative Entity Has One OR More Attributes in Addition To The Identifier.
  - The Associative Entity Participates in One OR More Relationships Independent of The Entities Related in The Associated Relationship.

Notation

Attributes

- An Attribute is A Property OR Characteristic of An Entity Type That is of Interest To The Organization.
- Some Types of Relationships May Also Have Attributes, That is An Associative Entity Has Attributes Describing The Relationship Among The Entities.
- While Naming Attributes, We Should Use An Initial Capital Letter Followed By Lowercase Letters.
- If An Attribute Name Consists of More Than One Word, We Should Use A Space Between The Words And We Should Start Each Word With A Capital Letter.
- In An E-R Diagram, We Represent An Attribute By Placing its Name in The Entity it Describes.
- Each Entity OR Instance of An Entity Type Potentially Has A Value Associated With Each of The Attributes of That Entity.
- An Attribute That Must Be Present For Each Entity Instance is Called A Required OR Mandatory Attribute, Whereas An Attribute That May Not Have A Value is Called An Optional Attribute.
- In E-R Diagramming Notations, Required OR Mandatory Attributes Are Prefixed "*" And Optional Attributes With "o", OR Required OR Mandatory Attributes in Boldface, And Optional Attributes in Normal Font.
- An Attribute Without A Value is Said To Be "NULL".

Notation



Illustrative Example

STUDENT
- Student ID
- Student Name
- Home Address
- Phone Number
- Major

AUTOMOBILE
- Vehicle ID
- Color
- Weight
- Horsepower

EMPLOYEE
- Employee ID
- Employee Name
- Payroll Address
- Skill

SkyEss Techno Solutions Pvt. Ltd.

Simple OR Atomic Attribute

- Simple OR Atomic Attribute is An Attribute That Cannot Be Broken Down into Smaller Components That Are Meaningful For The Organization.

Illustrative Examples

Vehicle

- Vehicle ID
- Color
- Weight
- Horse Power

Composite Attributes

- Some Attributes Can Be Broken Down into Meaningful Component Parts Called Detailed Attributes, Such Attributes Are Called As Composite Attributes.
- The Decision To Subdivide An Attribute into its Component Parts Depends on The Need To Refer To These Individual Components in Business OR Not.
- Composite Attributes if Not Analyzed Properly Can Be Misunderstood As Multivalued Attributes.

Illustrative Examples

Name

- First Name
- Middle Name
- Last Name

Address

- House Number
- Street Name
- City Name

Phone

- Country Code
- STD Code
- Actual Phone Number

Single Valued Attribute

- A Single Valued Attribute is An Attribute That Manages Only One Value For The Attribute in The Entity Instance.

Illustrative Example

- Student DOB
- Employee HireDate

Multivalued Attribute

- A Multivalued Attribute is An Attribute That May Take on More Than One Value For A Given Entity OR Relationship Instance.
- Multivalued Attribute is Represented With Curly Brackets Around The Attribute Name.
- A Multivalued Attribute, May Occur Multiple Times With Values For Each of Record That is Identified in The Entity Instance.

Illustrative Example

- Employee Skills
- Student Hobbies

Stored Value Attribute

- A Stored Value Attribute is An Attribute That Records The Values of The Business System That Are Direct Values Identified in The Business Process.

Illustrative Example

- Course Fees
- Salary Fixed
- Commission Percentage

Derived Value Attribute

- A Derived Value Attribute is An Attribute Whose Values Can Be Calculated From Related Attribute Values.
- A Derived Value Attribute Also Can Be Some Data That is Not Directly Available in The Database.
- We Indicate A Derived Attribute in An E-R Diagram By Using Square Brackets Around The Attribute Name.
- In Certain Situations, The Value of An Derived Value Attribute Can Be Derived From Attributes in Related Entities.

Illustrative Example

- Students Age
- Employees Experience
- Employees Gross Salary

Identifier Attribute

- An Identifier Attribute is An Attribute OR Combination of Attributes Whose Value Distinguishes Individual Instances of An Entity Type And An Identifier Attribute Also Should Be Mandatory By Nature.
- No Two Instances of The Entity Type May Have The Same Value For The Identifier Attribute.
- An Identifier Attribute is Underlined in The E-R Diagram.

Illustrative Example

- Employee ID
- Student ID

Composite Identifier

- A Composite Identifier is An Identifier That Consists of A Composite Attribute.
- The Composite Identifier Identifies The State of Uniqueness in A Combination To Identify Uniquely Individual Occurrences.

Illustrative Example

- Train ID, Train Number, Train Date
- Exam ID, Exam Name, Exam Date
- Flight ID, Flight Name, Flight Date

Criteria For Selecting Identifiers

- Choose An Identifier That Will Not Change its Value Over The Life of Each Instance of The Entity Type.
- Choose An Identifier Such That For Each Instance of The Entity, The Attribute is Guaranteed To Have Valid Values And Not Be NULL OR Unknown.
- If The Identifier is A Composite Attribute, Then Make Sure That All Parts of The Identifier Will Have Valid Values.
- Avoid The Use of So-Called Intelligent Identifiers OR Keys, Whose Structure Indicates Classifications, Locations, And So On.
- Consider Substituting Single-Attribute Surrogate Identifiers For Large Composite Identifiers.
- Identifiers Should Not Be Re-Allocatable To Other Entity Instances After An Instance is Deleted.

Modeling Relationships in E-R Model

SkyEss Techno Solutions Pvt. Ltd.

