**Конова Е. В.**

**Working with remote servers**

***git branch name: jbrowser***

**Questions**

**[0.4] What are computer ports on a high level?**

A port in computer networking is how a computer can use a single physical network connection to handle many incoming and outgoing requests by assigning a port number to each. The numbers go from 0 to 65535, which is a 16-bit number.

Some of these port numbers are specifically defined and always associated with a specific type of service -- for example, File Transfer Protocol (FTP) is always port number 21 and Hypertext Transfer Protocol web traffic is always port 80. These are called well-known ports and go from 0 to 1023.

The numbers from 1024 to 49151 are called registered ports and can be registered with the Internet Assigned Numbers Authority for a specific use. The numbers 49152 to 65535 are unassigned, can be used by any type of service and are called dynamic ports, private ports or ephemeral ports.

**How many ports are there on a typical computer?**

В каждом компьютере, сервере, маршрутизаторе существует ровно 65 535 портов. Конечно, используются они не все и есть свободные «адреса».

С 0 до 1023-го – это зарезервированные порты для систем как Windows, так и Linux. Далее с 1024 по 49151 идут свободно используемые входы. То есть их могут использовать отдельные приложения, утилиты или даже системные службы. Некоторые программы могут одновременно использовать один и тот же номер. Остальные порты, можно сказать – находятся в свободном полете и могут использоваться или не использоваться по усмотрению ОС или пользователя.

**[0.4] What is the difference between http, https, ssh, and other protocols? In what sense are they similar? Name default ports for several data transfer protocols.**

Both ssh and HTTP are protocols to communicate between client and server. Following are the basic difference between SSH and HTTP.

SSH means “Secure Shell”. It has a built-in username/password authentication system to establish a connection. It uses Port 22 to perform the negotiation or authentication process for connection. Authentication of the remote system is done by the use of public-key cryptography and if necessary, it allows the remote computer to authenticate users.

HTTP (Http is the same as https, but the second one is safer) means HyperText Transfer Protocol. HTTP is the underlying protocol used by the World Wide Web and this protocol defines how messages are formatted and transmitted, and what actions Web servers and browsers should take in response to various commands.

**Default ports for several data transfer protocols:**

SMTP – 25

HTTP - 80

HTTPS - 443

FTP - 20, 21

RDP – 3389

SSH - 22

DNS - 53

DHCP - 67, 68

**[0.4] Explain briefly: (1) what is IP, (2) what IPs are called 'white'/public, (3) and what happens when you enter 'google.com' into the web browser.**

1. An IP address is a string of numbers separated by periods. IP addresses are expressed as a set of four numbers — an example address might be 192.158.1.38. Each number in the set can range from 0 to 255. So, the full IP addressing range goes from 0.0.0.0 to 255.255.255.255.

2**. Public IP address**

These are public (global) addresses that are used on the Internet. A public IP address is an IP address that is used to access the Internet. Public IP addresses can be routed on the Internet, unlike private addresses.

**Private IP address**

Private (internal) addresses are not routed on the Internet, and no traffic can be sent to them from the Internet; they are only supposed to work within the local network.

3. **what happens when you enter 'google.com' into the web browser**

1. You type a URL in your browser and press Enter
2. Browser looks up IP address for the domain
3. Browser initiates TCP connection with the server
4. Browser sends the HTTP request to the server
5. Server processes request and sends back a response
6. Browser renders the content

**[0.4] What is Nginx? How does it work on the high level? List several alternative web servers.**

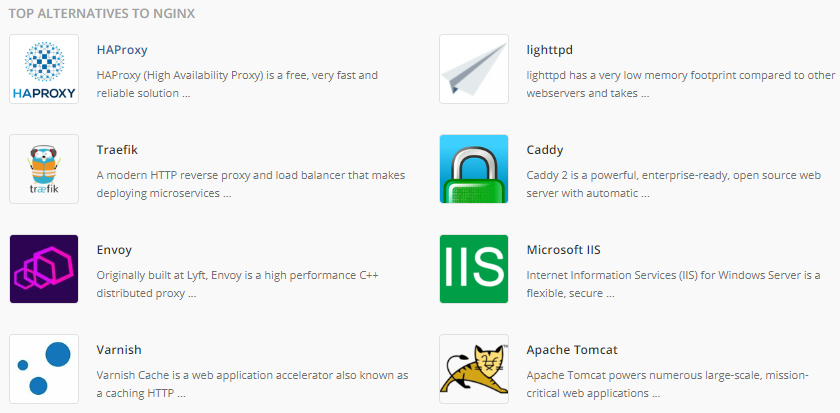
NGINX is open source software for web serving, reverse proxying, caching, load balancing, media streaming, and more.

