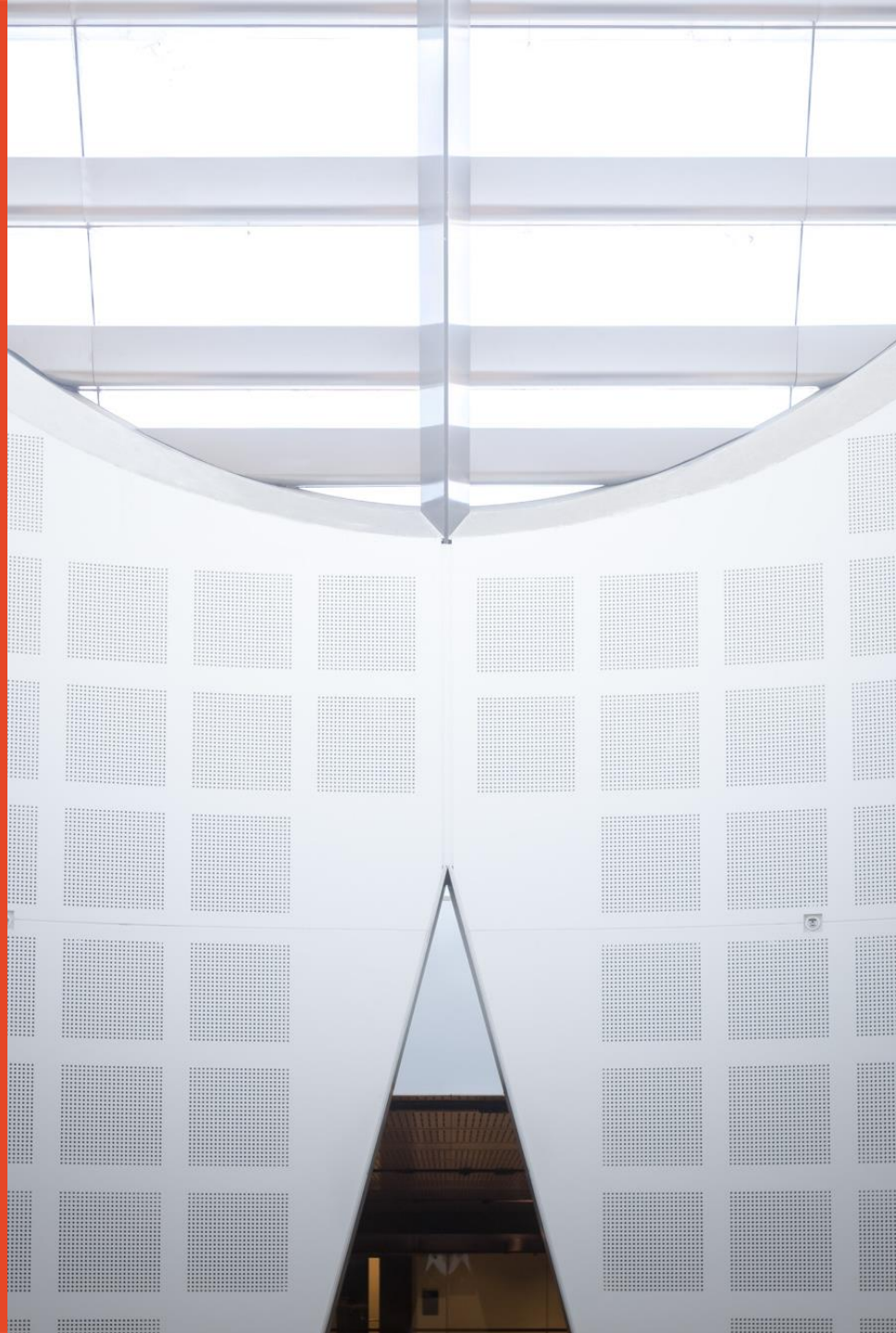
ISYS2110
Analysis and Design of Web Information Systems

Lecture 10
Data Design PART-B &
Systems Architecture Design

Semester 1, 2018

Dr Rabiul Hasan

THE UNIVERSITY OF SYDNEY



Recapture From Lecture 9

What we have covered on the topic: Data Design PART-A

- Data Design Concepts
- Database Environment
- Data Design Terms
- Entity Relationship Diagram

What Will We Do Today ?

- Lecture
  - Data Design Part-B (Normalization)
  - Issues to consider for a system architecture
  - Client/server architecture
  - System design specification

- Class activities
  - Critical Thinking / No Problem Solving Today
  - https://padlet.com
  - https://answergarden.ch

- Tutorial: ?
- Assessment ?
- Announcement (if any): We will have a guest speaker with us for about 30 minutes in today's lecture (3.30 pm)

Learning Objectives

- Explain the concept of normalization
- Provide a checklist of issues to consider when selecting a system architecture
- Explain client/server architecture, including tiers, cost-benefit issues, and performance

- Compare in-house ecommerce development with packaged solutions and service providers
- Describe the system design specification

Data Normalization

- **Normalization:** Process of creating table designs by assigning specific fields or attributes to each table in the database

- **Table design:** Specifies fields
  - Identifies the primary key in a particular table or file

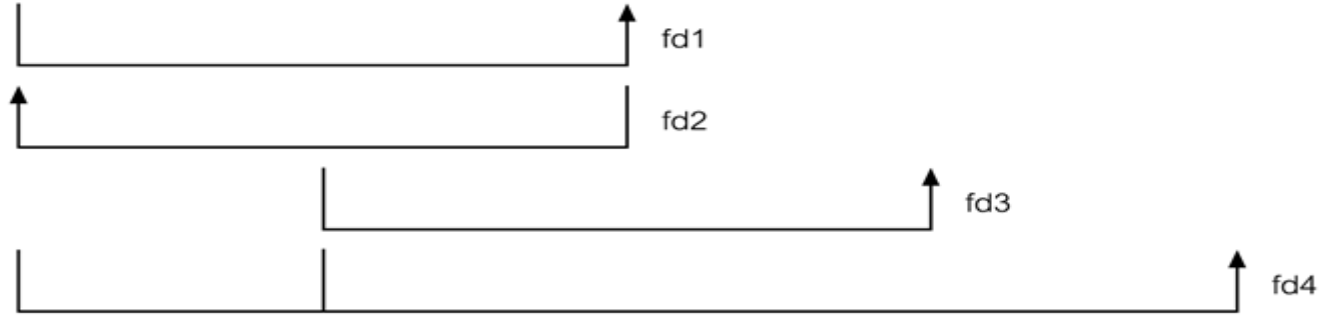
Data Normalization (Cont. 1)

- **Standard Notation Format**
  - Used to show a table's structure, fields, and primary key
  - The primary key field(s) is underlined
    - NAME (FIELD 1, FIELD 2, FIELD 3)
  - Recognition of repeating group fields is important
    - Repeating group: Set of one or more fields that can occur any number of times in a single record
    - Each occurrence would possess different values

Functional Dependencies

- Functional dependency describes relationship between attributes.
- For example, if A and B are attributes of table R, B is functionally dependent on A (A $\rightarrow$ B), if each value of A in R is associated with exactly one value of B in R

A	B	C	D	E
a	b	z	w	q
e	b	r	w	p
a	d	z	w	t
e	d	r	w	q
a	f	z	s	t
e	f	r	s	t



Adapted from Connolly and Begg (2010)

Transitive Dependencies

- Important to recognize a transitive dependency because its existence in a relation can potentially cause update anomalies.