- Relationships Act Like Glue With in The E-R Model And They Hold Together The Various Components of An E-R Model Upon The Rule of The Business System.
- A Relationship is An Association Representing An Interaction Among The Instances of One OR More Entity Types That is of Interest To The Organization.
- A Relationship is Identified With A Verb Phrase in The Clients Business Communication Process.
- Relationships And Their Characteristics Identified As Degree And Cardinality Represent Business Rules of The System.
- Relationships in A System Are Identified in Two Ways
  - Relationship Types
  - Relationship Instances

Relationship Type

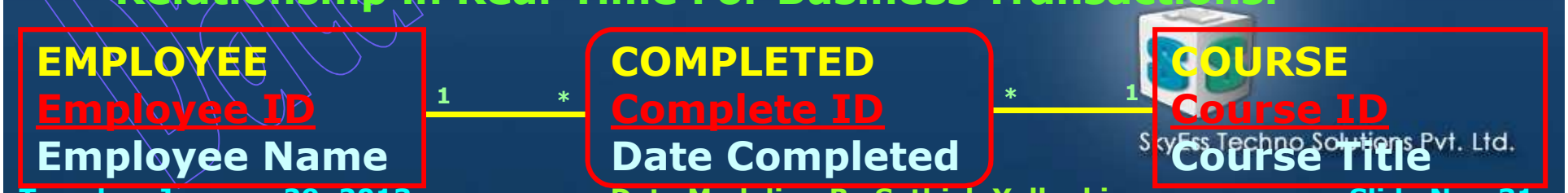
- A Relationship Type is A Meaningful Association Between OR Among Entity Types.
- The Relationship Type is The Verb Phrase That is Representing The Existential State of The Relation Between The Two Entities.
- A Relationship Type is Denoted By A Line Labeled With The Name of The Relationship in The E-R Model.

Relationship Instances

- A Relationship Instance is An Association Between OR Among Entity Instances, Where Each Relationship Instance Associates Exactly One Entity Instance From Each Participating Entity Type.
- Relationship Instances Are The Actual Physical Values That Are Part of The Business Data That Are Existing Through The Business Operations OR Transactions.

Attributes on Relationships

- Attributes May Be Associated With A Many-To-Many OR One-To-One Relationship.
- The Attributes on Relationship Are Collected By Representing An Entity That is Specially Defining The Corresponding Relationship Between The Entities.
- Entities That Actually Define The Data Being Collected Upon The Relationship Are Called As Associative Entities.
- As Per The Standards of The E-R Model Every One-To-One And Many-To-Many Relationship in The Business Environment Has To Be Converted To A Special Case of One-To-Many Relationship in Real-Time For Business Transactions.



Let Us Understand The Degree of a Relationship

SkyEss Techno Solutions Pvt. Ltd.

What is Meant By Degree of The Relationship?

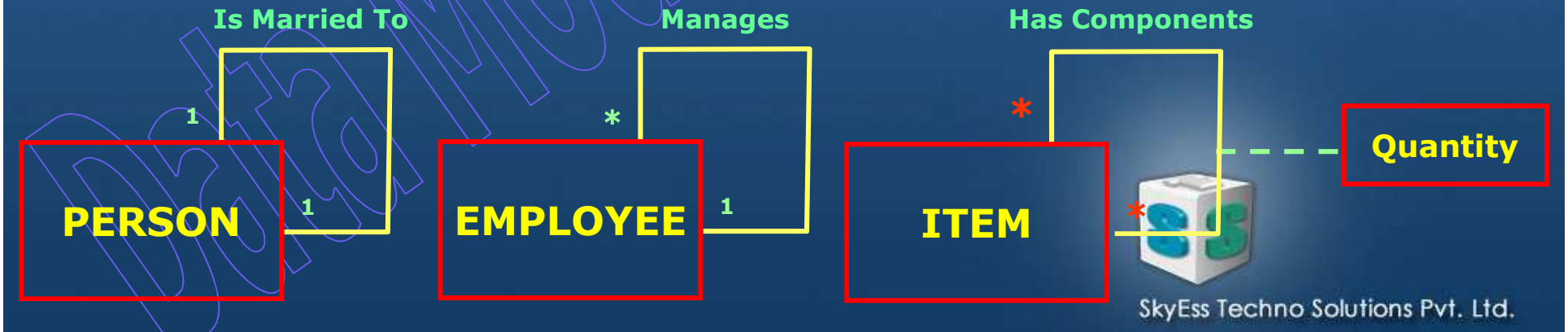
- The Degree of A Relationship is The Number of Entity Types That Participate in That Relationship.
- The Types of Relationship Degrees in E-R Models
  - Unary Which is of Degree 1
  - Binary Which is of Degree 2
  - Ternary Which is of Degree 3
  - N-Ary Which is of More Than 3

Note

- Data Model Represents A Specific Situation, Not A Generalization in The Business Model.

Unary Relationship

- A Unary Relationship is A Relationship Between The Instances of A Single Entity Type.
- Unary Relationships Are Also Called Recursive Relationships.



Binary Relationship

- A Binary Relationship is A Relationship Between The Instances of Two Entity Types.
- Binary Relationship is The Most Common Type of Relationship Encountered in Data Modeling Process.

Illustration 1

- One Employee is Assigned With Only One Parking Place

Illustration 2

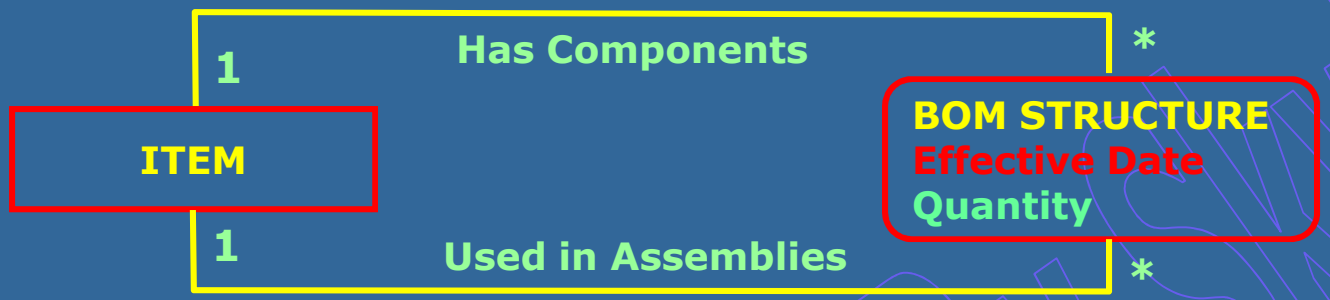
- One Product Line May Contain Many Products.

Illustration 3

- One Student May Register For More Than One Course.

