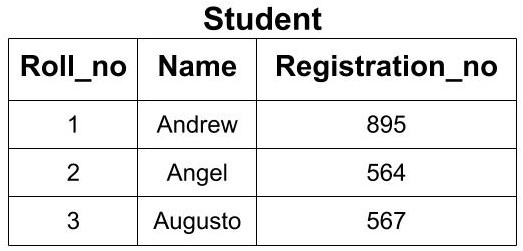
|  |  |  |  |
| --- | --- | --- | --- |
|  | Super Key | Candidate Key | Primary Key |
| Definition | A set of attributes that uniquely identifies each tuple in a table. | A minimal super key; a set of attributes that uniquely identifies each tuple and no proper subset of it has this property. | A chosen candidate key; the main method for accessing and identifying records in a table. |
| Uniqueness | Ensures uniqueness but may include redundant attributes. | Ensures uniqueness without any redundancy. | Uniquely identifies each tuple in a table without redundancy. |
| Number | Can be multiple in a table. | Can be multiple but usually, only one is chosen as the primary key. | Only one per table. |
| Minimality | Not necessarily minimal; may have more attributes than required. | Minimal; no proper subset of it can uniquely identify tuples. | A specific candidate key chosen as the primary means of identification. |
| Purpose | Identifies each tuple but may not be the primary means of identification. | Potential primary keys; uniquely identify tuples without redundancy. | Main method for accessing and identifying records in a table. |
| Constraints | No constraints. | Uniqueness constraint. | Uniqueness and not-null constraints. |
| Selection | Not selected; a conceptual construct. | Identified among super keys. | Chosen from candidate keys. |
| Modifiability | Can be modified freely. | Can be modified but should retain uniqueness. | Should not be modified once established to maintain integrity. |
| Database Constraints | Not subject to database constraints. | Subject to uniqueness constraint. | Subject to uniqueness and not-null constraints. |
| Database Performance | No direct impact on performance. | Improves performance due to uniqueness. | Optimizes performance by providing a unique identifier for records. |
| Clarity | May not be explicitly defined. | Explicitly defined and identified. | Clearly specified as the primary means of identification. |
| Flexibility | Flexible; can change with requirements. | Provides a level of stability; chosen from available options. | Provides stability; essential for maintaining data integrity. |
| Enforcement | No enforced constraint. | Enforced uniqueness constraint. | Enforced uniqueness and not-null constraints. |
| Example | {Employee ID, Email} | {Employee ID}, {Email} | {Employee ID} (chosen from candidate keys) |

**Super Key**

A super key or ***simply key***is a combination of all possible attribute which can uniquely identify the rows(tuples) in a table.

This means that a super key may have some extra attribute which isn't necessary for uniquely identifying the rows in the table.

***Example:***In the given Student Table we can have the following keys as the super key.



1. {Roll\_no}
2. {Registration\_no}
3. {Roll\_no, Registration\_no},
4. {Roll\_no, Name}
5. {Name, Registration\_no}
6. {Roll\_no, Name, Registration\_no}

All the above keys are able to uniquely identify each row. So, each of these keys is super key. Here you can see that by using Roll\_no only, we can uniquely identify the rows but if you are making a super key, then you will try to find all the possible cases of keys that can be used to identify data uniquely.

**Candidate Key**

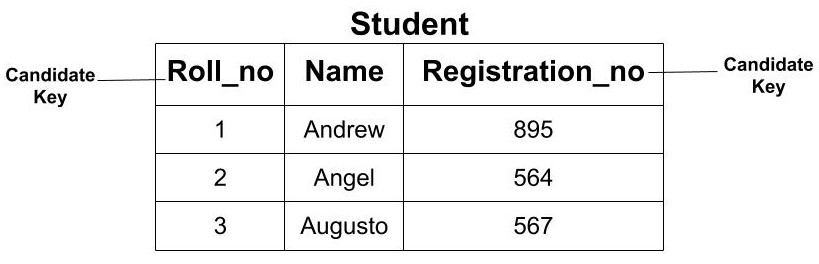
* A candidate key is a ***minimal super key***or a super key with no redundant attribute.
* It is called a minimal super key because we select a candidate key from a set of super key such that selected candidate key is the minimum attribute required to uniquely identify the table.
* It is selected from the set of the super key which means that all candidate keys are super key.
* Candidate Keys are not allowed to have NULL values.