**How does it work on the high level?**

NGINX uses a predictable process model that is tuned to the available hardware resources:

* The master process performs the privileged operations such as reading configuration and binding to ports, and then creates a small number of child processes (the next three types).
* The cache loader process runs at startup to load the disk‑based cache into memory, and then exits. It is scheduled conservatively, so its resource demands are low.
* The cache manager process runs periodically and prunes entries from the disk caches to keep them within the configured sizes.
* The worker processes do all of the work! They handle network connections, read and write content to disk, and communicate with upstream servers.

**List several alternative web servers.**



**[0.4] What is SSH, and for what is it typically used? Explain two ways to authenticate in an SSH server in detail.**

SSH or Secure Shell is a network communication protocol that enables two computers to communicate (c.f http or hypertext transfer protocol, which is the protocol used to transfer hypertext such as web pages) and share data. An inherent feature of ssh is that the communication between the two computers is encrypted meaning that it is suitable for use on insecure networks.

**for what is it typically used?**

SSH is often used to "login" and perform operations on remote computers but it may also be used for transferring data.

**Explain two ways to authenticate in an SSH server in detail.**

**Using SSH Key for authentication**

The SSH public key authentication has four steps:

1. Generate a private and public key, known as the key pair. The private key stays on the local machine.

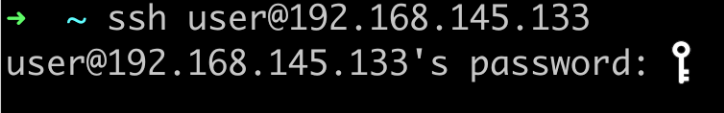
2. Add the corresponding public key to the server.

3. The server stores and marks the public key as approved.

4. The server allows access to anyone who proves the ownership of the corresponding private key.

The model assumes the private key is secured. Adding a passphrase to encrypt the private key adds a layer of security good enough for most user-based cases. For automation purposes, key management software and practices apply since the private key stays unprotected otherwise.

**Username and password combination** is the most popular authentication method to SSH server and is normally enabled by default. Username and password authentication is also the most familiar method as it's widely used everywhere else.



**Work**

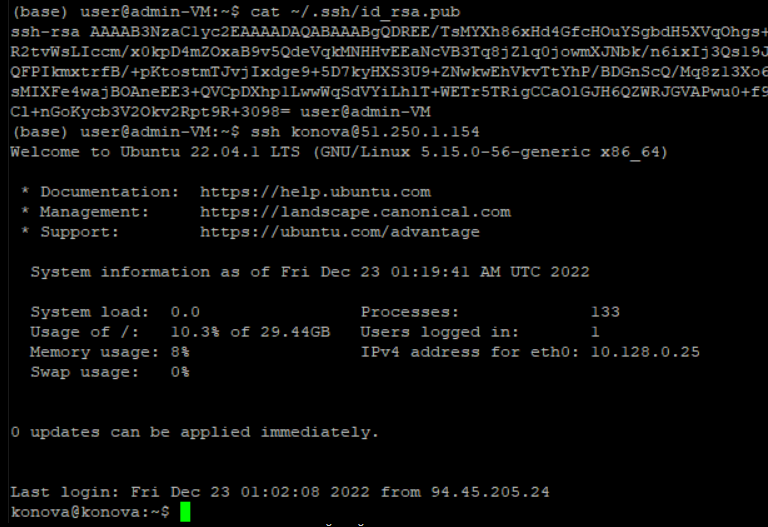
**Remote Server:**

* **[2] Create a new virtual machine in the Yandex/Mail/etc cloud (order at least 10GB of free disk space). Generate SSH key pair and use it to connect to your server.**

