

CS-310 Scalable Software Architectures

Lecture 2:

HTTP and Web Servers

Steve Tarzia

Last Time: Types of Scaling

- A software **service** is a program that runs continuously, giving responses to requests.
- **Scalability** is the ability of a service to grow to handle many concurrent users (ideally an arbitrarily large number).
- Two approaches to scaling that are useful in different scenarios:
- **Vertical scaling** is upgrading your machine(s).
 - The simplest and most efficient way of scaling... but there is a ceiling.
- **Horizontal scaling** is adding more machines.
 - Coordinating a cluster of machines is complicated, but it's necessary for global scale and massive throughput.

How are **services** different than programs?

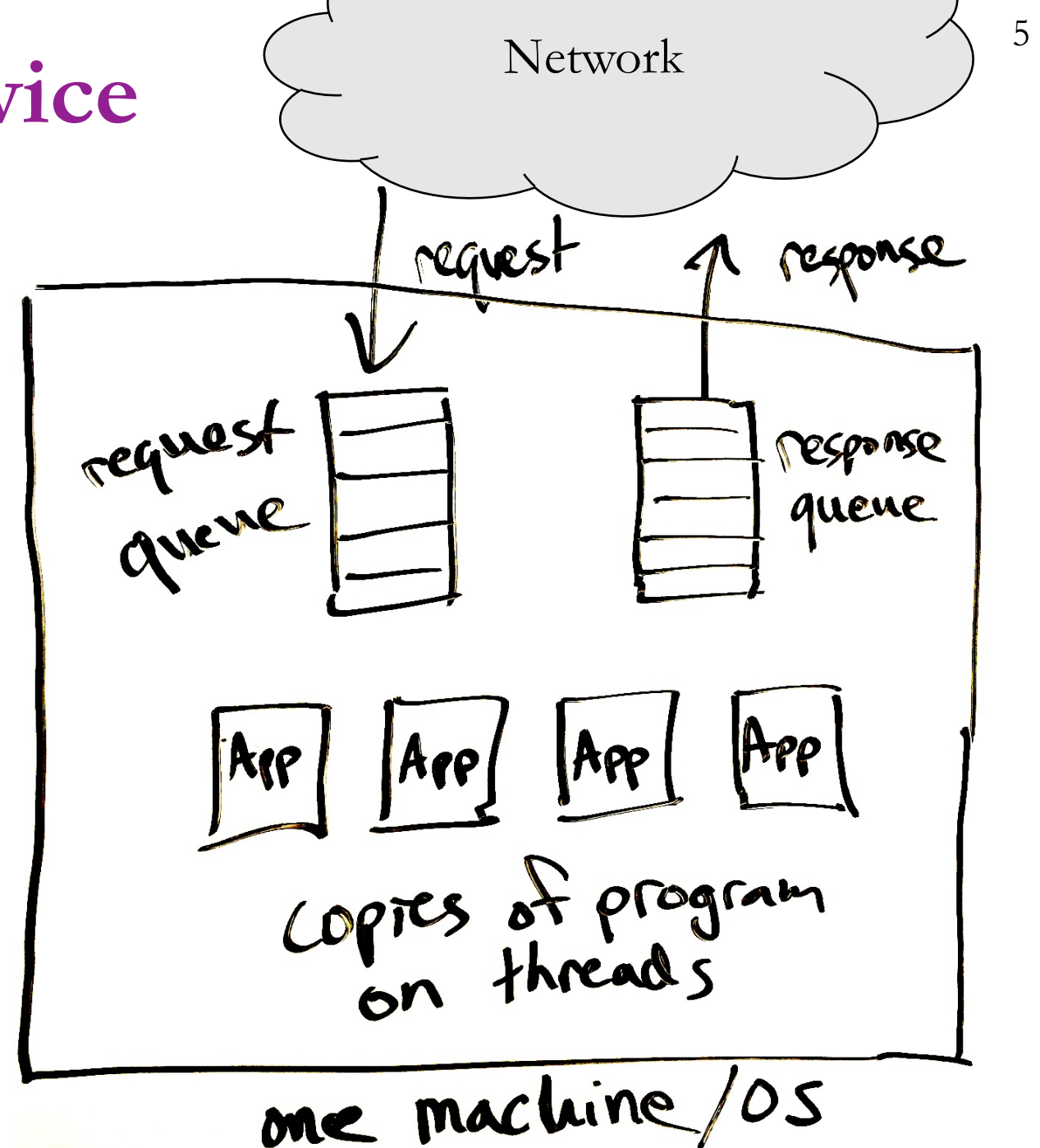


Basic Service Definition

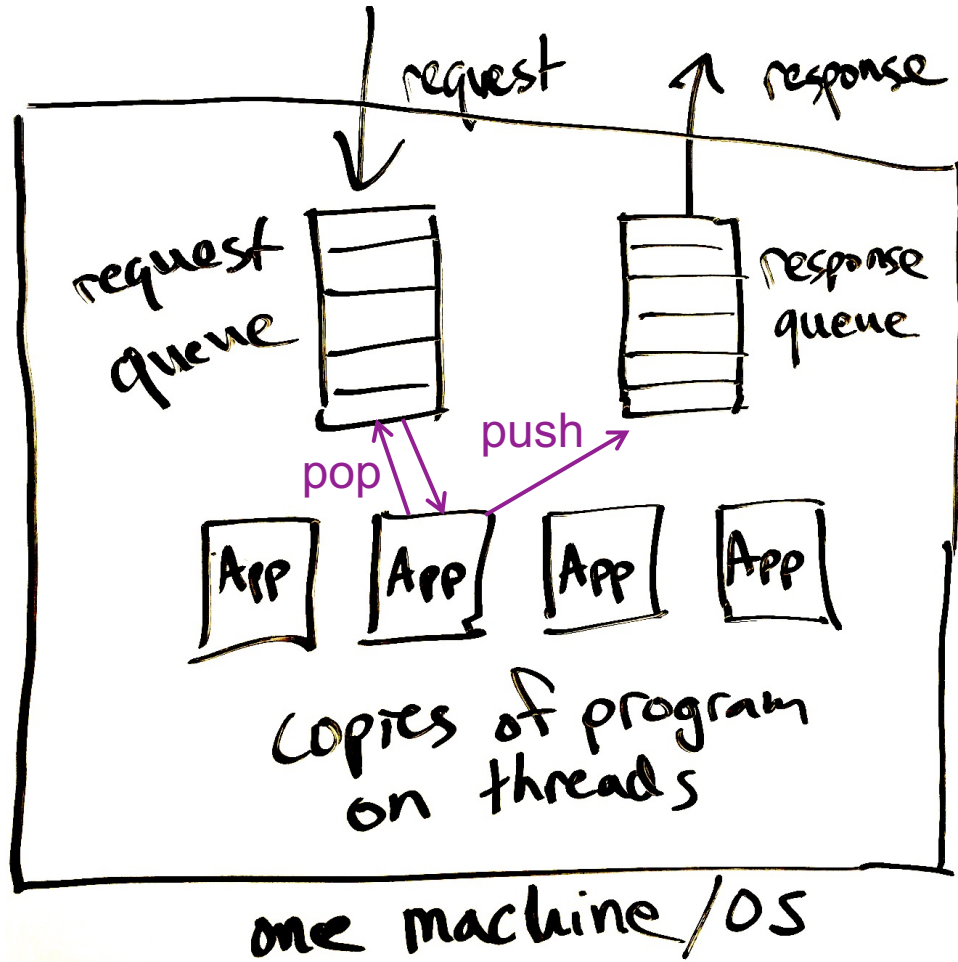
- In the theory of computation, a computer program (Turing Machine) takes a symbolic **input** and returns a symbolic **output**. Then it stops.
