COM1013
INTRODUCTION TO COMPUTER SCIENCE

Lecturer: Begüm MUTLU BİLGE, PhD

begummutlubilge+com1013@gmail.com (recommended)
bmbilge@ankara.edu.tr

Database Systems

Today's technology is capable of storing extremely

large amounts of data, but such data collections are **useless** unless we are able to **extract** those particular items of **information** that are pertinent to the task at hand.

Database Fundamentals

The term **database** refers to a collection of data that is

- **multidimensional** in the sense that **internal links** between its entries make the information accessible from a variety of perspectives.

Traditional file system, **flat file**:

- **A one-dimensional storage system**, meaning that it presents its information from a single point of view

The Significance of Database Systems

Each application tended to be implemented as a separate system with its own collection of data.

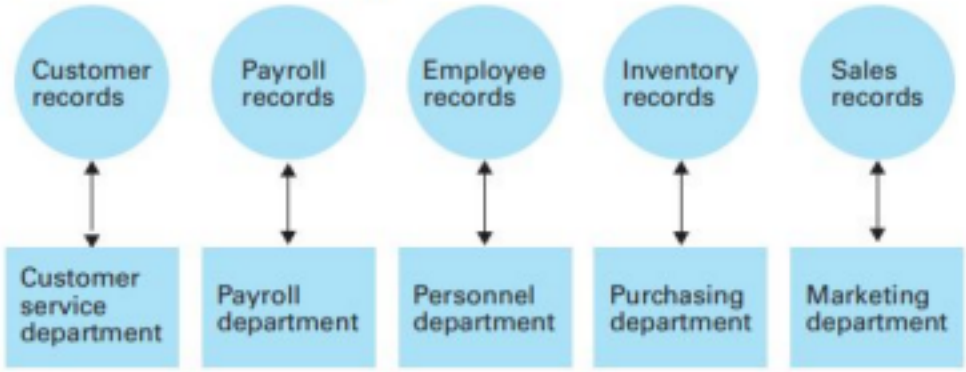
Many different but related items were stored in separate systems

- Duplicated information

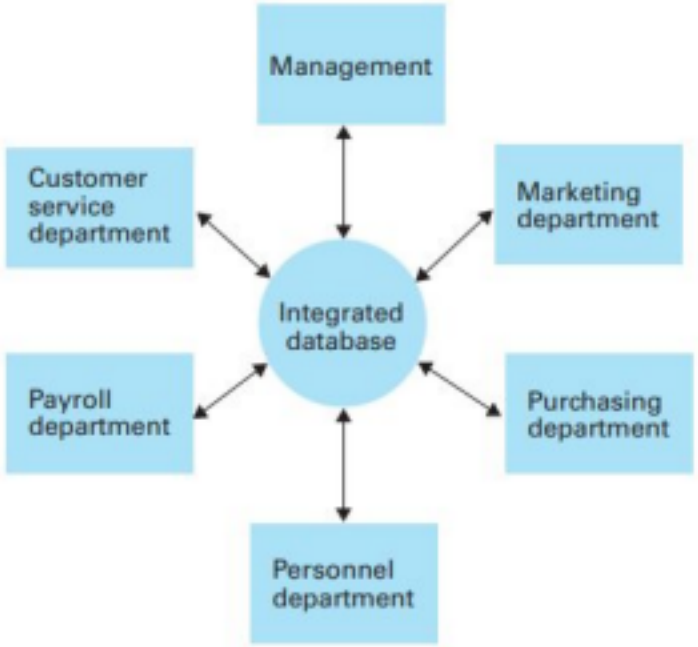
database systems emerged as a means of integrating the information stored and maintained by a particular organization

A file versus database organization

a. File-oriented information system



b. Database-oriented information system



The Significance of Database Systems

Database systems have become the underlying technology that supports many of the more popular sites on the World Wide Web

Google, eBay, and Amazon provide an interface between clients and databases.

- To respond to a client's request,
  - the server interrogates a database,
  - organizes the results in the form of a Web page, and ○ sends that page to the client.

The Role of Schemas

Danger: the potential of **sensitive data** being accessed

by **unauthorized personnel**.

- Someone placing an order at a company's website should not have access to the company's financial data;
- An employee in a company's benefits department may need access to the company's employee records but should not have access to the corporation's inventory or sales records.

Thus the ability to **control access to the information** in the database is as important as the ability to **share** it.

