Final Project: Technology Articles Search Engine Built on Hadoop

Stu.Name: Kangning Yuan   **Stu.ID:** MC451362

0. Introduction

0.1 Directory Structure

```
search_engine/
├── app.py
├── input.json
├── inverted_index.json
├── search_index.py
├── templates/
│   ├── index.html
│   └── results.html
└── mapreduce/
    ├── mapper.py
    ├── reducer.py
    └── run.sh
```

- **input.json**
    - This is our database stored in JSON format, containing information such as article ID, link, author, and the original text.
    - The database can be obtained through public datasets or web scraping. In this project, we referenced the open-source [Webz.io News Dataset Repository](#), integrating a portion of technology-related news articles into one `input.json` file.
- **inverted_index.json**
    - This is our index file built based on the inverted index.
- **search_index.py**
    - The search functionality API, implemented in Python.
- **app.py**
    - The search engine, implemented using Python and the Flask library.
- **templates/**
    - Templates for the Web UI interface, using HTML.
- **mapreduce/**
    - Contains `mapper.py` and `reducer.py`, used to generate the inverted index when dealing with a large database using the MapReduce framework.
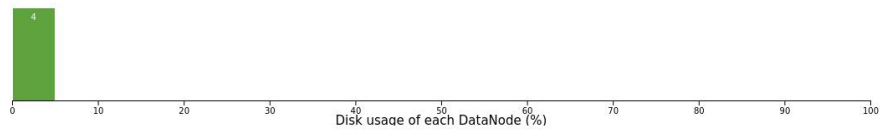
1. Environment Setup

Our project runs on a Hadoop cluster consisting of four Ubuntu nodes (node1 as the master and node2~4 as the slaves).
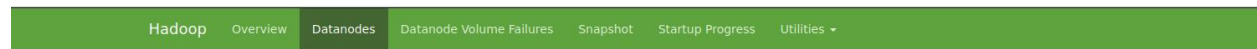
Datanode Information

In service | Down | Decommissioning | Decommissioned | Decommissioned & dead

Entering Maintenance | In Maintenance | In Maintenance & dead

Datanode usage histogram

Disk usage of each DataNode (%)

In operation

DataNode State: All   Show 25 entries   Search:

Node	Http Address	Last contact	Last Block Report	Used	Non DFS Used	Capacity	Blocks	Block pool used	Version
✔ /default-rack/node3:9866 (192.168.100.135:9866)	http://node3:9864	2s	262m	51.97 MB	10.45 GB	14.17 GB	69	51.97 MB (0.36%)	3.3.6
✔ /default-rack/node1:9866 (192.168.100.133:9866)	http://node1:9864	0s	115m	52.64 MB	10.85 GB	28.87 GB	90	52.64 MB (0.18%)	3.3.6
✔ /default-rack/node4:9866 (192.168.100.136:9866)	http://node4:9864	2s	302m	51.8 MB	9.91 GB	14.17 GB	73	51.8 MB (0.36%)	3.3.6
✔ /default-rack/node2:9866 (192.168.100.134:9866)	http://node2:9864	2s	144m	6.53 MB	10.72 GB	19.02 GB	79	6.53 MB (0.03%)	3.3.6

Showing 1 to 4 of 4 entries   Previous 1 Next

192.168.100.133:8088/cluster/nodes   Logged in as: dr.who

Nodes of the cluster

Cluster Metrics

Apps Submitted	Apps Pending	Apps Running	Apps Completed	Containers Running	Used Resources	Total Resources	Reserved Resources	Physical Mem Used %	Physical VCores Used %
23	0	0	23	0	<memory:0 B, vCores:0>	<memory:20 GB, vCores:8>	<memory:0 B, vCores:0>	68	0

Cluster Nodes Metrics

Active Nodes	Decommissioning Nodes	Decommissioned Nodes	Lost Nodes	Unhealthy Nodes	Rebooted Nodes	Shutdown Nodes
4	0	0	0	0	0	0

Scheduler Metrics

Scheduler Type	Scheduling Resource Type	Minimum Allocation	Maximum Allocation	Maximum Cluster Application Priority	Scheduler Busy %
Capacity Scheduler	[memory-mb (unit=Mi), vcores]	<memory:1024, vCores:1>	<memory:8192, vCores:2>	0	0

Show 20 entries   Search:

Node Labels	Rack	Node State	Node Address	Node HTTP Address	Last health-update	Health-report	Containers	Allocation Tags	Mem Used	Mem Avail	Phys Mem Used %	VCores Used	VCores Avail	Phys VCores Used %	Version
/ default-rack		RUNNING	node4:42681	node4:8042	Fri Dec 20 01:24:13 -0800 2024		0		0 B	4 GB	52	0	2	0	3.3.6
/ default-rack		RUNNING	node1:38533	node1:8042	Fri Dec 20 01:24:14 -0800 2024		0		0 B	8 GB	73	0	2	8	3.3.6
/ default-rack		RUNNING	node2:40145	node2:8042	Fri Dec 20 01:24:13 -0800 2024		0		0 B	4 GB	72	0	2	0	3.3.6
/ default-rack		RUNNING	node3:32785	node3:8042	Fri Dec 20 01:24:13 -0800 2024		0		0 B	4 GB	67	0	2	0	3.3.6