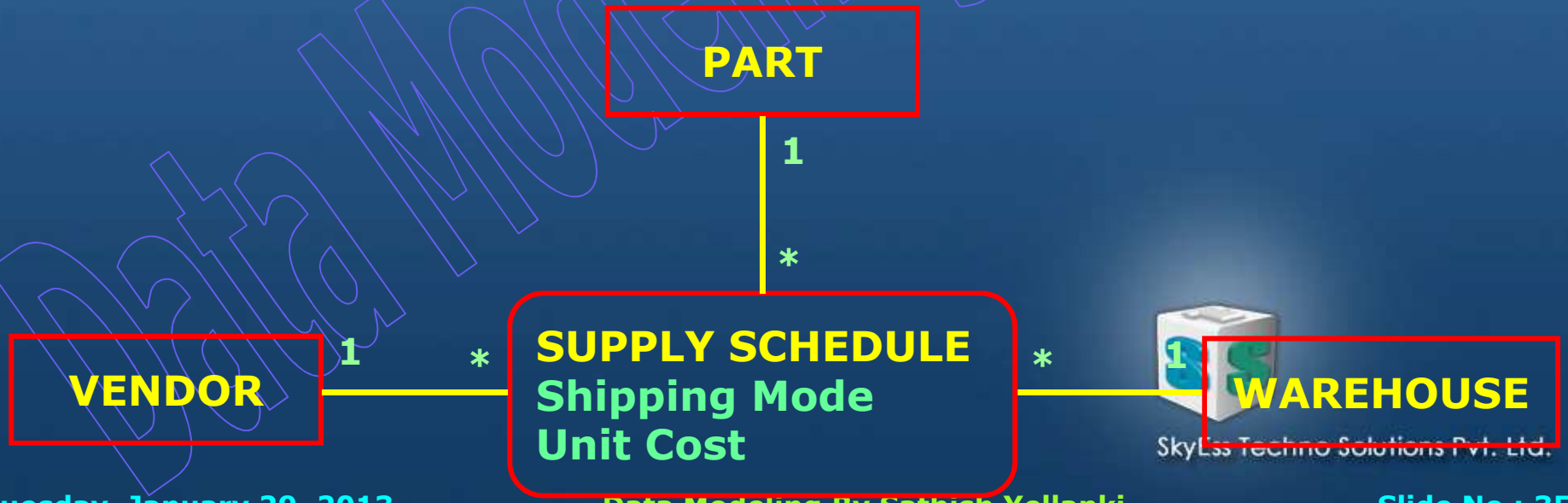
SkyEss Techno Solutions Pvt. Ltd.

Associative Entity



Ternary Relationship

- A Ternary Relationship is A Simultaneous Relationship Among The Instances of Three Entity Types.
- A Ternary Relationship is Not The Same As Three Binary Relationships.
- In Ternary Relationship We Should Convert All Ternary OR Higher Relationships To Associative Entities.



Let Us Kick Start With Normalization

SkyEss Techno Solutions Pvt. Ltd.

- Normalization is The Process of Successively Reducing Relations With Anomalies To Produce Smaller, Well-Structured Relations.
- Normalization is A Formal Process For Deciding Which Attributes Should Be Grouped Together in A Relation So That All Anomalies Are Removed.

Occasions That Usually Benefit From Using Normalization

- During Logical Database Design We Should Use Normalization Concepts As A Quality Check For The Relations That Are Obtained From Mapping E-R Diagrams.
- When Reverse-Engineering, As In Older Systems Many of The Tables And User Views in Older Systems Are Redundant And Subject To The Anomalies.

What Are The Goals of Normalization?

- Minimize Data Redundancy, Thereby Avoiding Anomalies And Conserving Storage Space
- Simplify The Enforcement of Referential Integrity Constraints
- Make it Easier To Maintain Data While Insert, Update, And Delete Operations Are Conducted.
- Provide A Better Design That is An Improved Representation of The Real World And A Stronger Basis For Future Growth.
- Keep The System on The Required Scope of Information And The Context of The Information.
- Provide A Modular Idea For The Entire System For Development

What Normalization Does Not Do?

- Normalization Makes No Assumptions About How Data Will Be Used in Displays, Queries, OR Reports.
- Normalization Places No Constraints on How Data Can OR Should Be Physically Stored OR, Does Not Dictated The Issues of Processing Performance.

Is Normalization is Based on Any Concept?

- Yes, Normalization is Based on The Concept of Normal Forms And Functional Dependencies, Which Define The Rules of The Business, Not Data Usage.

Is Normalization A Technique?

- Yes, Normalization is A Logical Data Modeling Technique Used To Ensure That The Business Data is Well Structured From An Organization-Wide View OR Enterprise Level View.

What is Meant By Normal Form?

- A Normal Form is A State of A Relation That Requires Certain Rules Regarding Relationships Between Attributes OR Functional Dependencies To Be Satisfied.

How To Understand Normalization?

- Normalization Can Be Accomplished And Understood in Stages, Each of The Stage Corresponds To A Normal Form.

What is The Final Result of Normalization?

- A Well Structured System With High Degree of Consistency And Integrity For Data Collection.

Let Us Catch Some Terminology And Jargon Before We Begin Normalization Process

SkyEss Techno Solutions Pvt. Ltd.

Functional Dependencies And Keys

- Up To The Boyce-Codd Normal Form, Normalization is Based on The Analysis of Functional Dependencies.
- A Functional Dependency is A Constraint Between Two Attributes OR Two Sets of Attributes.
- For Any Relation R, Attribute B is Functionally Dependent on Attribute A If, For Every Valid Instance of A, The Value of A Uniquely Determines The Value of B.
- The Functional Dependency of B on A is Represented By An Arrow, A → B.
- A Functional Dependency is Not A Mathematical Dependency Hence B Cannot Be Computed From A.
- As Per The Functional Dependency If We Know The Value of A, There Can Be Only One Value For B.
- An Attribute May Be Functionally Dependent on A Combination of Two OR More Attributes Rather Than on A Single Attribute.

Functional Dependency Representation

- SSN → Name, Address, Birthdate
- VIN → Make, Model, Color
- ISBN → Title, FirstAuthorName, Publisher
- EmpID, CourseTitle → DateCompleted

Determinants

- The Attribute on The Left Side of The Arrow in A Functional Dependency is Called A Determinant.

Candidate Keys

- A Candidate Key is An Attribute, OR Combination of Attributes, That Uniquely Identifies A Row in A Relation.

