



Web engineering

Developing Solutions for
Scalable Web Applications

Introduction

- Field of web engineering
- Object-oriented programming on the web
- Caching to reduce application load
- Database replication
- Distributed computing

Web engineering as a new discipline within computer science

- Concerned with unique concerns of scalability of today's web applications
- Traditionally, web applications were developed in a very ad-hoc manner
- Web engineering applies tested software engineering concepts to web application development

Object-oriented programming on the web

- Object-oriented design uses encapsulation to reduce the complexity of a software project – ideal for the web
- OOHDM is OO for the web, and encompasses conceptual, navigational, and abstract interface design, and implementation
- Applies OO concepts to developing rich user interfaces on the web

Caching to reduce load on the application

- Most common cache is reverse-proxy cache like Squid
- Can reduce application load with cache hit rates approaching 20-50%
- Caching brings data closer to user in order to increase performance
- Clustering increases cache hit rate
- In-memory object caching (memcached) can also be used to reduce load
- Only speeds up subsequent hits

Database replication

- Database servers didn't used to have failover, and blocked on lots of reads
- Replication scales out reads, improving performance and availability
- Master-slave replication ensure consistency
- Eventually consistent replication and sharding are also used

Distributed computing

- Virtual servers can dynamically respond to demand
- Sometimes demand is unpredictable, and sudden (disaster recovery web site)
- Distributed computing can also be used to process large quantities of data
- Companies are increasingly building large data warehouses that must be analyzed in real-time

Hadoop

- Suite of products for distributed computing
- Hadoop Distributed Filesystem allows you to transfer and manipulate large files on the web
- MapReduce allows you to perform transformations on data in parallel and in real time
- ZooKeeper exposes primitive data types that you can use in your application which are kept consistent across servers

Conclusion

- High school dropout using PHP can't build today's scalable web applications
- Web engineering allows us to use proven software engineering techniques to produce quality web applications that scale
- Object-oriented programming can be used to reduce the complexity of web applications

Conclusion (cont.)

- Caching can help us reduce load on our web application
- Database replication can help us scale reads on our application's data
- Distributed computing can help us dynamically respond to demand, and process massive quantities of data
- Careful design is necessary if we are to produce web applications that will keep pace with the increase of traffic on the web