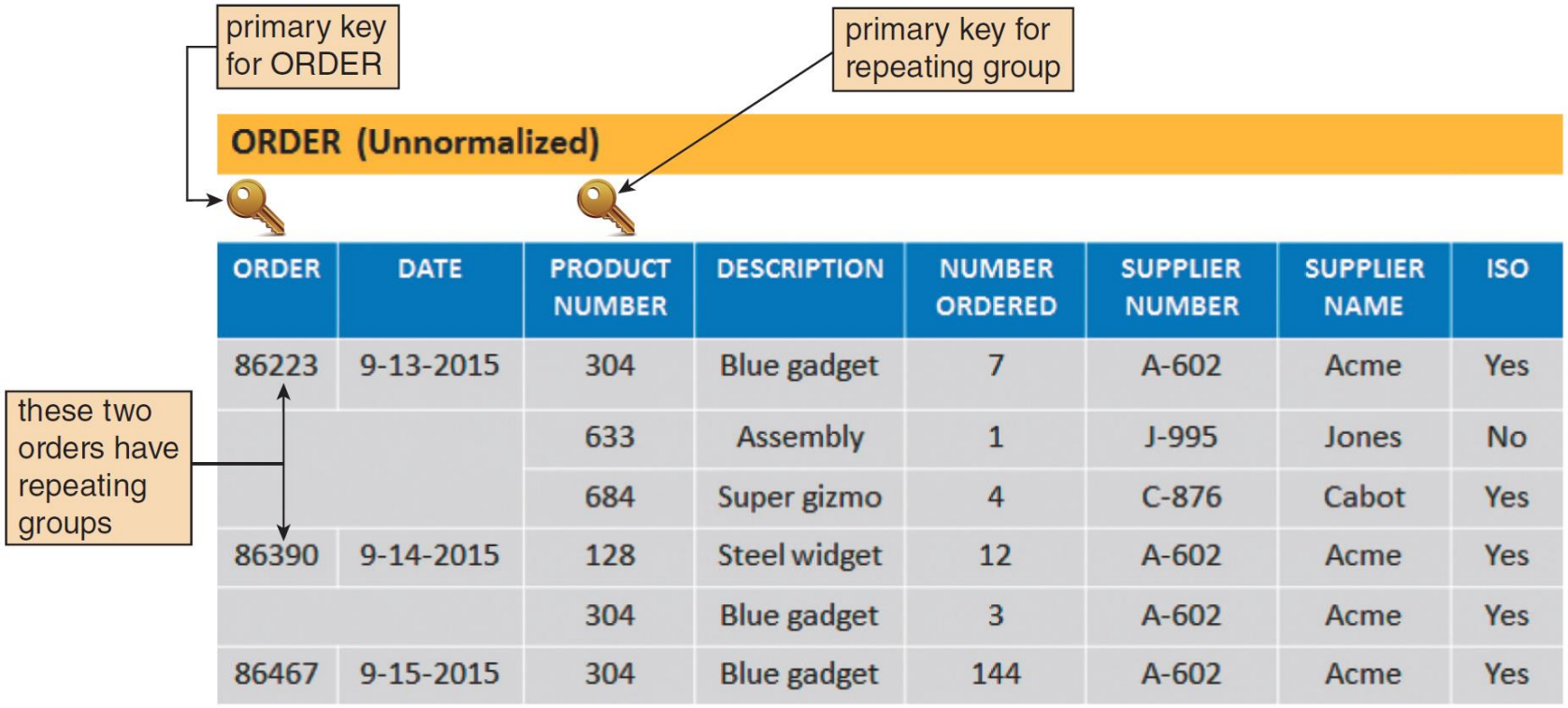
- Transitive dependency describes a condition where A, B, and C are attributes of a relation such that if A → B and B → C, then C is transitively dependent on A via B (provided that A is not functionally dependent on B or C).

Normalization Process -- Stages

- Unnormalized Form (UNF): A table that contains one or more repeating groups.
- First Normal Form (1NF): A table in which the intersection of each row and column contains one and only one value.

- Second Normal Form (2NF): A table that is in 1NF and every non-primary-key attribute is fully functionally dependent on the primary key.
- Third Normal Form (3NF): A table that is in 1NF and 2NF and in which no non-primary-key attribute is transitively dependent on the primary key.

Data Normalization (Cont. 2)



ORDER (Unnormalized)