Showing 1 to 4 of 4 entries   First Previous 1 Next Last

1.1 Installing Hadoop-Related Packages

```
kangning@node1:~$ hadoop version
Hadoop 3.3.6
Source code repository https://github.com/apache/hadoop.git -r 1be78238728da9266a4f88195058f08fd012bf9c
Compiled by ubuntu on 2023-06-18T08:22Z
Compiled on platform linux-x86_64
Compiled with protoc 3.7.1
From source with checksum 5652179ad55f76cb287d9c633bb53bbd
This command was run using /opt/hadoop/share/hadoop/common/hadoop-common-3.3.6.jar
kangning@node1:~$ jps
40497 NodeManager
40338 ResourceManager
40915 JobHistoryServer
81461 Jps
59029 NameNode
59434 SecondaryNameNode
59227 DataNode
kangning@node1:~$
```

```
kangning@node2:~$ hadoop version
Hadoop 3.3.6
Source code repository https://github.com/apache/hadoop.git -r 1be78238728da9266a4f88195058f08fd012bf9c
Compiled by ubuntu on 2023-06-18T08:22Z
Compiled on platform linux-x86_64
Compiled with protoc 3.7.1
From source with checksum 5652179ad55f76cb287d9c633bb53bbd
This command was run using /opt/hadoop/share/hadoop/common/hadoop-common-3.3.6.jar
kangning@node2:~$ jps
25444 DataNode
31700 Jps
16470 NodeManager
kangning@node2:~$
```

```
kangning@node3:~$ hadoop version
Hadoop 3.3.6
Source code repository https://github.com/apache/hadoop.git -r 1be78238728da9266a4f88195058f08fd012bf9c
Compiled by ubuntu on 2023-06-18T08:22Z
Compiled on platform linux-x86_64
Compiled with protoc 3.7.1
From source with checksum 5652179ad55f76cb287d9c633bb53bbd
This command was run using /opt/hadoop/share/hadoop/common/hadoop-common-3.3.6.jar
kangning@node3:~$ jps
23514 DataNode
28219 Jps
14508 NodeManager
kangning@node3:~$
```

```
kangning@node4:~$ hadoop version
Hadoop 3.3.6
Source code repository https://github.com/apache/hadoop.git -r 1be78238728da9266a4f88195058f08fd012bf9c
Compiled by ubuntu on 2023-06-18T08:22Z
Compiled on platform linux-x86_64
Compiled with protoc 3.7.1
From source with checksum 5652179ad55f76cb287d9c633bb53bbd
This command was run using /opt/hadoop/share/hadoop/common/hadoop-common-3.3.6.jar
kangning@node4:~$ jps
25251 DataNode
30102 Jps
16108 NodeManager
kangning@node4:~$
```

1.2 Installing Python Packages

```
kangning@node1:~$ python3 --version
Python 3.8.10
kangning@node1:~$ python3 -m pip --version
pip 20.0.2 from /usr/lib/python3/dist-packages/pip (python 3.8)
kangning@node1:~$
```

2. Building Inverted Index with Python and Hadoop Streaming

2.1 mapper.py - Mapper Program

2.2 reducer.py - Reducer Program

2.3 Ensure Scripts Have Execute Permissions

2.4 Running MapReduce Jobs with Hadoop Streaming

Before running, we can using the following script to check mapper.py and reducer.py

cat input.json | python3 /home/kangning/search_engine/test3/mapper.py | python3 /home/kangning/search_engine/test3/reducer.py

```
kangning@node1:~/search_engine/test3$ cat input.json | python3 /home/kangning/search_engine/test3/mapper.py | python3 /home/kangn
ing/search_engine/test3/reducer.py
```

Output would be like:

```
            "title": "Best Transcription Jobs That Pay",
            "author": "Christopher Murray"
        }
    ],
    "cryptoheadio": [
        {
            "uuid": "012028e5fdecf6e9b4178fa9731eec767dba69df",
            "title": "Bitfinex Hacker Ilya Lichtenstein Gets Five Years in Prison - Crypto Head",
            "author": "@CryptoHeadio"
        }
    ],
    "comunidade": [
        {
            "uuid": "84a13d18522856dfeea7a195d74aaadba077228d",
            "title": "Mais uma comunidade recebe wi-fi gratuito em Porto Alegre",
            "author": "Porto Alegre 24 horas"
        }
    ]
}
kangning@node1:~/search_engine/test3$
```

