

系统简介

在作业三中，我们将文件进行分块，并存储到 7 个服务器上进行测试。在分块前，我们对数据进行了一些预处理，使得服务器能够承载文件块大小，将与题目无关的数据进行了剔除。我们在分块时采取了尽量平均的做法，分割的地点当前块最后一个 xml 项结束的位置。在将分割块和备份块放到服务器后，我们制作了有关数据的 hash 表，根据服务器上所拥有的源文件块进行放入，以实现 hash 搜索。

索引文件示例：

```
<YearlyAuthorFrequency>
  <Year value="1976">
    <Author frequency="1" name="Jacques Maurin"/>
    <Author frequency="1" name="Chi Hau Chen"/>
    <!-- 更多作者 -->
  </Year>
  <!-- 更多年份 -->
</YearlyAuthorFrequency>
```

这种格式和源格式比较起来，清晰地将数据按年份和作者组织起来，易于解析和数据检索。通过年份和作者名作为关键字，可以快速定位到特定作者在特定年份的发表频次。

系统功能

负载均衡：

在切割好文件之后，一共有 7 块源文件。为了实现在 $N/2$ 的情况下仍然能够正确查询，我们一共需要对每块源文件创建三块副本。所有一共有 28 块文件需要上传到服务器。为了实现负载均衡，我们实现了一个分配文件的算法，来将副本和源文件上传到服务器且保证每个服务器上的内容大小基本一致。

Ps（此处取 7 块和 21 块只是为了验证结果方便，实际上可以分成任意多个源文件，比如 100 个源文件和 300 个副本上传，更能体现算法优势）

示例：

```
Welcome to Alibaba Cloud Elastic Compute Service !

Last login: Sun Jan  7 16:48:26 2024 from 111.187.11.32
root@iZbp114ud6h3jdolobm9u3Z:~# ls
ds_application  fileWare  hash  snap  test-docker.sh
root@iZbp114ud6h3jdolobm9u3Z:~#

Last login: Sun Jan  7 16:48:26 2024 from 111.187.11.32
root@iZbp114ud6h3jdolobm9u3Z:~# ls
ds_application  fileWare  hash  snap  test-docker.sh
root@iZbp114ud6h3jdolobm9u3Z:~# cd fileWare/
root@iZbp114ud6h3jdolobm9u3Z:~/fileWare# ls
1_1.xml  2_0.xml  5_0.xml  6_0.xml
root@iZbp114ud6h3jdolobm9u3Z:~/fileWare#
```

结果查询：

因为第四题我们制作了一个简单的前端来进行控制，所以第三题我们同样使用了 vue3 来作为一个查询的界面。在界面上可以模拟服务器的关闭与打开，同时进行全局扫描和索引扫描。我们的示例查询放在附录当中。

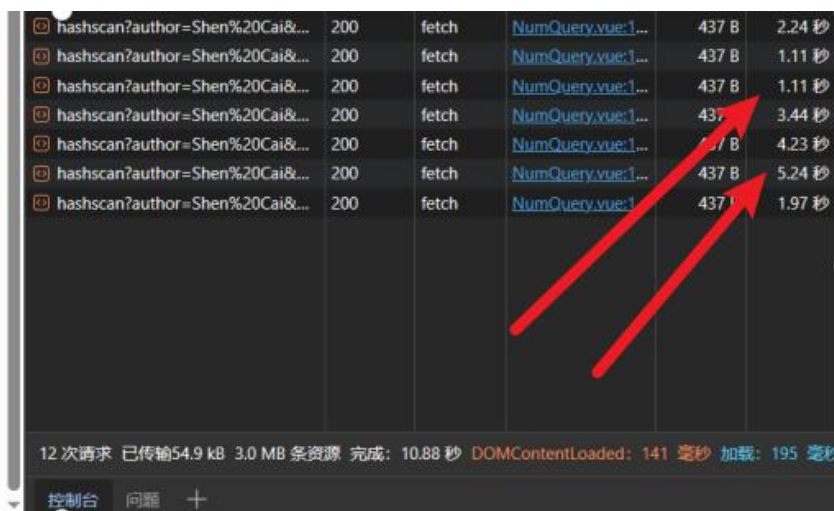


查询容错：

因为我们使用了副本块，所以支持查询容错，能够在 $N/2$ （在本题目中是 3）台服务关闭的情况下，仍然能查询到正确结果。

结果演示

查询演示：在 hash 搜索情况下，最快情况为 1.11 秒，最慢情况为 5.24 秒，这跟服务器的性能有关。因为我们有两台是阿里云的免费服务器，所以性能会低一点。



