My experiences hosting linux mirrors aa

By Shrirang Kahale

$ whoami

➜ Shrirang Kahale, also known as “albony” or “albonycal” on the internet.

➜ I have strong interest in Networking, Cybersecurity and Linux.

➜ Big supporter of the FOSS philosophy.

Table of **contents**

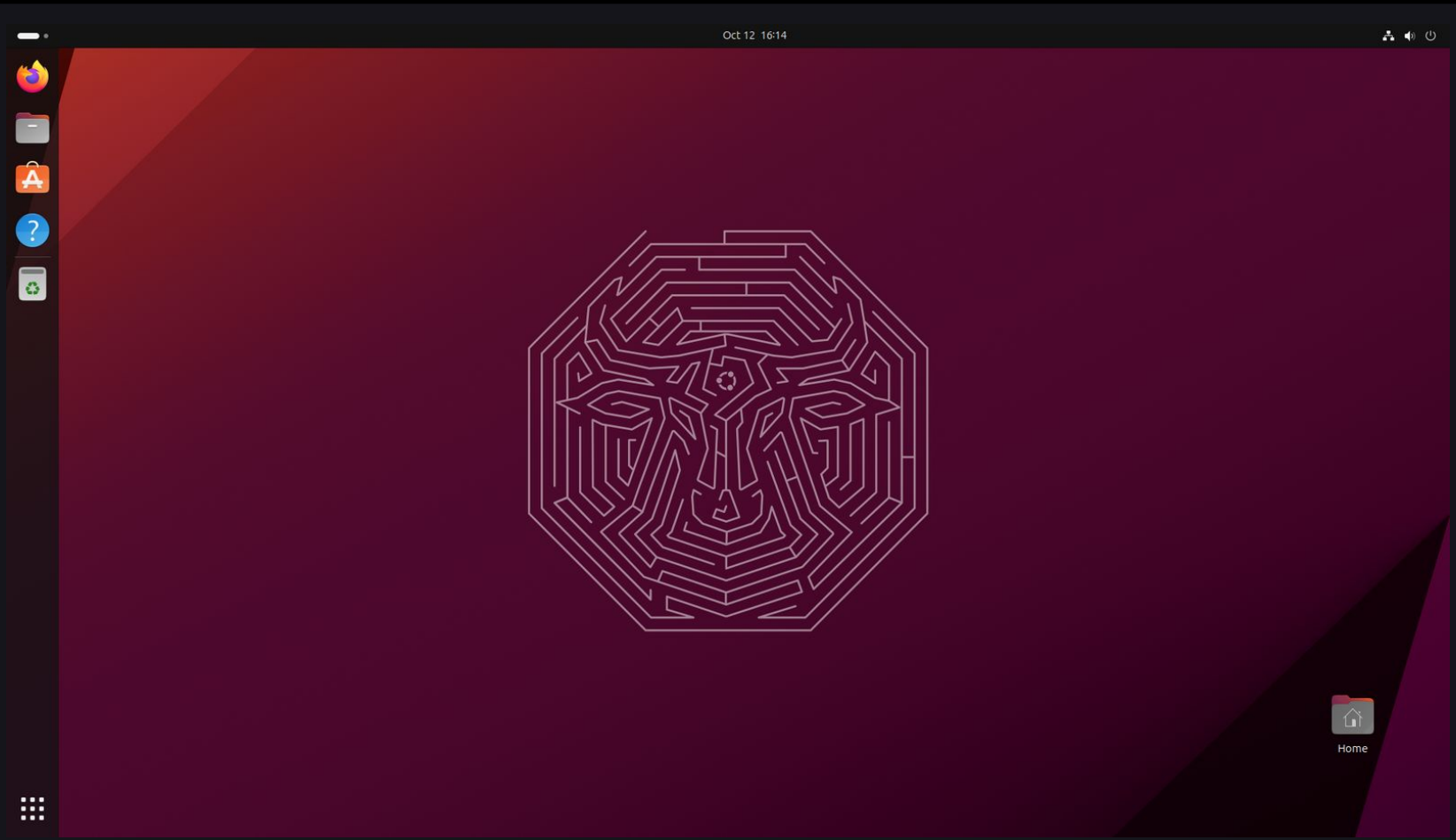
Shrirang Kahale
shrirangkahale.com

What is a Linux Distro?

Shrirang Kahale
shrirangkahale.com

→

- A Linux distribution (or distro) is an operating system made from a software collection that includes the Linux kernel and often a package management system.
- Examples Include Debian, Ubuntu, Archlinux, Fedora etc.
- Linux users usually obtain their operating system by downloading one of the Linux distributions.


Oct 12 16:14
Home

Package Managers

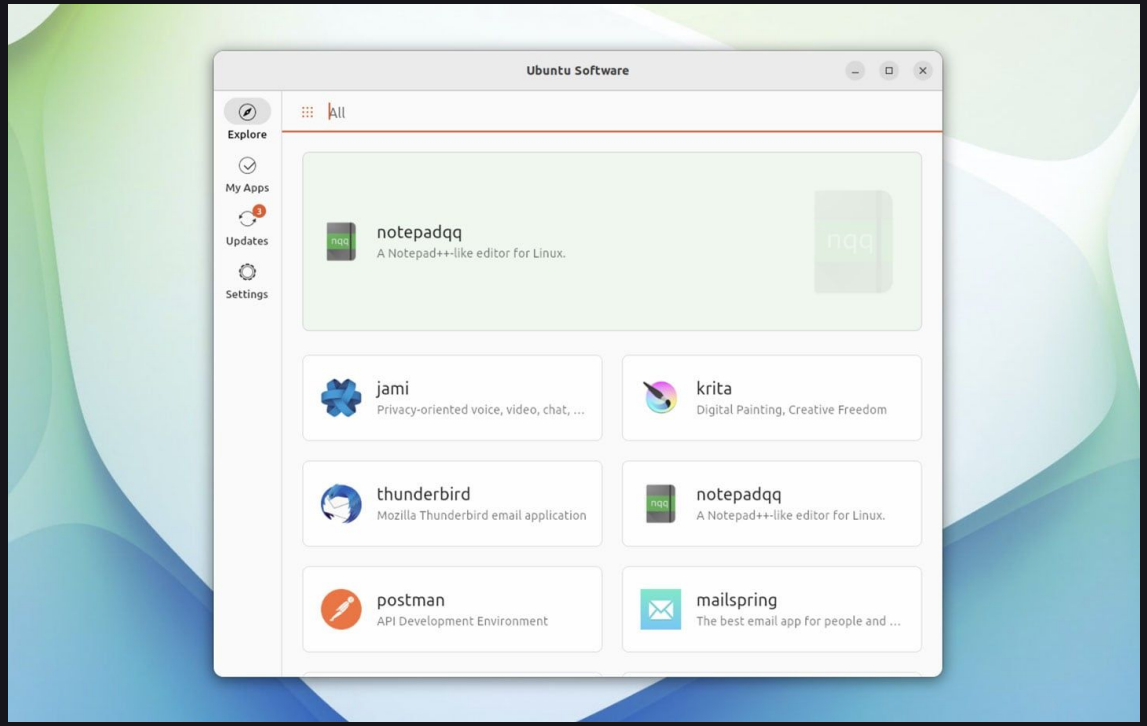
➜ Software tool for managing software packages

