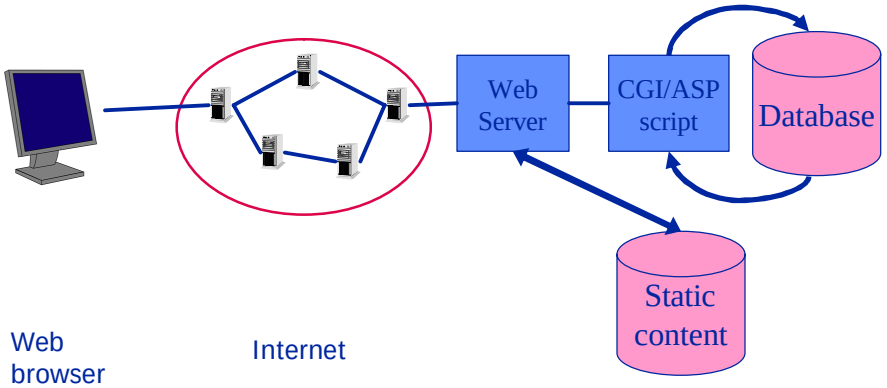
Under the hood of a commercial Web Site

- Issues
- Typical Site Architecture
- Case Study: Travelocity

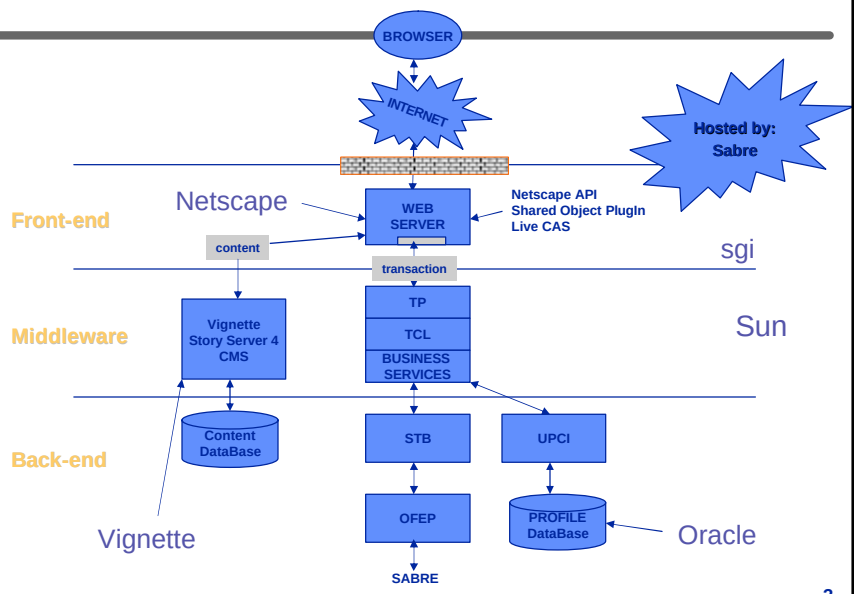
1

The story so far…



2

Travelocity Architecture



3

Issues for building mission-critical eCommerce sites

- **Content management**
  - Ensure that content development is done in a streamlined and orderly fashion
- **Security**
  - Physical security
  - Access control
- **Availability/Fault Tolerance**
  - Ensure the computer services remain available to users in the face of partial failures
- **Accuracy**
  - Ensure that multi-user access and system crashes leave data in an accurate state
- **Scalability**
  - Ensure that response time remains acceptable as site traffic grows

4

Building mission-critical eCommerce sites: Summary of Technologies

- Content management
  - Content Management Software
- Security
  - Physical security: Hosting
  - Access control: Firewalls
- Availability/Fault Tolerance
  - Replication
- Accuracy
  - Transaction Processing
- Scalability
  - Replication
  - Load Balancing
  - Web Caching

5

Issue #1: Large-scale content development

- Large number of authors contributing site content
- Diverse types of content (e.g., image, video, and other media files)
- Need for regular content posting and replacement (i.e., weekly sales promotions)
- Often one or more approvals are required before content is posted
- Some content needs to be personalized or tailored to match the needs and interests of a site visitor