1. Run Hadoop Streaming Job:

   you can run directly after modifying the script run.sh:

   bash mapreduce/run.sh

```
kangning@node1:~/search_engine/test3$ bash run.sh
packageJobJar: [/tmp/hadoop-unjar1430282263653639717/] [] /tmp/streamjob6367932392299335313.jar tmpDir=null
2024-12-19 21:57:08,828 INFO client.DefaultNoHARMFailoverProxyProvider: Connecting to ResourceManager at /192.168.100.133:8032
2024-12-19 21:57:08,917 INFO client.AHSProxy: Connecting to Application History server at /0.0.0.0:10200
2024-12-19 21:57:08,944 INFO client.DefaultNoHARMFailoverProxyProvider: Connecting to ResourceManager at /192.168.100.133:8032
2024-12-19 21:57:08,945 INFO client.AHSProxy: Connecting to Application History server at /0.0.0.0:10200
2024-12-19 21:57:09,076 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/kangning/
.staging/job_1734605893012_0023
2024-12-19 21:57:09,262 INFO mapred.FileInputFormat: Total input files to process : 1
2024-12-19 21:57:09,325 INFO mapreduce.JobSubmitter: number of splits:2
2024-12-19 21:57:09,420 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1734605893012_0023
2024-12-19 21:57:09,420 INFO mapreduce.JobSubmitter: Executing with tokens: []
2024-12-19 21:57:09,551 INFO conf.Configuration: resource-types.xml not found
2024-12-19 21:57:09,551 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
2024-12-19 21:57:09,595 INFO impl.YarnClientImpl: Submitted application application_1734605893012_0023
```

2. View Output Results:

localhost:9870/explorer.html#/user/kangning

Hadoop   Overview   Datanodes   Datanode Volume Failures   Snapshot   Startup Progress   Utilities ▾

Browse Directory

/user/kangning   Go!

Show 25 entries   Search:

	Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name	
☐	drwxrwxrwx	kangning	supergroup	0 B	Dec 19 01:13	0	0 B	QuasiMonteCarlo_1734599591030_698071242	🗑
☐	drwxrwxrwx	kangning	supergroup	0 B	Dec 19 01:45	0	0 B	QuasiMonteCarlo_1734601499847_1624812585	🗑
☐	drwxrwxrwx	kangning	supergroup	0 B	Dec 19 02:14	0	0 B	QuasiMonteCarlo_1734603243648_1336780038	🗑
☐	-rwxrwxrwx	kangning	supergroup	336 B	Dec 19 12:50	3	128 MB	app.py	🗑
☐	-rwxrwxrwx	kangning	supergroup	1.18 KB	Dec 19 12:49	3	128 MB	build_index.py	🗑
☐	-rwxrwxrwx	kangning	supergroup	32.34 MB	Dec 19 10:44	3	128 MB	input.json	🗑
☐	-rwxrwxrwx	kangning	supergroup	13.65 MB	Dec 19 12:46	3	128 MB	inverted_index.json	🗑
☐	-rwxrwxrwx	kangning	supergroup	1.64 KB	Dec 19 13:22	3	128 MB	search_index.py	🗑
☐	drwxrwxrwx	kangning	supergroup	0 B	Dec 19 01:11	0	0 B	wordcount	🗑

Showing 1 to 9 of 9 entries   Previous 1 Next

Hadoop, 2023.

Hadoop   Overview   Datanodes   Datanode Volume Failures   Snapshot   Startup Progress   Utilities ▾

Browse Directo

/user/kangning

Show 25 entries   Search:

File information - inverted_index.json   ✕

Download   Head the file (first 32K)   Tail the file (last 32K)

Block information --   Block 0 ▾

Block ID: 1073742094

Block Pool ID: BP-279830477-192.168.100.133-1734594973656

Generation Stamp: 1270

Size: 14314615

Availability:

- node1
- node3
- node4

File contents

          }
        ],
        "entrepreneur": [
          {
            "uuid": "fb94aad9a179c31694e2c551c70f7c1b557b48fe",
            "title": "Top award entrepreneur Secrets",
            "author": "Youngentrepreneuraward"
          },
          {
            "uuid": "945e05f1419928e0d0ab33b7ca8894b58287345d",
            "title": "US Tech Entrepreneur Sentenced to 5 Years for Crypto Laundering in Bitfinex Hack Case",
            "author": "Lore Apostol"
          }
        ],
        "secrets": [
          {
            "uuid": "fb94aad9a179c31694e2c551c70f7c1b557b48fe",
            "title": "Top award entrepreneur Secrets",

	Permission	Owner		
☐	drwxrwxrwx	kangning	_1734599591030_698071242	🗑
☐	drwxrwxrwx	kangning	_1734601499847_1624812585	🗑
☐	drwxrwxrwx	kangning	_1734603243648_1336780038	🗑
☐	-rwxrwxrwx	kangning		🗑
☐	-rwxrwxrwx	kangning		
☐	-rwxrwxrwx	kangning		
☐	-rwxrwxrwx	kangning		
☐	-rwxrwxrwx	kangning		
☐	drwxrwxrwx	kangning		

Showing 1 to 9 of 9 entries   Previous 1 Next

Close

Hadoop, 2023.