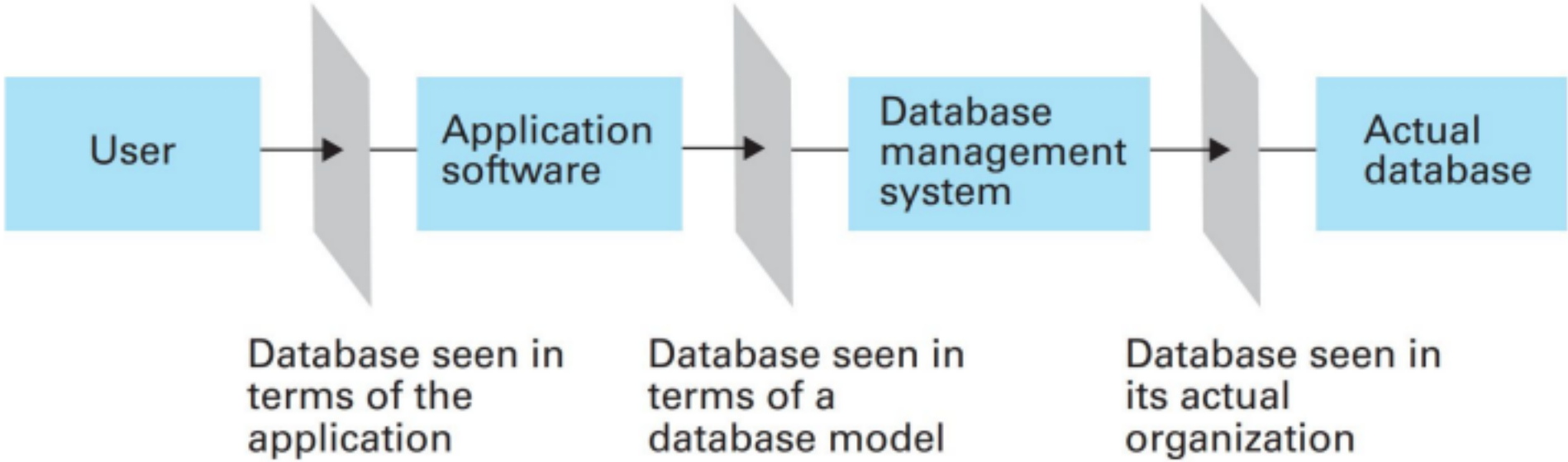
The Role of Schemas

To

Schemas

provide different users access to different information within a database, database systems often rely on **schemas** and **subschemas**.

Database Management Systems



The conceptual layers of a database implementation

Database Models

This conceptual view of the database is called a database model.

- the relational database model
- the object-oriented database model

The Relational Model

Empl Id	Name	Address	SSN
25X15	Joe E. Baker	33 Nowhere St.	111223333
34Y70	Cheryl H. Clark	563 Downtown Ave.	999009999
23Y34	G. Jerry Smith	1555 Circle Dr.	111005555
•	•	•	•
•	•	•	•
•	•	•	•

We want to include information about the jobs held by

the employees. HOW?

What to add: job title (secretary, office manager, floor supervisor), a job identification code (unique to each job), the skill code associated with each job, the department, and the period during which the employee held the job.

The Relational Model

Empl Id	Name	Address	SSN	Job Id	JobTitle	Skill Code	Dept	Start Date	Term Date
25X15	Joe E. Baker	33 Nowhere St.	111223333	F5	Floor manager	FM3	Sales	9-1-2009	9-30-2010
25X15	Joe E. Baker	33 Nowhere St.	111223333	D7	Dept. head	K2	Sales	10-1-2010	*
34Y70	Cheryl H. Clark	563 Downtown Ave.	999009999	F5	Floor manager	FM3	Sales	10-1-2009	*
23Y34	G. Jerry Smith	1555 Circle Dr.	111005555	S25X	Secretary	T5	Personnel	3-1-1999	4-30-2010
23Y34	G. Jerry Smith	1555 Circle Dr.	111005555	S26Z	Secretary	T6	Accounting	5-1-2010	*
•	•	•	•	•	•	•	•	•	•
•	•	•	•	•	•	•	•	•	•
•	•	•	•	•	•	•	•	•	•

What is your opinion about this design ?

The cause of all problems is that we have combined more than one concept in a single relation.

The Relational Model

EMPLOYEE relation

Empl Id	Name	Address	SSN
25X15	Joe E. Baker	33 Nowhere St.	111223333
34Y70	Cheryl H. Clark	563 Downtown Ave.	999009999
23Y34	G. Jerry Smith	1555 Circle Dr.	111005555
•	•	•	•
•	•	•	•
•	•	•	•