- Similarly, a computer **service** receives **requests** and returns a **response** for each request.
- However, a service can handle many concurrent, independent requests, which may be from different users.

From a Program to a Service

- A simple computer program can be translated into a service by:
 - Listening to requests that arrive from the **network**.
 - Running many copies of the program concurrently (using the OS features called *threads* or *processes*).
 - Using queues to store unhandled requests and unsent responses.
 - Queues allow competing threads to share a single network “socket” (one IP address and port).



Service thread

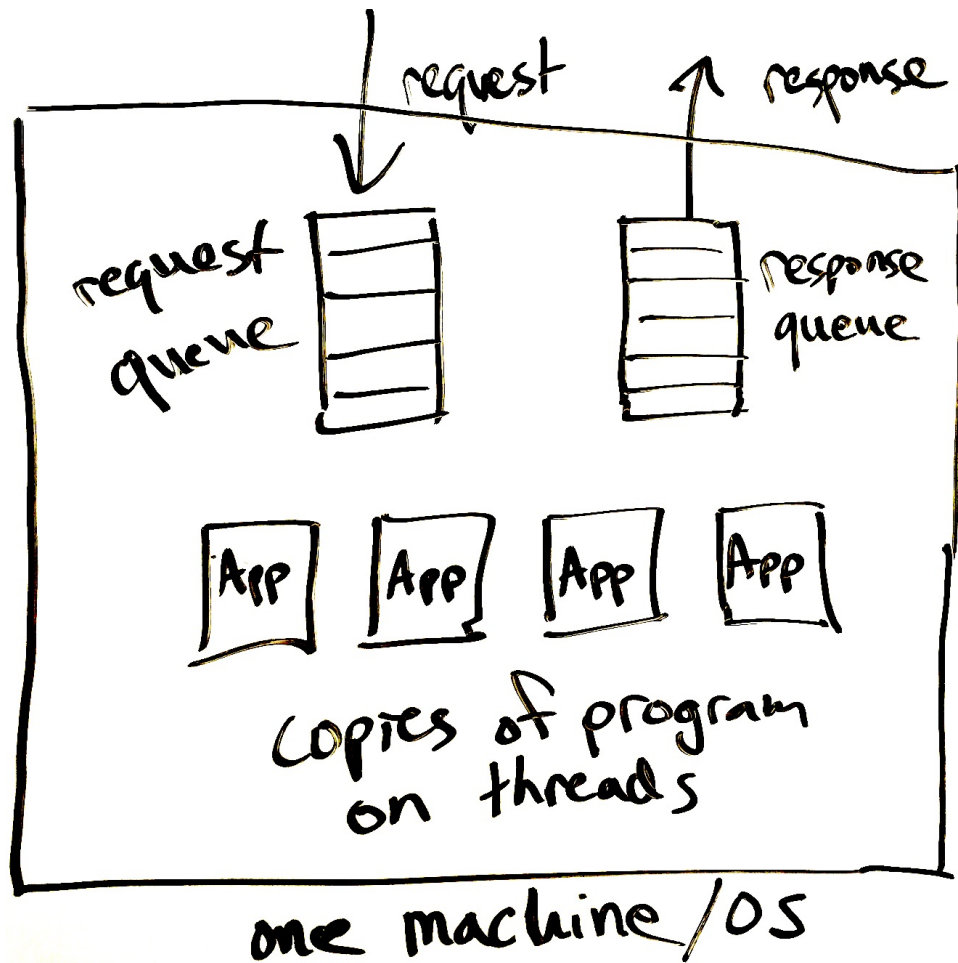


- The program on each thread runs an infinite loop:

```
while (true) {
    request = requestQ.pop();
    response = doWork(request);
    responseQ.push(response);
}
```