3. Implementing Search Functionality Using Flask Web Application

After constructing the inverted index with the MapReduce job, we can implement the search functionality using Flask. Ensure the database and index files are stored in HDFS, as the subsequent implementation relies on HDFS for storing these files.

3.1 search_index.py

3.2 app.py

3.3 HTML Templates

- templates/index.html
- templates/results.html

4. Start the Flask Application

You can start the Flask application to implement the frontend search functionality.

1. Ensure the Flask service scripts (app.py, search_index.py) and HTML template files (index.html and results.html) are in the appropriate directories (as shown in the directory structure at the beginning).
2. Ensure the relevant database and index files are in HDFS (input.json and inverted_index.json).
   - You can skip the indexing step and directly upload the provided database and index to your HDFS.
3. Start the Flask application in the background:

   nohup python3 app.py &

4. Open a browser and visit http://node_IP:5000/. You will see an input box and search button, allowing users to enter query terms and get results.
5. Enter the query terms (for example, AMD or Apple), and then click the "Search" button. You will see a list of articles related to the query. Each article will display its title, author, and a collapsible body text. Clicking on the article title will take you directly to the original link.
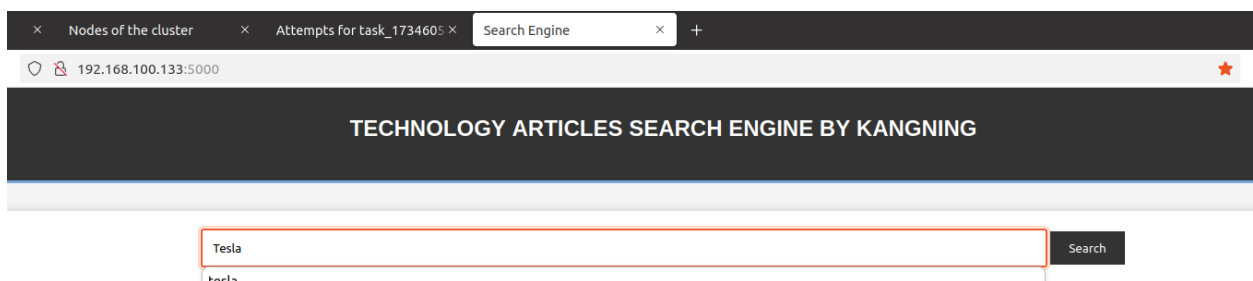
192.168.100.133:5000/search?q=Tesla

TECHNOLOGY ARTICLE SEARCH ENGINE BY KANGNING

Enter search keyword... Search

Back to Home

Rivian (RIVN) CEO says R2 will be its Tesla Model 3 moment

Author: Peter Johnson

Snippet: Although Rivian (RIVN) and Tesla have different products, they are more alike than you may think. Rivian's CEO RJ Scaringe said its upcoming R2 model will be its version of the Tesla Model 3. It may n

Show more

Xiaomi SU7 EV now outperforms Porsche and Tesla...

Author: Hyper Star Offer Upgrade User ID

Snippet: Users Online Now: 1,824 (Who's On?) Visitors Today: 653,143 Pageviews Today: 1,238,187 Threads Today: 439 Posts Today: 9,885 05:40 PM Directory Adv. Search Topics Forum Back to Forum Post New Thread

Show more

Xiaomi sorprende con su propio vehículo eléctrico: ¿Será el fin del reinado de Tesla? – FayerWayer

Author: Dannae Arias

Snippet: Dentro del ámbito tecnológico [hay más que sólo celulares](https://www.fayerwayer.com/moviles/2023/12/05/samsung-apple-xiaomi-y-honor-gobiernan-el-mercado-de-los-smartphones-segun-este-mapa-de-dominio

Show more

Tesla announces new partnership with Samsung over energy connectivity | Electrek

Author: Fred Lambert

Snippet: Tesla and Samsung have announced a new partnership over energy connectivity, which appears to be the reason behind Tesla's API update. Yesterday, we reported on Tesla updating their API to include Pow

Show more

You'll soon be able to control Samsung SmartThings devices from your Tesla or Hyundai

Author: Patrick O'Rourke

Snippet: Ahead of CES 2024, Samsung has partnered with Tesla and Hyundai to bring its SmartThings smart home platform to both companies' respective vehicles and devices. First, Samsung SmartThings users will s

Show more

Click "Show more" to expand:

