Database Analysis and Design

Madiba Hudson-Quansah

CONTENTS

Chapter 1

Entity-Relationship Modelling

Entity-Relationship modelling is a technique used to model the structure of a database. It is a graphical representation of the entities and their relationships to each other. The entities are the objects that are stored in the database, and the relationships are the connections between the entities. The ER model is used to design the database schema, which is the blueprint for the database.

1.1 Entity Types

> **Definition 1.1.1: Entity Type**
>
> A group of objects with the same properties, which are identified as having an independent existence. Usually drawn on an ER diagram as a rectangle, with the name of the entity type written inside the rectangle.
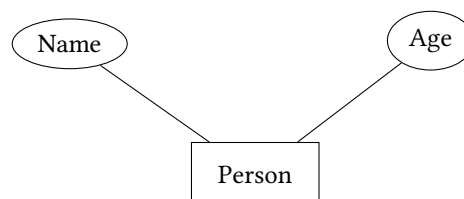
Person
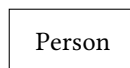
> **Definition 1.1.2: Entity Occurrence**
>
> A uniquely identifiable object of an entity type.

1.2 Attributes

> **Definition 1.2.1: Attribute**
>
> A property of an entity type. Attributes are the characteristics of the entity type that are stored in the database. Attributes are usually drawn as ovals, with the name of the attribute written inside the oval.

Name   Age

Person

1.3 Relationships

A set of meaningful associations among entity types. Relationships are usually drawn as diamonds, with the name of the relationship type written inside the diamond.

Person — Has — Address
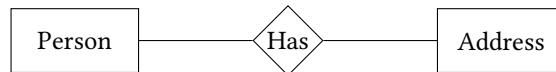
Chapter 2

Normalization

> **Definition 2.0.1: Normalization**
>
> A technique of producing a set of relations with desirable properties, given the data requirements of an application / enterprise.

The purpose of normalization is to identify a suitable set of relations that support the data requirements of an enterprise. The characteristics of a normalized database are:

- The minimal number of attributes necessary to support the data requirements of the enterprise.
- Attributes with a close logical relationship (i.e. functional dependency) are found in the same relation.
- Minimal redundancy, with each attribute represented only once, with important exception of attributes that form all or part of foreign keys, which are necessary for joining relations.

2.1 Data Redundancy and Anomalies

2.2 Functional Dependencies

2.3 Normal Forms

Three normal forms are commonly used in database design:

- First Normal Form (1NF)
- Second Normal Form (2NF)
- Third Normal Form (3NF)

Subsequent normal forms exist, but are not commonly used in practice, for example:

- Boyce-Codd Normal Form (BCNF)
- Fourth Normal Form (4NF)
- Fifth Normal Form (5NF)

2.3.1 Unnormalized Form (UNF)

A relation is in Unnormalized Form (UNF) if it contains repeating groups or multi-valued (non-atomic) attributes. In removing repeating groups in a relation, there are two common approaches:

- By entering the appropriate data in the empty columns of rows containing the repeating data.
- By placing the repeating data, along with a copy of the original key attributes, in a separate relation.

2.3.2 First Normal Form (1NF)

> **Definition 2.3.1: First Normal Form**
>
> A relation in which the intersection of each row and column contains one and only one value. I.e. complete atomicity.

By removing repeating groups, a relation is moved from Unnormalized Form to the First Normal Form (1NF). If a relation is in 1NF, it:

- Has atomic attributes.
- Attributes must be of the same data type.
- It has a primary key.
- There are no repeating attributes.

2.3.3 Second Normal Form (2NF)

> **Definition 2.3.2: Second Normal Form**
>
> A relation that is in 1NF and every non-primary-key attribute is fully functionally dependent on the primary key.

Second Normal form applies to relations that have composite keys i.e. the relations with primary comprised of more than one attribute. A relation with a single attribute primary key is automatically in 2NF.

Normalization from 1NF to 2NF involves:

- Removal of partial dependencies.
- Creation of new relations for the attributes that are partially dependent on the primary key, where the primary key of the original relation is included as a foreign key in the new relation.

2.3.4 Third Normal Form (3NF)

> **Definition 2.3.3: Third Normal Form**
>
> A relation that is in 2NF in which no non-primary-key attribute is transitively dependent on the primary key.