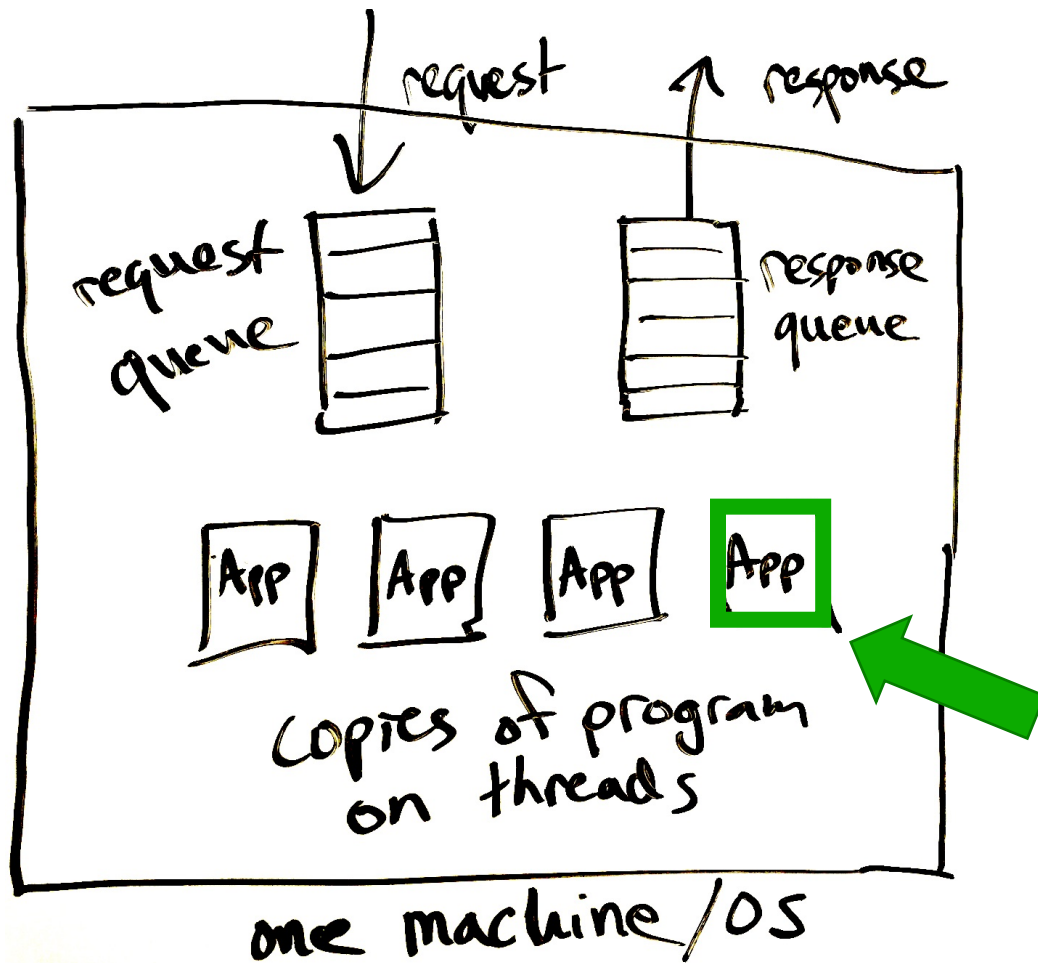
- `pop()` waits if the queue is empty.
- `push()` might wait if queue is too full.
- A thread that waits is also said to **block**.

Concurrency



- Many requests can be processed at the same time (concurrently).
- This allows many CPU cores to be used in parallel.
- The threaded design is also helpful even if there is only one CPU core, because the app may block to request data from disk or over the network.
 - This is called IO (input/output).
- While one thread is waiting for the IO to complete, another can use the CPU.

Web/HTTP server frameworks



- **Web/HTTP server** software provides this basic framework.
 - For example:
 - Java: Jetty, Tomcat
 - Python: Flask, Django
 - Javascript: Node.js
 - Static files: Apache httpd, Nginx
- Just plug in the app code.

Hyper Text Transport Protocol (HTTP)

- HTTP is a client-server data exchange protocol
- It was invented for web browsers to fetch pages from web servers
- **Request** specifies:
 - A human-readable header with: *URL*, *method*, (plus some optional headers)
 - An optional *body*, storing raw data (bytes).
- **Response** includes:
 - A human-readable header with *response code*, (plus some optional headers)
 - An optional *body*

Request:

GET /doc/test.html HTTP/1.1

Host: www.test101.com

Accept: image/gif, image/jpeg, */*

Accept-Language: en-us

Accept-Encoding: gzip, deflate

User-Agent: Mozilla/4.0

Content-Length: 35

bookId=12345&author=Tan+Ah+Teck

Request Line

Request Headers

Request
Message
Header

A blank line separates header & body

Request Message Body
(optional for GET)

Response:

HTTP/1.1 200 OK

Date: Sun, 08 Feb xxxx 01:11:12 GMT

Server: Apache/1.3.29 (Win32)

Last-Modified: Sat, 07 Feb xxxx

ETag: "0-23-4024c3a5"