192.168.100.133:5000/search?q=Tesla

TECHNOLOGY ARTICLE SEARCH ENGINE BY KANGNING

Enter search keyword... Search

Back to Home

Rivian (RIVN) CEO says R2 will be its Tesla Model 3 moment

Author: Peter Johnson

Full Text: Although Rivian (RIVN) and Tesla have different products, they are more alike than you may think. Rivian's CEO RJ Scaringe said its upcoming R2 model will be its version of the Tesla Model 3. It may not look like a Tesla Model 3, but Rivian's R2 will help shape the EV startup's future. Like Tesla, Rivian started with more expensive flagship models. Tesla released the Roadster in 2008, which led to the Model S in 2012. Several years later, Tesla announced its first mass-market electric car, the Model 3 . Although Tesla went through what CEO Elon Musk described as "production hell" ramping up the Model 3, the work paid off. In 2018, Tesla's Model 3 became the best-selling EV globally, holding the title until the Model Y topped it in 2020. Tesla's Model 3 will be among the top-selling vehicles globally (gas or electric) in 2023 as it continues to see strong demand. Rivian has been ramping production all year. Despite several rivals cutting EV targets, Rivian raised its production goal following strong Q3 results. The EV maker now aims to build 52,000 vehicles this year. That's more than double the 24,337 it produced last year. Rivian R1T (Source: Rivian) Rivian R2 will be its Tesla Model 3 moment Like Tesla, Rivian plans to expand into new markets. The automaker chose Clayco as its partner earlier this month to build its massive $5 billion EV plant in Georgia. The new facility will be home to Rivian's R2 vehicles. Scaringe previously said R2 will start at around $40,000 to $50,000. Rivian's R2 will be a mid-size electric SUV. On the How I Built This podcast with Guy Raz, Scaringe recently spoke about ramping production and expanding the brand, as well as other plans for the future. Rivian R1S (Source: Rivian) When asked if R2 will be Rivian's version of the Tesla Model 3, Scaringe said that "strategically, it's very similar." Scaringe explained that Rivian started with its flagship R1 products that are intentionally priced higher for their features, capabilities, and more. Rivian will take the "essence of the brand" represented in those products and build it into more affordable models. Rivian production at its Normal, Ill facility (Source: Rivian) The Rivian R2 is expected to be unveiled for the first time next year before launching in 2026. Construction on the Georgia site will begin soon, with grading work finishing up through the end of the year. An official groundbreaking ceremony is expected in early 2024. Once the first stage is complete, Rivian expects to be able to build 200,000 EVs a year. By 2030, the EV maker looks to double it. Electrek's Take Rivian continues outpacing rivals despite the R1T electric truck and R1S SUV costing around $70,000. The company also has a major lifeline with its commercial electric van. E-commerce giant Amazon announced in 2019 that it would deploy 100,000 Rivian EDVs by 2030. It has already deployed over 10,000. The EV maker recently ended its exclusivity with Amazon, signing its second partner in AT&T earlier this month. Rivian drivers will also get access to Tesla's supercharger network. Starting early next year, Rivian will provide an adapter, and eventually, it will move to the NACS connector. The EV maker is taking another page from Tesla with its own charging network. Rivian is focusing on reliability and uptime like the EV leader. Scaringe said having several different charging providers can help adoption. It plans to open its network to other brands like Tesla. Scaringe said it has around 60 Adventure Network sites live, but there are hundreds more to build. Add Electrek to your Google News feed.

Show less

Xiaomi SU7 EV now outperforms Porsche and Tesla...

Author: Hyper Star Offer Upgrade User ID

Snippet: Users Online Now: 1,824 (Who's On?) Visitors Today: 653,143 Pageviews Today: 1,238,187 Threads Today: 439 Posts Today: 9,885 05:40 PM Directory Adv. Search Topics Forum Back to Forum Post New Thread

Click title to get the link of original text.

