1. Database Concepts

8   Databases

8.1   Database Concepts

Candidates should be able to:	Notes and guidance
Show understanding of the limitations of using a file-based approach for the storage and retrieval of data	
Describe the features of a relational database that address the limitations of a file-based approach	
Show understanding of and use the terminology associated with a relational database model	Including entity, table, record, field, tuple, attribute, primary key, candidate key, secondary key, foreign key, relationship (one-to-many, one-to-one, many-to-many), referential integrity, indexing
Use an entity-relationship (E-R) diagram to document a database design	
Show understanding of the normalisation process	First Normal Form(1NF), Second Normal Form (2NF) and Third Normal Form (3NF)
Explain why a given set of database tables are, or are not, in 3NF	
Produce a normalised database design for a description of a database, a given set of data, or a given set of tables	

- Limitations of a file-based approach
    - Data redundancy - the same data is stored many times
    - Program-data dependence - the program that access the data has to be re-written if any changes are made to the structure of the files
    - Data dependency - changes to data means changes to programs accessing the data
    - Data inconsistency - when data is updated in one place, it is not updated everywhere
    - Hard to perform complex queries - a new program has to be written each time
    - Lack of data privacy - user views cannot easily be implemented
- Features of a relational database
    - Multiple tables are linked together
        - reduces redundancy
        - increases integrity
        - data only needs to be updated once
        - associated data will be automatically updated
    - Program-data independence

- the structure of the data can change and does not affect program
- the structure of programs can change and does not affect data
- the data can be accessed by any appropriate program
- Allows concurrent access to data
  - by the use of record locking
  - by restricting over-writing changes
- Different user can be given different access rights
  - which improves security
- Different users can be given different views of the data
  - so data privacy is maintained
- More complex searches and queries can be executed

- Entity
  Anything that can have data stored about it (e.g. a person, a place)
- Table
  A group of similar data, in a database, with rows for each instance of an entity and columns for each attribute
- Record
  A row in a table in a database
- Field
  A column in a table in a database
- Tuple
  One instance of an entity, which is represented by a row in a table
- Attribute
  A column in a table = field
- Primary key
  Unique identifier (an attribute) for a table, a special case of candidate keys
- Candidate key
  An attribute of smallest set of attributes in a table where no tuple has the same value
- Secondary key
  A candidate key that is an alternative to the primary key
- Foreign key
  A set of attributes in one table that refers to the primary key in another table
- Relationship
  Situation in which one table in a database has a foreign key that refers to a primary key in another table in the database

- Referential integrity

  Property of a database that does not contain any values of a foreign key that are not matched to the corresponding primary key
- Indexing
  - Data structure built from one or more columns in a database table
  - Helps to efficiently retrieve records from database files
- Normalization
  - First Normal Form (1NF)
    - No repeated groups of attributes
    - All attributes should be atomic (single value, non-divisible)
    - No duplicate rows
  - Second Normal Form (2NF)
    - No partial dependencies (some attribute only depends on part of a composite key)
  - Third Normal Form (3NF)
    - No non-key dependencies
    - No transitive dependencies