Accept-Ranges: bytes

Content-Length: 35

Connection: close

Content-Type: text/html

<h1>My Home page</h1>

Status Line

Response Headers

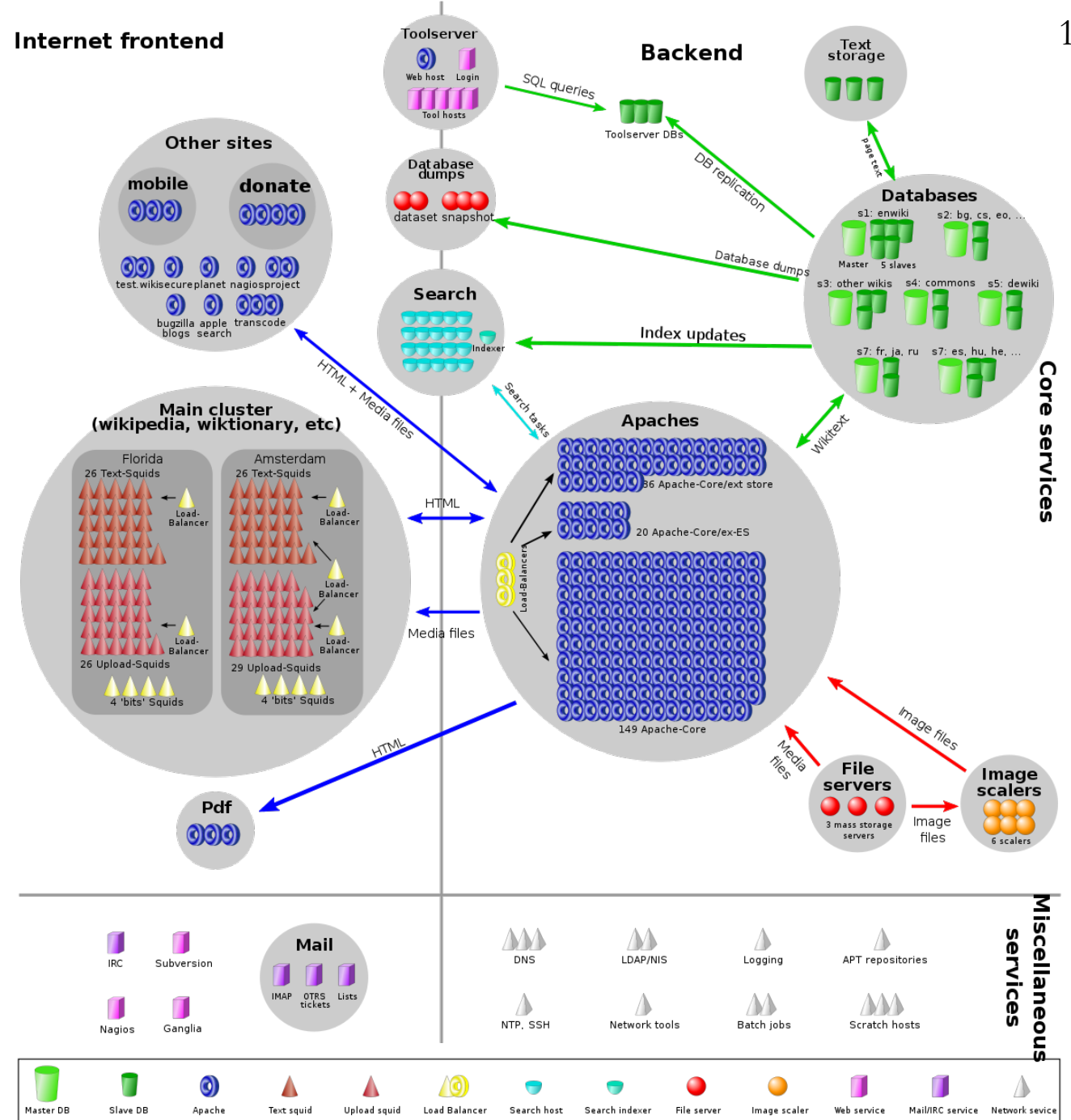
Response
Message
Header

A blank line separates header & body

Response Message Body

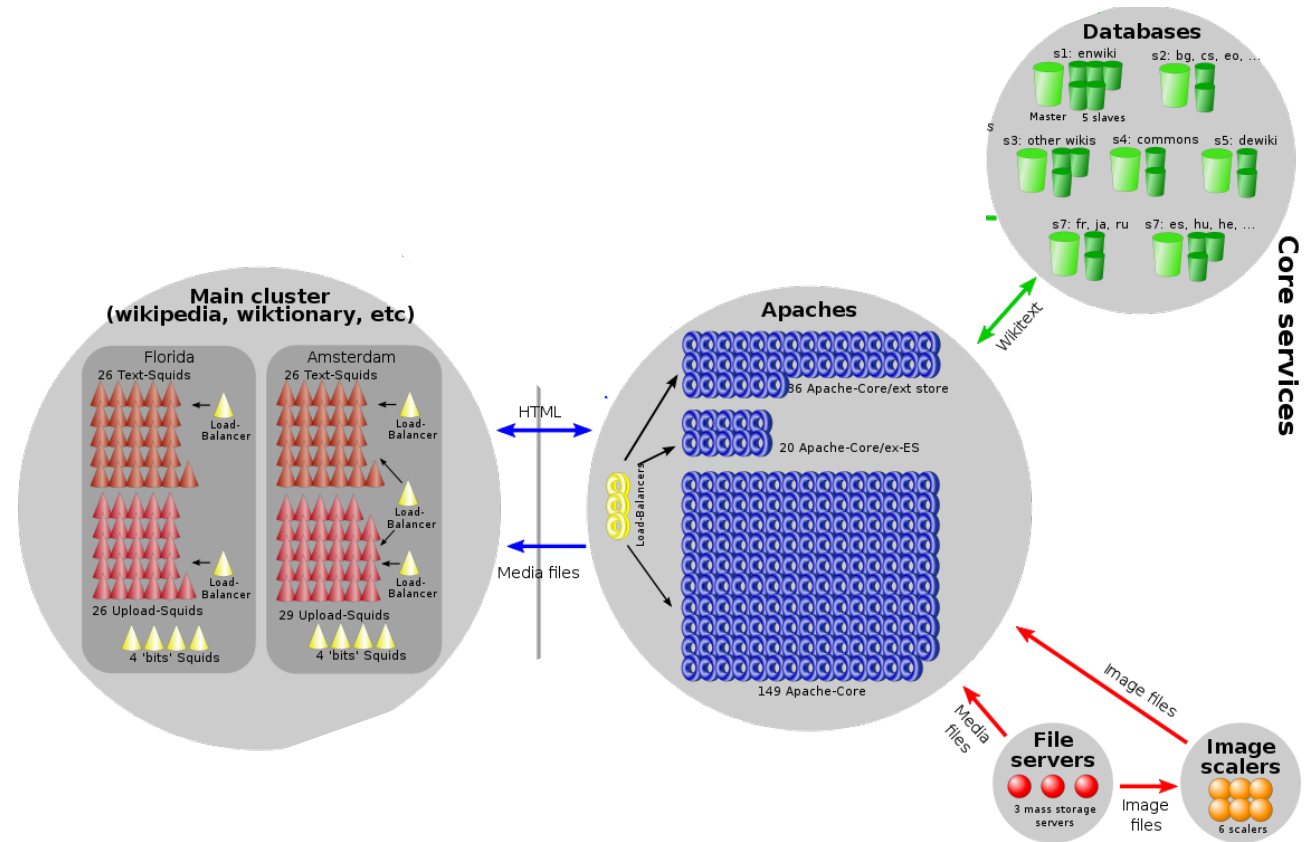
Wikipedia architecture

- Main application is MediaWiki