TECHNOLOGY ARTICLE SEARCH ENGINE BY KANGNING

Enter search keyword...   Search

Back to Home

Rivian (RIVN) CEO says R2 will be its Tesla Model 3 moment

Author: Peter Johnson

Full Text: Although Rivian (RIVN) and Tesla have different products, they are more alike than you may think. Rivian's CEO RJ Scaringe said its upcoming R2 model will be its version of the Tesla Model 3. It may not look like a Tesla Model 3, but Rivian's R2 will help shape the EV startup's future. Like Tesla, Rivian started with more expensive flagship models. Tesla released the Roadster in 2008, which led to the Model S in 2012. Several years later, Tesla announced its first mass-market electric car, the Model 3 . Although Tesla went through what CEO Elon Musk described as "production hell" ramping up the Model 3, the work paid off. In 2018, Tesla's Model 3 became the best-selling EV globally, holding the title until the Model Y topped it in 2020. Tesla's Model 3 will be among the top-selling vehicles globally (gas or electric) in 2023 as it continues to see strong demand. Rivian has been ramping production all year. Despite several rivals cutting EV targets, Rivian raised its production goal following strong Q3 results. The EV maker now aims to build 52,000 vehicles this year. That's more than double the 24,337 it produced last year. Rivian R1T (Source: Rivian) Rivian R2 will be its Tesla Model 3 moment Like Tesla, Rivian plans to expand into new markets. The automaker chose Clayco as its partner earlier this month to build its massive $5 billion EV plant in Georgia. The new facility will be home to Rivian's R2 vehicles. Scaringe previously said R2 will start at around $40,000 to $50,000. Rivian's R2 will be a mid-size electric SUV. On the How I Built This podcast with Guy Raz, Scaringe recently spoke about ramping production and expanding the brand, as well as other plans for the future. Rivian R1S (Source: Rivian) When asked if R2 will be Rivian's version of the Tesla Model 3, Scaringe said that "strategically, it's very similar." Scaringe explained that Rivian started with its flagship R1 products that are intentionally priced higher for their features, capabilities, and more. Rivian will take the "essence of the brand" represented in those products and build it into more affordable models. Rivian production at its Normal, Ill facility (Source: Rivian) The Rivian R2 is expected to be unveiled for the first time next year before launching in 2026. Construction on the Georgia site will begin soon, with grading work finishing up through the end of the year. An official groundbreaking ceremony is expected in early 2024. Once the first stage is complete, Rivian expects to be able to build 200,000 EVs a year. By 2030, the EV maker looks to double it. Electrek's Take Rivian continues outpacing rivals despite the R1T electric truck and R1S SUV costing around $70,000. The company also has a major lifeline with its commercial electric van. E-commerce giant Amazon announced in 2019 that it would deploy 100,000 Rivian EDVs by 2030. It has already deployed over 10,000. The EV maker recently ended its exclusivity with Amazon, signing its second partner in AT&T earlier this month. Rivian drivers will also get access to Tesla's supercharger network. Starting early next year, Rivian will provide an adapter, and eventually, it will move to the NACS connector. The EV maker is taking another page from Tesla with its own charging network. Rivian is focusing on reliability and uptime like the EV leader. Scaringe said having several different charging providers can help adoption. It plans to open its network to other brands like Tesla. Scaringe said it has around 60 Adventure Network sites live, but there are hundreds more to build. Add Electrek to your Google News feed.

Show less

Xiaomi SU7 EV now outperforms Porsche and Tesla...
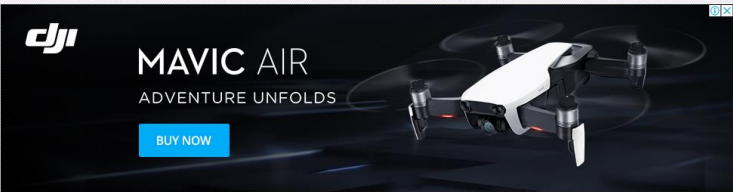
Author: Hyper Star Offer Upgrade User ID

https://electrek.co/2023/12/28/rivian-r2-tesla-model-3-moment/

RIVIAN   RIVIAN R2

Rivian (RIVN) CEO says R2 will be its Tesla Model 3 moment as more affordable EVs are needed

