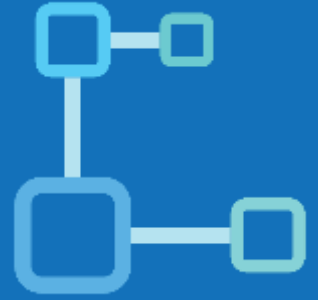# DATABASE

netlink
digital solutions

www.netlink.com

Normalization

Normalization is the process of creating small, stable, yet flexible and adaptive data structures from complex groups of data and minimizes repeated data groups.

Before Normalization

Order_ Number	Order_ Date	Part_ Number	Part_ Name	Unit_ Price	Part_ Quantity	Supplier_ Number	Supplier_ Name	Supplier_ Street	Supplier_ City	Supplier_ State	Supplier_ Zip

Normalization

Normalization of Database

- NORMALIZATION is a database design technique that organizes tables in a manner that reduces redundancy (useless) and dependency of data.

- To avoid insertion, update & deletion anomaly.

- Normalization divides larger tables into smaller tables and links them using relationships. (Primary Key and Foreign Key)

Normal forms #

There are six normal forms, but we will only look at the first four, which are:

1. First normal form (1NF)
2. Second normal form (2NF)
3. Third normal form (3NF)
4. Boyce-Codd normal form (BCNF)

First Normal Form (1NF)

First Normal Form (1NF)

➤ A relation will be 1NF if it contains an atomic value.

➤ It states that an attribute of a table cannot hold multiple values. It must hold only single-valued attribute.

➤ First normal form disallows the multi-valued attribute, composite attribute, and their combinations.

Example: Relation EMPLOYEE is not in 1NF because of multi-valued attribute EMP_PHONE.

EMPLOYEE table:

EMP_ID	EMP_NAME	EMP_PHONE	EMP_STATE
14	John	7272826385, 9064738238	UP
20	Harry	8574783832	Bihar
12	Sam	7390372389, 8589830302	Punjab

The decomposition of the EMPLOYEE table into 1NF has been shown below:

EMP_ID	EMP_NAME	EMP_PHONE	EMP_STATE
14	John	7272826385	UP
14	John	9064738238	UP
20	Harry	8574783832	Bihar
12	Sam	7390372389	Punjab
12	Sam	8589830302	Punjab

Second Normal Form (2NF)

- In the 2NF, relational must be in 1NF.
- In the second normal form, all non-key attributes are fully functional dependent on the primary key

Example: Let's assume, a school can store the data of teachers and the subjects they teach. In a school, a teacher can teach more than one subject.

TEACHER table

TEACHER_ID	SUBJECT	TEACHER_AGE
25	Chemistry	30
25	Biology	30
47	English	35
83	Math	38
83	Computer	38

In the given table, non-prime attribute TEACHER_AGE is dependent on TEACHER_ID which is a proper subset of a candidate key. That's why it violates the rule for 2NF.

TEACHER_DETAIL table:

TEACHER_ID	TEACHER_AGE
25	30
47	35
83	38

TEACHER_SUBJECT table:

TEACHER_ID	SUBJECT
25	Chemistry
25	Biology
47	English
83	Math
83	Computer

Third Normal Form (3NF)

➢ A relation will be in 3NF if it is in 2NF and not contain any transitive partial dependency.

➢ 3NF is used to reduce the data duplication. It is also used to achieve the data integrity.

➢ If there is no **transitive dependency** for non-prime attributes, then the relation must be in third normal form.

Candidate key: {EMP_ID}

Non-prime attributes: In the given table, all attributes except EMP_ID are non-prime.

EMPLOYEE_DETAIL table:

EMP_ID	EMP_NAME	EMP_ZIP	EMP_STATE	EMP_CITY
222	Harry	201010	UP	Noida
333	Stephan	02228	US	Boston
444	Lan	60007	US	Chicago
555	Katharine	06389	UK	Norwich
666	John	462007	MP	Bhopal

Third Normal Form (3NF)

EMPLOYEE table:

EMP_ID	EMP_NAME	EMP_ZIP
222	Harry	201010
333	Stephan	02228
444	Lan	60007
555	Katharine	06389
666	John	462007

EMPLOYEE_ZIP table:

EMP_ZIP	EMP_STATE	EMP_CITY
201010	UP	Noida
02228	US	Boston
60007	US	Chicago
06389	UK	Norwich
462007	MP	Bhopal

Boyce Codd normal form (BCNF)

Boyce Codd normal form (BCNF)

➢ BCNF is the advance version of 3NF. It is stricter than 3NF.

➢ A table is in BCNF if every functional dependency X → Y, X is the super key of the table.

➢ For BCNF, the table should be in 3NF, and for every FD, LHS is super key.