在直接搜索的情况下，查询结果为 20.898 秒，最快的为 7.89 秒，也与服务器的水平有关。

关闭DS7

Shen Cai

2017

2018

全局扫描查询

哈希索引查询

查询结果: 4

查询耗时: 20.898s

ws	101	websocket	chunk-vendors.js...	0 B	挂起
directscan?author=Shen%20Cai&...	200	fetch	NumQuery.vue:1...	440 B	8.11 秒
directscan?author=Shen%20Cai&...	200	fetch	NumQuery.vue:1...	440 B	8.20 秒
directscan?author=Shen%20Cai&...	200	fetch	NumQuery.vue:1...	440 B	7.89 秒
directscan?author=Shen%20Cai&...	200	fetch	NumQuery.vue:1...	440 B	20.89 秒
directscan?author=Shen%20Cai&...	200	fetch	NumQuery.vue:1...	439 B	8.65 秒
directscan?author=Shen%20Cai&...	200	fetch	NumQuery.vue:1...	440 B	9.07 秒
directscan?author=Shen%20Cai&...	200	fetch	NumQuery.vue:1...	440 B	8.25 秒

12 次请求 已传输3.9 kB 3.0 MB 条资源 完成: 30.99 秒 DOMContentLoaded: 433 毫秒 加载: 497 毫秒

控制台 问题 +

容错查询：未关闭服务器的情况

关闭DS7

Peter A. Ivey

2000

2005

全局扫描查询

哈希索引查询

查询结果: 5

查询耗时: 19.339s

ws	101	websocket	chunk-vendors.js...	0 B	挂起
directscan?author=Peter%20A.%2...	200	fetch	NumQuery.vue:1...	439 B	8.17 秒
directscan?author=Peter%20A.%2...	200	fetch	NumQuery.vue:1...	440 B	8.14 秒
directscan?author=Peter%20A.%2...	200	fetch	NumQuery.vue:1...	439 B	19.33 秒
directscan?author=Peter%20A.%2...	200	fetch	NumQuery.vue:1...	439 B	8.33 秒
directscan?author=Peter%20A.%2...	200	fetch	NumQuery.vue:1...	440 B	8.03 秒
directscan?author=Peter%20A.%2...	200	fetch	NumQuery.vue:1...	439 B	7.79 秒
directscan?author=Peter%20A.%2...	200	fetch	NumQuery.vue:1...	439 B	8.98 秒