➜ Simplifies:
- Installation
- Updates
- Removal

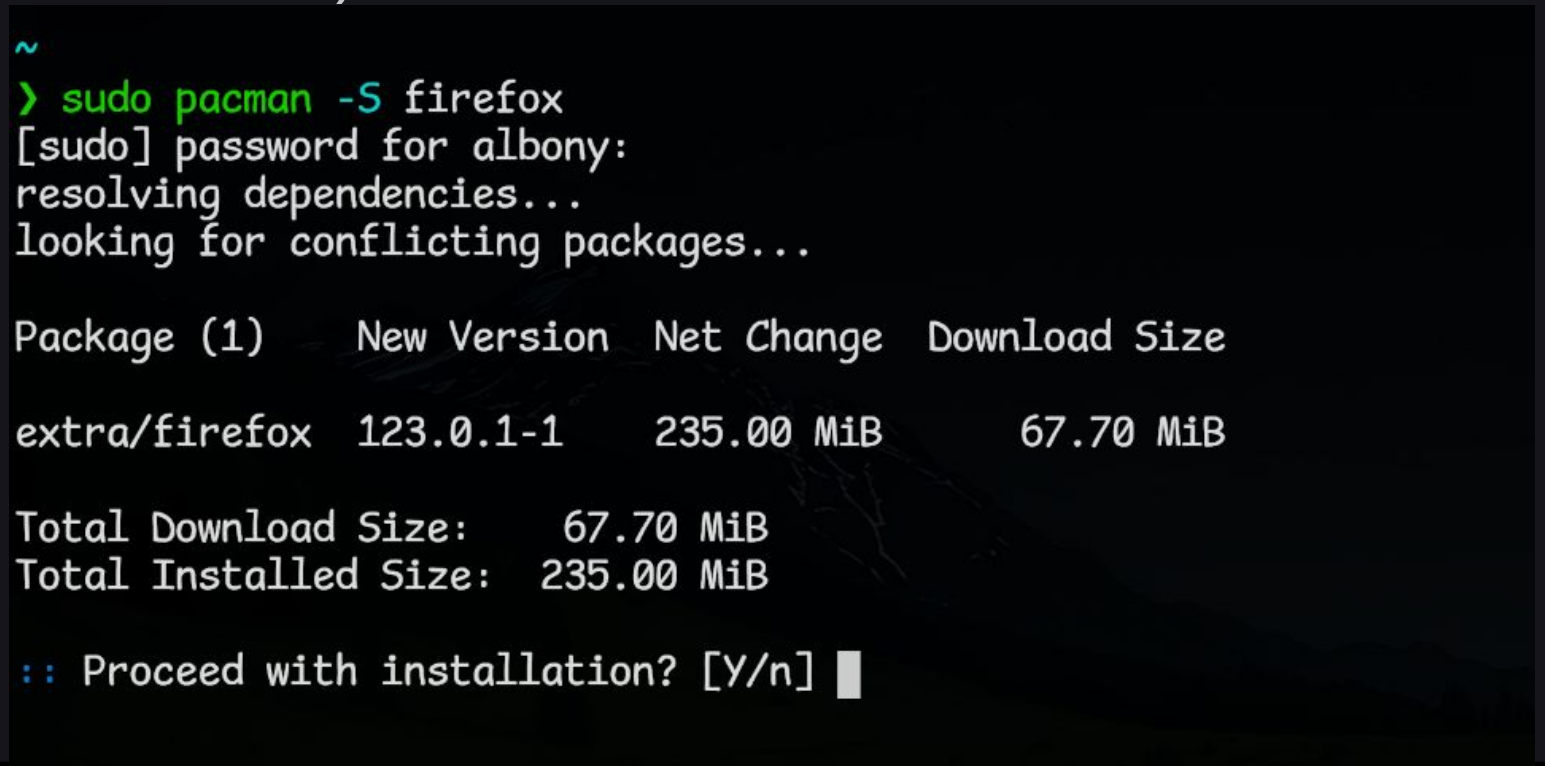
➜ Handles dependencies automatically
➜ Interacts with online repositories

Examples: APT, DNF, Pacman

Ubuntu App Store

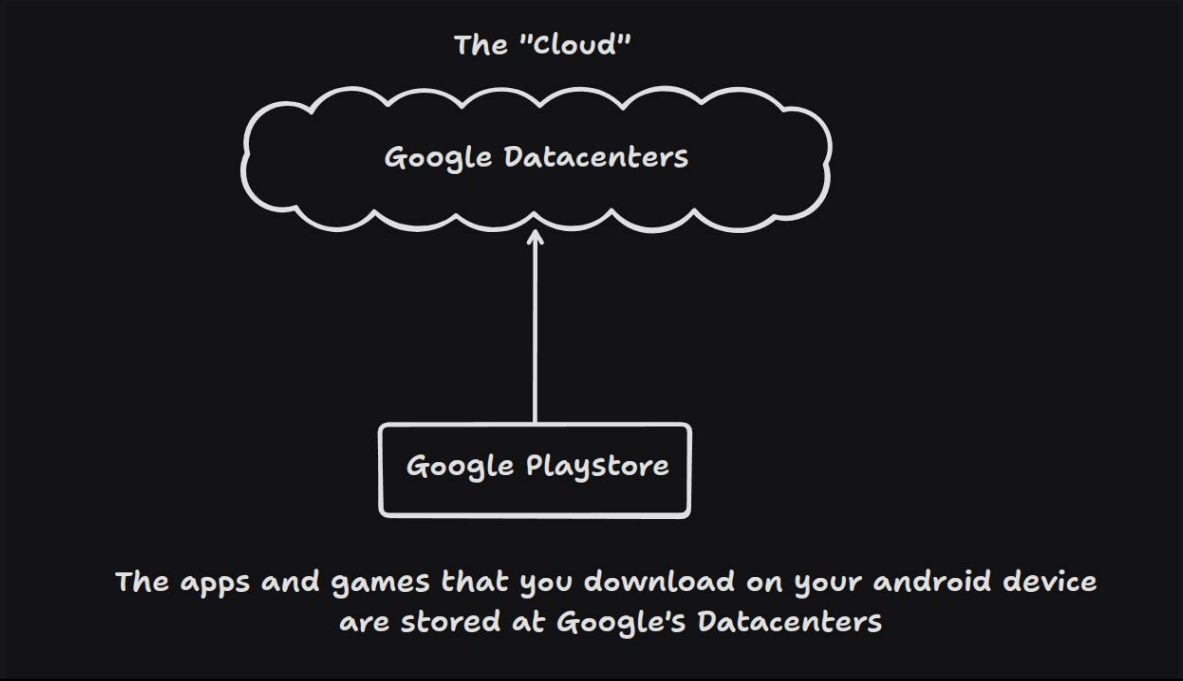


Pacman (command line interface)



Where are those packages stored?

Let us consider the following analogy



But what about linux.. ?

➜ Budget is nowhere close to Google’s or Microsoft’s

➜ Commercial distros like RHEL or Ubuntu exist.

➜ Most distros rely on volunteer run servers or “**mirrors**”

What is a mirror?