- 70% PHP, 30% JavaScript
- Databases are MariaDB (SQL)
- https://meta.m.wikimedia.org/wiki/Wikimedia_servers

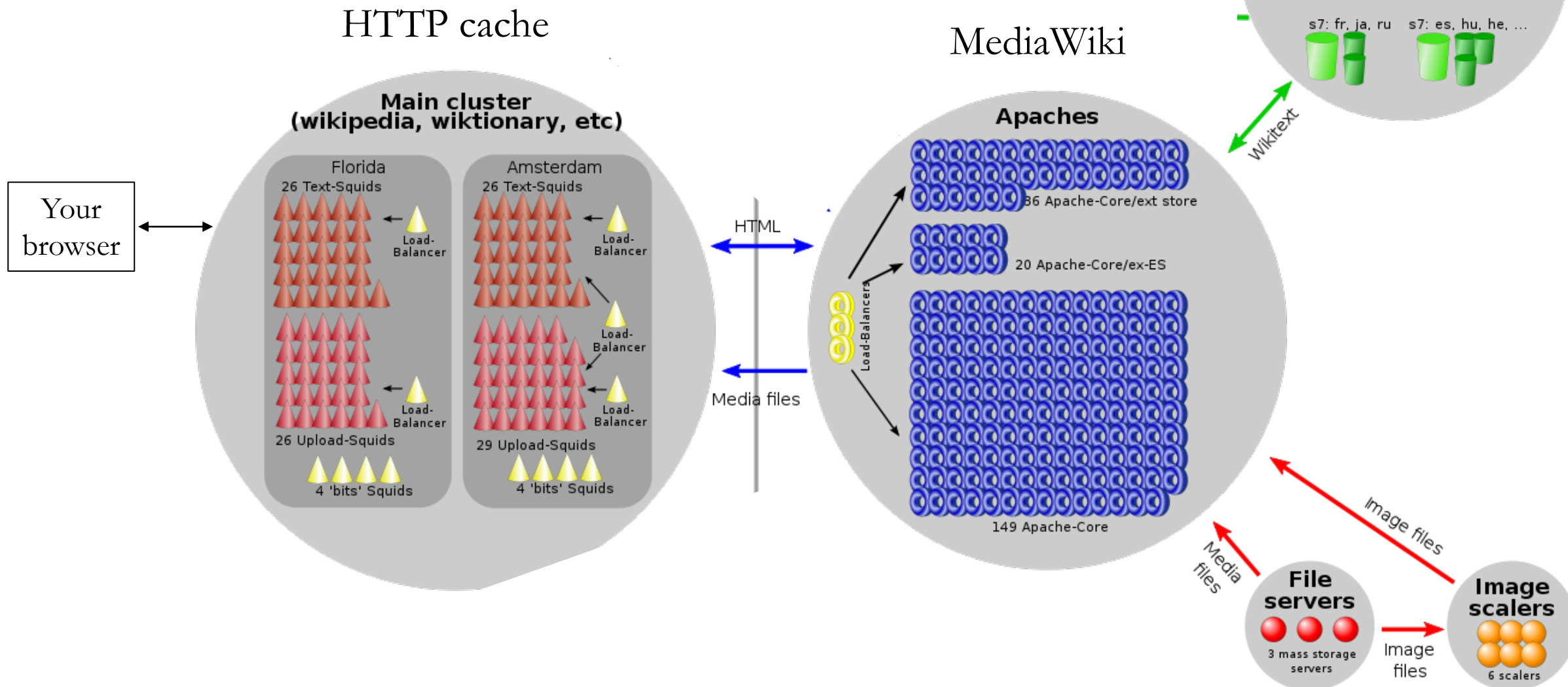


Key parts

- ▲ **Squid:** Caching HTTP proxy on frontend.
- **Apache httpd:** web servers running the main application (MediaWiki)
- **SQL databases** for wiki text, etc.
- **File servers** for images.