6

Additional issues

- Often support for multiple languages and time zones is required
- Content presentation must consistently conform to branding and appearance standards
- Version archiving and an audit trail
- Content must be viewable across a variety of browsing devices, not just PCs

7

Example: Content update workflow

- Automatic support for content management workflows

8

Solution: Content management systems

- Still a fledgling application category
  - Lots of Web authoring tools claim to be but aren't!
- Relatively small adoption from site developers
  - More than 50% of sites use manual methods
  - Market leader (Vignette) only has 8% of market!!!
- Hefty price tag
  - Vignette comes to more than $0.5M
- Lots of growth potential, but also room for better products

9

Representative Vendors

- CardoNet
- Interwoven
- OnDisplay
- Poet
- Versifi
- Vignette

10

Issue #2: Network and Physical Security

11

What a firewall does

- Hides the address of the network by making it appear that all transmissions originate from the firewall.
- Passes outgoing traffic without screening, while hiding the network address.
- Blocks all data not specifically requested by a legitimate user of the network.
- Screens data for source and destination address so you receive data from only trusted locations like people on your approved guest list.
- Screens the contents of data packets for known hacker attacks

12

Types of firewalls

- Packet filter: Looks at each packet entering or leaving the network and accepts or rejects it based on user-defined rules.
- Application gateway: Applies security mechanisms to specific applications, such as FTP and Telnet servers.
- Proxy server: Intercepts all messages entering and leaving the network. The proxy server effectively hides the true network addresses

13

Web Hosting

- Ensure 24x7 site operation
- Provide access to network bandwidth
- Provide physical site security

14

Self-manage or Host?

- Hosting services are on the rise
  - Concentrated technology expertise
  - Scalability and performance issues
  - Security issues
- Hosters are an uneven lot
- Hosting is not a commodity

15

The various flavors of hosting

- Simple hosting
  - Examples: Sprint, UUNET
- Collocation hosting
  - Examples: Frontier GlobalCenter, Sprint, UUNET
- Managed hosting
  - Examples: GTE, UUNET, IBM, Qwest
- Full-service hosting
  - Examples: IBM, Qwest

16

There is no single right answer

- Select level of hosting based on what kind of company you are
  - eCommerce pioneer → collocation hosting
    - Yahoo, Amazon
  - basic eCommerce presence → simple hosting
    - Century 21
  - some in-house expertise → managed hosting
    - Land'sEnd, Vanguard
  - little in-house expertise → full-service hosting
    - Amtrak, General Motors

17

Issue #3: Ensuring Availability and Fault Tolerance

- Why do computers crash?
  - Hardware errors
  - Operating system errors
  - Application errors
  - Human errors
- Use redundancy to restore normal operation after crashes
  - Data redundancy
  - Active Replication

18

Data Replication

- Keep several copies of same data (replicas)
- If one server is down, query next server
- Can improve response when load is heavy
- Problem: How to synchronize replicas?

19

Active Replication

- Establish redundant copies of vital programs and servers
  - process groups
  - every group member operates on its own replica
- Every message is processed by all group members
  - members remain in mutually consistent states
- If one member fails, other members can still respond

20

How can we do even better?

21

Issue #4: System Crashes might corrupt data

- Transfer money from savings to checking
  - x=sav; sav = x+1
  - w=chk; chk = w -1
- What happens if system crashes in the middle
  - x=sav; sav = x+1
  - SYSTEM CRASH
- Need to have a way to undo the effects of partially completed logical operations

22

Solution: Transaction Model

- Transaction: a process that possibly changes the database state
  - Self-contained, indivisible set of accesses to the database
  - May involve several reads and writes
  - All-or-nothing execution
  - Example: Transfer money is a single transaction
    - x=sav; sav = x+1; w=chk; chk = w -1