Properties of A Candidate Key

Unique Identification

- For Every Row, The Value of The Key Must Uniquely Identify That Row.
- Each Nonkey Attribute is Fully Functionally Dependent on The Candidate Key.

Non-Redundancy

- No Attribute in The Key Can Be Deleted Without Destroying The Property of Unique Identification.

Note

- A Candidate Key is Also A Primary Key As Per The Relation That is Identified in The Business System.

Functional Dependency Diagram

Illustration 1 : Employee 01 Relation

SkyEss Techno Solutions Pvt. Ltd.

Let Us Visualize The Steps in Normalization

SkyEss Techno Solutions Pvt. Ltd.

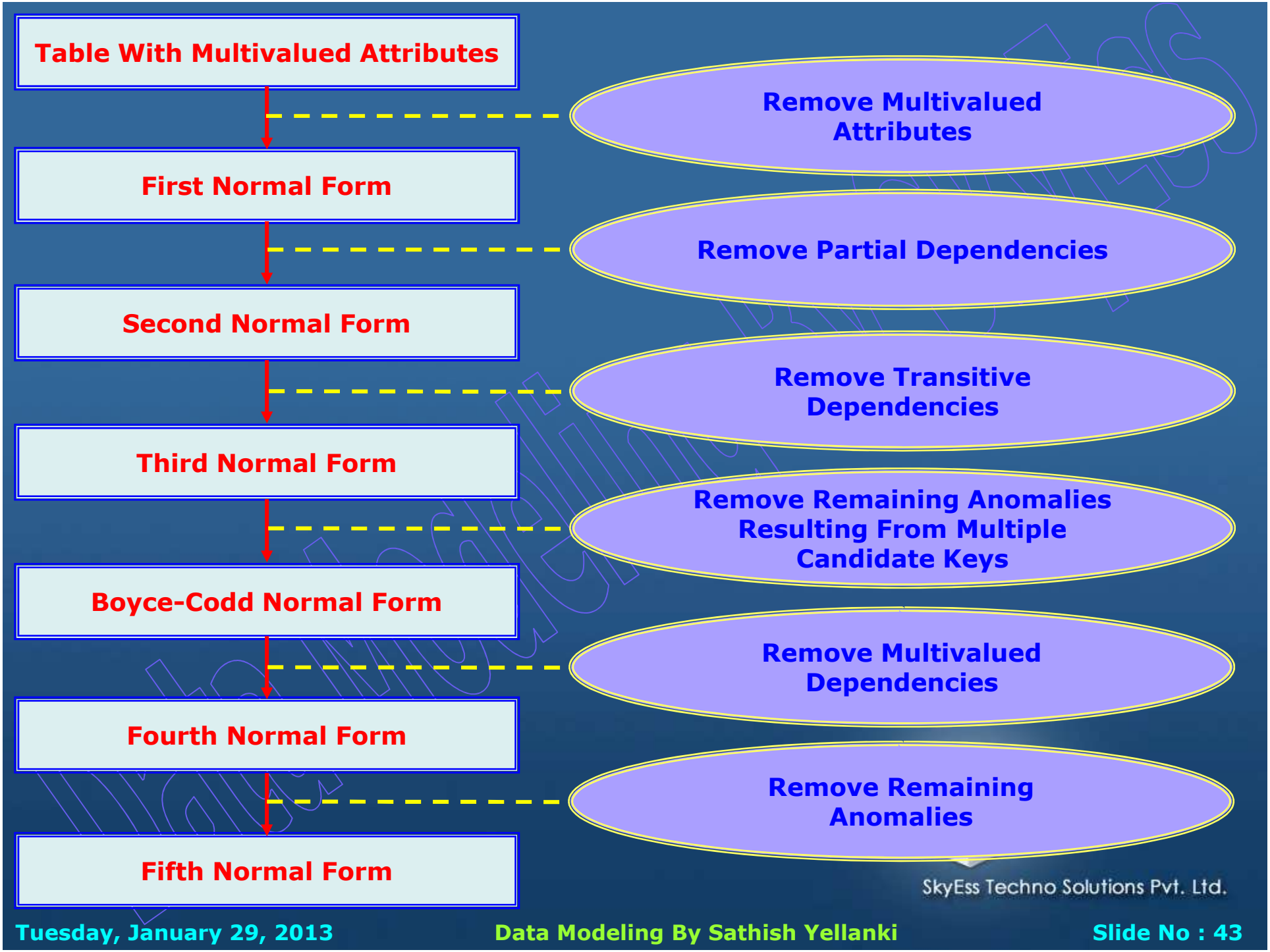


Table With Multivalued Attributes
First Normal Form
Second Normal Form
Third Normal Form
Boyce-Codd Normal Form
Fourth Normal Form
Fifth Normal Form
Remove Multivalued Attributes
Remove Partial Dependencies
Remove Transitive Dependencies
Remove Remaining Anomalies Resulting From Multiple Candidate Keys
Remove Multivalued Dependencies
Remove Remaining Anomalies
SkyEss Techno Solutions Pvt. Ltd.

Let Us Apply The Steps in Normalization

SkyEss Techno Solutions Pvt. Ltd.

Step 0 : Represent the View in Tabular Form

- The First Step Which is Preliminary To Normalization is To Represent The User View As A Single Table, OR Relation, With The Attributes Recorded As Column Headings.
- Sample Data Should Be Recorded in The Rows of The Table, Including Any Repeating Groups That Are Present in The Data.

SkyEss Techno Solutions Pvt. Ltd.

Step 1 : Convert to First Normal Form

- A Relation is in First Normal Form if The Domain of Each Attribute Contains Only Atomic Values, And The Value of Each Attribute Contains Only A Single Value From That Domain.

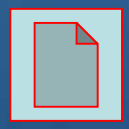
SkyEss Techno Solutions Pvt. Ltd.

Rules To Accept The Data in First Normal Form

- The Actual Data Should Not Have Any Repeating Groups in The Relation. That is With in The Relation For Every Single Fact There Should Be A Single Fact At The Intersection of Each Row And Column of The Table.
- A Primary Key Has Been Defined, Which Uniquely Identifies Each Row in The Relation.

