Back To Module Home

Basic Building Blocks for Modern System Design
0% completed

Introduction to Building Blocks

Domain Name System

Load balancers

Cache

Databases

Key-value Store

Content Delivery Network (CDN)

Sequencer

Distributed Monitoring

Distributed Cache

Distributed Messaging Queue

Pub-sub

Rate Limiter

Blob Store

Distributed Search

System Design: The Distributed Search

Requirements of a Distributed Search System's Design

Indexing in a Distributed Search

Design of a Distributed Search

Scaling Search and Indexing

Evaluation of a Distributed Search's Design

Distributed Task Scheduler

Sharded Counters

Conclusion

Mark Module as Completed

System Design: The Distributed Search

Learn about how a search system works and understand our high-level plan for designing a distributed search system.

We'll cover the following

- Why do we need a search system?
- What is a search system?
- How will we design a distributed search system?

Why do we need a search system?#

Nowadays, we see a search bar on almost every website. We use that search bar to pick out relevant content from the enormous amount of content available on that website. A search bar enables us to quickly find what we're looking for. For example, there are plenty of courses present on the Educative website. If we didn't have a search feature, users would have to scroll through many pages and read the name of each course to find the one they're looking for.

Let's take another example. There are billions of videos uploaded and stored on YouTube. Imagine if YouTube didn't provide us with a search bar. How would we find a specific video among the millions of videos that have been posted on YouTube over the years? It would take months to navigate through all of those videos and find the one we need. Users find it challenging to find what they're looking for simply by scrolling around.

Search bar

Search engines are an even bigger example. We have billions of websites on the Internet. Each website has many web pages and there is plenty of content on each of these web pages. With so much content, the Internet would practically be useless without search engines, and users would end up lost in a sea of irrelevant data. Search engines are, essentially, filters for the massive amount of data available. They let users quickly obtain information that is of true interest without having to sift through too many unnecessary web pages.

Behind every search bar, there is a search system.

What is a search system?#

A **search system** is a system that takes some text input, a search query, from the user and returns the relevant content in a few seconds or less. There are three main components of a search system, namely:

- A **crawler**, which fetches content and creates <u>documents</u>.
- An **indexer**, which builds a searchable index.

- A **searcher**, which responds to search queries by running the search query on the *index* created by the *indexer*.

The three phases of a search system

Note: We have a separate chapter dedicated to the explanation of the crawler component. In this chapter, we'll focus on indexing.

How will we design a distributed search system?#

We divided the design of a distributed search system into five lessons:

1. **Requirements:** In this lesson, we list the functional and non-functional requirements of a distributed search system. We also estimate our system's resources, such as servers, storage, and the bandwidth needed to serve a number of queries.

2. **Indexing:** This lesson provides us with background knowledge on the process of indexing with the help of an example. After discussing indexing, we also look into a centralized architecture of distributed search systems.

3. **Initial design:** This lesson consists of the high-level design of our system, its API, and the details of the indexing and searching process.

4. **Final design:** In this lesson, we evaluate our previous design and revamp it to make it more scalable.

5. **Evaluation:** This lesson explains how our designed distributed search system fulfills its requirements.

Let's start by understanding the requirements of designing a distributed search system.

Back

Quiz on the Blob Store's Design

Next

Requirements of a Distributed Search…

Mark as Completed

Report an Issue