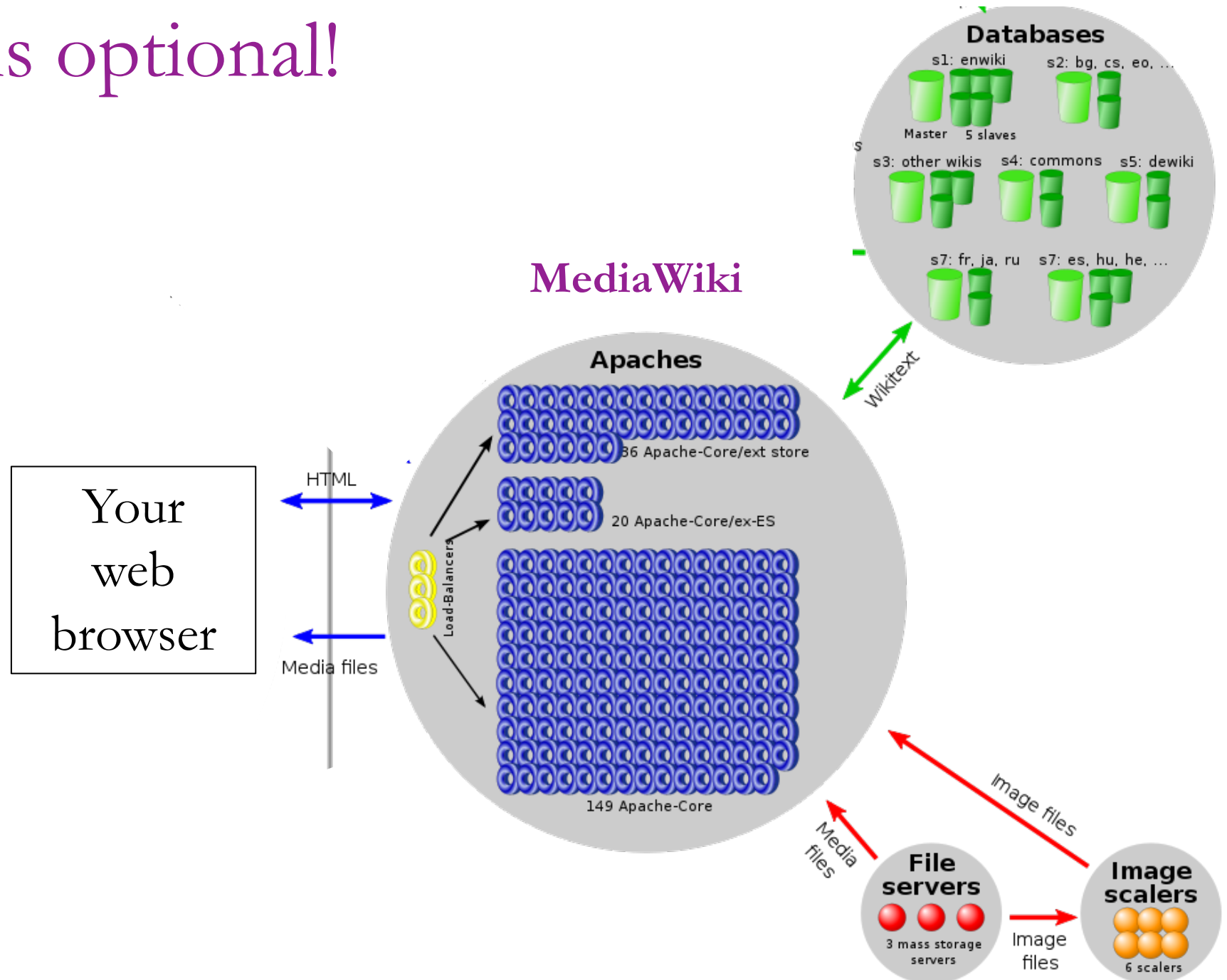


Key parts



Caching layer is optional!

- We'll come back to caching in the next lecture



MediaWiki application

- Runs PHP code in Apache Httpd web framework.
- **Input** request:
 - An HTTP request from the browser.
 - For example, **GET /wiki/Embioptera**
- **Output** response:
 - An HTTP response understandable to the browser.
 - Usually an **HTML document**, sometimes an image, etc.

MediaWiki's main task is to generate article HTML



The screenshot shows the Wikipedia article for **Embioptera**. At the top, there's a navigation bar with the user's name (Starzia), a bell icon, a monitor icon, and links for Talk, Sandbox, Preferences, Beta, Watchlist, Contributions, and Log out. Below this is a search bar and a navigation bar with links for Article, Talk, Read, Edit source, View history, and a star icon. A banner for "Wiki Loves Monuments" is displayed, encouraging users to photograph a monument to help Wikipedia. The main heading is "Embioptera", followed by the text "From Wikipedia, the free encyclopedia". The article text describes the order Embioptera, commonly known as webspinners or footspinners, as a small group of mostly tropical and subtropical insects. It mentions that the order has also been called Embiodea or Embiidina. The article states that more than 400 species in 11 families have been described, with the oldest known fossils dating back to the mid-Jurassic. It notes that species are very similar in appearance, having long, flexible bodies, short legs, and only males having wings. The article also mentions that webspinners are gregarious, living subsocially in galleries of fine silk. A sidebar on the right provides a temporal range (199–0 Ma) and a scientific classification table. The table lists the Kingdom as Animalia, Phylum as Arthropoda, Class as Insecta, Infraclass as Neoptera, and Order as Embioptera. A table of contents is also visible at the bottom left, listing "1 Name and etymology" and "2 Evolution".

WIKIPEDIA
The Free Encyclopedia

Main page
Contents
Featured content
Current events
Random article
Donate to Wikipedia
Wikipedia store

Interaction
Help
About Wikipedia
Community portal
Recent changes
Contact page

Tools
What links here
Related changes
Upload file
Special pages
Permanent link
Page information
Wikidata item
Cite this page

In other projects
Wikimedia Commons
Wikispecies

Article [Talk](#) [Read](#) [Edit source](#) [View history](#) [☆](#)

 Wiki Loves Monuments: Competition closes soon. Photograph a monument, help Wikipedia and win!
[Learn more](#)

Embioptera

From Wikipedia, the free encyclopedia

The **order Embioptera**, commonly known as **webspinners** or **footspinners**,^[2] are a small group of mostly **tropical** and **subtropical insects**, classified under the **subclass Pterygota**. The order has also been called **Embiodea** or **Embiidina**.^[3] More than 400 species in 11 families have been described, the oldest known fossils of the group being from the mid-**Jurassic**. Species are very similar in appearance, having long, flexible bodies, short legs, and only males having wings.