➜ *Server or repository that hosts copies of Linux distributions and their associated packages.*

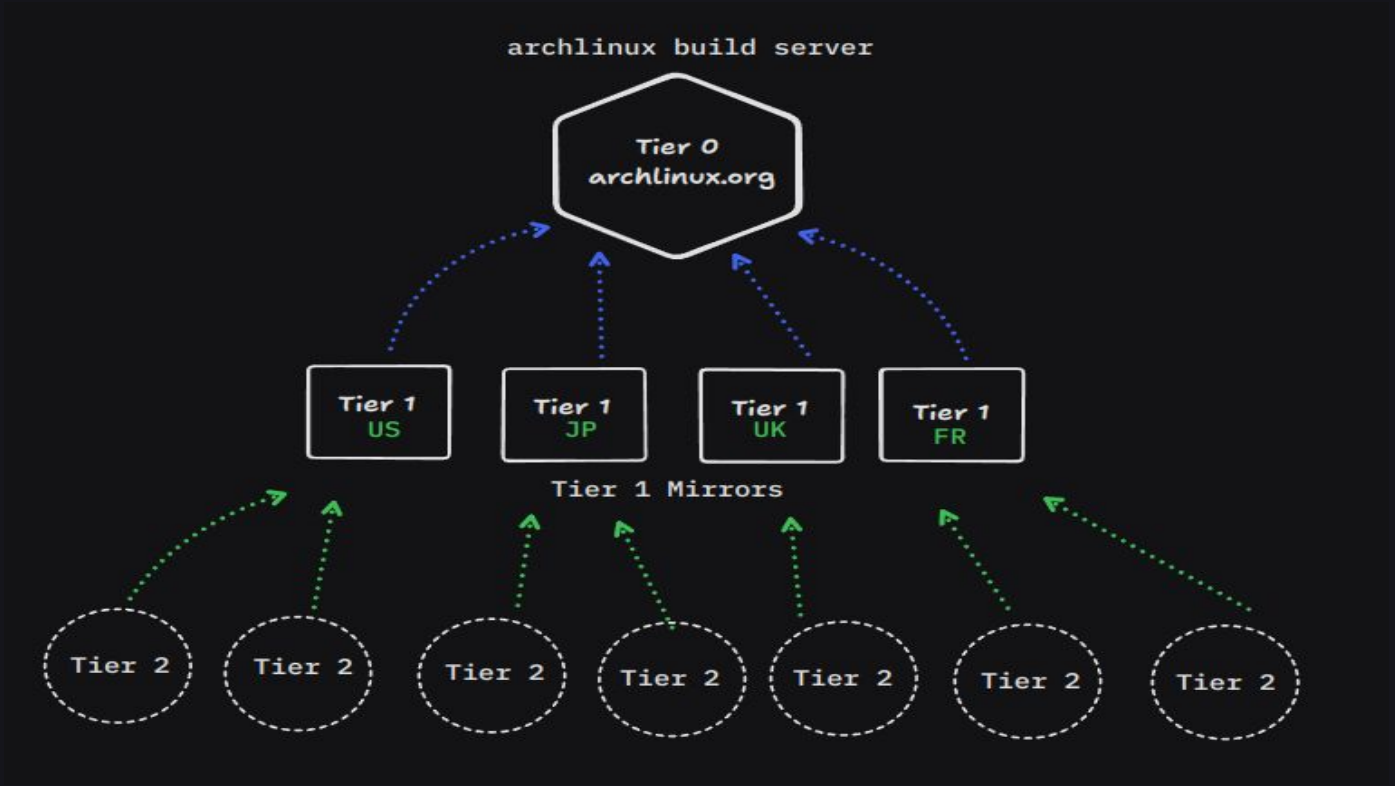
➜ Mirrors are used to distribute the load of downloading Linux distributions and updates among multiple servers.

➜ Users can choose a mirror close to their location to download files faster and more reliably.

➜ Mirrors are synchronized regularly to ensure they have the latest updates and packages.

➜ They play a crucial role in providing redundancy and scalability for distributing Linux software globally.

Archlinux mirror architecture



Mirror Presence in India

Let us take a look at mirror presence of popular distros in India

Distribution Name	Number of mirrors present in India
Debian	1
Archlinux	10 (including mine)
Fedora	2 (only EPEL edition)
Manjaro	2 (including mine)
Ubuntu	12

Why does it matter?

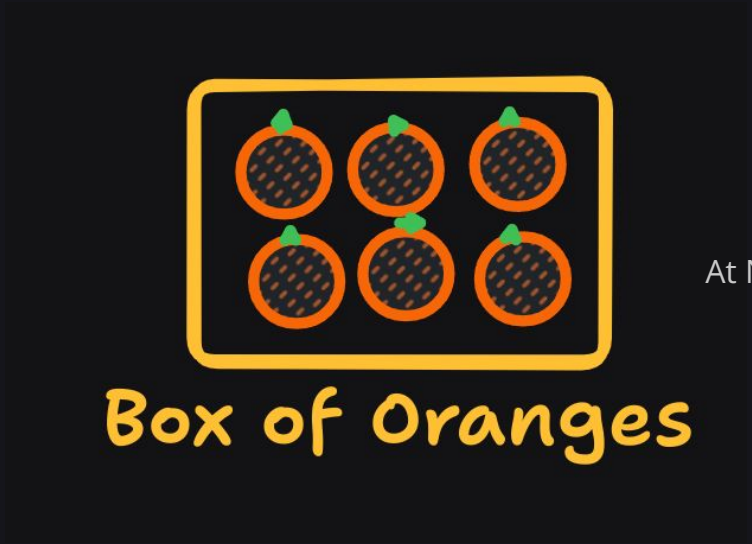
➜ Local Linux mirrors (generally) reduce latency and offer faster download speed.

➜ Using a local mirror conserves international bandwidth which is expensive.

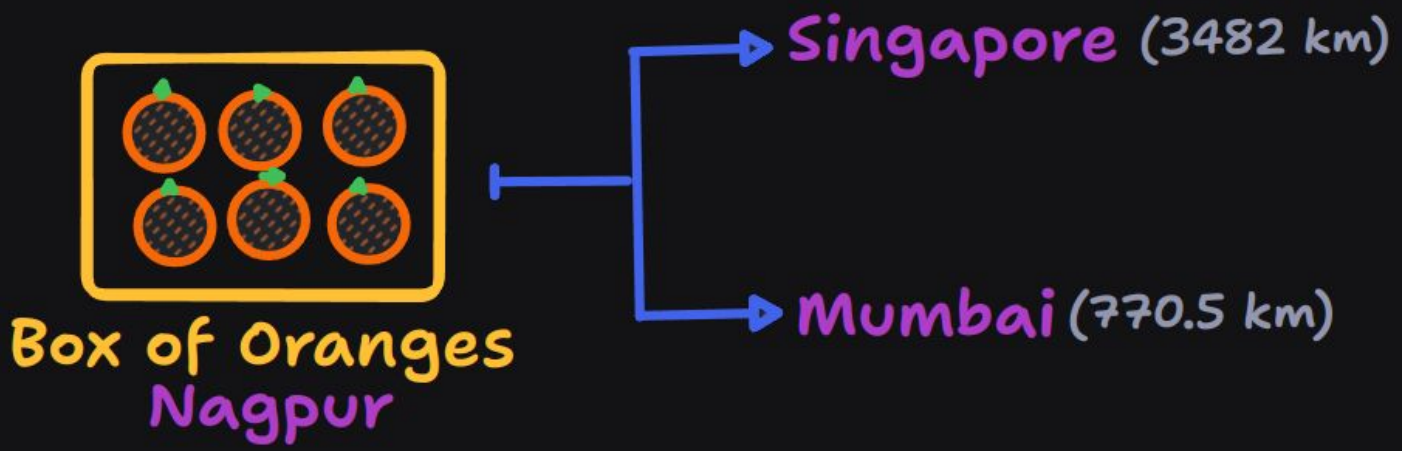
➜ More reliable? Using a mirror outside the country adds more points of failure along the path, so a local one might be more reliable.