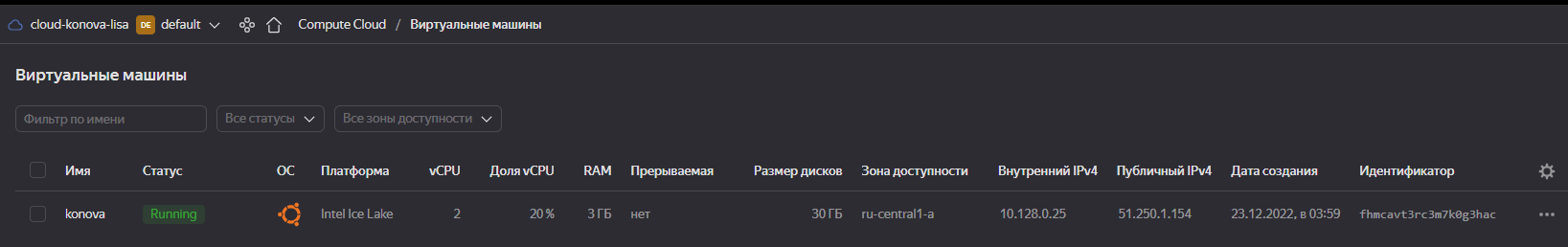
Generate SSH key using command below:

*ssh-keygen -t rsa*

And show the key and connect to the server using ssh:

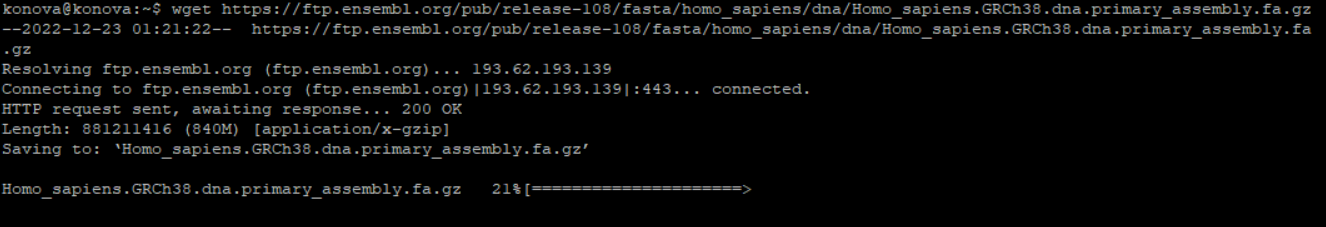


Before login we created a VM in Yandex.Cloud:



* **[1] Download the latest human genome assembly (GRCh38) from the Ensemble FTP server ([fasta](https://ftp.ensembl.org/pub/release-108/fasta/homo_sapiens/dna/Homo_sapiens.GRCh38.dna.primary_assembly.fa.gz" \t "_blank),**[**GFF3**](https://ftp.ensembl.org/pub/release-108/gff3/homo_sapiens/Homo_sapiens.GRCh38.108.gff3.gz)**). Index the fasta using samtools (samtools faidx) and GFF3 using tabix.**

I used wget for downloading, one of the examples is below:

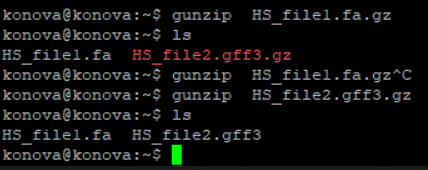


I have already installed tabix for this moment. I used the command *sudo apt-get install -y tabix*

Rename files:

*mv Homo\_sapiens.GRCh38.108.gff3.gz HS\_file2.gff3.gz*

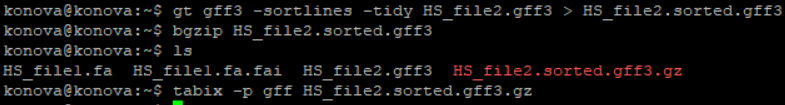
and unzip:



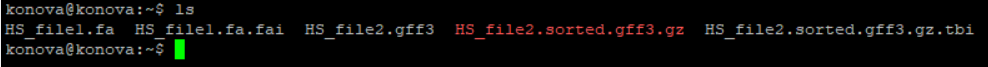
Index the fasta file using samtools faidx:



And gff3 using tabix. We need to sort file and zip it. After this, we can work with file.



Lets look at the result:



* **[1] Select and download BED files for three ChIP-seq and one ATAC-seq experiment from the ENCODE (use one tissue/cell line). Sort, bgzip, and index them using tabix.**

I used files from my ML homework.

Lets download all using *wget*, *unzip, sort, zip* again and index using tabix. Look at final files (yellow):

