

# **Distributed Web Crawler**

## **COMP 599 - Graduate Seminar, Spring 2019**

Rihan Pereira, MSCS

Department of Computer Science  
California State University, Channel Islands

April 9, 2019

## Web Crawling

The vision

Distributed web crawler

Distributed Indexing

Presentation tier

Thoughts on infrastructure setup

Motivation for this project

References

Example

Single node crawler (block diagram)

Indexer

Challenges faced when crawling

# Web Crawling

---

type something to see how the font looks

## Web Crawling

The vision

Distributed web crawler

Distributed Indexing

Presentation tier

Thoughts on infrastructure setup

Motivation for this project

References

## Example

Single node crawler (block diagram)

Indexer

Challenges faced when crawling

# Example

## Web Crawling

The vision

Distributed web crawler

Distributed Indexing

Presentation tier

Thoughts on infrastructure setup

Motivation for this project

References

Example

**Single node crawler (block diagram)**

Indexer

Challenges faced when crawling

# Single node crawler (block diagram)

---

## Web Crawling

The vision

Distributed web crawler

Distributed Indexing

Presentation tier

Thoughts on infrastructure setup

Motivation for this project

References

Example

Single node crawler (block diagram)

## Indexer

Challenges faced when crawling

# Indexer

## Web Crawling

The vision

Distributed web crawler

Distributed Indexing

Presentation tier

Thoughts on infrastructure setup

Motivation for this project

References

Example

Single node crawler (block diagram)

Indexer

**Challenges faced when crawling**

# Challenges faced when crawling

---

Web Crawling

**The vision**

Distributed web crawler

Distributed Indexing

Presentation tier

Thoughts on infrastructure setup

Motivation for this project

References

# The vision

---

Web Crawling

The vision

**Distributed web crawler**

Distributed Indexing

Presentation tier

Thoughts on infrastructure setup

Motivation for this project

References

# Distributed web crawler

---



Web Crawling

The vision

Distributed web crawler

**Distributed Indexing**

Presentation tier

Thoughts on infrastructure setup

Motivation for this project

References

# Distributed Indexing

---

Web Crawling  
The vision  
Distributed web crawler  
Distributed Indexing  
**Presentation tier**  
Thoughts on infrastructure setup  
Motivation for this project  
References

# Presentation tier

---

# Thoughts on infrastructure setup

---

- Web Crawling
  - The vision
  - Distributed web crawler
  - Distributed Indexing
  - Presentation tier
- Thoughts on infrastructure setup
- Motivation for this project**
- References

---



Martin Kleppmann,  
*Designing Data Intensive applications, 2017*



James Hamilton,  
*On Designing and Deploying Internet-Scale Services, 2007.*



Twitter,  
*Automatic Management of Partitioned, Replicated Search Services, 2011.*



<http://blog.acolyer.org/>,  
*The morning Paper.*



<http://aosabook.org>,  
*The architecture of open source applications.*



<http://book.mixu.net/distsys>,  
*Distributed Systems: fun and profit.*

Thank you! Questions ?