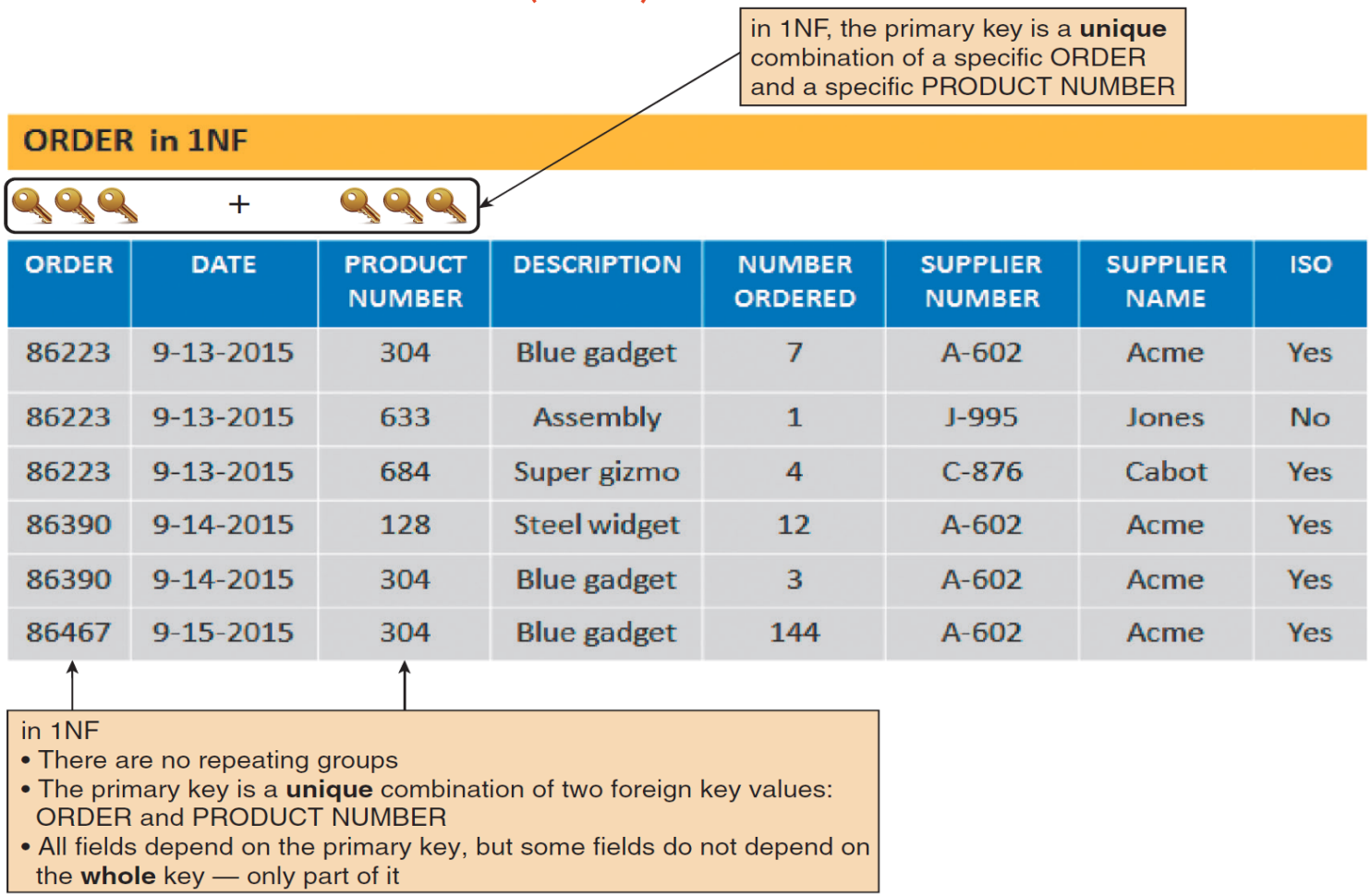
ORDER	DATE	PRODUCT NUMBER	DESCRIPTION	NUMBER ORDERED	SUPPLIER NUMBER	SUPPLIER NAME	ISO
86223	9-13-2015	304	Blue gadget	7	A-602	Acme	Yes
		633	Assembly	1	J-995	Jones	No
		684	Super gizmo	4	C-876	Cabot	Yes
86390	9-14-2015	128	Steel widget	12	A-602	Acme	Yes
		304	Blue gadget	3	A-602	Acme	Yes
86467	9-15-2015	304	Blue gadget	144	A-602	Acme	Yes

In the ORDER table design, two orders have repeating groups that contain several products. ORDER is the primary key for the ORDER table, and PRODUCT NUMBER serves as a primary key for the repeating group. Because it contains repeating groups, the ORDER table is unnormalized.

Data Normalization (Cont. 3)

- **UNF to1NF**
  - Nominate an attribute or group of attributes to act as the key for the unnormalized table.
  - **Identify and remove** the repeating group by entering appropriate data into the empty columns of rows containing the repeating data.

Data Normalization (Cont. 4)



ORDER	DATE	PRODUCT NUMBER	DESCRIPTION	NUMBER ORDERED	SUPPLIER NUMBER	SUPPLIER NAME	ISO
86223	9-13-2015	304	Blue gadget	7	A-602	Acme	Yes
86223	9-13-2015	633	Assembly	1	J-995	Jones	No
86223	9-13-2015	684	Super gizmo	4	C-876	Cabot	Yes
86390	9-14-2015	128	Steel widget	12	A-602	Acme	Yes
86390	9-14-2015	304	Blue gadget	3	A-602	Acme	Yes
86467	9-15-2015	304	Blue gadget	144	A-602	Acme	Yes

The ORDER table as it appears in 1NF. The repeating groups have been eliminated. Notice that the repeating group for order 86223 has become three separate records, and the repeating group for order 86390 has become two separate records. The 1NF primary key is a combination of ORDER and PRODUCT NUMBER, which uniquely identifies each record.

Data Normalization (Cont. 5)