Webspinners are gregarious, living **subsocially** in galleries of fine **silk** which they spin from glands on their forelegs. Members of these colonies are often related females and their offspring; males do not feed and die soon after mating. Males of some species have wings and are able to disperse, whereas the females remain near where they were hatched. Newly-mated females may vacate the colony and found a new one nearby. Others may emerge to search for a new food source to which the galleries can be extended, but in general, the insects rarely venture from their galleries.

Contents [\[hide\]](#)

- [Name and etymology](#)
- [Evolution](#)

Embioptera
Temporal range: 199–0 Ma

PreЄ **Є** **O** **S** **D** **C** **P** **T** **J** **K** **PgN**

Jurassic – Recent



Adult winged male *Oligotoma saundersii*

Scientific classification 

Kingdom: **Animalia**
Phylum: **Arthropoda**
Class: **Insecta**
Infraclass: **Neoptera**
Order: **Embioptera**

How?

- Get corresponding wiki text from DB.
- Translate wiki text to HTML.
- Add wrapping content and banners.
- Add user-specific page header, based on cookies in request.

MediaWiki's main task is to generate article HTML



The screenshot shows the Wikipedia article for **Embioptera**. At the top, there is a banner for "Wiki Loves Monuments: Competition closes soon. Photograph a monument, help Wikipedia and win!" with a "Learn more" link. The article title "Embioptera" is prominently displayed. Below the title, it states "From Wikipedia, the free encyclopedia". The main text describes the order Embioptera, commonly known as **webspinners** or **footspinners**, as a small group of mostly **tropical** and **subtropical insects**, classified under the subclass **Pterygota**. It mentions that the order has also been called **Embiodea** or **Embiidina**. More than 400 species in 11 families have been described, with the oldest known fossils from the mid-Jurassic. Species are very similar in appearance, having long, flexible bodies, short legs, and only males having wings. A sidebar on the right provides additional information, including a temporal range from 199–0 Ma, a geological time scale bar, a photograph of an adult winged male *Oligotoma saundersii*, and a scientific classification table.

Embioptera

From Wikipedia, the free encyclopedia

The **order Embioptera**, commonly known as **webspinners** or **footspinners**,^[2] are a small group of mostly **tropical** and **subtropical insects**, classified under the subclass **Pterygota**. The order has also been called **Embiodea** or **Embiidina**.^[3] More than 400 species in 11 families have been described, the oldest known fossils of the group being from the mid-Jurassic. Species are very similar in appearance, having long, flexible bodies, short legs, and only males having wings.

Webspinners are gregarious, living **subsocially** in galleries of fine **silk** which they spin from glands on their forelegs. Members of these colonies are often related females and their offspring; males do not feed and die soon after mating. Males of some species have wings and are able to disperse, whereas the females remain near where they were hatched. Newly-mated females may vacate the colony and found a new one nearby. Others may emerge to search for a new food source to which the galleries can be extended, but in general, the insects rarely venture from their galleries.

Contents ^[hide]

- Name and etymology**
- Evolution**

Embioptera
Temporal range: 199–0 Ma

PreЄЄЄSDCPTJKPgN

Jurassic – Recent



Adult winged male *Oligotoma saundersii*

Scientific classification

Kingdom:	Animalia
Phylum:	Arthropoda
Class:	Insecta
Infraclass:	Neoptera
Order:	Embioptera

How?

- Get corresponding wiki text from DB.
- Translate wiki text to HTML.
- Add wrapping content and banners.
- Add user-specific page header, based on cookies in request.