Rule 1 : Removing The Repeating Groups

- Remove Repeating Groups From A Table By Filling Relevant Data Values Into The Vacant Cells of The Table.



Rule 2 : Select The Primary Key

- Find The Determinants in The Relation And Their Functional Dependencies.
- In The Example Data We Have Four Determinants And Each Determinant Having Its Own Functional Dependencies.

Step 2 : Convert to Second Normal Form

- **A Relation is in Second Normal Form if it is in First Normal Form And Contains No Partial Functional Dependencies.**

Rules To Accept The Data in Second Normal Form

- A Relation Should Be in First Normal Form.
- Every Nonkey Attribute Should Be Fully Functionally Dependent on The Primary Key.

What is A Partial Functional Dependency?

- A Functional Dependency in Which One OR More Nonkey Attributes Are Functionally Dependent on Part of The Primary Key Columns But Not All of The Primary Key Columns.

Steps To Eliminate Partial Functional Dependency

Step 1 → Creating A New Relation

- Create A New Relation For Each Primary Key Attribute OR Combination of Attributes, That is A Determinant in A Partial Dependency. The Identified Determinant is The Primary Key in The New Relation.

Step 2 → Move The Nonkey Attributes

- Move The Non Key Attributes That Are Dependent on The Identified Primary Key Attribute OR Attributes From The Old Relation To The New Relation.

Final Effect of Removing Partial Functional Dependencies

- Removal of Each Partial Functional Dependency Results in The Formation of Two New Relations.
- And Each Relation Should Be Identified For A Situation of Fully Functional Dependency.

Final Conclusion

- A Relation That is in First Normal Form Will Be Accepted To Be in Second Normal Form if Any One of The Following Conditions OR Rules Are Applied
  - The Primary Key Consists of Only One Attribute, By Definition There Cannot Be A Partial Dependency In Such A Relation.
  - No Nonkey Attributes Exist in The Relation, Hence All of The Attributes in The Relation Are Components of The Primary Key, And There Are No Functional Dependencies in Such A Relation.
  - Every Nonkey Attribute is Functionally Dependent on The Full Set of Primary Key Attributes.

Step 3 : Convert to Third Normal Form

- A Relation is in Third Normal Form if it is in Second Normal Form And No Transitive Dependencies Exist.

Rules To Accept The Data in Third Normal Form

- A Relation Should Be in Second Normal Form.
- The Relation Should Not Contain Any Transitive Dependencies.

What is Transitive Dependency?

- A Transitive Dependency in A Relation is A Functional Dependency Between The Primary Key And One OR More Nonkey Attributes That Are Dependent on The Primary Key Via Another Nonkey Attribute.

Steps To Eliminate Transitive Dependency

Step 1 → Creating A New Relation With Determinant

- For Each Nonkey Attribute OR Set of Attributes That is A Determinant in A Relation, Create A New Relation.
- The Relation That is Created Should Make That Attribute OR Set of Attributes The Primary Key of The New Relation.

Step 2 → Move The Functionally Dependent Attributes

- Move All of The Attributes That Are Functionally Dependent on The Primary Key of The New Relation From The Old To The New Relation.

Step 3 → Set The Required Foreign Keys

- Leave The Attribute That Serves As A Primary Key in The New Relation in The Old Relation To Serve As A Foreign Key, Such That it Allows Us To Associate The Two Relations.

What is The Effect of Transitive Dependencies?

- Transitive Dependencies Create Unnecessary Redundancy That May Lead To insert, Delete And Update Anomalies.

The Final Effect of Eliminating Transitive Dependencies

- Removal of Transitive Dependencies Results in The Control of Redundant Data.
- The Control of Redundancy Gives The Advantage of Saving The Memory Space As Well of Consistency.
- The Overall Integrity of The System Increases There By Increasing The Quality.
- The System Becomes Free of Update, Insertion, And Deletion Anomalies.

SkyEss Techno Solutions Pvt. Ltd.

Step 4 : Convert to Boyce-Codd Normal Form

- A Relation is in Boyce-Codd Normal Form (BCNF) if And Only if Every Determinant in The Relation is A Candidate Key.

Converting A Relation To Boyce-Codd Normal Form

- A Relation That is in 3NF But Not Yet Been Applied For BCNF Can Be Converted To Relations in BCNF Using A Two-Step Process.

Step 1 → Make The Determinant A Component of Primary Key

- The Relation Should Be Modified Such That The Determinant in The Relation Which is Not A Candidate Key Becomes A Component of The Primary Key of The Revised Relation.
- The Attribute That is Functionally Dependent on That Determinant Should Becomes A Nonkey Attribute.

Step 2 → Decomposing The Relation

- Decompose The Relation To Eliminate The Partial Functional Dependency, Which Results in Two Relations.
- The Decomposed Relations Will Be in Third Normal Form.

Point of Concentration

- The New Relations That Are Created Will Be in BCNF As There is Only One Candidate Key That is The Primary Key, For Each Relation.
- As The Relation Has Only One Candidate Key Hence Becomes The Primary Key Satisfying The Rule of Third Normal Form.
- In Practical Sense Third Normal Form And Boyce-Codd Normal Form Are Almost Equivalent.
- Boyce-Codd Normal Form is Essential Only To Eliminate The Anomalies Not Dealt in Third Normal Form.