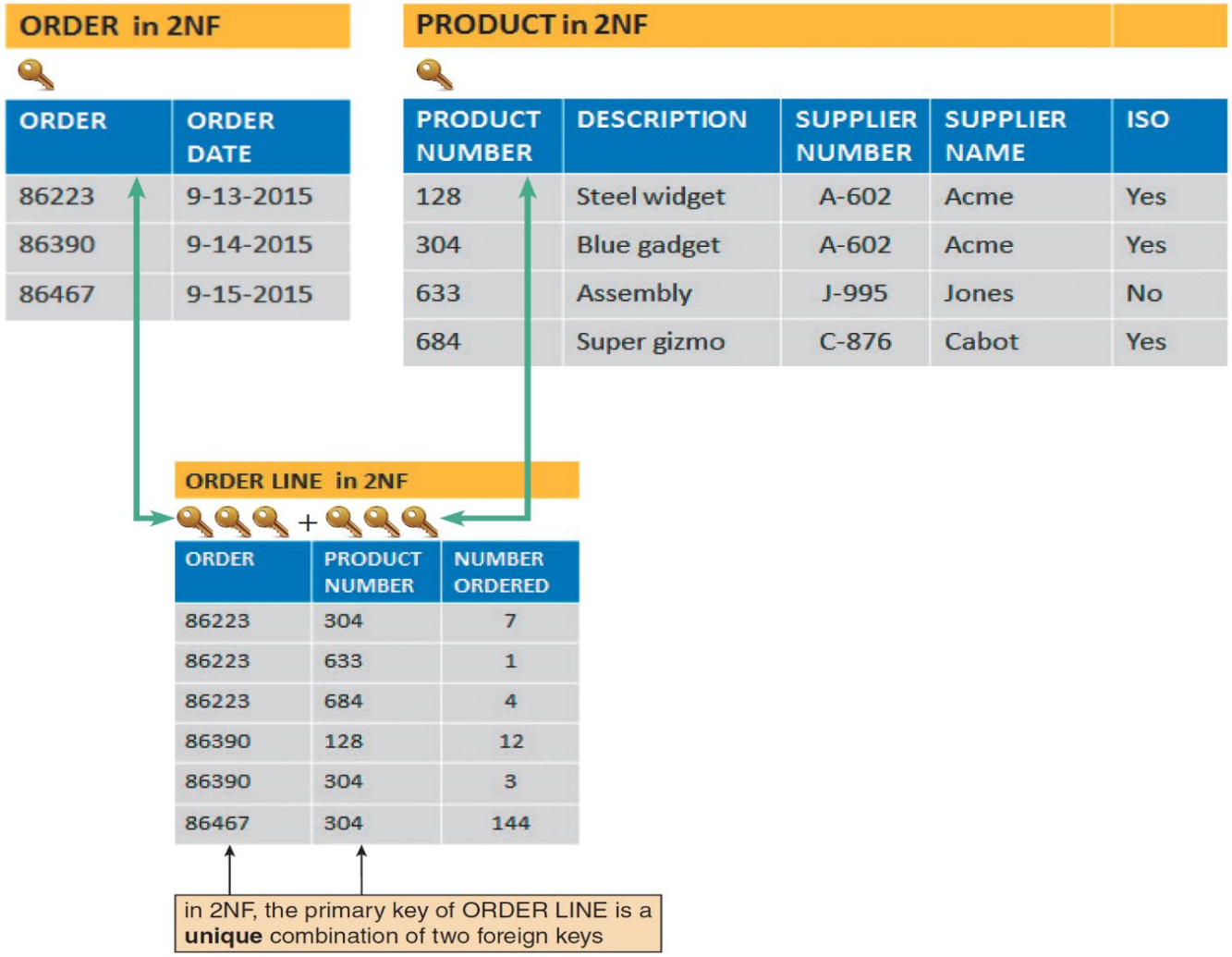
- **1NF to 2NF**

  - Identify the primary key for the 1NF relation.

  - Identify the functional dependencies in the relation.

  - If partial dependencies exist on the primary key remove them by placing then in a new relation along with a copy of their determinant.

Data Normalization (Cont. 6)



ORDER, PRODUCT, and ORDER LINE tables in 2NF. All fields are functionally dependent on the primary key.

Data Normalization (Cont. 7)

- **2NF to 3NF**

  - Identify the primary key in the 2NF relation.

  - Identify functional dependencies in the relation.

  - If transitive dependencies exist on the primary key remove them by placing them in a new relation along with a copy of their dominant.

Data Normalization (Cont. 8)

PRODUCT in 2NF

PRODUCT NUMBER	DESCRIPTION	SUPPLIER NUMBER	SUPPLIER NAME	ISO
128	Steel widget	A-602	Acme	Yes
304	Blue gadget	A-602	Acme	Yes
633	Assembly	J-995	Jones	No
684	Super gizmo	C-876	Cabot	Yes

PRODUCT in 3NF

PRODUCT NUMBER	DESCRIPTION	SUPPLIER NUMBER
128	Steel widget	A-602
304	Blue gadget	A-602
633	Assembly	J-995
684	Super gizmo	C-876

SUPPLIER in 3NF

SUPPLIER NUMBER	SUPPLIER NAME	ISO
A-602	Acme	Yes
J-995	Jones	No
C-876	Cabot	Yes

When the PRODUCT table is transformed from 2NF to 3F, the result is two separate tables: PRODUCT and SUPPLIER. Note that in 3NF, all fields depend on the key, the whole key, and nothing but the key!

More Examples of Normalization

- **Crossroads College**

STUDENT (Unnormalized)

STUDENT NUMBER	STUDENT NAME	TOTAL CREDITS	GPA	ADVISOR NUMBER	ADVISOR NAME	OFFICE	COURSE NUMBER	CREDIT HOURS	GRADE
1035	Linda	47	3.60	49	Smith	B212	CSC151	4	B
							MKT212	3	A
							ENG101	3	B
							CHM112	4	A
							BUS105	2	A
3397	Sam	29	3.00	49	Smith	B212	ENG101	3	A
							MKT212	3	C
							CSC151	4	B
4070	Kelly	14	2.90	23	Jones	C333	CSC151	4	B
							CHM112	4	C
							ENG101	3	C
							BUS105	2	C

The STUDENT table is unnormalized because it contains a repeating group that represents the courses each student has taken.

More Examples of Normalization

The STUDENT table in 1NF. Notice that the primary key has been expanded to include STUDENT NUMBER and COURSE NUMBER.

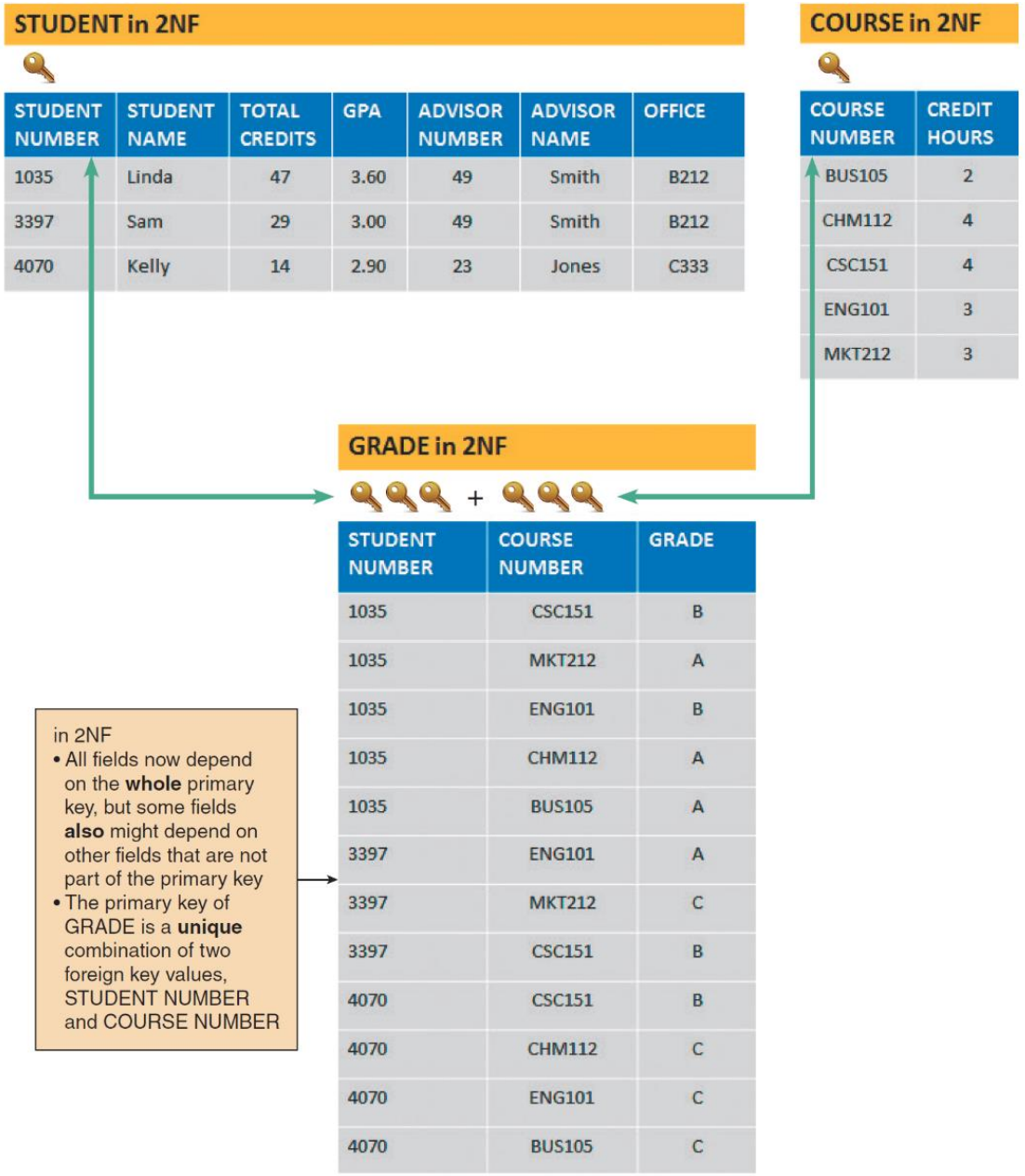
in 1NF, the primary key is a **unique** combination of a specific STUDENT NUMBER and a specific COURSE NUMBER