Peter Johnson | Dec 28 2023 - 3:01 pm PT | 61 Comments

Featured
from electrek

Tesla (TSLA) sales in Europe are down 14% year-to-date and it's time to worry

Use Spaces to differentiate between keywords
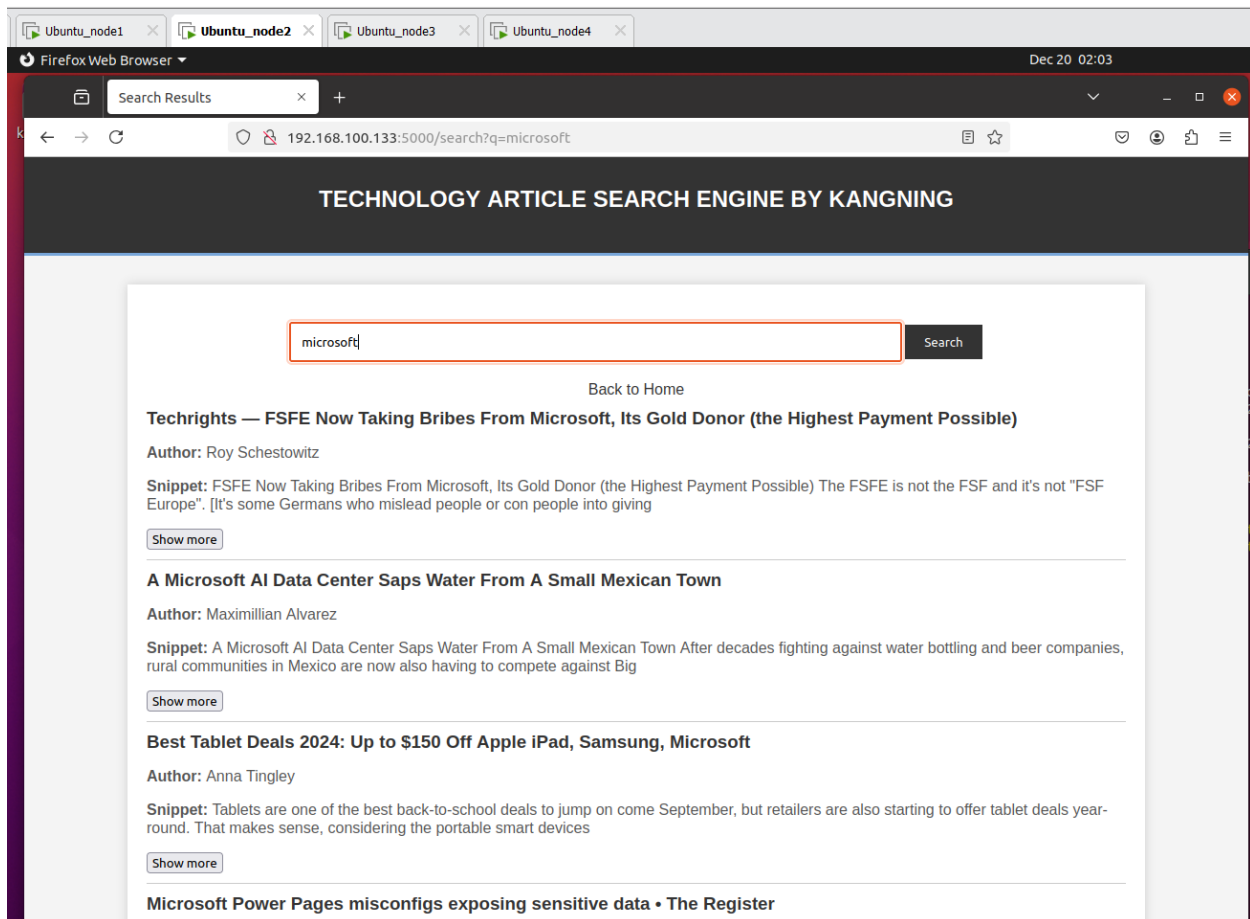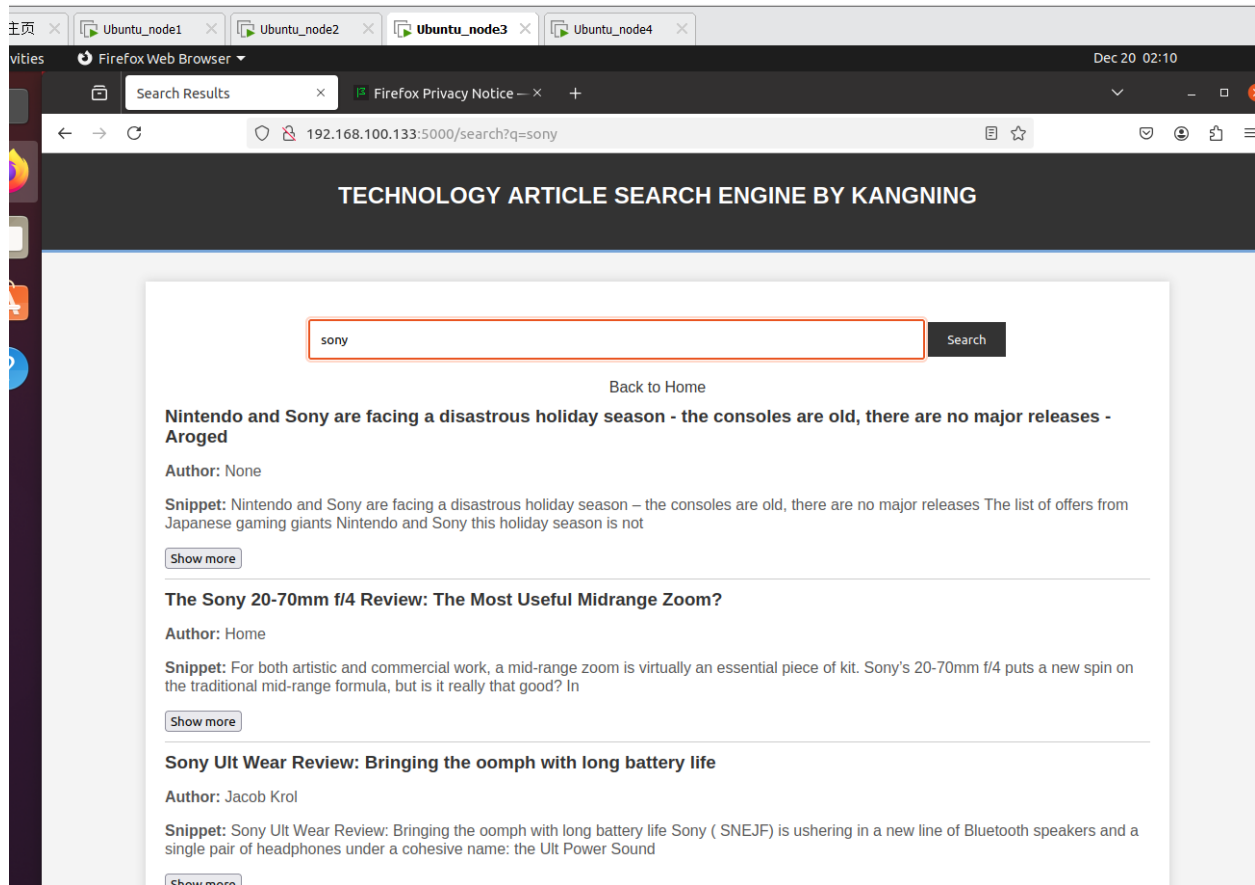
Also, the search engine can also be access by other node in the cluster(search engine running in node1):