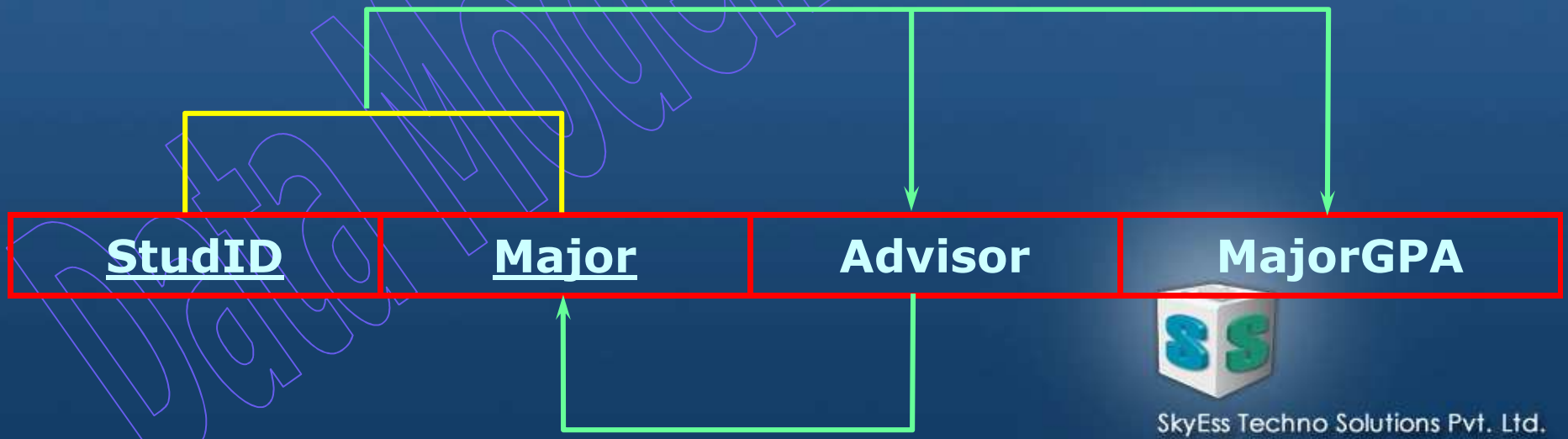
Illustrative Example

SkyEss Techno Solutions Pvt. Ltd.

Student Advisor Data in Third Normal Form

StudID	Major	Advisor	MajorGPA
123	Physics	Hawking	4.0
123	Music	Mahler	3.3
456	Literature	Michener	3.2
789	Music	Bach	3.7
678	Physics	Hawking	3.5

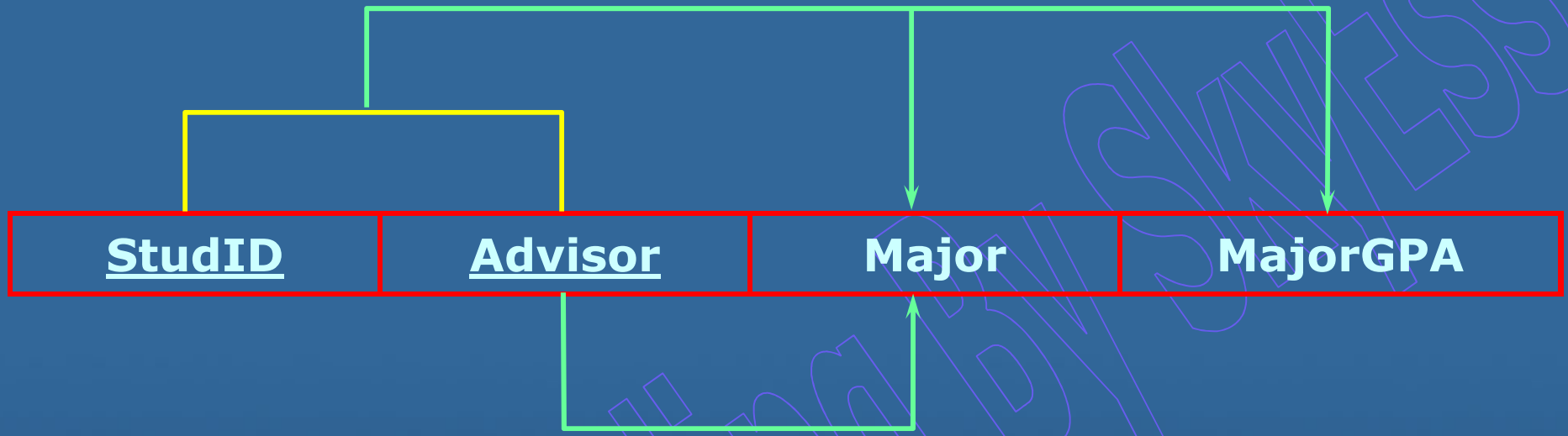
Functional Dependencies in Student Advisor Data



Converting a Relation to Boyce-Codd Normal Form

SkyEss Techno Solutions Pvt. Ltd.

Revised Student Advisor Relation

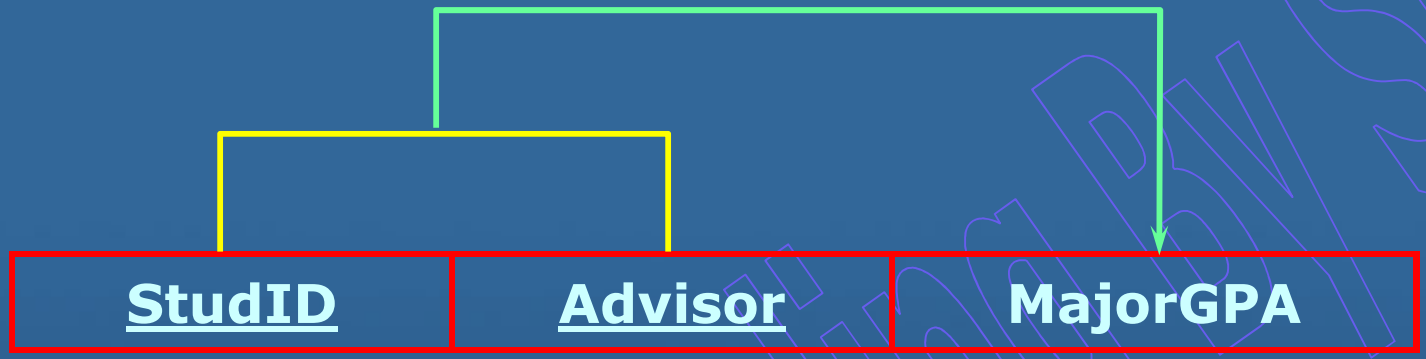


Student Advisor Data in Revised Form

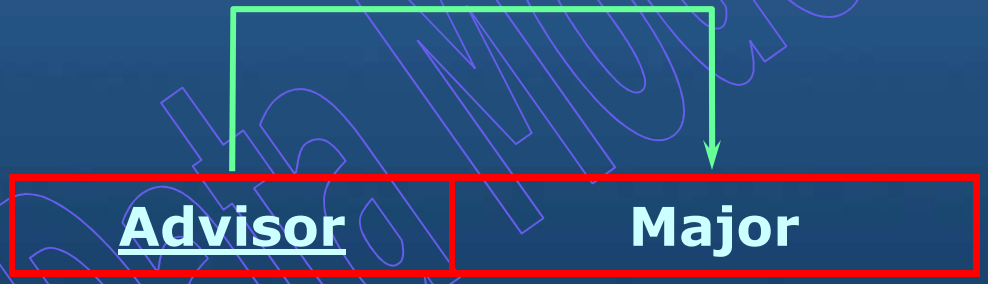
StudID	Advisor	Major	MajorGPA
123	Hawking	Physics	4.0
123	Mahler	Music	3.3
456	Michener	Literature	3.2
789	Bach	Music	3.7
678	Hawking	Physics	3.5

Relations of Student Advisor in Boyce-Codd Normal Form

Relation 01



Relation 02



SkyEss Techno Solutions Pvt. Ltd.