23

Using Transactions for Crash Recovery

- Transactions are atomic (indivisible)
  - Either executes completely or not at all
  - Any transaction that has not yet issued any writes may restart without causing any damage.
  - Once a transaction starts writing, it should do all of its writes
    - Even if it crashes in the middle
- Transactions are durable
  - If the transaction is "committed" (starts writing)
    - It should finish, even if machine crashes partway through

24

Implementation: DO-UNDO-REDO Log

- Keep a log of all database writes ON DISK
  - transaction id; data item; new value
    - (Tj; x=25) (Ti; y=56)
  - But don't write to the database yet
- At the end of transaction execution
  - Add "commit <transaction id>" to the log
  - Do all the writes to the database
  - Add "complete <transaction id>" to the log
  - Now it's OK to release the locks
- Restart after a crash by redoing the log
  - Any write for a committed but uncompleted transaction gets written again
    - What if the value was already written?
  - Any write for a non-committed or completed transaction is ignored

25

Distributed Transactions

- Distributed transaction may write data to several sites
  - Transfer money between accounts on separate computers
  - Update several copies of a database
- Want to write all data or cancel the transaction
  - Transaction Manager program may crash
  - Data sites may crash
  - Network may temporarily stop sending messages
- Needs more complicated protocols
  - out of the scope of this class

26

Issue #5 Scalability: Why you should care

- "... the No.1 reason that customers got fed up and took their business elsewhere was technical problems, including unacceptably slow response times."
  - Fortune magazine, November 8, 2001
- "...28% of Netizens that encountered glitches, left the site never to return"
  - Business Week, November 1, 2001