But the world isn’t ideal..

➜ To understand this better, first we need to understand how networks work. I won’t be going in the details, so let’s understand this via an analogy.

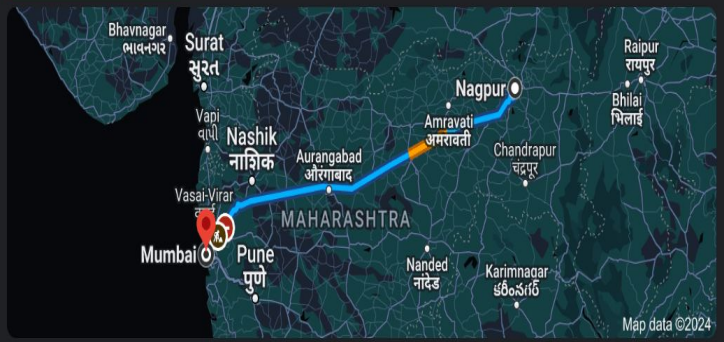


At Nagpur



If you want to send the box to both places, which one will be faster?

The answer is that it depends..



11 hr 5 min (770.5 km) via Hindu Hrudaysamrat Balasaheb Thackeray Maharashtra Samruddhi Mahamarg

Directions

5 hr 30 min
BOM–SIN

1 hr 25 min
NAG–BOM

So let’s say total 7 hours, which is faster than going to Nagpur → Mumbai via road

Nagpur to Singapore latency

```
~
❯ ping 9.9.9.9
PING 9.9.9.9 (9.9.9.9) 56(84) bytes of data.
64 bytes from 9.9.9.9: icmp_seq=1 ttl=59 time=72.3 ms
64 bytes from 9.9.9.9: icmp_seq=2 ttl=59 time=72.2 ms
64 bytes from 9.9.9.9: icmp_seq=3 ttl=59 time=72.1 ms
64 bytes from 9.9.9.9: icmp_seq=4 ttl=59 time=72.2 ms
64 bytes from 9.9.9.9: icmp_seq=5 ttl=59 time=72.1 ms
^C
--- 9.9.9.9 ping statistics ---
5 packets transmitted, 5 received, 0% packet loss, time 4006ms
rtt min/avg/max/mdev = 72.113/72.188/72.264/0.059 ms

~ took 4s
❯ ping 1.1.1.1
PING 1.1.1.1 (1.1.1.1) 56(84) bytes of data.
64 bytes from 1.1.1.1: icmp_seq=1 ttl=58 time=15.6 ms
64 bytes from 1.1.1.1: icmp_seq=2 ttl=58 time=15.3 ms
64 bytes from 1.1.1.1: icmp_seq=3 ttl=58 time=15.5 ms
64 bytes from 1.1.1.1: icmp_seq=4 ttl=58 time=15.5 ms
64 bytes from 1.1.1.1: icmp_seq=5 ttl=58 time=15.1 ms
64 bytes from 1.1.1.1: icmp_seq=6 ttl=58 time=15.8 ms
^C
--- 1.1.1.1 ping statistics ---
6 packets transmitted, 6 received, 0% packet loss, time 5008ms
rtt min/avg/max/mdev = 15.073/15.482/15.845/0.241 ms
```

Nagpur to Mumbai latency

Geographically closer doesn’t always mean it’s closer on the network layer

➜ Let’s take an example of connectivity between Vi (Vodafone Idea) and Airtel to understand this better.

India (Airtel) ➜ Germany (HE.net) ➜ UK (Level3) ➜ India (Vi) with RTT of more than 200ms

mirror.albony.in

➜ I started hosting it in 2022

➜ Around **20k daily users**

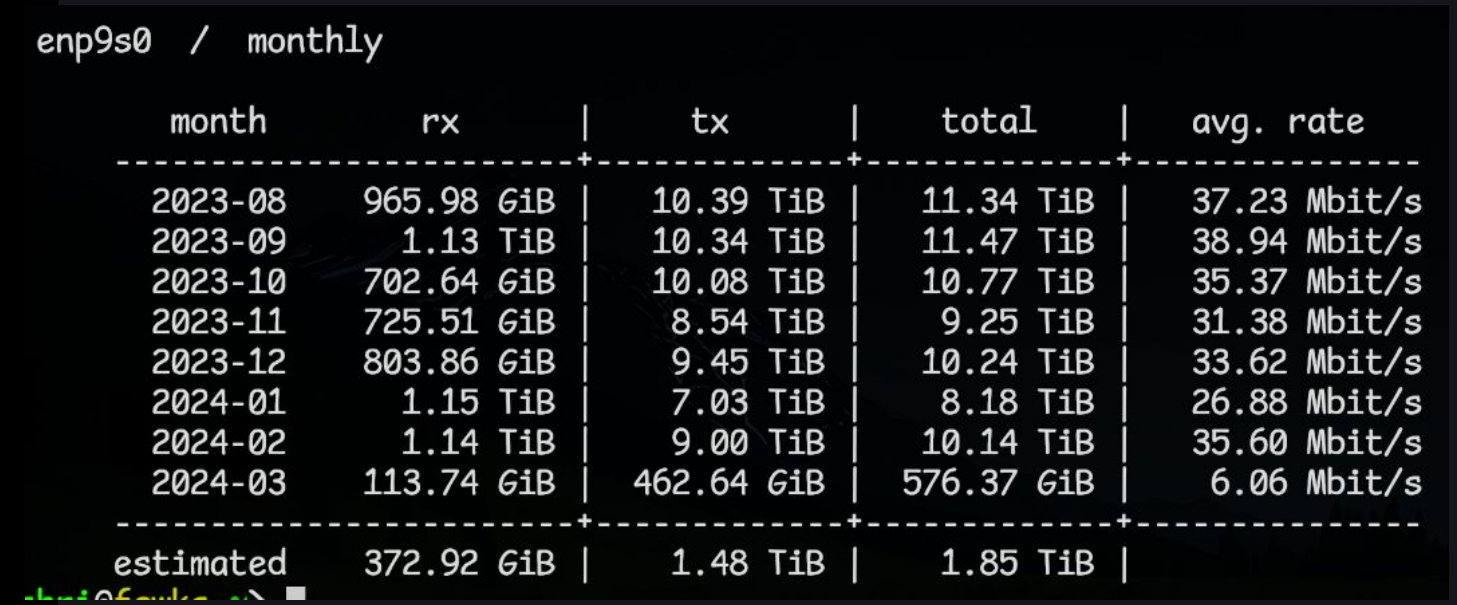
➜ More than **15TB** traffic / month

➜ That's ~ **180TB / year**!