JOB relation

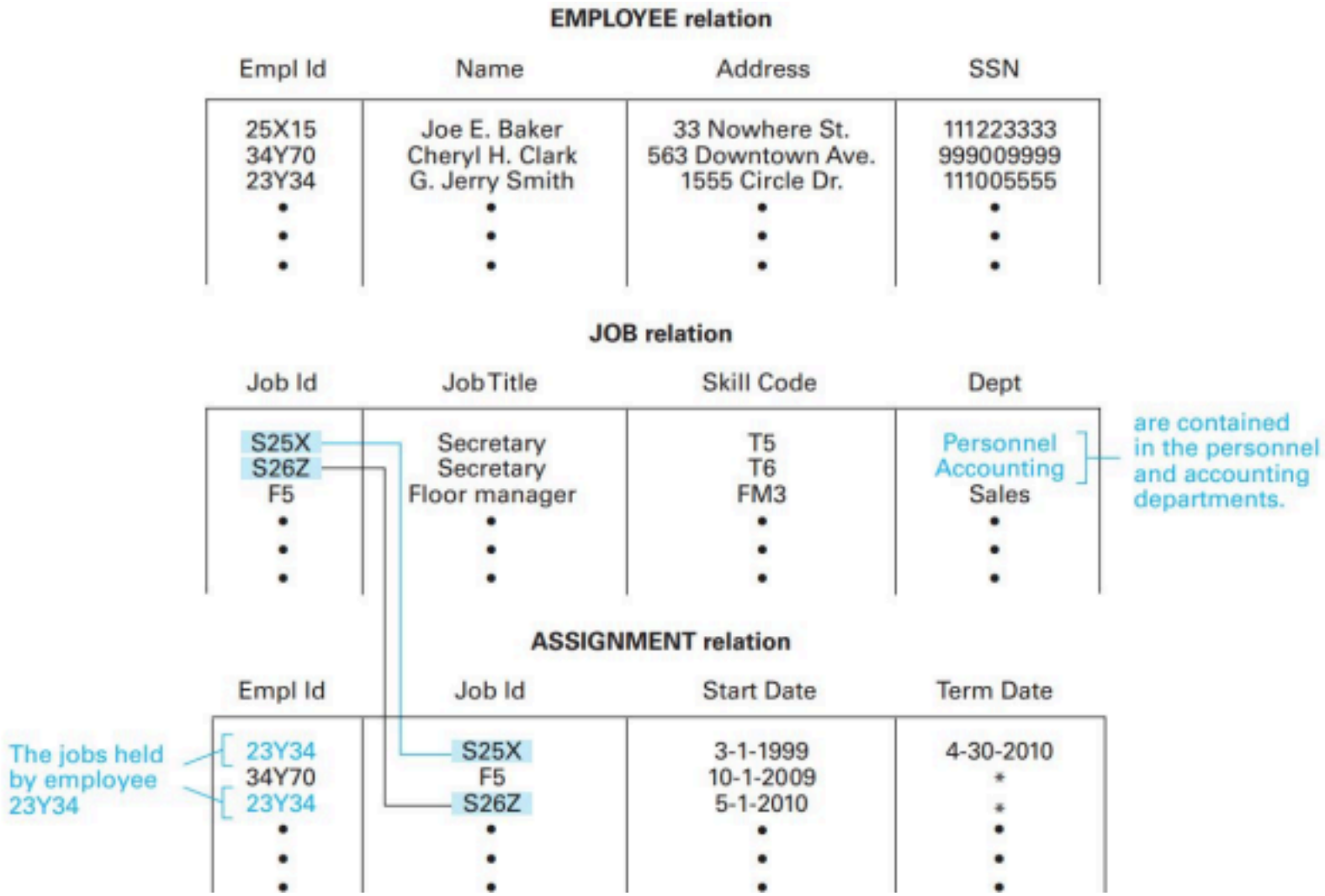
Job Id	JobTitle	Skill Code	Dept
S25X	Secretary	T5	Personnel
S26Z	Secretary	T6	Accounting
F5	Floor manager	FM3	Sales
•	•	•	•
•	•	•	•
•	•	•	•

ASSIGNMENT relation

Empl Id	Job Id	Start Date	Term Date
23Y34	S25X	3-1-1999	4-30-2010
34Y70	F5	10-1-2009	*
23Y34	S26Z	5-1-2010	*
•	•	•	•
•	•	•	•
•	•	•	•

The Relational Model

Finding the departments in which employee **23Y34** has worked.



The Relational Model

Dividing information into various relations is not always as trouble-free.

The Relational Model

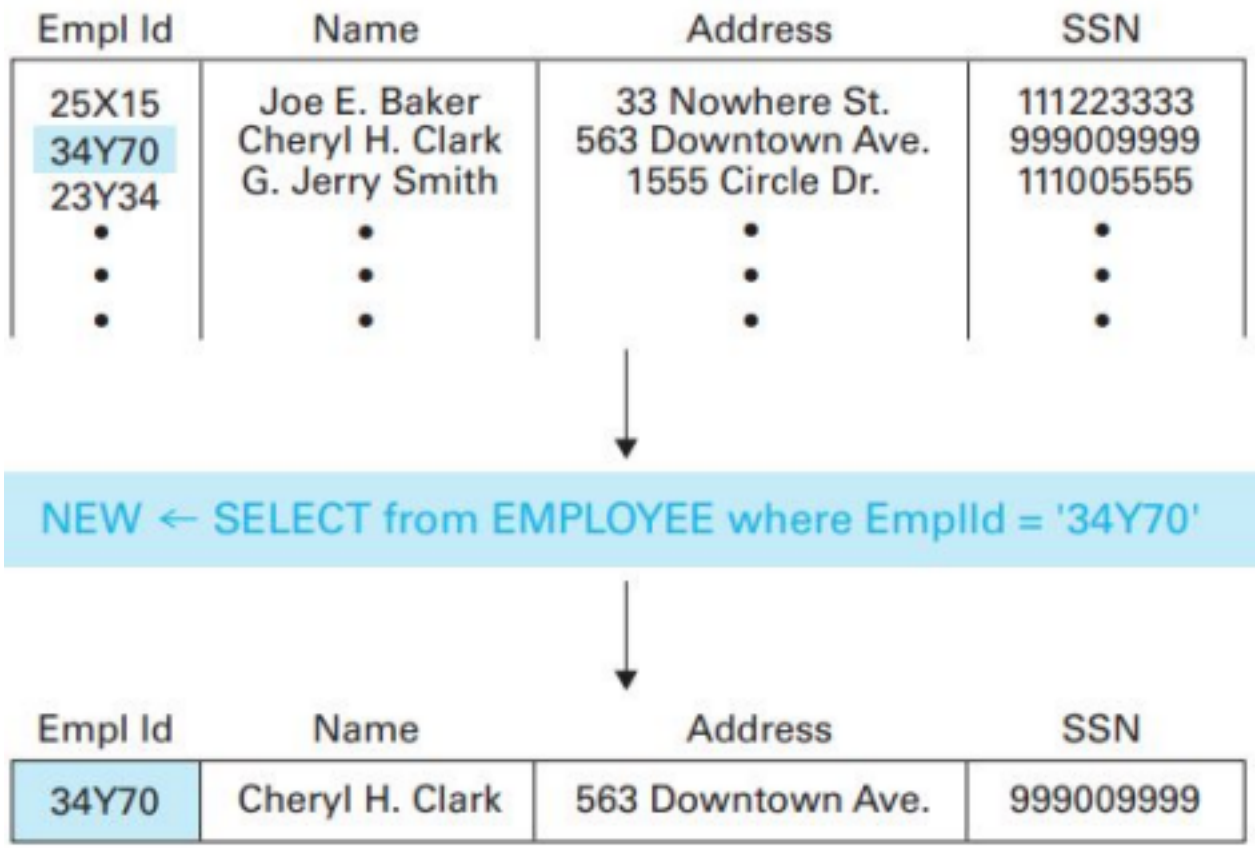
Sometimes dividing a relation into smaller relations causes the **loss of information**, and sometimes it **does not**.