MediaWiki's main task is to generate article HTML

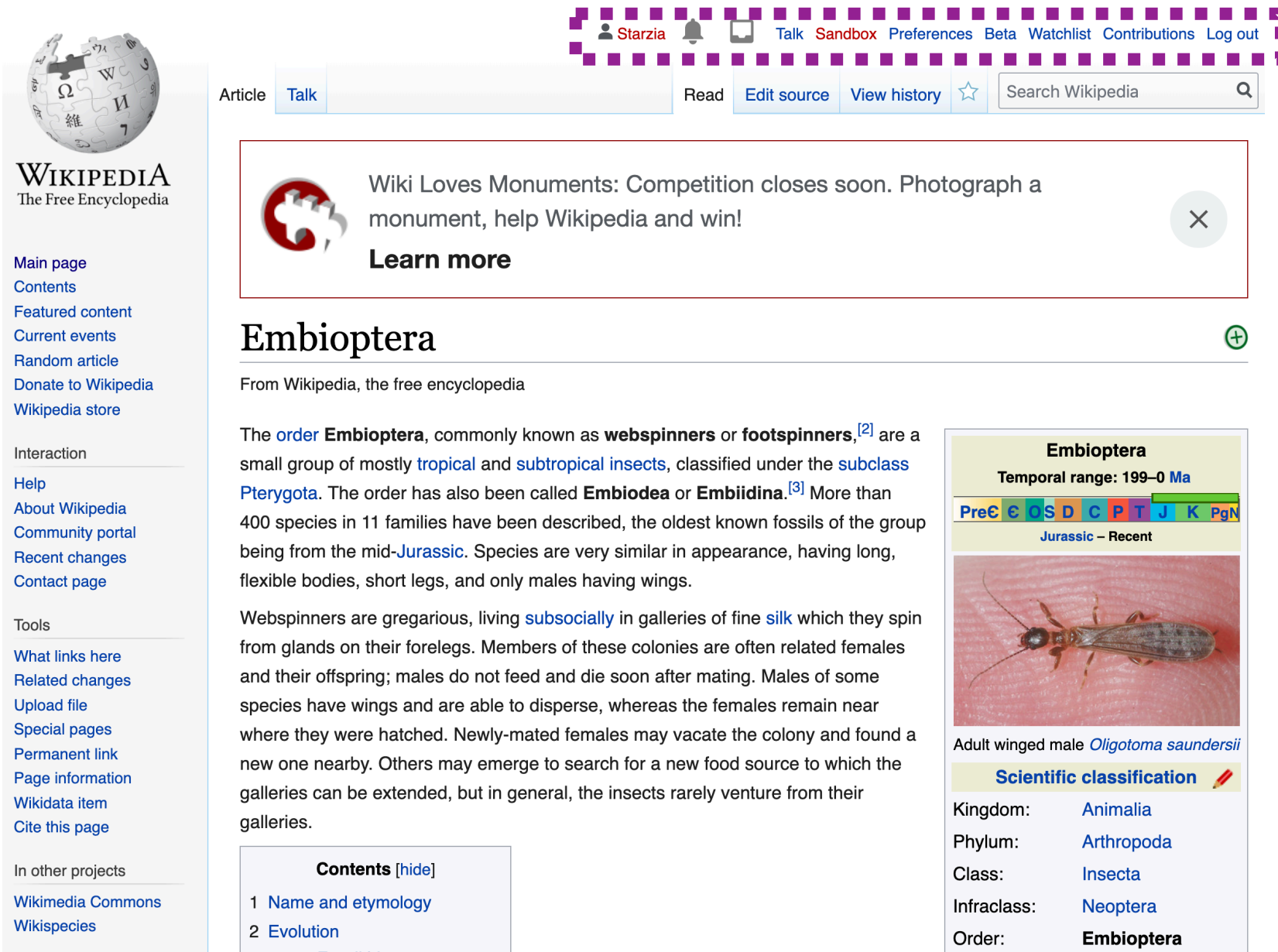


The screenshot shows a Wikipedia article page for "Embioptera". The page is framed by a purple dashed border. At the top, there is a navigation bar with links for "Talk", "Sandbox", "Preferences", "Beta", "Watchlist", "Contributions", and "Log out". Below this is a search bar and a "Search Wikipedia" button. The main content area features a banner for "Wiki Loves Monuments" with a "Learn more" link. The article title "Embioptera" is prominently displayed, followed by a sub-header "From Wikipedia, the free encyclopedia". The main text describes the order Embioptera, commonly known as webspinners or footspinners, and their characteristics. A sidebar on the right provides a "Scientific classification" table, listing the Kingdom (Animalia), Phylum (Arthropoda), Class (Insecta), Infraclass (Neoptera), and Order (Embioptera). The sidebar also includes a "Contents" section with links to "Name and etymology" and "Evolution".

How?

- Get corresponding wiki text from DB.
- Translate wiki text to HTML.
- Add wrapping content and banners.
- Add user-specific page header, based on cookies in request.

MediaWiki's main task is to generate article HTML



The screenshot shows a Wikipedia article for **Embioptera**. The top navigation bar includes links for [Starzia](#), [Talk](#), [Sandbox](#), [Preferences](#), [Beta](#), [Watchlist](#), [Contributions](#), and [Log out](#). The article title is **Embioptera**, and the subtitle is "From Wikipedia, the free encyclopedia". The main text describes the order Embioptera, commonly known as **webspinners** or **footspinners**, as a small group of mostly **tropical** and **subtropical insects**, classified under the subclass **Pterygota**. It mentions that the order has also been called **Embiodea** or **Embiidina**. The text states that more than 400 species in 11 families have been described, with the oldest known fossils of the group being from the mid-Jurassic. Species are very similar in appearance, having long, flexible bodies, short legs, and only males having wings. A sidebar on the right provides a **Scientific classification** for the order, listing the Kingdom as **Animalia**, Phylum as **Arthropoda**, Class as **Insecta**, Infraclass as **Neoptera**, and Order as **Embioptera**. The sidebar also includes a **Temporal range** of 199–0 Ma and a **Geological time scale** diagram. The main content area includes a **Contents** table of contents with links to [Name and etymology](#) and [Evolution](#).

How?

- Get corresponding wiki text from DB.
- Translate wiki text to HTML.
- Add wrapping content and banners.
- Add user-specific page header, based on cookies in request.

Code walkthroughs

- (Bitbucket links will not work for students, refer to video)
- <https://stevetarzia.com/>
 - <https://bitbucket.org/starzia/www.stevetarzia.com/src/master/index.php?mode=view&spa=0&at=master&fileviewer=file-view-default>
- <https://stevetarzia.com/xmas/>
 - <https://bitbucket.org/starzia/www.stevetarzia.com/src/master/xmas/index.php?mode=view&spa=0&at=master&fileviewer=file-view-default>
- <https://stevetarzia.com/listen/>
 - <https://bitbucket.org/starzia/www.stevetarzia.com/src/master/listen/index.php?mode=view&spa=0&at=master&fileviewer=file-view-default>
- <https://gunmemorial.org/donate>
 - <https://bitbucket.org/starzia/gunmemorial/src/master/victim-portal/src/main/webapp/donate.jsp?mode=view&spa=0&at=master&fileviewer=file-view-default>
- <https://gunmemorial.org/sitemap.txt>
- <https://gunmemorial.org/sitemap.txt?startYear=2020&endYear=2020>
 - <https://bitbucket.org/starzia/gunmemorial/src/master/victim-portal/src/main/java/org/gunmemorial/web/servlet/SiteMapServlet.java?mode=view&spa=0&at=master&fileviewer=file-view-default>

Recap

- Showed that web server frameworks let you translate a simple program into a multi-threaded service with concurrency.
- Introduced HTTP as the most common type of service.
 - Client **requests** a document (specified in path/url)
 - Server sends document in the **response**.
- High-level overview of Wikipedia's architecture.

Open questions:

- What's caching and why is it possible?
- What's the purpose of the database and how to make it scalable?