➜ 1G bits/s bandwidth

➜ 99.1% uptime (since last 2 years)

➜ UPS backup in case of power outage



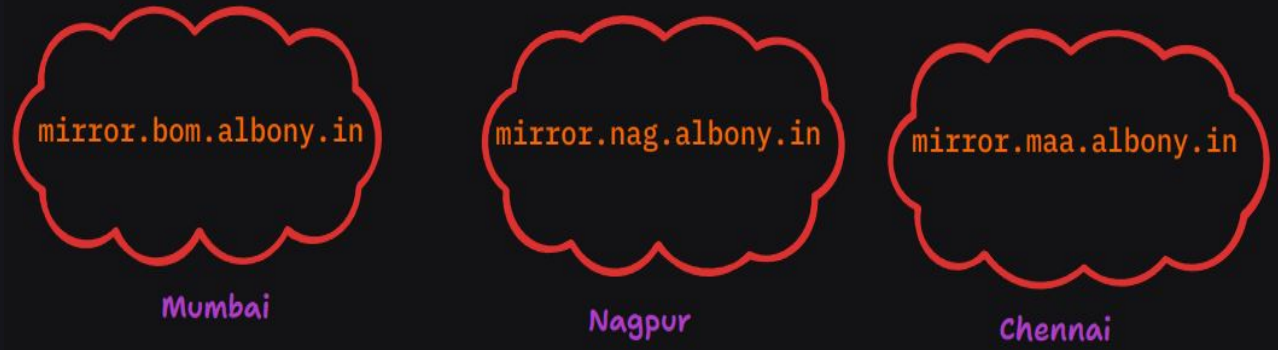
```
enp9s0  /  monthly

     month        rx      |     tx      |    total    |   avg. rate
  ------------------------+-------------+-------------+---------------
    2023-08    965.98 GiB |   10.39 TiB |   11.34 TiB |   37.23 Mbit/s
    2023-09      1.13 TiB |   10.34 TiB |   11.47 TiB |   38.94 Mbit/s
    2023-10    702.64 GiB |   10.08 TiB |   10.77 TiB |   35.37 Mbit/s
    2023-11    725.51 GiB |    8.54 TiB |    9.25 TiB |   31.38 Mbit/s
    2023-12    803.86 GiB |    9.45 TiB |   10.24 TiB |   33.62 Mbit/s
    2024-01      1.15 TiB |    7.03 TiB |    8.18 TiB |   26.88 Mbit/s
    2024-02      1.14 TiB |    9.00 TiB |   10.14 TiB |   35.60 Mbit/s
    2024-03    113.74 GiB |  462.64 GiB |  576.37 GiB |    6.06 Mbit/s
  ------------------------+-------------+-------------+---------------
    estimated  372.92 GiB |    1.48 TiB |    1.85 TiB |
```

Month traffic data from one of my servers.

enp9s0 / daily

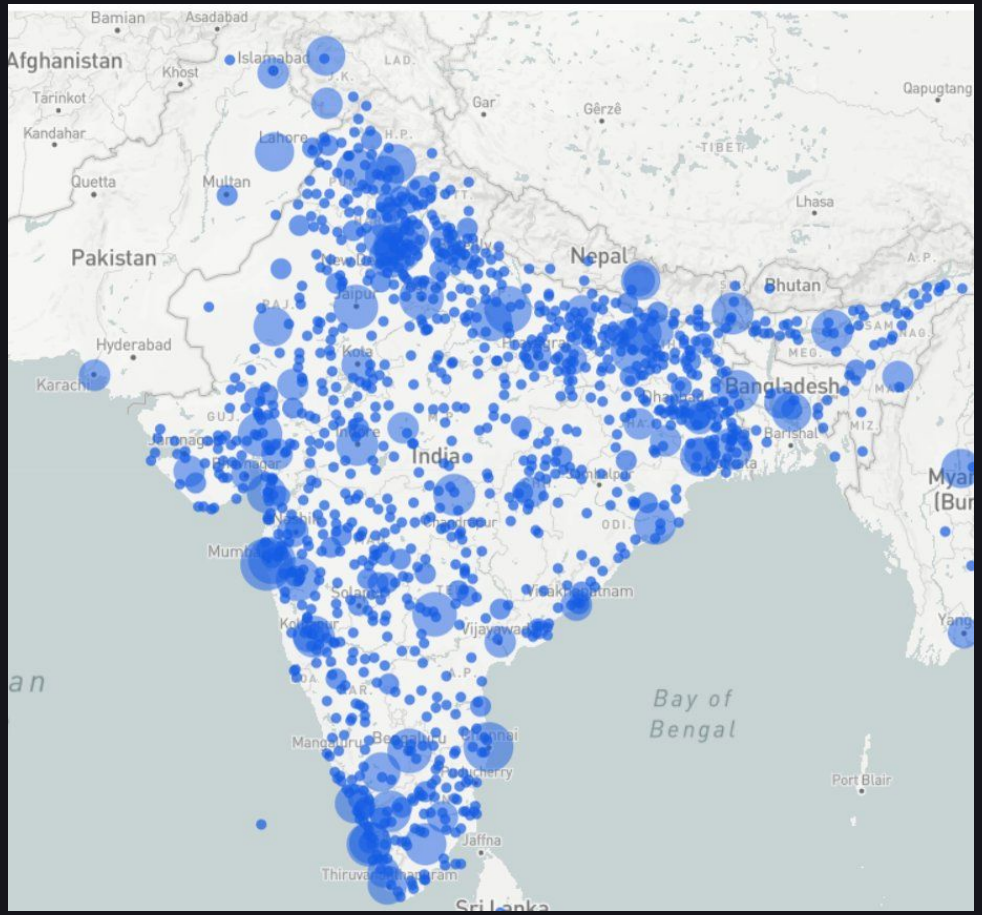
day	rx	tx	total	avg. rate
2024-02-10	43.79 GiB	173.27 GiB	217.07 GiB	21.58 Mbit/s
2024-02-11	29.83 GiB	371.12 GiB	400.95 GiB	39.86 Mbit/s
2024-02-12	58.03 GiB	277.40 GiB	335.43 GiB	33.35 Mbit/s
2024-02-13	23.86 GiB	250.88 GiB	274.73 GiB	27.31 Mbit/s
2024-02-14	36.13 GiB	329.15 GiB	365.28 GiB	36.32 Mbit/s
2024-02-15	45.74 GiB	354.15 GiB	399.89 GiB	39.76 Mbit/s
2024-02-16	37.76 GiB	270.07 GiB	307.83 GiB	30.60 Mbit/s
2024-02-17	42.93 GiB	307.81 GiB	350.74 GiB	34.87 Mbit/s
2024-02-18	23.17 GiB	269.96 GiB	293.13 GiB	29.14 Mbit/s
2024-02-19	43.07 GiB	570.14 GiB	613.21 GiB	60.97 Mbit/s
2024-02-20	26.40 GiB	364.87 GiB	391.27 GiB	38.90 Mbit/s
2024-02-21	41.67 GiB	504.40 GiB	546.07 GiB	54.29 Mbit/s
2024-02-22	40.35 GiB	596.03 GiB	636.38 GiB	63.27 Mbit/s
2024-02-23	44.80 GiB	542.55 GiB	587.34 GiB	58.39 Mbit/s
2024-02-24	41.33 GiB	315.88 GiB	357.20 GiB	35.51 Mbit/s
2024-02-25	18.60 GiB	373.09 GiB	391.69 GiB	38.94 Mbit/s
2024-02-26	47.34 GiB	249.73 GiB	297.07 GiB	29.53 Mbit/s
2024-02-27	30.56 GiB	265.59 GiB	296.15 GiB	29.44 Mbit/s
2024-02-28	28.07 GiB	287.21 GiB	315.27 GiB	31.34 Mbit/s
2024-02-29	35.18 GiB	277.76 GiB	312.94 GiB	31.11 Mbit/s
2024-03-01	77.67 GiB	268.20 GiB	345.87 GiB	34.39 Mbit/s
2024-03-02	31.83 GiB	191.03 GiB	222.86 GiB	22.16 Mbit/s

Daily traffic data from Nagpur Server.