12 次请求 已传输3.9 kB 3.0 MB 条资源 完成: 25.72 秒 DOMContentLoaded: 434 毫秒 加载: 492 毫秒

控制台 问题 +

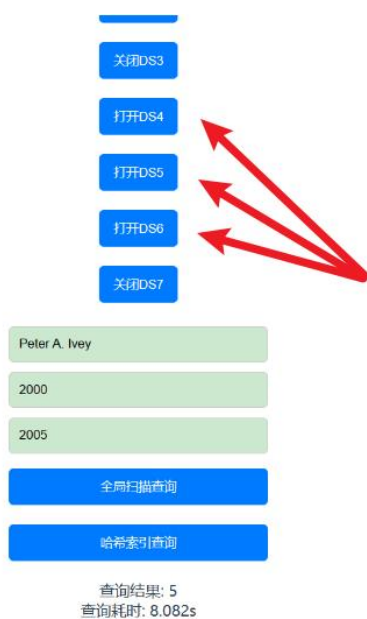
关闭服务器的情况



名称	状态	类型	发起程序	大小	时间	备注
localhost	304	document	其他	201 B	2 毫秒	
chunk-vendors.js	304	script	浏览器	204 B	7 毫秒	
app.js	304	script	浏览器	203 B	9 毫秒	
favicon.ico	304	vnd.micro...	其他	202 B	3 毫秒	
ws	101	websocket	chunk-vendors.js...	0 B	挂起	
off	200	fetch	NumQuery.vue:82	376 B	16 毫秒	
off	200	fetch	NumQuery.vue:82	374 B	15 毫秒	
off	200	fetch	NumQuery.vue:82	378 B	12 毫秒	
directscan?author=Peter%20A.%2...	200	fetch	NumQuery.vue:1...	439 B	7.98 秒	
directscan?author=Peter%20A.%2...	200	fetch	NumQuery.vue:1...	440 B	8.08 秒	
directscan?author=Peter%20A.%2...	200	fetch	NumQuery.vue:1...	439 B	7.98 秒	
directscan?author=Peter%20A.%2...	200	fetch	NumQuery.vue:1...	440 B	7.74 秒	

12 次请求 已传输 3.7 kB 3.0 MB 各资源 完成: 21.03 秒 DOMContentLoaded: 139 毫秒 加载: 192 毫秒

另一个案例：直接搜索的情况下，用时比哈希搜索明显长一些（直接搜索 8 秒左右，哈希搜索 3 秒左右）



名称	状态	类型	发起程序	大小	时间	备注
localhost	304	document	其他	201 B	2 毫秒	
chunk-vendors.js	304	script	浏览器	204 B	7 毫秒	
app.js	304	script	浏览器	203 B	9 毫秒	
favicon.ico	304	vnd.micro...	其他	202 B	3 毫秒	
ws	101	websocket	chunk-vendors.js...	0 B	挂起	
off	200	fetch	NumQuery.vue:82	376 B	16 毫秒	
off	200	fetch	NumQuery.vue:82	374 B	15 毫秒	
off	200	fetch	NumQuery.vue:82	378 B	12 毫秒	
directscan?author=Peter%20A.%2...	200	fetch	NumQuery.vue:1...	439 B	7.98 秒	
directscan?author=Peter%20A.%2...	200	fetch	NumQuery.vue:1...	440 B	8.08 秒	
directscan?author=Peter%20A.%2...	200	fetch	NumQuery.vue:1...	439 B	7.98 秒	
directscan?author=Peter%20A.%2...	200	fetch	NumQuery.vue:1...	440 B	7.74 秒	

12 次请求 已传输 3.7 kB 3.0 MB 各资源 完成: 21.03 秒 DOMContentLoaded: 139 毫秒 加载: 192 毫秒

打开DS5

打开DS6

打开DS7

Peter A. Ivey

2000

2005

全局扫描查询

哈希索引查询

查询结果: 5

查询耗时: 3.602s

名称	状态	类型	发起程序	大小	时间
localhost	304	document	其他	201 B	2 毫秒
chunk-vendors.js	304	script	(索引)8	204 B	5 毫秒
app.js	304	script	(索引)8	203 B	7 毫秒
favicon.ico	304	vnd.micr...	其他	202 B	2 毫秒
ws	101	websocket	chunk-vendors.js...	0 B	挂起
off	200	fetch	NumQuery.vue:82	374 B	16 毫秒
off	200	fetch	NumQuery.vue:82	378 B	11 毫秒
off	200	fetch	NumQuery.vue:82	372 B	37 毫秒
hashscan?author=Peter%20A.%2...	200	fetch	NumQuery.vue:1...	437 B	911 毫秒
hashscan?author=Peter%20A.%2...	200	fetch	NumQuery.vue:1...	437 B	930 毫秒
hashscan?author=Peter%20A.%2...	200	fetch	NumQuery.vue:1...	437 B	3.60 秒
hashscan?author=Peter%20A.%2...	200	fetch	NumQuery.vue:1...	437 B	1.84 秒

12 次请求 已传输 3.7 kB 3.0 MB 条资源 完成: 14.68 秒 DOMContentLoaded: 135 毫秒 加载: 199 毫秒

控制台 问题 +