STUDENT in 1NF

STUDENT NUMBER	STUDENT NAME	TOTAL CREDITS	GPA	ADVISOR NUMBER	ADVISOR NAME	OFFICE	COURSE NUMBER	CREDIT HOURS	GRADE
1035	Linda	47	3.60	49	Smith	B212	CSC151	4	B
1035	Linda	47	3.60	49	Smith	B212	MKT212	3	A
1035	Linda	47	3.60	49	Smith	B212	ENG101	3	B
1035	Linda	47	3.60	49	Smith	B212	CHM112	4	A
1035	Linda	47	3.60	49	Smith	B212	BUS105	2	A
3397	Sam	29	3.00	49	Smith	B212	ENG101	3	A
3397	Sam	29	3.00	49	Smith	B212	MKT212	3	C
3397	Sam	29	3.00	49	Smith	B212	CSC151	4	B
4070	Kelly	14	2.90	23	Jones	C333	CSC151	4	B
4070	Kelly	14	2.90	23	Jones	C333	CHM112	4	C
4070	Kelly	14	2.90	23	Jones	C333	ENG101	3	C
4070	Kelly	14	2.90	23	Jones	C333	BUS105	2	C

in 1NF
- There are no repeating groups
- The primary key is a **unique** combination of two foreign key values: STUDENT NUMBER and COURSE NUMBER
- All fields depend on the primary key, but some fields do not depend on the **whole** key — only part of it

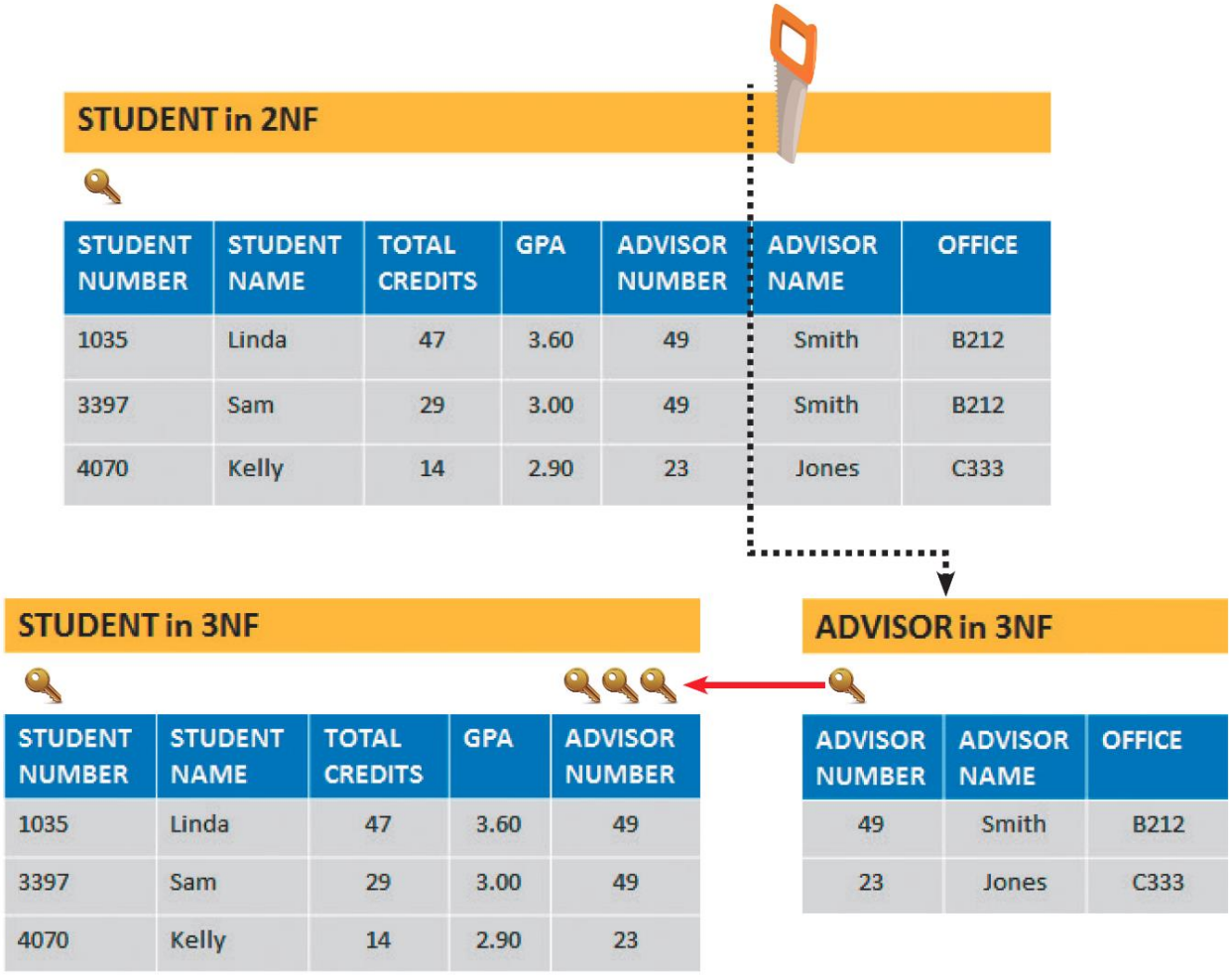
More Examples of Normalization



The STUDENT, COURSE, and GRADE tables in 2NF. Notice that all fields are functionally dependent on the entire primary key of their respective tables.

More Examples of Normalization

STUDENT, ADVISOR, COURSE, and GRADE tables in 3NF. When the STUDENT table is transformed from 2NF to 3NF, the result is two tables: STUDENT and ADVISOR.



Question?

- For Crossroads College, after performing 3NF, how many tables to build and what are they?