Serves ~20TB traffic per month and handle ~20k users daily.


104 COUNTRIES
2733 CITIES
Copy URL
New Map
Worldwide 17207 IPs
India 7543 IPs
United States 1934 IPs
North Atlantic Ocean
Indian Ocean
Pacific Ocean
Greenland
Iceland
Svalbard
Canada
Russia
China
Kazakhstan
Mongolia
Australia
Brisbane
Perth
Melbourne
New Zealand
Papua New Guinea
Madagascar
Namibia
South Africa
Angola
Ethiopia
Kenya
Tanzania
Nigeria
Mali
Niger
Senegal
Algeria
Libya
Egypt
Tunisia
Sweden
Norway
Ukraine
Turkmenistan
Yemen
Oman
Kinshasa
Venezuela
Chile
Bolivia
Peru
Brazil
Cuba

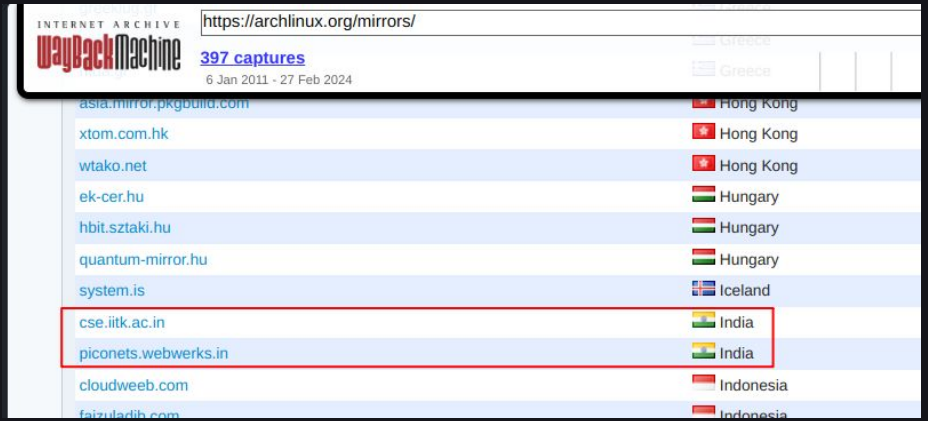


Note that this map is 1 year old
And the location is approximate based on geoip, I don't store this information anymore and this was purely collected for analytical purpose.

Currently hosts mirrors for 10 projects.

Why did I start hosting a mirror?

Archived copy of archlinux's mirrorlist from 2021[#1]

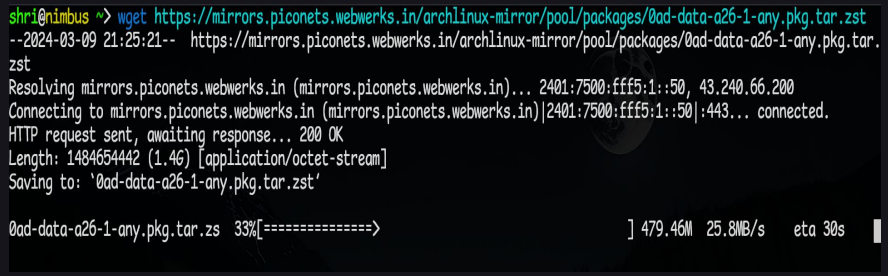


Only two Indian mirrors were present.

#1 https://web.archive.org/web/20211214035259/https://archlinux.org/mirrors/

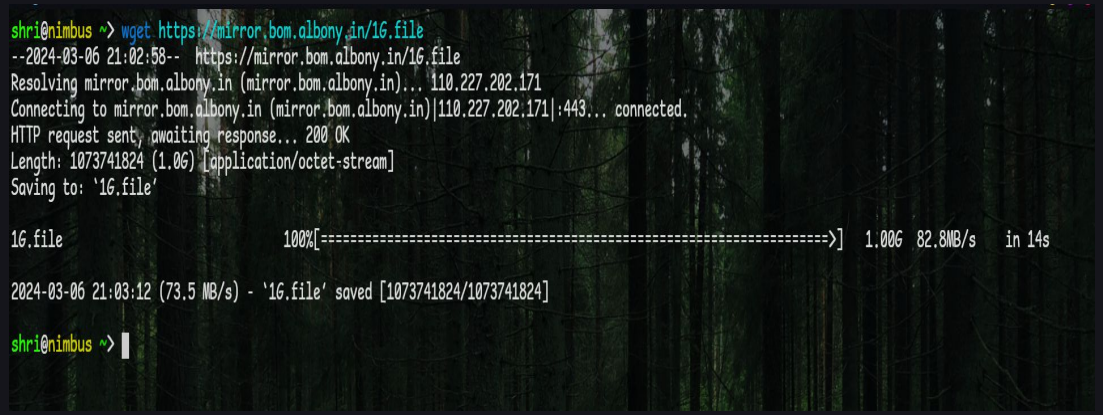
Speedtests!

Downloads from IITK mirror ~600KB/s (**4.8M**)



Downloads from piconet mirror ~25Mb/s (**200M**)

Apart from this, both of them have outdated state, i.e they don’t sync up that often with the Tier1 mirror.



Downloads from my Mumbai node ~**650M** (82Mb/s)

mirror.bom.albony.in

```
emerald :: ~ 130 » wget http://mirror.maa.albony.in/1G.file -O /dev/null

--2024-02-28 12:12:45--  http://mirror.maa.albony.in/1G.file
Resolving mirror.maa.albony.in (mirror.maa.albony.in)... 122.165.76.23
Connecting to mirror.maa.albony.in (mirror.maa.albony.in)|122.165.76.23|:80... connected.
HTTP request sent, awaiting response... 200 OK
Length: 1073741824 (1.0G) [application/octet-stream]
Saving to: ‘/dev/null’

/dev/null                          100%[=================================================================>]   1.00G  16.9MB/s    in 1m 48s

2024-02-28 12:14:33 (9.50 MB/s) - ‘/dev/null’ saved [1073741824/1073741824]
```

(136 Mbits/s) from ~Delhi (100M plan)

```
wget http://mirror.nag.albony.in/1G.file
--2024-02-25 22:33:10--  http://mirror.nag.albony.in/1G.file
Resolving mirror.nag.albony.in (mirror.nag.albony.in)... 122.169.102.70
Connecting to mirror.nag.albony.in (mirror.nag.albony.in)|122.169.102.70|:80... connected.
HTTP request sent, awaiting response... 200 OK
Length: 1073741824 (1.0G) [application/octet-stream]
Saving to: ‘1G.file’

1G.file                            100%[=================================================>]   1.00G  23.5MB/s    in 49s

2024-02-25 22:34:00 (20.7 MB/s) - ‘1G.file’ saved [1073741824/1073741824]
```

BSNL Kerala 192 Mbits/s (200M plan)

```
anurag@host01 ~> wget http://mirror.maa.albony.in/1G.file -O /dev/null
--2024-03-03 18:43:22--  http://mirror.maa.albony.in/1G.file
Resolving mirror.maa.albony.in (mirror.maa.albony.in)... 122.165.76.23
Connecting to mirror.maa.albony.in (mirror.maa.albony.in)|122.165.76.23|:80... connected.
HTTP request sent, awaiting response... 200 OK
Length: 1073741824 (1.0G) [application/octet-stream]
Saving to: '/dev/null'

/dev/null                          40%[==================================>                                  ] 411.49M  12.4MB/s    eta 52s
```