Student Advisor in Boyce-Codd Normal Form With Data

Student Data

StudID	Advisor	MajorGPA
123	Hawking	4.0
123	Mahler	3.3
456	Michener	3.2
789	Bach	3.7
678	Hawking	3.5

Advisor Data

Advisor	Major
Hawking	Physics
Mahler	Music
Michener	Literature
Bach	Music

SkyEss Techno Solutions Pvt. Ltd.

Step 5 : Convert to Fourth Normal Form

- **A Relation is in Fourth Normal Form if The Relation is in BCNF And Contains No Multivalued Dependencies.**

SkyEss Techno Solutions Pvt. Ltd.

What is A Multivalued Dependency?

- The Type of Dependency That Exists When There Are At Least Three Attributes Like A, B, C in A Relation, With A Well-Defined Set of B And C Values For Each A Value, But Those B And C Values Are Independent of Each Other.

How To Remove Multivalued Dependency?

- To Remove The Multivalued Dependency From A Relation, We Divide The Relation into Two New Relations.
- Each of These New Tables Should Contain Two Attributes That Have A Multivalued Relationship in The Original Relation.

Steps To Achieve Fourth Normal Form

Step 1 → **Identify The Relation To Be Boyce-Codd Normal Form**

- Analyze The Relation That it is First Satisfying The Rule of Boyce-Codd Normal Form Having The Required Key Columns.

Step 2 → **Decomposing The Relation**

- Decompose The Relation To Eliminate The Multivalued Dependency Columns, Which Results in Two Relations.
- Set The New Relations With A Set of Key Values To Maintain Them in Fourth Normal Form.

Point of Concentration

- The Fourth Normal Form is a Situation That Arises Only When A Boyce-Codd Normal Form Process is Applied.
- The Fourth Normal Form May Make The System To Achieve Partial Dependency Problems Leading To Fifth Normal Form.

Illustrative Example

SkyEss Techno Solutions Pvt. Ltd.

A Table With A Violation of Fourth Normal Form

Course	Teacher	Hire Date	Text	Copyright
BIT340	Moore	1992	340 Course Pack	2007
BIT340	Moore	1992	Access	2007
BIT340	Ravishankar	1986	340 Course Pack	2007
BIT340	Ravishankar	1986	Access	2007
BIT301	Moore	1992	301 Course Pack	2007
BIT301	Moore	1992	Excel	2007
BIT301	Moore	1992	Being Digital	2005
BIT301	Walls	1993	301 Course Pack	2007
BIT301	Walls	1993	Excel	2007
BIT301	Walls	1993	Being Digital	2005

Identified Relations in Third Normal Form

Teacher → HireDate
Text → Copyright
Course, Teacher, Text

SkyEss Techno Solutions Pvt. Ltd.

The Identified Multivalued Dependencies

- One Course Can Have One OR More Number of Teachers.
- One Course Can Have One OR More Number of Text Books.

Identified Relations

- Course → Teacher
- Course → Text

Identified Relations To Fix into Fourth Normal Form

- Teacher → HireDate
- Text → Copyright
- Course —» Teacher
- Course —» Text

Solution Process

Teacher Data

Teacher	HireDate
Moore	1992
Ravi Shankar	1986
Walls	1993

Course Table

Course	Teacher
BIT340	Moore
BIT340	Ravi Shankar
BIT301	Moore
BIT301	Walls

Book Table

Text	Copyright
340 Course Pack	2007
Access	2007
301 Course Pack	2008
Excel	2007
Being Digital	2005

Course Text Table

Course	Text
BIT340	340 Course Pack
BIT340	Access
BIT301	301 Course Pack
BIT301	Excel
BIT301	Being Digital

Final Presentation of Relations

Teacher Data

Teacher	HireDate

Course Table

Course	Teacher

Book Table

Text	Copyright

Course Text Table

Course	Text

SkyEss Techno Solutions Pvt. Ltd.

Step 6 : Convert to Fifth Normal Form

- **A Relation is in Fifth Normal Form OR Project-Join Normal Form if it is in Fourth Normal Form And it Cannot Have A Lossless Decomposition into Any Number of Smaller Tables.**

SkyEss Techno Solutions Pvt. Ltd.

- Fifth Normal Form Deals With Cases Where Information Can Be Reconstructed From Smaller Pieces of Information That Can Be Maintained With Less Redundancy.
- Second, Third, And Fourth Normal Forms Can Also Serve The Purpose of Reconstruction From Smaller Pieces, But Fifth Normal Form Generalizes To Cases That Are Not Covered By The Other Normal Forms.

Properties of Fifth Normal Form

- Anomalies Can Occur in Relations With Fourth Normal Form if The Primary Key Have Three OR More Fields.
- Fifth Normal Form is Based on The Concept of Join Dependencies, if The Join Dependencies of A Relation Cannot Be Decomposed in Fourth Normal Form Then We Should Decompose The Relation into Fifth Normal Form.
- Fifth Normal Form is Essential in Any Table That May Represent A Many-To-Many-To-Many Relationship Among The Different Attributes.
- A Fifth Normal Form May Split A Single Table into Three OR More Number of Table in One Single Stage.

What is Meant By Pair Wise Cyclic Dependency?

- We Always Need To Know Two Values At A Time i.e. Pair Wise.
- For Any One Value We Need To Know The Other Two Attributes Which Are Cyclic.