- Such relational characteristics are important design considerations.

- The goal is
  - To identify the relational characteristics that can lead to problems in database design and
  - To find ways to reorganize those relations to remove these problematic characteristics.

Relational Operations

SELECT: One operation we might want to perform on a relation is [redacted] possessing certain characteristics and to place these selected tuples in a new



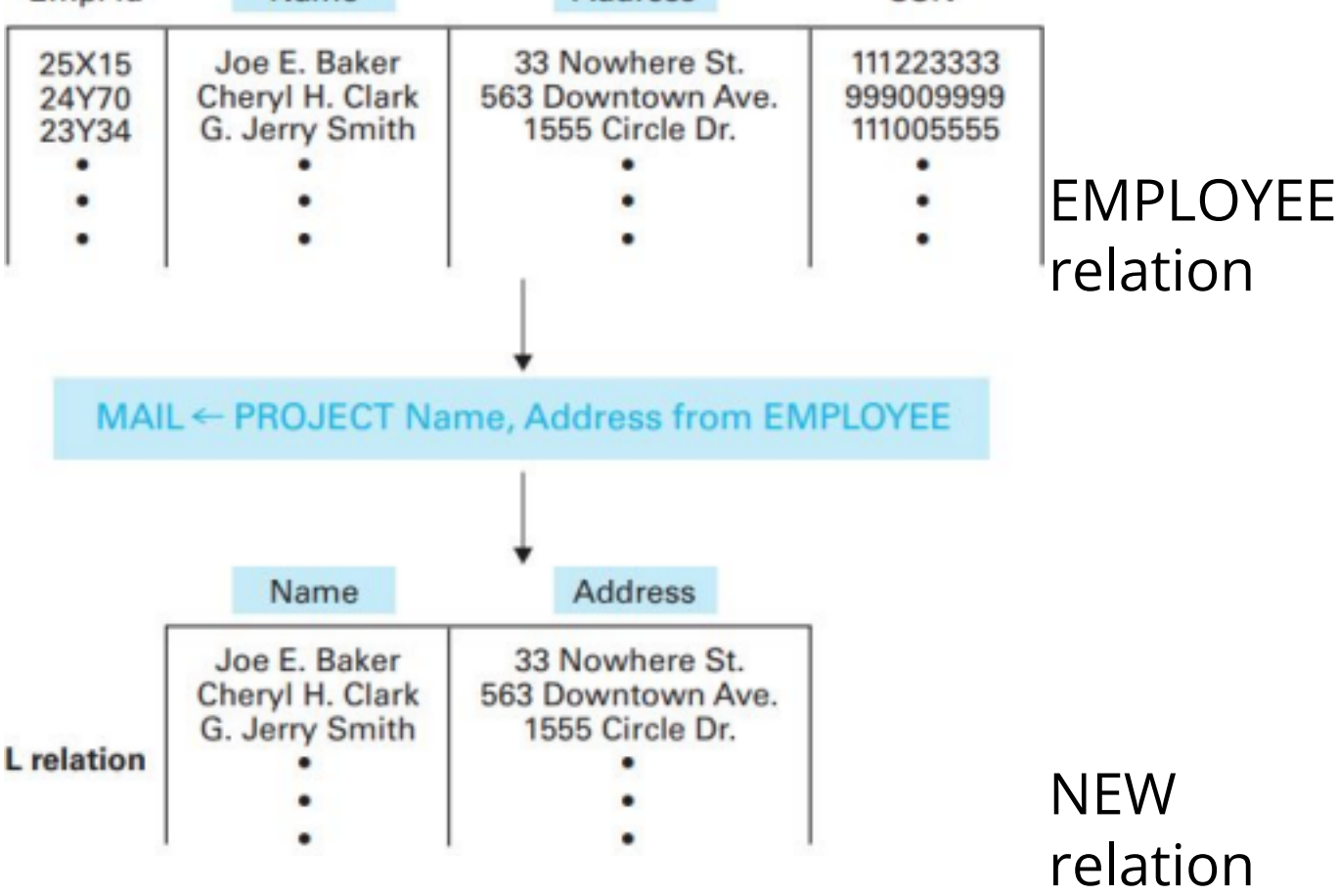
relation.