From Rohtak 100 Mbit/s

```
anurag@host01 ~/tmp> wget -O /dev/null https://mirror.albony.xyz/cachylinux/llvm-bolt-17.tar.zst
--2024-02-21 17:44:06--  https://mirror.albony.xyz/cachylinux/llvm-bolt-17.tar.zst
Resolving mirror.albony.xyz (mirror.albony.xyz)... 104.21.86.189, 172.67.136.25, 2606:4700:3033::ac43:8819, ...
Connecting to mirror.albony.xyz (mirror.albony.xyz)|104.21.86.189|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 785839576 (749M) [application/octet-stream]
Saving to: '/dev/null'

/dev/null                          100%[=====================================================>] 749.43M  26.5MB/s    in 18s

2024-02-21 17:44:24 (40.9 MB/s) - '/dev/null' saved [785839576/785839576]

anurag@host01 ~/tmp>
```

From Rohtak 212 Mbits/s

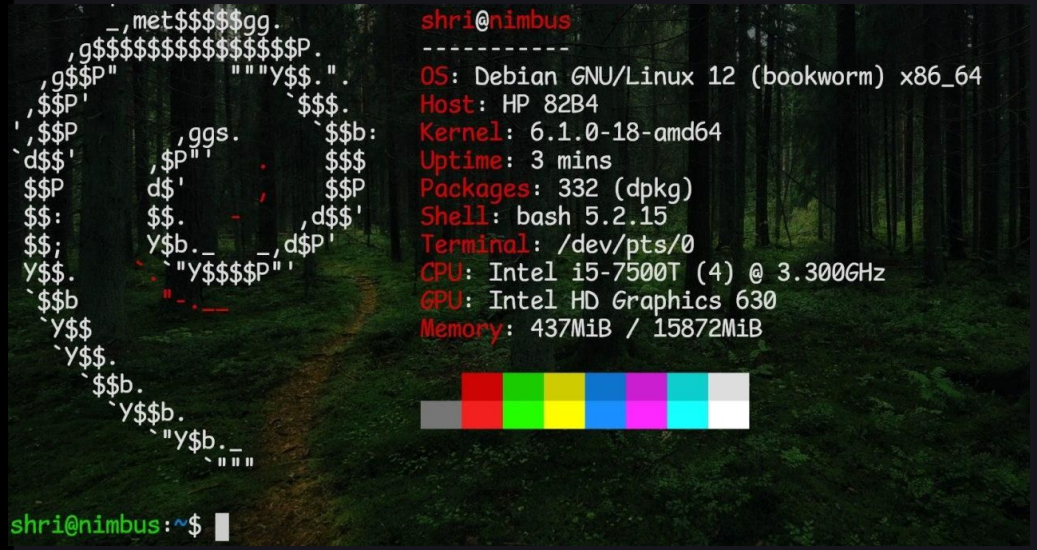
How do my servers look like?

You might think they look like this

But instead..

My setup when I started hosting the mirror

New server!

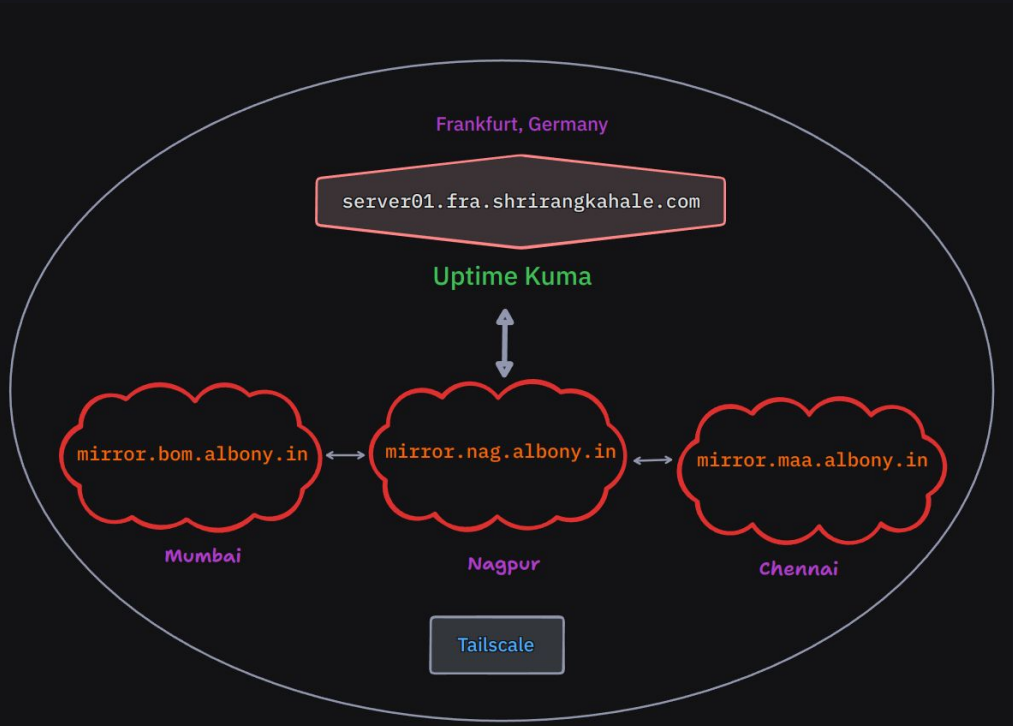


Has 1x 1TB SSD and 1x 1TB HDD

= **2TB Storage**

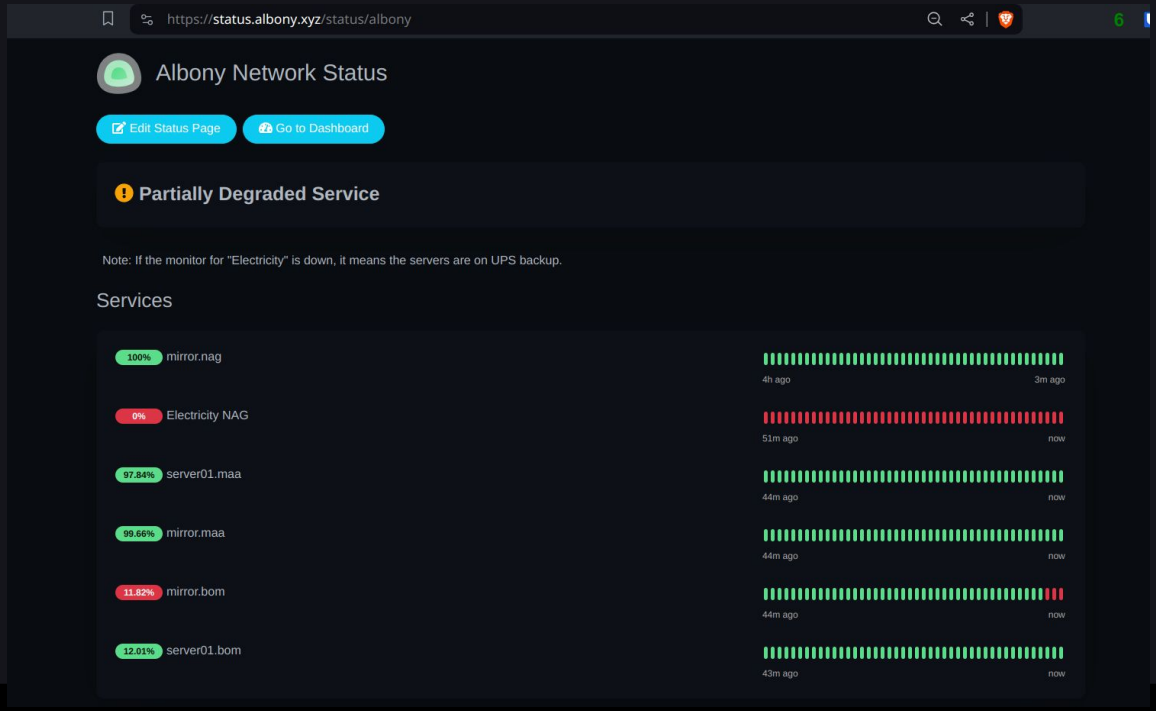
Monitoring

If my server goes down, how do I know?