Illustrative Example

SkyEss Techno Solutions Pvt. Ltd.

A Table With A Violation of Fifth Normal Form

Salesman Name	Brand Name	Product Type
Jack Schneider	Acme	Vacuum Cleaner
Jack Schneider	Acme	Breadbox
Willy Loman	Robusto	Pruning Shears
Willy Loman	Robusto	Vacuum Cleaner
Willy Loman	Robusto	Breadbox
Willy Loman	Robusto	Umbrella Stand
Louis Ferguson	Robusto	Vacuum Cleaner
Louis Ferguson	Robusto	Telescope
Louis Ferguson	Acme	Vacuum Cleaner
Louis Ferguson	Acme	Lava Lamp
Louis Ferguson	Nimbus	Tie Rack

Identified Relations

Salesman —» Brand

Brand —» Product

Salesman, Brand —» Product

SkyEss Techno Solutions Pvt. Ltd.

Solution Steps

SkyEss Techno Solutions Pvt. Ltd.

Salesman Table

SalesmanID	Salesman Name
1000	Jack Schneider
1001	Willy Loman
1002	Louis Ferguson

Brand Table

BrandID	Brand Name
100	Acme
101	Robusto
102	Nimbus

Products Table

ProductID	Product Type
200	Vacuum Cleaner
201	Breadbox
202	Lava Lamp
203	Pruning Shears
204	Umbrella Stand
205	Telescope
206	Tie Rack

SkyEss Techno Solutions Pvt. Ltd.

Salesman Products Table

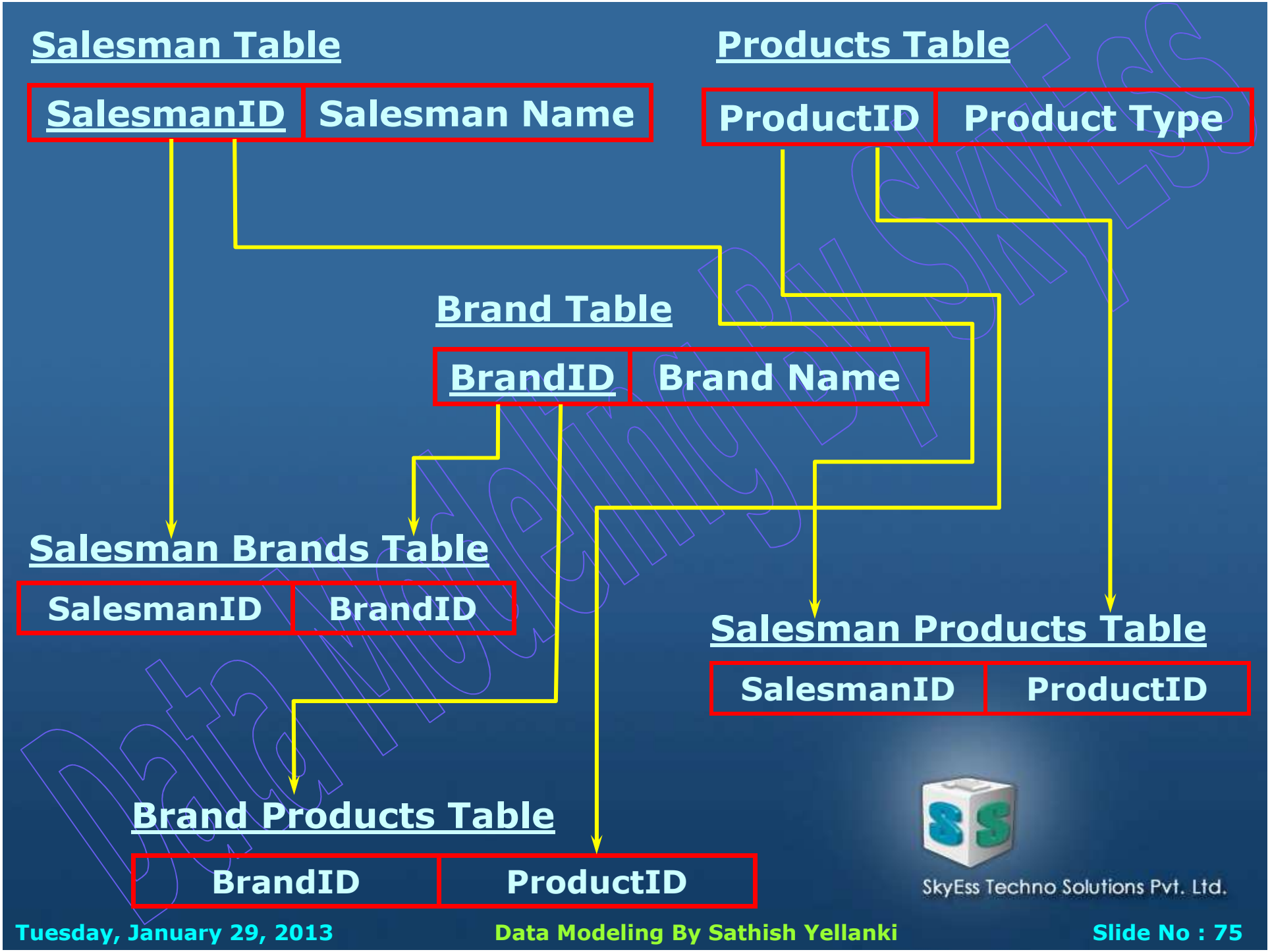
SalesmanID	ProductID
1000	200
1000	201
1001	203
1001	200
1001	201
1001	204
1002	205
1002	200
1002	202
1002	206

Salesman Brands Table

SalesmanID	BrandID
1000	100
1001	101
1002	101
1002	100
1002	102

Brand Products Table

BrandID	ProductID
100	200
100	201
100	202
101	203
101	200
101	201
101	204
101	205
102	206


Salesman Table
SalesmanID
Salesman Name
Products Table
ProductID
Product Type
Brand Table
BrandID
Brand Name
Salesman Brands Table
SalesmanID
BrandID
Salesman Products Table
SalesmanID
ProductID
Brand Products Table
BrandID
ProductID
SkyEss Techno Solutions Pvt. Ltd.