Class Exercise -- Normalization

- Assume that a patient is registered at only one surgery and he/she may have more than one appointment on a given day. All the schedules have been fixed for the whole days and week.

staffNo	dentistName	patNo	patName	date	time	surgeryNo
S1011	Tony Smith	P100	Gillian White	12-Sep-08	10.00	S15
S1011	Tony Smith	P105	Jill Bell	12-Sep-08	12.00	S15
S1024	Helen Pearson	P108	Ian MacKay	12-Sep-08	10.00	S10
S1024	Helen Pearson	P108	Ian MacKay	14-Sep-08	14.00	S10
S1032	Robin Plevin	P105	Jill Bell	14-Sep-08	16.30	S15
S1032	Robin Plevin	P110	John Walker	15-Sep-08	18.00	S13

Class Exercise -- Normalization

staffNo	dentistName	patNo	patName	date	time	surgeryNo

Task1: Identify the functional dependencies (e.g. Primary key, Candidate key, Partial dependency, Transitive dependency) for the table provided

Class Exercise -- Normalization

staffNo	dentistName	patNo	patName	date	time	surgeryNo

Task2: Describe the table structure from 1NF to 3NF.

1NF:

2NF:

3NF:

System Architecture and Design

Architecture Checklist

- Issues that influence the architecture choice
  - Corporate organization and culture
  - Enterprise resource planning (ERP)
  - Initial and total cost of ownership (TCO)
  - Scalability
  - Web integration
  - Legacy system interface requirements
  - Processing options
  - Security issues
  - Corporate portals

Architecture Checklist (Cont. 1)

- **Corporate Organization and Culture**
  - A successful system performs well in a company's organization and culture
- **Enterprise resource planning (ERP)**
  - Objective – To establish a company-wide strategy for using IT that includes a specific architecture, standards for data, processing, network, and user interface design

Oracle offers ERP solutions as a cloud-based service.

Source: Oracle

Architecture Checklist (Cont. 2)

- Companies are extending internal ERP systems to their suppliers and customers, using supply chain management (SCM)

- **Initial Cost and Total Cost Ownership (TCO)**
  - TCO includes tangible purchases, fees, and contracts called hard costs
  - TCO analysis answers questions about the validity, effectiveness, and new trends in systems planning
    - TCO analysis may affect the initial cost and TCO for a proposed system

Architecture Checklist (Cont. 3)

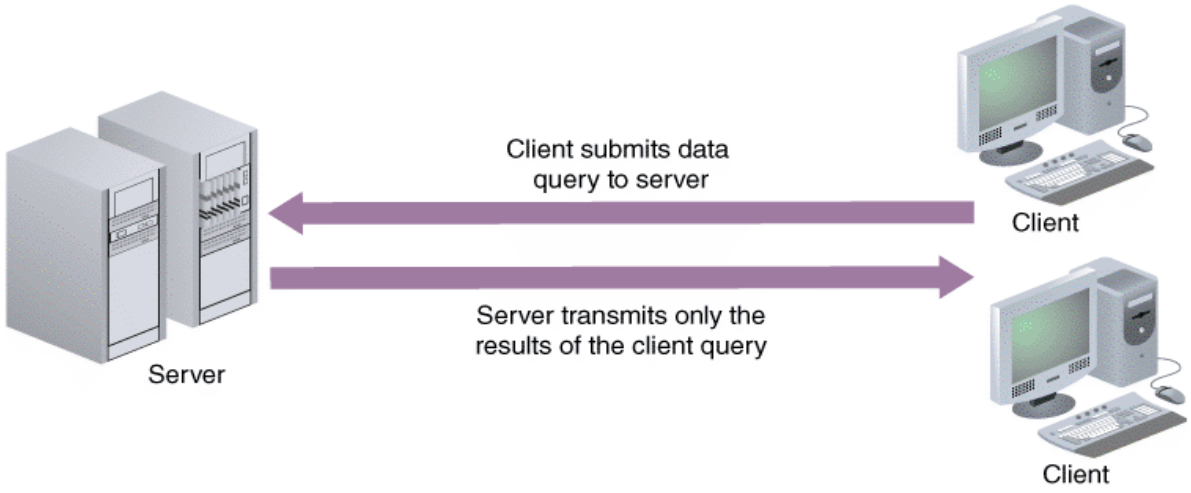
- **Scalability** (Extensibility)
  - A system's ability to expand, change, or downsize easily to meet the changing needs of a business enterprise
- **Web Integration**
  - A **web-centric** architecture enables a company to integrate new **applications** into its ecommerce strategy
- **Legacy Systems**
  - A new system might have to interface with old systems
    - Involves analysis of data formats and compatibility

Architecture Checklist (Cont. 4)

- **Processing Options**
  - Systems can process data online or in batches
- **Security Issues**
  - Analysts must consider security issues and how the company will address them
- **Corporate Portals**
  - Provide access for customers, employees, suppliers, and the public
  - A well-designed portal can:
    - Integrate with various other systems
    - Provide a consistent look and feel across organizational divisions

Client/Server Designs

- **Client/Server Architecture**
  - Includes systems that divide processing between one or more networked clients and a central server
    - Client handles the entire user interface
    - Server stores data and provides data access and database management functions



In a client/server design, data is stored and usually processed on the server.