**If the subset of the candidate key is a super key, then that candidate key is not a valid candidate key.**

***Example:***In the above example, we had 6 super keys but all of them cannot become a candidate key. Only those super keys would become a candidate key which have no redundant attributes.

1. **{Roll\_no}:**This key doesn't have any redundant or repeating attribute. So, it can be considered as a candidate key.
2. **{Registration\_no}:**This key also doesn't have any repeating attribute. So, it can be considered as a candidate key.
3. **{Roll\_no, Registration\_no}:**This key cannot be considered as a candidate key because when we take the subset of this key we get two attributes i.e Roll\_no or Registration\_no. Each of these attributes is the candidate key. So, it is not a minimal super key. Hence, this key is not a candidate key.
4. **{Roll\_no, Name}:**This key cannot be considered as a candidate key because when we take the subset of this key we get two attributes i.e. Roll\_no or Name. Roll\_no is a candidate key. So, it is not a minimal super key. Hence, this key is not a candidate key.
5. **{Name, Registration\_no}:**This key cannot be considered as a candidate key because when we take the subset of this key we get two attributes i.e Registration\_no or Name. Registration\_no is a candidate key. So, it is not a minimal super key. Hence, this key is not a candidate key.
6. **{Roll\_no, Name, Registration\_no}:**This key cannot be considered as a candidate key because when we take the subset of this key we get three attributes i.e Roll\_no, Registration\_no and Name. Two of these attributes i.e Roll\_no and Registration\_no are the candidate key. So, it is not a minimal superkey. Hence, this key is not a candidate key.

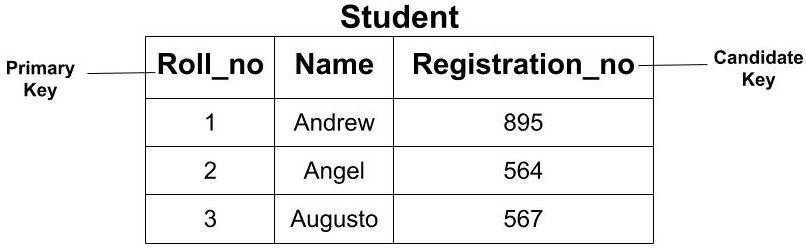
So, from the above discussion, we conclude that we can have only 2 out of above 6 super keys as the candidate key. i.e. (Roll\_no) and(Registration\_no).



**Primary Key**

* The primary key is the minimal set of attributes which uniquely identifies any row of a table. It is selected from a set of candidate keys.
* ***Any candidate key can become a primary key*.**It depends upon the requirements and is done by the Database Administrator (DBA).
* The primary key cannot have a NULL value.
* It cannot have a duplicate value.

***Example:***In the above example, we saw that we have two candidate keys i.e (Roll\_no) and (Registration\_no). From this set, we can select any key as the primary key for our table. It depends upon our requirement. Here, if we are talking about class then selecting ‘Roll\_no’ as the primary key is more logical instead of ‘Registrartion\_no’.



**Define Database. List and explain the applications of Database.**

DB is a organized collection of structured data which is typically stored electronically on a computer device. DB is designed to efficiently manage, retrieve and manipulate data.

A database is a structured collection of data that is organized, stored, and managed in a way that allows for efficient retrieval, manipulation, and querying of data. It typically consists of tables, each containing rows (records) and columns (fields) that represent specific entities or relationships between entities.

Databases are managed by database management systems (DBMS), which is software that allows creation, definition and manipulation of database, allowing users to store, process and analyze data easily. DBMS provides us with an interface or a tool, to perform various operations like creating database, storing data in it, updating data, creating tables in the database and a lot more.

DBMS also provides protection and security to the databases. It also maintains data consistency in case of multiple users.

**Here are some examples of popular DBMS used:**

* MySql
* Oracle
* SQL Server
* IBM DB2
* PostgreSQL
* Amazon SimpleDB (cloud based) etc.