EMPLOYEE
relation

NEW
relation

Relational Operations

PROJECT: In contrast to the SELECT operation, which extracts rows from a relation, the PROJECT operation extracts columns

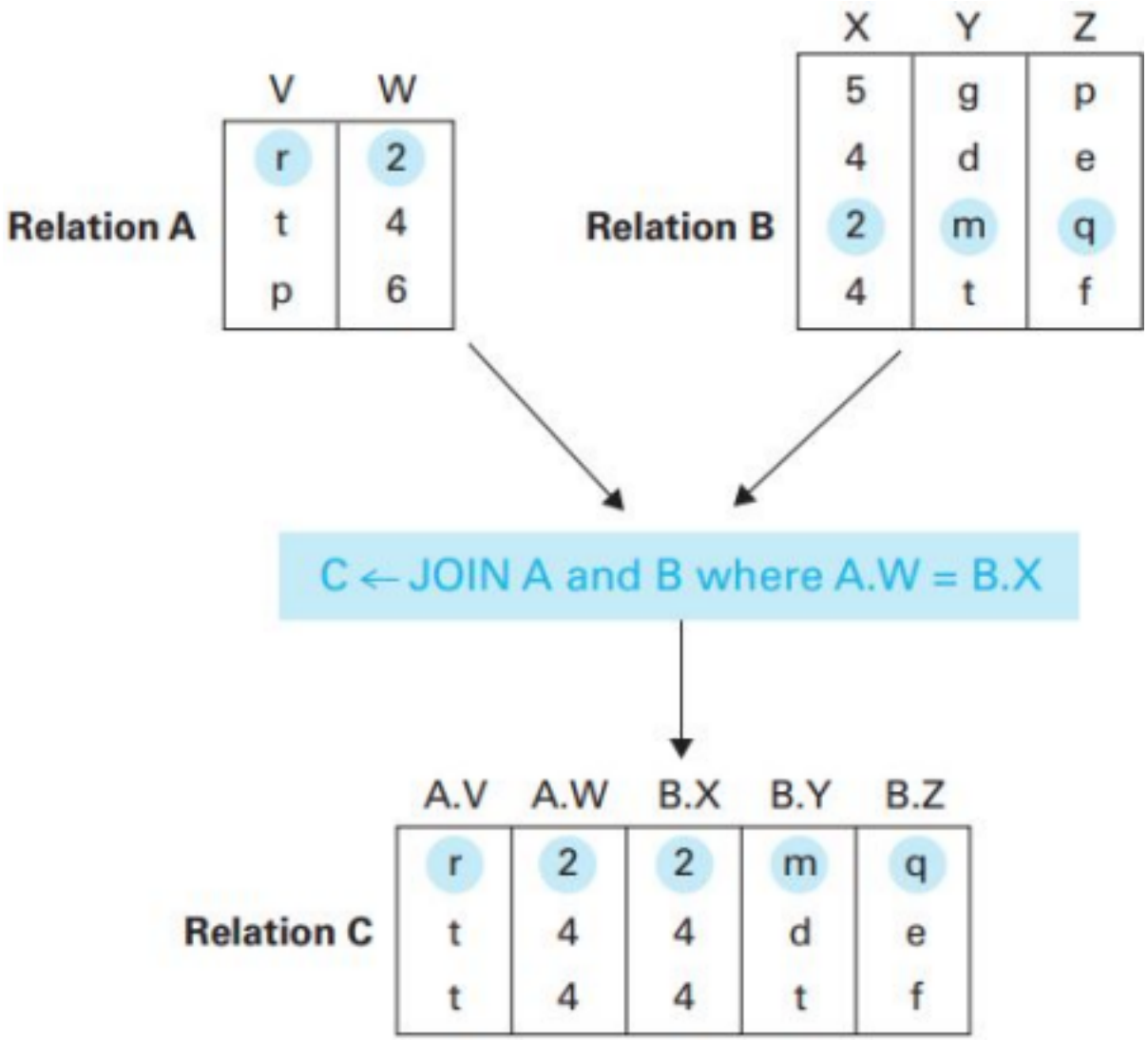


EMPLOYEE relation

NEW relation

Relational Operations

JOIN: It is used to combine different relations into one



relation.

