

The Relational Database

Relational Databases are designed as Tables with defined relationships among them

- Table: A two-dimensional structure that stores data in rows and columns
 - Rows correspond to a specific item or idea; columns provide information about that item or idea

FLIGHT_ID	TAIL_NUMBER	FLIGHT_NUMBER	DEPARTURE_AIRPORT	ARRIVAL_AIRPORT	DEPARTURE_TIME	ARRIVAL_TIME	PASSENGER_COUNT
10011	N123AA	DAL135	KATL	KORD	6/20/14 12:00 PM	6/20/14 2:00 PM	165
10012	N234BB	AAL246	KDFW	KLAX	6/20/14 12:01 PM	6/20/14 4:31 PM	117
10013	N345CC	UAL357	KJFK	KSFO	6/20/14 12:02 PM	6/20/14 5:14 PM	109
10014	N456DD	FFT468	KDEN	KSTL	6/20/14 12:03 PM	6/20/14 2:12 PM	103
10015	N567EE	TRS579	KMIA	KORF	6/20/14 12:04 PM	6/20/14 2:41 PM	110
10016	N678FF	AWE680	KPIT	KPHX	6/20/14 12:05 PM	6/20/14 4:59 PM	93
10017	C-GASC	ACA791	KDET	CYYZ	6/20/14 12:06 PM	6/20/14 1:18 PM	114
...

- Rows = Records
- Columns = Fields = Attributes

The Relational Database

- Each table needs to have a Primary Key, which is a column or set of columns that define a unique row of the table
- A Natural Key is a key that can be constructed using data elements that occur 'naturally' in the table
- A Composite Key is a key defined by more than one field
- A Surrogate Key is a column added to a table that is not derived from naturally occurring data fields
 - Common example is just a number that increments with each new row
 - Usually does not have meaning outside the database

The Relational Database

 **Surrogate Primary Key**

FLIGHT_ID	TAIL_NUMBER	FLIGHT_NUMBER	DEPARTURE_AIRPORT	ARRIVAL_AIRPORT	DEPARTURE_TIME	ARRIVAL_TIME	PASSENGER_COUNT
10011	N123AA	DAL135	KATL	KORD	6/20/14 12:00 PM	6/20/14 2:00 PM	165
10012	N234BB	AAL246	KDFW	KLAX	6/20/14 12:01 PM	6/20/14 4:31 PM	117
10013	N345CC	UAL357	KJFK	KSFO	6/20/14 12:02 PM	6/20/14 5:14 PM	109
10014	N456DD	FFT468	KDEN	KSTL	6/20/14 12:03 PM	6/20/14 2:12 PM	103
10015	N567EE	TRS579	KMIA	KORF	6/20/14 12:04 PM	6/20/14 2:41 PM	110
10016	N678FF	AWE680	KPIT	KPHX	6/20/14 12:05 PM	6/20/14 4:59 PM	93
10017	C-GASC	ACA791	KDET	CYYZ	6/20/14 12:06 PM	6/20/14 1:18 PM	114
...