**Applications of Database:**

1. Enterprise Information SHAMO

* Sales: For customer, product, and purchase information.
* Accounting: For payments, receipts, account balances, assets and other accounting information.
* Human resources: For information about employees, salaries, payroll taxes and benefits, and for generation of paychecks.
* Manufacturing: For management of the supply chain and for tracking production of items in factories, inventories of items in warehouses and stores, and orders for items.
* Online retailers: For sales data noted above plus online order tracking, generation of recommendation lists, and maintenance of online product evaluations.

1. Banking and Finance BFC

* Banking: For customer information, accounts, loans, and banking transactions.
* Credit card transactions: For purchases on credit cards and generation of monthly statements.
* Finance: For storing information about holdings, sales, and purchases of financial instruments such as stocks and bonds; also for storing real-time market data to enable online trading by customers and automated trading by the firm.

1. Universities:

* For student information, course registrations, and grades (in addition to standard enterprise information such as human resources and accounting).
* Educational management systems (EMS) or learning management systems (LMS) provide platforms for online learning, course delivery, student assessments, and collaboration among educators and students.
* Improve administrative efficiency, support personalized learning experiences, enable data-driven decision-making for educational planning and policy-making, and enhance communication and collaboration among stakeholders.

1. Airlines:

* For reservations and schedule information. Airlines were among the first to use databases in a geographically distributed manner.
* Airlines rely on databases to manage flight schedules, passenger reservations, ticketing, crew assignments, aircraft maintenance, and operational logistics.
* Airline reservation systems and flight operations databases support real-time updates, seat availability, flight planning, and customer service interactions.
* Ensure efficient flight operations, optimize resource allocation, minimize scheduling conflicts and disruptions, enhance customer service quality, and facilitate revenue management strategies.

1. Telecommunication:

* Telecommunication companies utilize databases to manage subscriber data, call detail records (CDR), network configurations, billing information, and service provisioning.
* Databases support subscriber management, network monitoring, fault detection, performance analysis, and service delivery across various communication technologies.
* Improve network reliability and performance, enhance customer experience, enable targeted marketing and service offerings, support billing and revenue assurance processes, and facilitate compliance with regulatory requirements.

1. Geographic Information Systems (GIS):

* GIS databases store and manage spatial data, including maps, geographic features, satellite imagery, and attribute data related to locations.
* GIS systems analyze spatial relationships, support urban planning, environmental management, emergency response, natural resource management, and geographic research.
* Enable spatial analysis and visualization, support informed decision-making for land use planning and development, facilitate disaster management and emergency response efforts, and enhance environmental conservation and resource allocation.

1. Social Media and Networking Platforms:

* Social media platforms utilize databases to store user profiles, connections, posts, comments, multimedia content, and interaction data.
* Social media databases support user engagement, content recommendation, targeted advertising, trend analysis, sentiment analysis, and social network analysis.
* Enhance user engagement and retention, personalize content and advertisements, facilitate community building and networking, generate insights into user behavior and preferences, and support data-driven marketing strategies.

1. Healthcare Information Systems:

* Healthcare organizations use databases to manage patient records, medical histories, diagnostic tests, prescriptions, treatment plans, and healthcare facility operations.
* Electronic health record (EHR) systems, hospital information systems (HIS), and health information exchanges (HIE) support clinical decision-making, patient care coordination, and regulatory compliance.
* Improve patient safety and quality of care, support interoperability and information exchange among healthcare providers, enhance care coordination and continuity, enable population health management and disease surveillance, and facilitate research and public health initiatives.

**Advantages of DB**

* Data Centralization: Centralizes data storage for consistency and integrity.
* Data Security: Provides features for secure data access and protection.
* Data Sharing: Facilitates collaboration and real-time data updates.
* Scalability: Scales to handle growing data volumes and user loads.
* Query and Analysis: Supports efficient data retrieval and analysis.
* Recovery and Backup: Offers mechanisms for data recovery and backup.

**Disadvantages of DB**

* Cost: Requires significant investment in implementation and maintenance.
* Complexity: Demands specialized technical expertise for management.
* Security Risks: Exposes data to threats like breaches and cyberattacks.
* Vendor Lock-In: Ties organizations to specific vendors or technologies.
* Performance Challenges: May encounter bottlenecks and scalability issues.
* Data Redundancy: Can lead to redundancy and inconsistency if not managed properly.