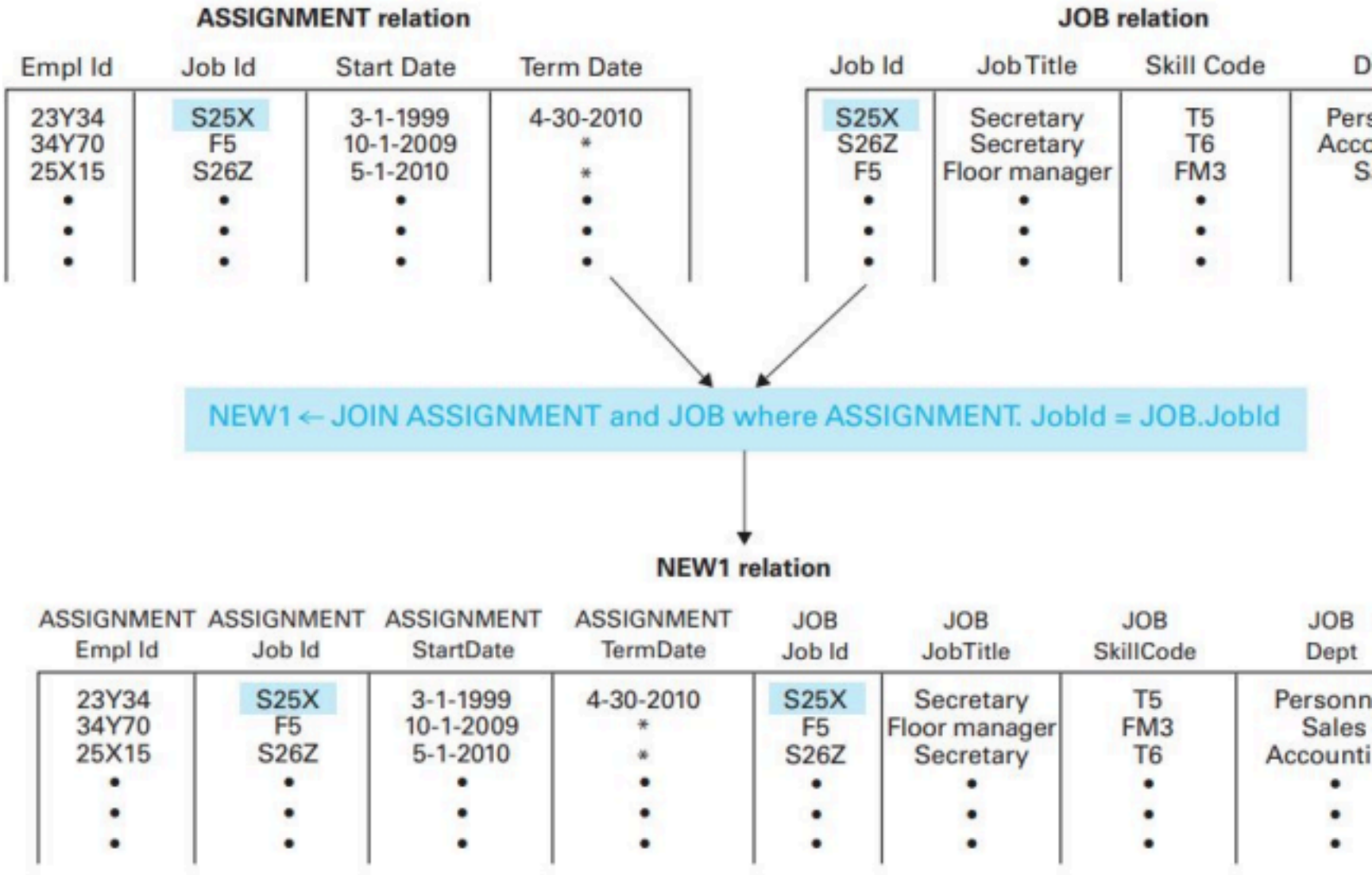
Relational Operations

JOIN: The JOIN of two relations produces a new relation

whose attributes consist of the attributes from the original relations.

Relational Operations

JOIN



SQL: Structured Query Language

```
SELECT Name, Address
FROM Employee;
```

produces a listing of all employee names and addresses contained in the relation Employee.

Note that this is **merely** a PROJECT operation.

SQL

```
SELECT EmplId, Name, Address, SSNum
FROM Employee
WHERE Name = 'Cheryl H. Clark';
```

produces all the information from the tuple associated

with Cheryl H. Clark in the Employee relation.

This is essentially a SELECT operation.

SQL

```
SELECT Name, Address
FROM Employee
WHERE Name = 'Cheryl H. Clark';
```

produces the name and address of Cheryl H. Clark as contained in the Employee relation.

This is a combination of SELECT and PROJECT operations.

SQL

```
SELECT Employee.Name, Assignment.StartDate
FROM Employee, Assignment
WHERE Employee.EmplId = Assignment.EmplId;
```

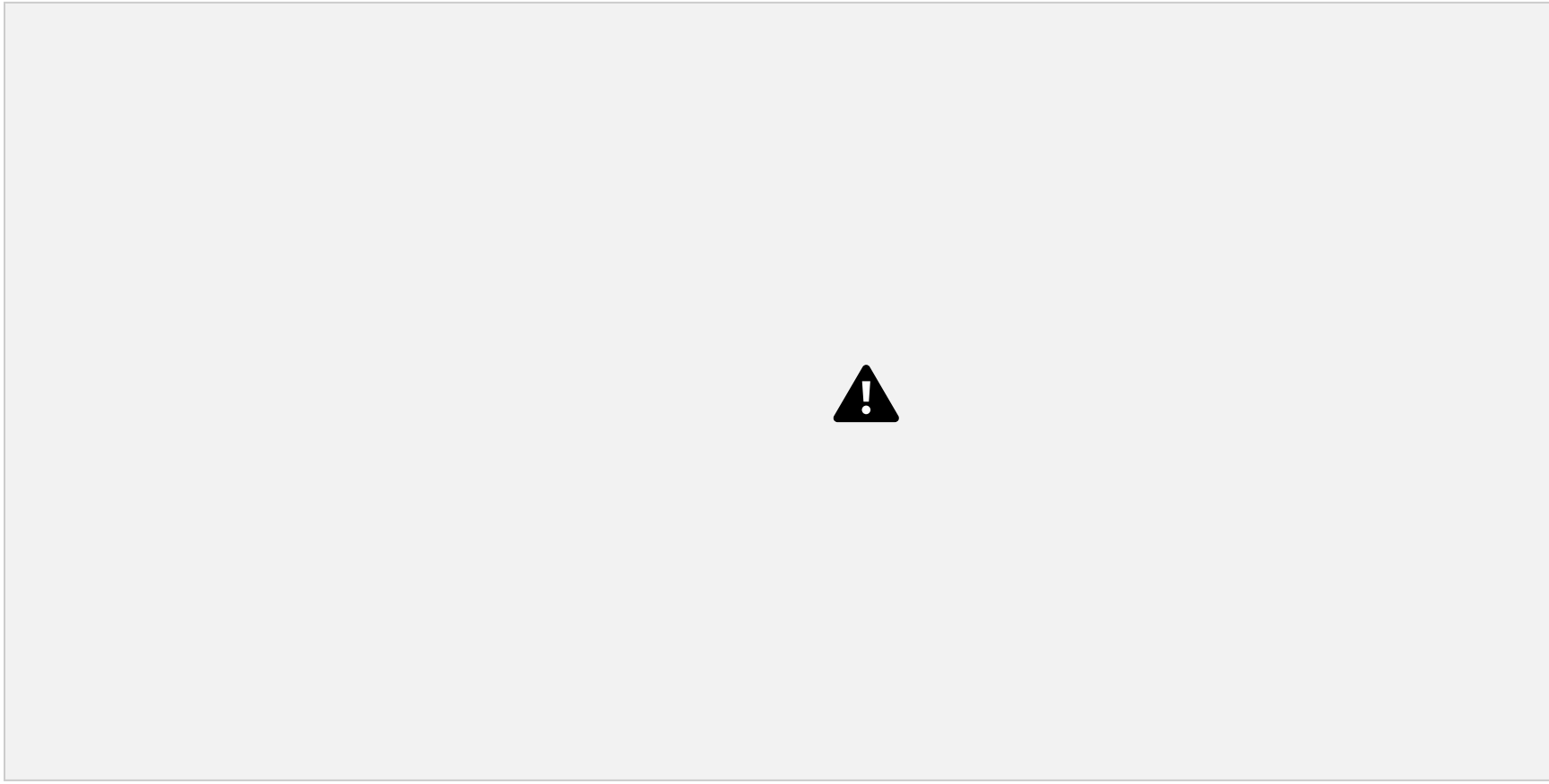
produces a listing of all employee names and their dates of initial employment.

