Reasons for Database Normalization

Reasons for Database Normalization

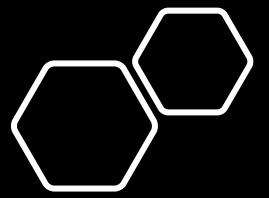
Three Main Reasons

- Minimize duplicate data
- Minimize or avoid data modification issues
- Simplify queries

SalesStaff						
EmployeeID	**SalesPerson**	**SalesOffice**	**OfficeNumber**	**Customer1**	**Customer2**	**Customer3**
1003	Mary Smith	Chicago	312-555-1212	Ford	GM	
1004	John Hunt	New York	212-555-1212	Dell	HP	Apple
1005	Martin Hap	Chicago	312-555-1212	Boeing		

Unnormalized Table

- The first thing to notice is this table serves many purposes including:
  - Identifying the organization's salespeople
  - Listing the sales offices and phone numbers
  - Associating a salesperson with an sales office
  - Showing each salesperson's customers

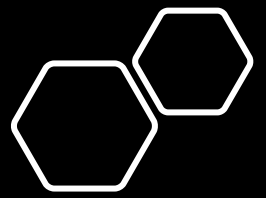


Reason #1- Data Duplication

SalesStaff						
EmployeeID	**SalesPerson**	**SalesOffice**	**OfficeNumber**	**Customer1**	**Customer2**	**Customer3**
1003	Mary Smith	Chicago	312-555-1212	Ford	GM	
1004	John Hunt	New York	212-555-1212	Dell	HP	Apple
1005	Martin Hap	Chicago	312-555-1212	Boeing		

Data Duplication Issues

- Duplicated information presents two problems:
  - It increases storage and decrease performance.
  - It becomes more difficult to maintain data changes.

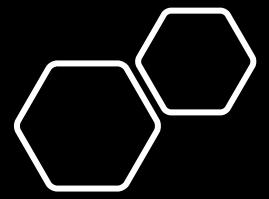


Reason #2- Data Anomalies

EmployeeID	SalesPerson	SalesOffice	OfficeNumber	Customer1	Customer2	Customer3
1003	Mary Smith	Chicago	312-555-1212	Ford	GM	
1004	John Hunt	New York	212-555-1212	Dell	HP	Apple
1005	Martin Hap	Chicago	312-555-1212	Boeing		
???	**???**	**Atlanta**	**312-555-1212**			

Insert Anomaly

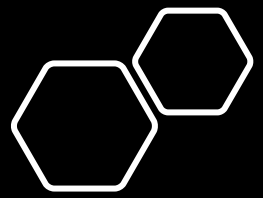
There are facts we cannot record until we know information for the entire row. In our example we cannot record a new sales office until we also know the salesperson. Why? Because in order to create the record, we need provide a primary key. In our case this is the EmployeeID.



EmployeeID	SalesPerson	SalesOffice	OfficeNumber	Customer1	Customer2	Customer3
1003	Mary Smith	Chicago	312-555-1212	Ford	GM	
1004	John Hunt	New York	212-555-1212	Dell	HP	Apple
1005	Martin Hap	Chicago	312-555-1212	Boeing		

Update Anomaly

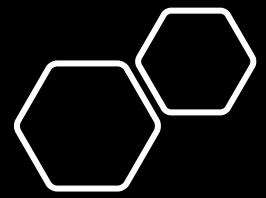
The same information is recorded in multiple rows. For instance, if the office number changes, then there are multiple updates that need to be made. If these updates are not successfully completed across all rows, then an inconsistency occurs



EmployeeID	SalesPerson	SalesOffice	OfficeNumber	Customer1	Customer2	Customer3
1003	Mary Smith	Chicago	312-555-1212	Ford	GM	
~~1004~~	~~John Hunt~~	~~New York~~	~~212-555-1212~~	~~Dell~~	~~HP~~	~~Apple~~
1005	Martin Hap	Chicago	312-555-1212	Boeing		

Deletion Anomaly

Deletion of a row can cause more than one set of facts to be removed. For instance, if John Hunt retires, then deleting that row causes us to lose information about the New York office.



Reason #3 – Query Issues

SalesStaff						
EmployeeID	**SalesPerson**	**SalesOffice**	**OfficeNumber**	**Customer1**	**Customer2**	**Customer3**
1003	Mary Smith	Chicago	312-555-1212	Ford	GM	
1004	John Hunt	New York	212-555-1212	Dell	HP	Apple
1005	Martin Hap	Chicago	312-555-1212	Boeing		

Search and Sort Issues

```
SELECT SalesOffice
FROM SalesStaff
WHERE Customer1 = ‘Ford’ OR
      Customer2 = ‘Ford’ OR
      Customer3 = ‘Ford’
```