In the above table Functional dependencies are as follows:
- EMP_ID → EMP_COUNTRY
- EMP_DEPT → {DEPT_TYPE, EMP_DEPT_NO}

Candidate key: {EMP-ID, EMP-DEPT}

The table is not in BCNF because neither EMP_DEPT nor EMP_ID alone are keys.
To convert the given table into BCNF, we decompose it into three tables:

EMPLOYEE table:

EMP_ID	EMP_COUNTRY	EMP_DEPT	DEPT_TYPE	EMP_DEPT_NO
264	India	Designing	D394	283
264	India	Testing	D394	300
364	UK	Stores	D283	232
364	UK	Developing	D283	549

Boyce Codd normal form (BCNF)

Functional dependencies:

1.EMP_ID → EMP_COUNTRY

2.EMP_DEPT → {DEPT_TYPE, EMP_DEPT_NO}

Candidate keys:

For the first table: EMP_ID
For the second table: EMP_DEPT
For the third table: {EMP_ID, EMP_DEPT}

EMP_COUNTRY table:

EMP_ID	EMP_COUNTRY
264	India
264	India

EMP_DEPT table:

EMP_DEPT	DEPT_TYPE	EMP_DEPT_NO
Designing	D394	283
Testing	D394	300
Stores	D283	232
Developing	D283	549

EMP_DEPT_MAPPING table:

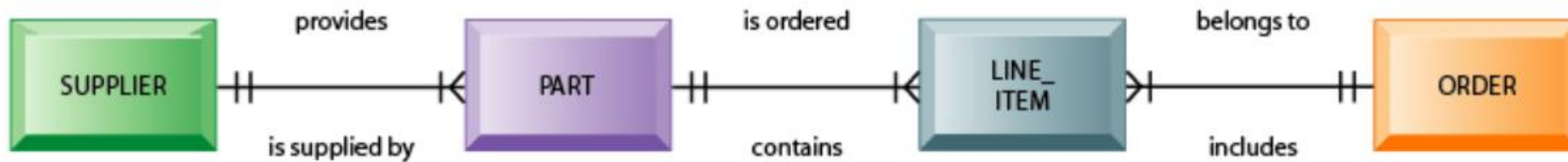| EMP_ID	EMP_DEPT
D394	283
D394	300
D283	232
D283	549

AN UNNORMALIZED RELATION FOR ORDER

An unnormalized relation contains repeating groups. For example, there can be many parts and suppliers for each order. There is only a one-to-one correspondence between Order_Number and Order_Date.
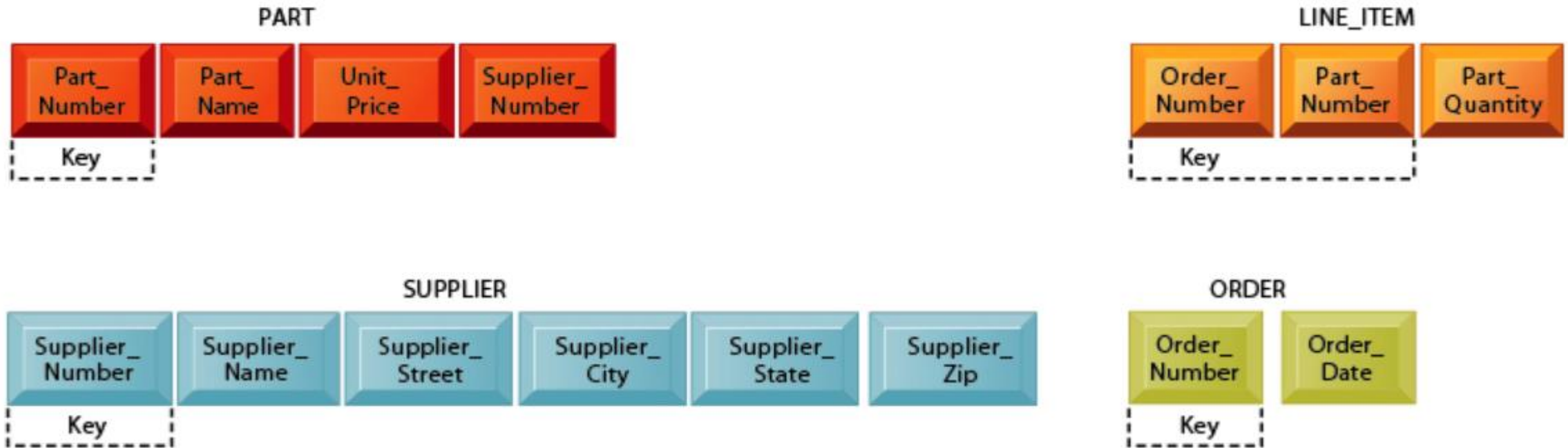
NORMALIZED TABLES CREATED FROM ORDER

After normalization, the original relation ORDER has been broken down into four smaller relations. The relation ORDER is left with only two attributes and the relation LINE_ITEM has a combined, or concatenated, key consisting of Order_Number and Part_Number

netlink

digital solutions

Thank you

www.netlink.com