Note that this is the result of JOINing the relations Employee and Assignment and then SELECTing and PROJECTing the appropriate tuples and attributes as identified in the where and select clauses

SQL

SQL encompasses statements for **defining** the structure of relations, **creating** relations, and

modifying the contents of relations as well as performing queries.



Traditional File Structures

Sequential Files

- Audio files, video files, files containing programs, and files containing textual documents

Indexed Files

Hash Files

Sequential File

A file that is accessed in a **serial manner** from its beginning to its end as though the information in the file were arranged in **one long row**.

In fact, **most of the files** created by a **typical personal computer** user are sequential files.

For instance, when a spreadsheet is saved, its

information is encoded and stored as a sequential file from which the spreadsheet application software can reconstruct the spreadsheet.

Sequential File

Indexed Files

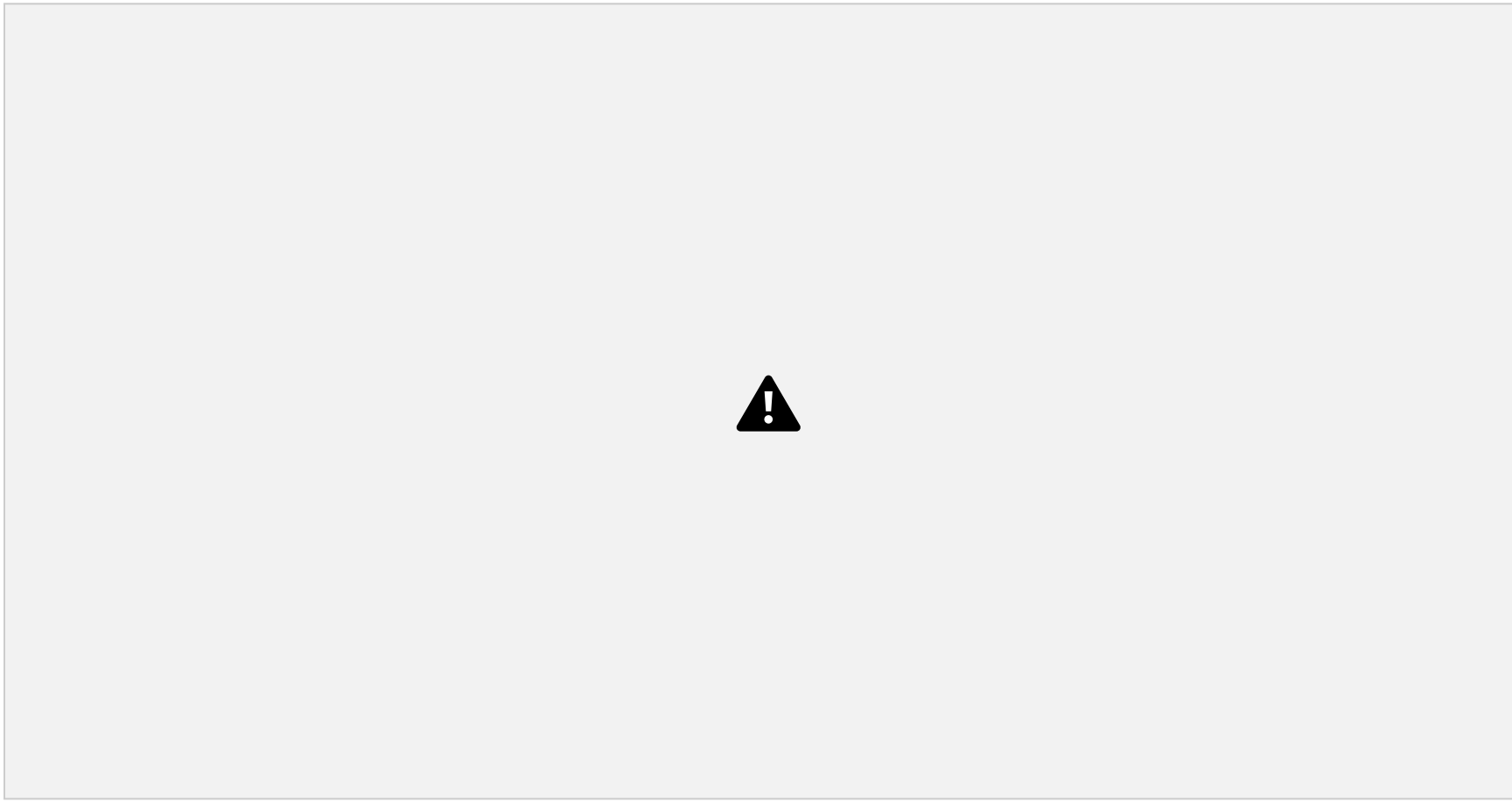
Sequential files are ideal for storing data that will be processed **in the order** in which the file's entries are stored.