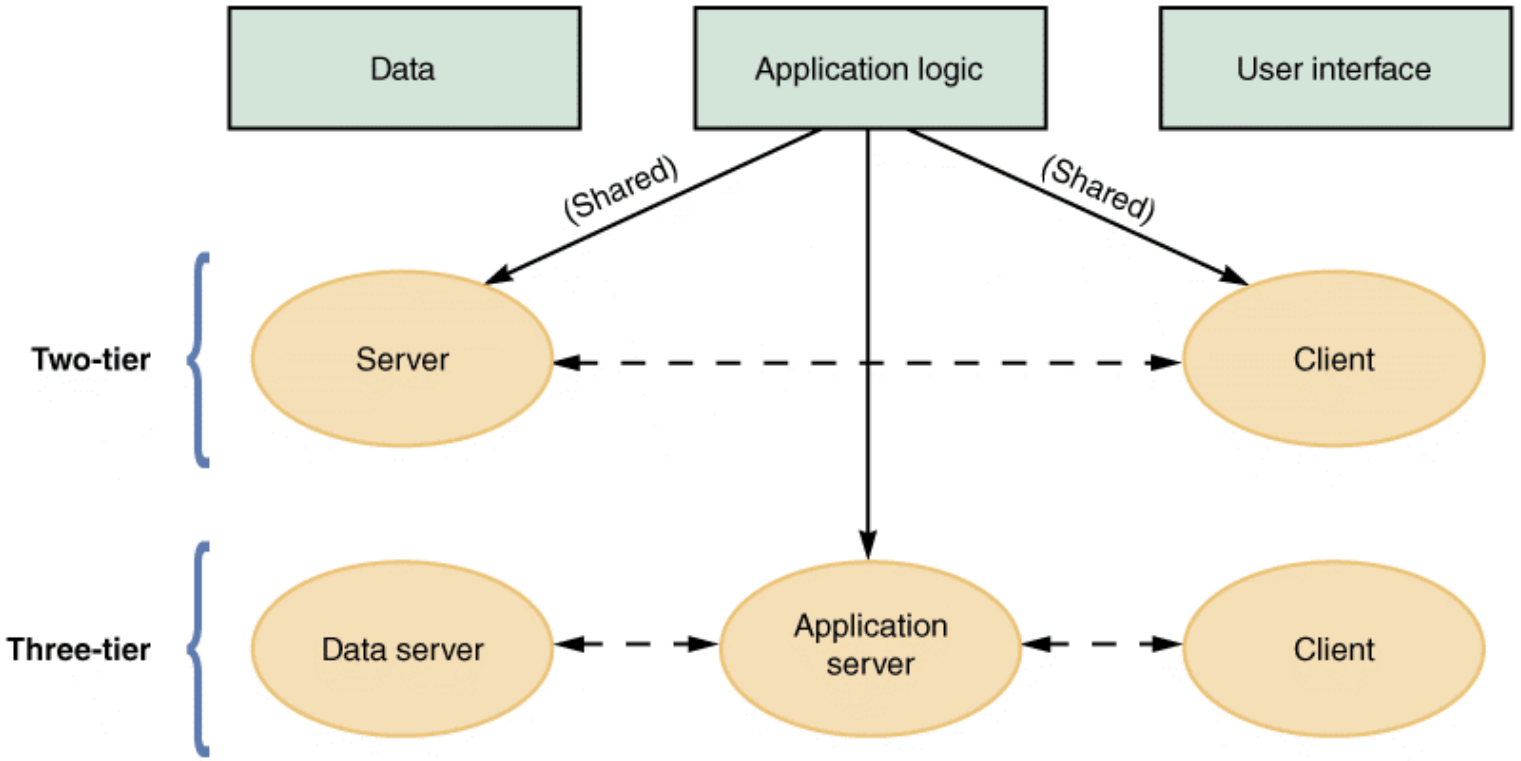
Client/Server Designs (Cont. 2)

- **The Client's Role**
  - Client/server relationship must specify how the processing will be divided between the client and the server
  - **Fat client (thick client)** design: Locates all or most of the application processing logic at the client
  - **Thin client** design: Locates all or most of the processing logic at the server
    - Provides better performance as the program code resides on the server

Client/Server Designs (Cont. 3)

- **Client/Server Tiers**
  - **Two-tier design**
    - User interface resides on the client
    - Data resides on the server
    - Application logic can run either on the server or on the client, or be divided between the client and the server
  - **Three-tier (*n*-tier) design**
    - User interface runs on the client
    - Data is stored on the server
    - Has a middle layer between the client and server
      - Processes the client requests and translates them into data access commands
    - Considered an application server

Client/Server Designs (Cont. 4)



Characteristics of two-tier versus three-tier client/server design.

Client/Server Designs (Cont. 5)

Architecture		Data	Application Logic	User Interface
Central data processing center	Server	X	X	X
	Client			
Central server with remote terminals	Server	X	X	
	Client			X
Stand-alone client	Server			
	Client	X	X	X
Two-tier client/server	Server	X	X	
	Client		X	X
Three-tier client/server	Data server	X		
	Application server		X	
	Client			X

The location of the data, the application logic, and the user interface depend on the type of architecture.

Client/Server Designs (Cont. 6)

- **Middleware**
  - Enables communication between the tiers
  - Referred to as **glueware**
    - Used to connect two or more software components in a federated system architecture
  - Integrates legacy systems and Web-based and/or cloud applications
  - Represents the slash in the term client/server
- **Cost-Benefit Issues**
  - Client/server systems offer the best combination of features to meet information system requirements

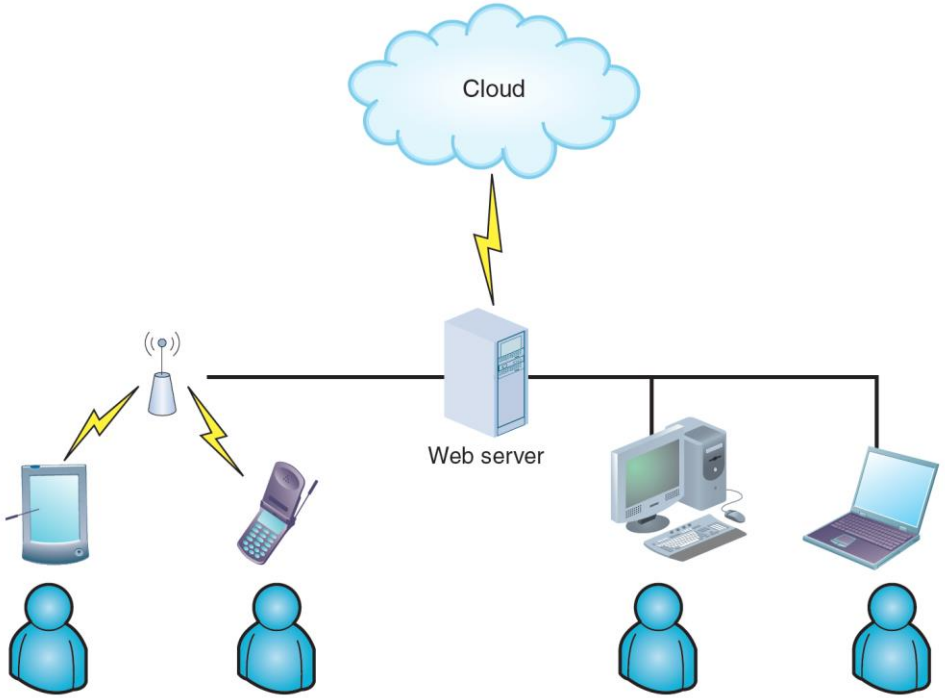
Internet-based architecture

- In an Internet-based architecture, the entire user interface is provided by the web server in the form of HTML documents

- Shifting the responsibility for the interface from the client to the server simplifies data transmission and results in lower hardware cost and complexity

Internet-based architecture

- **Cloud Computing**
  - The concept envisions a cloud of remote computers providing a total online software and data environment that is hosted by third parties
  - Eliminates compatibility issues and provides scaling on demand



Cloud computing

Ecommerce Architecture

- **In-House Solutions**
  - Benefits
    - A unique website, with a look and feel consistent with the company's other marketing efforts
    - Complete control over the organization of the site
    - A scalable structure to handle increases in sales and product offerings in the future
    - More flexibility to modify and manage the site
    - The opportunity to integrate the firm's web-based business systems with its other information systems

Ecommerce Architecture (Cont. 1)

Guidelines for companies developing ecommerce strategies.