27

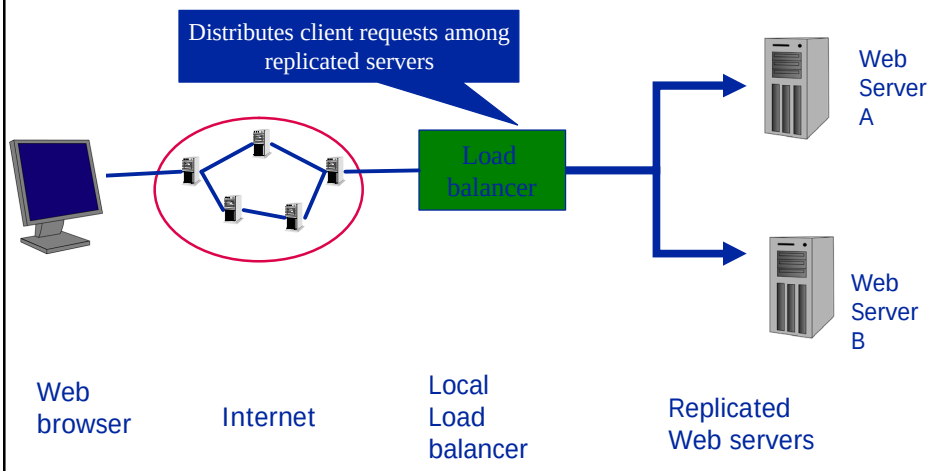
Technological Alternatives

- Local load balancer
- Site mirroring
- Network caching
- Content routing

28

Load balancing solutions

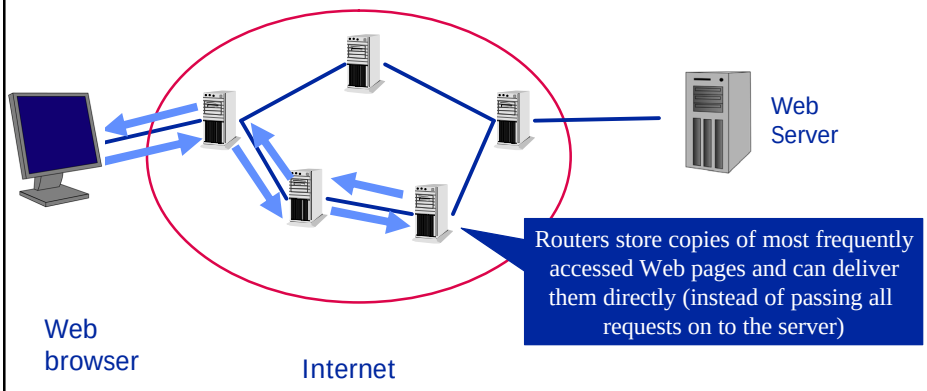
Companies: Cisco, HydraWeb



29

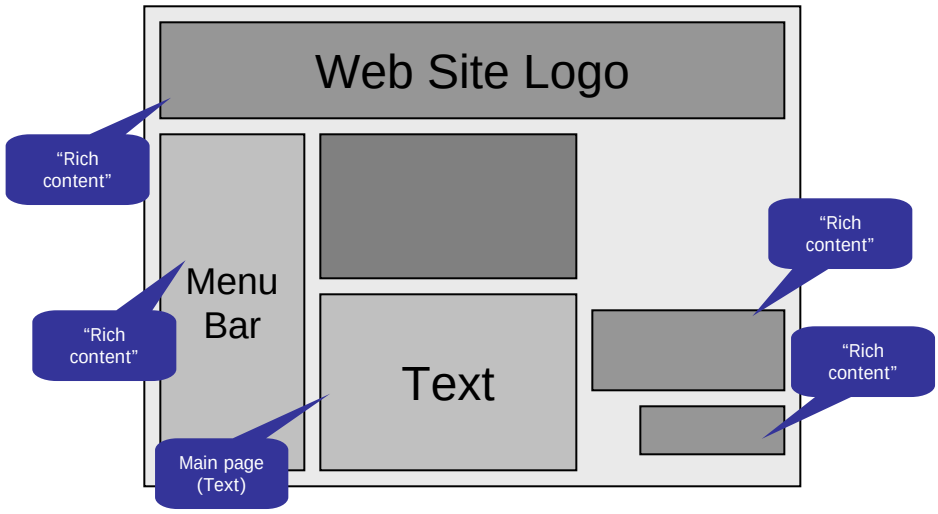
Network caching solutions

Companies: Inktomi, CacheFlow



30

Intelligent content routing



31

Intelligent content routing

Companies: Akamai, Sandpiper

32

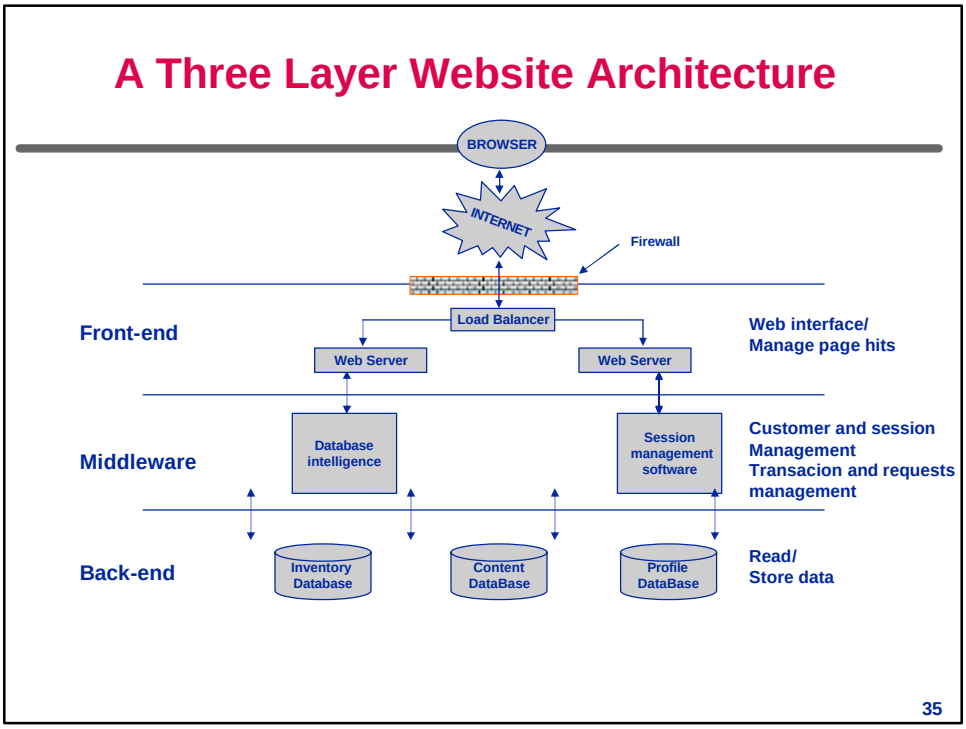
Akamai Current Network

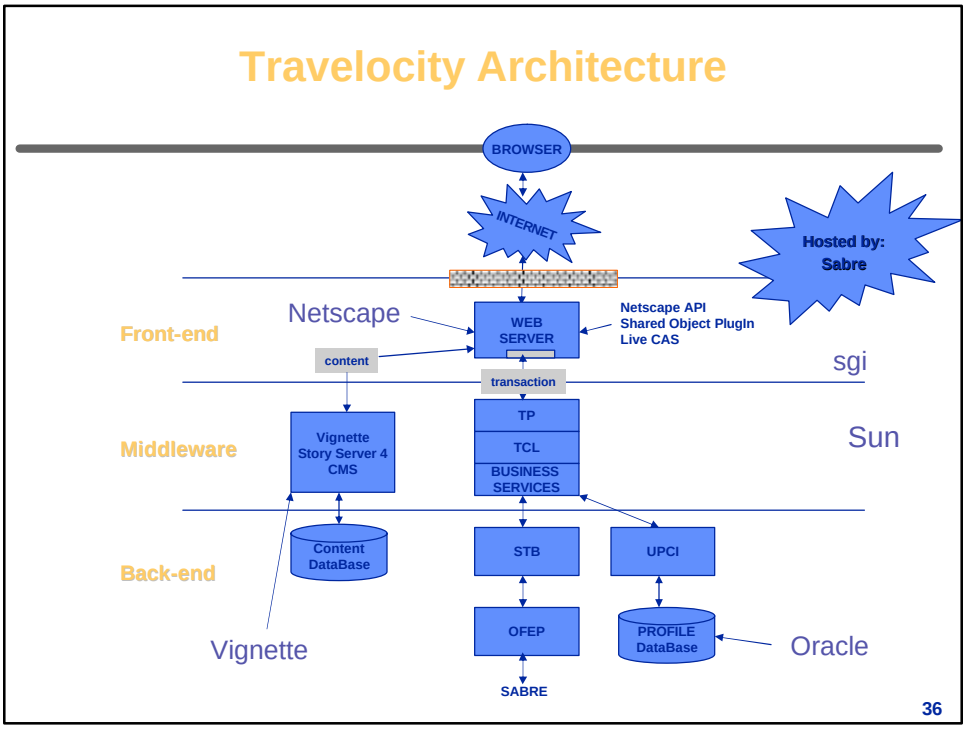
33

Building mission-critical eCommerce sites: Summary of Technologies

- Content management
  - Content Management Software
- Security
  - Physical security: Hosting
  - Access control: Firewalls
- Availability
  - Server Replication
- Accuracy
  - Transaction Processing
- Scalability
  - Replication
  - Load Balancing
  - Web Caching

34


A Three Layer Website Architecture
BROWSER
INTERNET
Firewall
Load Balancer
Front-end
Web Server
Web Server
Web interface/
Manage page hits
Middleware
Database intelligence
Session management software
Customer and session Management
Transacion and requests management
Back-end
Inventory Database
Content DataBase
Profile DataBase
Read/
Store data
35


Travelocity Architecture
BROWSER
INTERNET
Hosted by:
Sabre
Netscape
WEB SERVER
Netscape API
Shared Object PlugIn
Live CAS
Front-end
content
transaction
sgi
Middleware
Vignette Story Server 4 CMS
TP
TCL
BUSINESS SERVICES
Sun
Back-end
Content DataBase
STB
UPCI
OFEP
PROFILE DataBase
Oracle
Vignette
SABRE
36