However, such files are inefficient when records within the file must be retrieved in an unpredictable order.

In such situations what is needed is a way to identify the location of the desired logical record quickly.

Indexed Files

A popular solution is to use an index for the file. Such a file system is called an **indexed file**.



Hash Files

Issue in indexing: The expense of index

maintenance. A hash system can be summarized as follows:

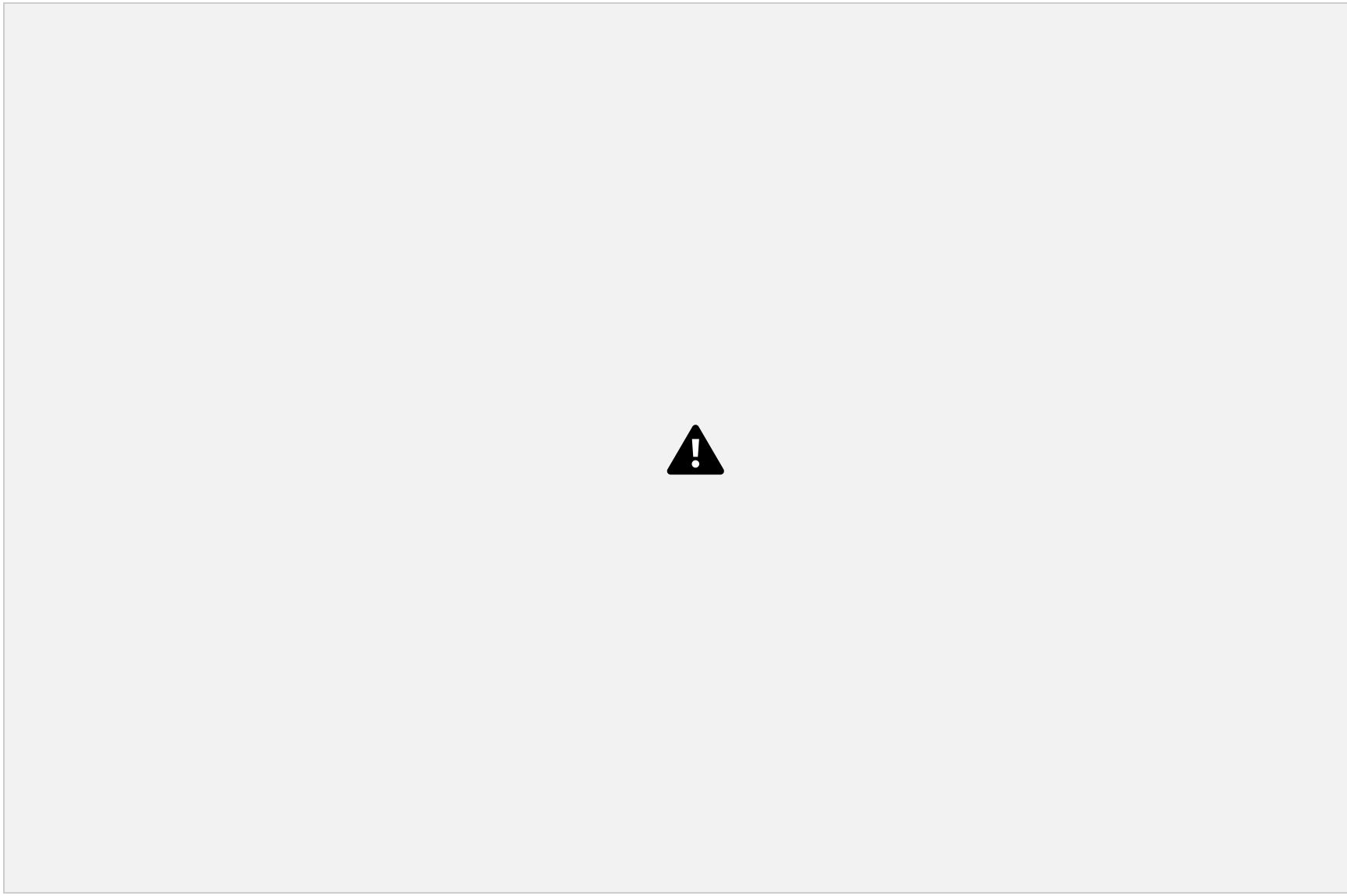
- The data storage space is divided into several sections, called **buckets**, each of which is capable of holding **several records**.
- The records are dispersed among the buckets according to an algorithm (**hash function**) that converts key values into bucket numbers

Hash Files

Hashing the key field value 25X3Z to one of 41 buckets.

Hash Files

The rudiments of a hashing system



Research Themes

1. Object Oriented Databases
2. Database Management Systems, DBMSs 3. Database Security, Threats and Defense Mechanisms
4. SQL and NoSQL
5. Data Analytics

COM1013
INTRODUCTION TO COMPUTER SCIENCE

Lecturer: Begüm MUTLU BİLGE, PhD

begummutlubilge+com1013@gmail.com (recommended)
bmbilge@ankara.edu.tr