 **Composite Natural Primary Key**

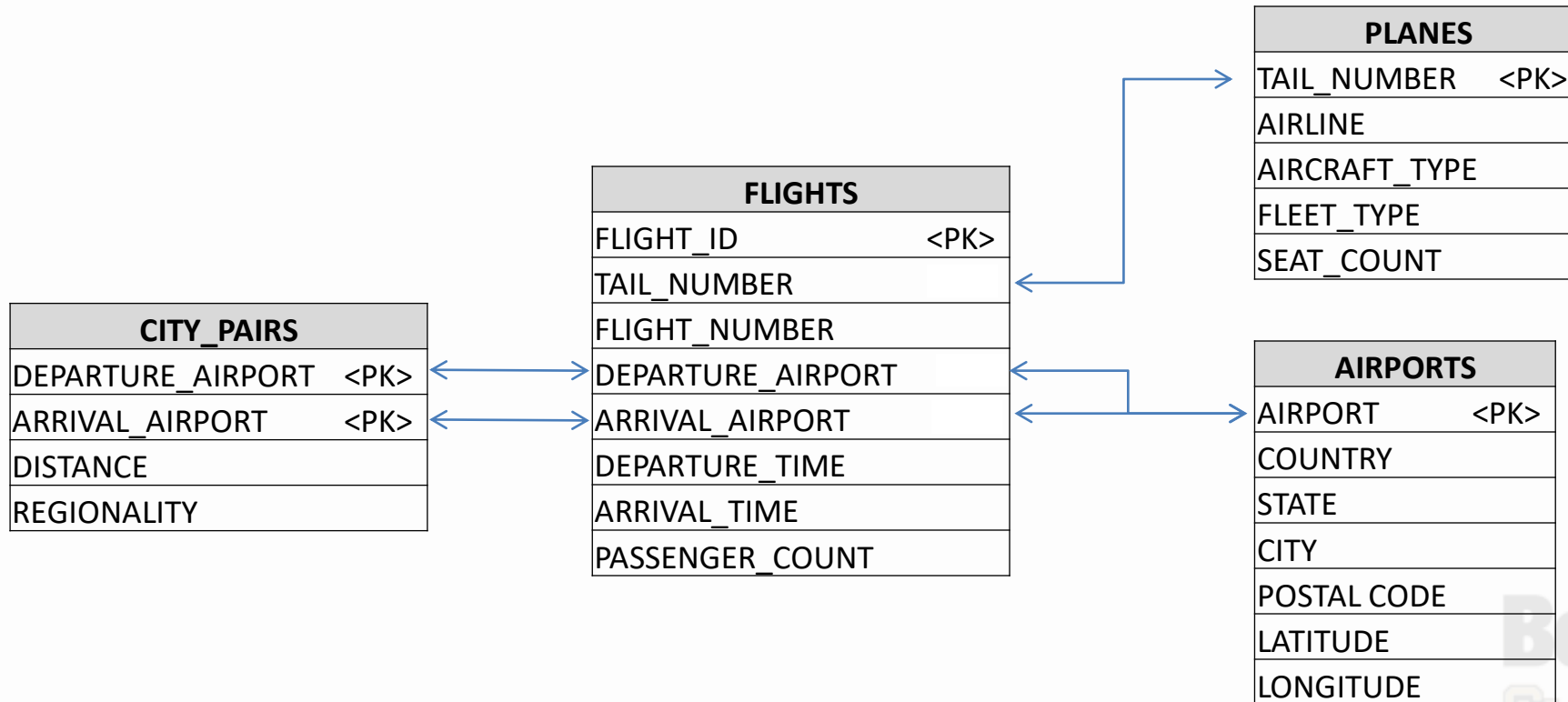
FLIGHTS
FLIGHT_ID
TAIL_NUMBER
FLIGHT_NUMBER
DEPARTURE_AIRPORT
ARRIVAL_AIRPORT
DEPARTURE_TIME
ARRIVAL_TIME
PASSENGER_COUNT

FLIGHTS
FLIGHT_ID <PK>
TAIL_NUMBER
FLIGHT_NUMBER
DEPARTURE_AIRPORT
ARRIVAL_AIRPORT
DEPARTURE_TIME
ARRIVAL_TIME
PASSENGER_COUNT

The Relational Database

How are relationships between tables defined?

- A Foreign Key is an attribute or attributes that establishes a logical link between tables
- It usually associates a field of one table with the primary key of another table or tables



The Relational Database

Normalization

- The process of making data less redundant by separating distinct ideas into separate tables and relating them to each other using keys
- A database is considered 'Normalized' when it reaches something called 'Third Normal Form (3NF)'
 - The definition of 3NF is fairly technical, but has been paraphrased as follows:
 - "[Every] non-key [attribute] must provide a fact about the key, the whole key, and nothing but the key."

Order_No	Date	Customer_ID	Product_Name	Product_Type
12345	3/4/2012	1034678	Gravity	DVD
12346	3/5/2012	1347659	The Hobbit	Book
12347	3/7/2012	1783258	Guitar World	Magazine
12348	3/9/2012	2348910	Wall Street Journal	Newspaper

- The above would NOT be 3NF since it contains the Product_Type attribute which is describing the Product, not the Order, which is the Primary Key

The Relational Database

So why are relational databases so popular?

- Allows data to be grouped logically around discrete ideas
- Minimizes the amount of duplicate data stored in a database
- Minimizes the number of places where changes to data need to be made
- Improves performance in highly transactional systems where lots of updates and additions are made
- Is incredibly flexible in terms of how data can be queried and extracted

What's the downside?

- The more Normalized the database becomes, the more complex extraction and analysis becomes, as more tables need to be joined together
- Requires more effort to integrate fundamentally new data domains, as they must be architected to fit within with the current database model

Denormalization, Data Marts & Cubes

- **The more often we find ourselves joining the same tables together for repeated operations, the more it makes sense to ‘denormalize’ information**
 - Tradeoff is less efficient storage for more efficient interrogation
 - Usually accomplished in ‘downstream’ structures or through database ‘Views’ that execute joins in the background
- **‘Data Marts’ are specialized data structures that are constructed for a specific function or business need**
 - Usually contains a subset of all available data
 - Can be more normalized or more denormalized, depending on need
- **A ‘Cube’ can be thought of as an N-dimensional table**

Recap

- Relational Databases
- Uniqueness and Primary Keys: Natural, Surrogate, Composite
- Foreign Keys and Table Linkages
- Normalization and 3NF
- Denormalization