Guidelines for In-house Ecommerce Site Development
Analyze the company's business needs and develop a clear statement of your goals. Consider the experience of other companies with similar projects.
Obtain input from users who understand the business and technology issues involved in the project. Plan for future growth, but aim for ease of use.
Determine whether the IT staff has the necessary skills and experience to implement the project. Consider training, additional resources, and the use of consultants if necessary.
Consider integration requirements for existing legacy systems or enterprise resource planning. Select a physical infrastructure carefully, so it will support the application, now and later.
Develop the project in modular form so users can test and approve the functional elements as you go along.
Connect the application to existing in-house systems and verify interactivity.
Test every aspect of the site exhaustively. Consider a preliminary rollout to a pilot group to obtain feedback before a full launch.

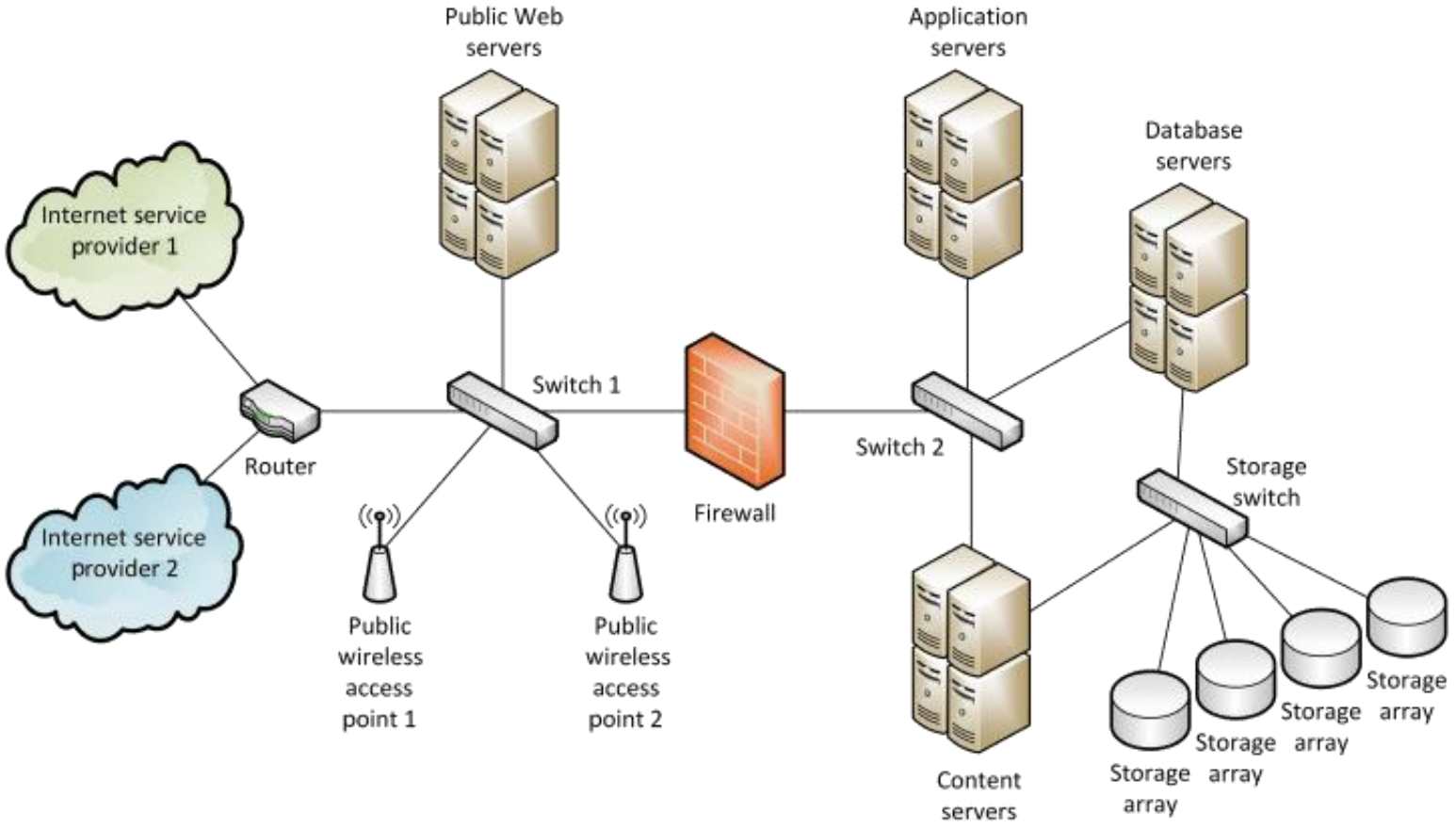
Ecommerce Architecture (Cont. 2)

- **Packaged Solutions**
  - Viable alternative for medium- to large-sized firms
  - Less complex than an in-house effort

- **Service Providers**
  - Application service provider (ASP) - Provides applications or access to applications by charging a fee
    - Many ASPs offer full-scale Internet business services for companies that decide to outsource functions

Network Diagram for System Architecture

- How the application software is deployed across the hardware and system software



Adapted from Satzinger et al., 2016

System Design Completion

- System architecture marks the end of the systems design phase of the SDLC

- Final activities in the systems design phase
  - Preparing a system design specification
  - Obtaining user approval
  - Delivering a presentation to management

System Design Completion (Cont. 1)

- **System Design Specification**
  - Document that presents the complete design for a new information system
    - Contains detailed costs, staffing, and scheduling for completing the next SDLC phase
  - Used as a baseline to measure the operational system

System Design Completion (Cont. 2)

- Sections in a system design specification
  - Management summary
  - System components
  - System environment
  - Implementation requirements
  - Time and cost estimates
  - Additional material

System Design Completion (Cont. 3)

- **User Approval**
  - Users must review and approve the interface design, report and menu designs, data entry screens, source documents, and other areas of the system that affect them
    - Ensures that approvals are obtained as and when required
    - Keeps the users involved with the system's development
    - Provides feedback that can be used to guide efforts
  - System design specification should be reviewed by other IT department members as well

System Design Completion (Cont. 4)

- **Presentation to management**
  - Provide an opportunity to explain the system, answer questions, consider comments, and secure final approval
    - The first presentation is to the systems analysts, programmers, and technical support staff members who will be involved in future project phases or operational support for the system
    - Next presentation is to the department managers and users from departments affected by the system
    - Final presentation is delivered to management
  - Management will reach a decision based on the presentation

Lecture Summary

- Normalization is a process for avoiding problems in data design
- ERP establishes an enterprise-wide strategy for IT resources and specific standards for data, processing, network, and user interface design
- A system architecture requires servers and clients
  - Client/server architecture divides processing between one or more clients and a central server
- A thick client design places all or most of the application processing logic at the client
- A thin client design places all or most of the processing logic at the server

Lecture Summary

- Client/server designs can be two- or three-tier
- The Internet has had an enormous impact on system architecture

- The system design specification presents the complete systems design for an information system and is the basis for the presentations that complete the systems design phase

Announcement (if any)

Q &A?

Thanks everyone !