Subsection Systems and methods of data processing

INFORMATION TELECOMMUNICATIONS OF PUSHCHINO RADIO ASTRONOMY OBSERVATORY, ASTRO SPACE CENTER OF LEBEDEV PHYSICAL INSTITUTE

D.V. Dumsky[1], E.A. Isaev[1,2,3], V.A. Samodurov[1,2], S.F. Likhachev[4], M.V. Shatskaya[4], M.A. Kitaeva[1], A.Yu. Zaytcev[3], I.L. Ovchinnikov[1], V.V. Kornilov[2,3]

[1] Pushchino Radio Astronomy Observatory ASC LPI, Pushchino, Russia, *dumsky@prao.ru*
[2] National research university Higher school of economics, Moscow, Russia, *is@itaec.ru*
[3] Institute of Mathematical Problems of Biology, Russian Academy of Sciences, Pushchino, Russia,
[4] Astro Space Center LPI, Moscow, Russia

ABSTRACT. Buffer data center was created in the territory of the Pushchino Radio Astronomy Observatory three years ago. The necessity of its creation was caused by the high requirements to the speed and quality of the transmission large amounts of scientific and telemetry data received by tracking station RT-22 from the space radio telescope of the international project “Radioastron”. The transfer of this data is carried out over a long distance over 100 km from the Pushchino to Moscow center of processing and storage ASC FIAN. And now we use the data center as a center of local network of the Observatory.

Key words: Telecommunications: networks: virtualization.

1. Network

Buffer data center (BDC) was housed in the Pushchino Radio Astronomy Observatory (PRAO) polygon building in a server room isolated from external environment with cooling, raised floors, racks for network and computer equipment and has a sufficient number of UPS devices for uninterruptible power supply switches and servers in case of power failure. Observatory network communication center was moved from main building to the BDC this year (Fig. 1). And it is now used not only as the communication node for gigabit Ethernet channel to the Moscow data center and “Radioastron” buffer data storage. Along with the move of the central node of PRAO LAN some modernization of network equipment working on the second and third level network protocols took place. As the main gateway to the Internet and a network of Pushchino Research Center today we use the router Mikrotik RB1100AHx2 which replaced the old Linux router. This device provides not only the routing, but also the function of a stateful firewall and DHCP server for local network of the observatory and has many other features such as traffic filtering and shaping, dynamic routing, network monitoring and so on.

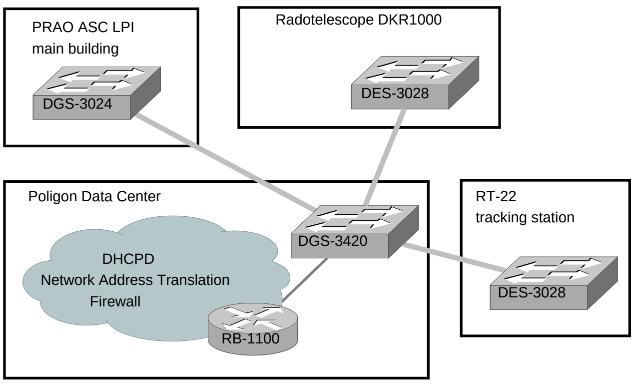

Figure 1: PRAO network diagram.

On distribution level we use managed Layer 2+ Gigabit switches D-link DGS-3420 with the ability to move in the future 10 gigabit Ethernet.

2. Local servers

Two servers mounted in the BDC we use as a virtualization platform based on GNU/Linux system. It helps us to isolate different network services such as DNS, NTP, WWW, email, sip and database from each other. One service - one virtual container. We did this to increase the safety and reliability of these services.

On one of the servers we use virtualization based on an open platform OpenVz. And on another – system virtualization based on KVM (Kernel-based Virtual Machine) because it is more powerful (Intel Xeon 22.53-2.80 GHz and 24 GB of RAM, Adaptec RAID controller, 3.6 Tb disk space) and support hardware virtualization. KVM unlike Openvz allows one to run virtual servers running different operating systems, not only Linux. Last year we tried to use a virtualization system based on Linux container (LXC), but because it was unstable, we were forced to go to the KVM virtualization.

We have two file storage servers in the data center. One of which is used for backup storage project "Radioastron" (20 Tb disk space) and second storage data of radio astronomy observations PRAO and of its processing (48 Tb). All servers are connected to a local network with two or more Ethernet cards to increase the speed of information exchange between servers and a local area network, and also for increase redundancy. This is possible thanks to the ability of our Gigabit switch D-link DGS-3420 to combine multiple twisted-pair or fiber Ethernet links into one fault-tolerant and load balanced logical link. In Linux systems this feature is called bonding, in switches its commonly referred to as a port channel or link aggregation. Either way, its using the LACP (802.1ad) protocol for that.

PROSPECTS OF CLOUDY TECHNOLOGIES IN THE SOLUTION OF THE TASK OF THE ANALYSIS OF LARGE VOLUMES OF THE DATA OBTAINED IN ASTRONOMICAL SUPERVISION

E.A. Isaev [1,2], V.V. Kornilov [2,3], V.A. Samodurov [1,2], A.A. Tarasova [4]

[1] Pushchino Radio Astronomy Observatory ASC LPI, Pushchino, Russia, *is@itaec.ru*
[2] National research university Higher school of economics, Moscow, Russia
[3] Institute of Mathematical Problems of Biology RAS, Pushchino, Russia
[4] FIAN, Moscow, Russia

ABSTRACT. In modern astronomy the problem of the big data obtained during scientific experiments is very actual. One of the perspective approaches to the solution of the problem of processing of superlarge volumes of experimental data in real time is the use of the technology of cloud computing which assumes ensuring remote dynamic access of users to services, computing resources and appendices on the Internet.

Key words: Big Data, Cloud Computing

The success in the development of modern computer technologies and electronics led to possibility of creation of scientific experimental installations of the new generation which characteristic is the large volumes of the data obtained during scientific experiments. Especially it is actual for astronomy and astrophysics. Devices for astronomical observing allow to obtain data with more and more high resolution, the observations of astronomical objects are conducted not only in visible light, and in all range of an electromagnetic range, thus the only observation which can last from several seconds to several minutes, gives from several megabytes to several gigabytes of information. Such astronomical projects as "Radioastron", "Millimetron", "Kvazar-KVO", "RATAN-600", PAN-STARRS, SDSS, LOFAR, ASKAP, SKA and others are capable to generate tens and hundreds gigabytes of the supervision given for every second. So, for example, the data archive of the Telescope of Hubble for 15 years of its work is about 25 Tb [1]; the Large Survey Telescope (LSST) with a diameter of mirror of 8.4 meters and a 3 Gigapixels matrix will make 30 Tb of data only for one night, and the full volume of supervisory archive is estimated at 200 Petabyte [2].

The data obtained during experiments should be able to be stored, processed, transferred and analyzed to receive the new knowledge from these data. Enormous volumes of these data and high speed of their increasing do these tasks rather difficult for the effective decision.

The term Big data usually use for data sets with sizes beyond the ability of commonly used software tools to capture, curate, manage, and process the data within a tolerable elapsed time. In 2012, Gartner did the definition of Big data as follows: "Big data is high volume, high velocity, and/or high variety information assets that require new forms of processing to enable enhanced decision making, insight discovery and process optimization." [3]. Additionally, the new V: "Veracity" and "Visualization" now are often added to describe Big data [4].