Uptime Kuma

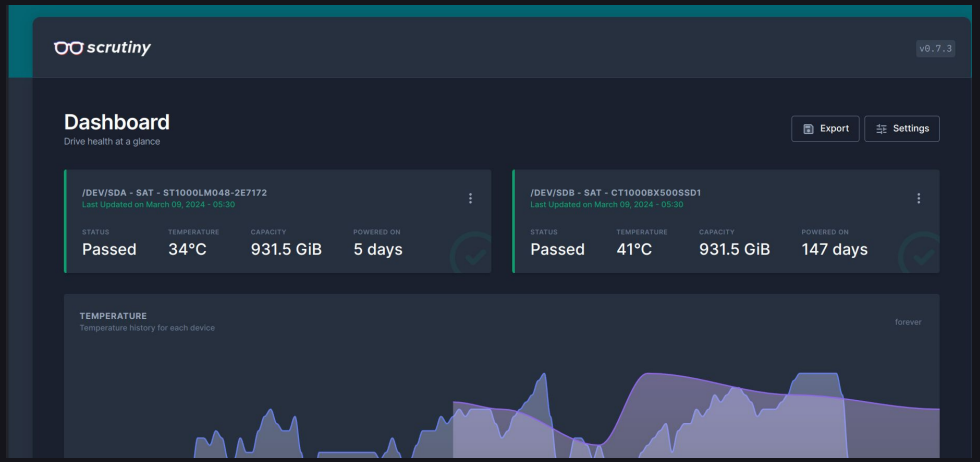
Hosted on a VPS at Frankfurt

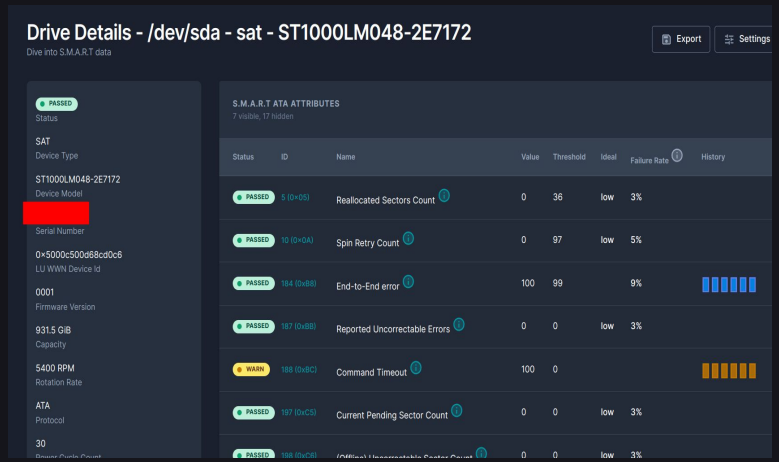


Notification

Message from monitoring bot indicating that the Mumbai node is down:

Storage Drive Monitoring





➜ Gathers data via S.M.A.R.T once a day
➜ Can be used to send warnings using email if the drive fails a test,

https://en.wikipedia.org/wiki/Self-Monitoring,_Analysis_and_Reporting_Technology

Causes of downtime

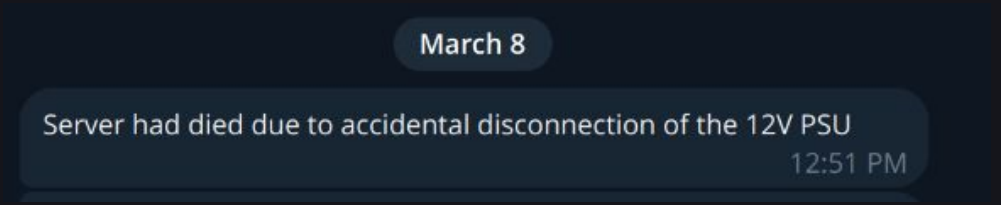
➜ Mostly due to storage drive failure

➜ Monitoring isn’t reliable for drives. (4 dead drives so far)

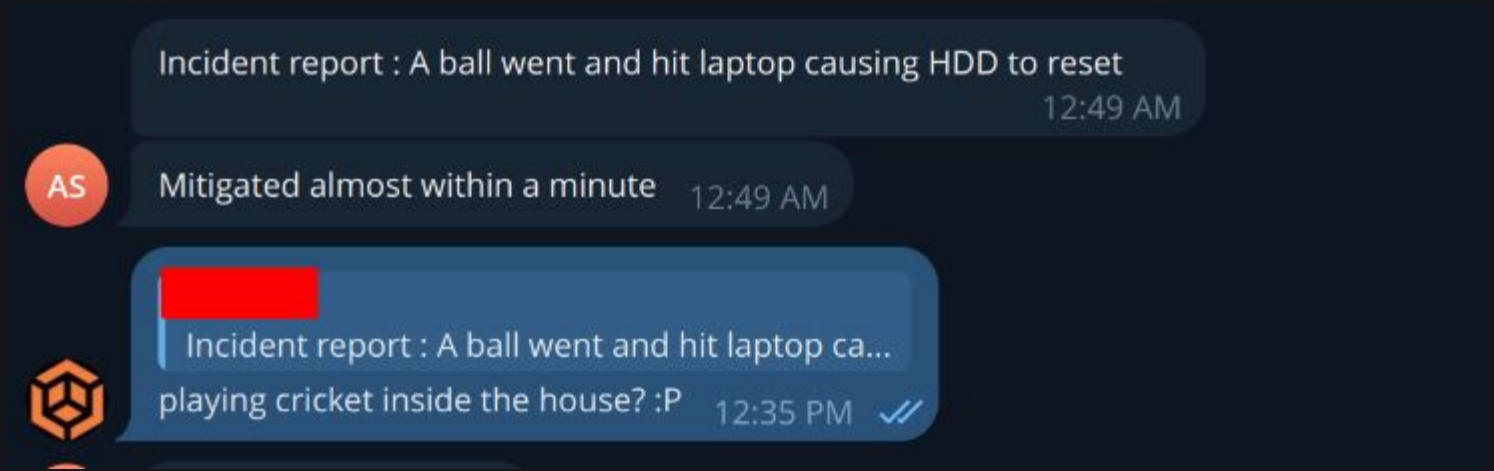
➜ Power outages (UPS exhausted)

➜ Accidental disconnection of some cable.

➜ ISP outage (rare)



And then there are things like this..



(Message from the guy hosting the Chennai Node)

Downsides of hosting a server at home

Any Questions?

Thank You!

Contact:

Website: shrirangkahale.com
Email: me@shrirangkahale.com

Twitter: @albonycal