(node1: 192.168.100.133   node2: 192.168.100.134

node3: 192.168.100.135   node4: 192.168.100.136)

```
kangning@node1:~$ sudo arp-scan --interface=ens33 --localnet
Interface: ens33, type: EN10MB, MAC: 00:0c:29:b0:0b:50, IPv4: 192.168.100.133
Starting arp-scan 1.9.7 with 256 hosts (https://github.com/royhills/arp-scan)
192.168.100.1    00:50:56:c0:00:08       VMware, Inc.
192.168.100.2    00:50:56:ea:e0:62       VMware, Inc.
192.168.100.134 00:0c:29:2b:de:e6       VMware, Inc.
192.168.100.135 00:0c:29:06:c2:8b       VMware, Inc.
192.168.100.136 00:0c:29:20:3d:43       VMware, Inc.
192.168.100.254 00:50:56:fc:a2:ae       VMware, Inc.

6 packets received by filter, 0 packets dropped by kernel
Ending arp-scan 1.9.7: 256 hosts scanned in 1.923 seconds (133.13 hosts/sec). 6 responded
kangning@node1:~$ host node1
node1.localdomain has address 192.168.100.133
Host node1.localdomain not found: 3(NXDOMAIN)
Host node1.localdomain not found: 3(NXDOMAIN)
kangning@node1:~$ host node2
node2.localdomain has address 192.168.100.134
Host node2.localdomain not found: 3(NXDOMAIN)
Host node2.localdomain not found: 3(NXDOMAIN)
kangning@node1:~$ host node3
node3.localdomain has address 192.168.100.135
Host node3.localdomain not found: 3(NXDOMAIN)
Host node3.localdomain not found: 3(NXDOMAIN)
kangning@node1:~$ host node4
node4.localdomain has address 192.168.100.136
Host node4.localdomain not found: 3(NXDOMAIN)
Host node4.localdomain not found: 3(NXDOMAIN)
kangning@node1:~$
```

Ubuntu_node1   Ubuntu_node2   Ubuntu_node3   Ubuntu_node4

Firefox Web Browser ▼   Dec 20 02:03

Search Results

192.168.100.133:5000/search?q=microsoft

TECHNOLOGY ARTICLE SEARCH ENGINE BY KANGNING

microsoft   Search

Back to Home

Techrights — FSFE Now Taking Bribes From Microsoft, Its Gold Donor (the Highest Payment Possible)

Author: Roy Schestowitz

Snippet: FSFE Now Taking Bribes From Microsoft, Its Gold Donor (the Highest Payment Possible) The FSFE is not the FSF and it's not "FSF Europe". [It's some Germans who mislead people or con people into giving

Show more

A Microsoft AI Data Center Saps Water From A Small Mexican Town

Author: Maximillian Alvarez

Snippet: A Microsoft AI Data Center Saps Water From A Small Mexican Town After decades fighting against water bottling and beer companies, rural communities in Mexico are now also having to compete against Big

Show more

Best Tablet Deals 2024: Up to $150 Off Apple iPad, Samsung, Microsoft

Author: Anna Tingley

Snippet: Tablets are one of the best back-to-school deals to jump on come September, but retailers are also starting to offer tablet deals year-round. That makes sense, considering the portable smart devices

Show more

Microsoft Power Pages misconfigs exposing sensitive data • The Register

5. Summary

Using the Hadoop framework, we accomplished the following two main parts:

1. **Constructing the Inverted Index (Based on Hadoop MapReduce):**
   - We used **Hadoop Streaming** and **Python** to write `mapper.py` and `reducer.py`, processing JSON data stored in HDFS, extracting keywords, and generating the inverted index.
   - Leveraged Hadoop's distributed computing capabilities to map keywords to article info from JSON data, built the inverted index, and saved the results in HDFS.
2. **Creating a Web Search Service Using Flask (Based on HDFS):**
   - Built a simple Web application using **Flask** that allows users to enter query terms and return matching articles based on the inverted index.
   - Displayed detailed information about the articles (like `title`, `author`, body text, and link) on the search results page.