УДК 550.34.034

РАЗВИТИЕ ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ В КАЗАХСТАНСКОМ НАЦИОНАЛЬНОМ ЦЕНТРЕ ДАННЫХ (KNDC)

Гордиенко Д.Д., Комаров И.И., Смирнов Ю.А.

Институт геофизических исследований, Курчатов, Казахстан

Рассмотрена реализация плана по совершенствованию, модернизации и введению новых программно-технических средств в Казахстанском национальном центре данных с целью повышения качества сбора, обработки и хранения данных в соответствии с современными требованиями и тенденциями развития информационных технологий.

Введение

В 1999 г. на базе Института геофизических исследований в г. Алматы был создан Центр сбора и обработки специальной сейсмической информации, исполняющий в настоящее время функции Казахстанского национального центра данных (KNDC) в Международной системе мониторинга (в рамках Договора о всеобъемлющем запрещении ядерных испытаний – ДВЗЯИ) [1].

Основными задачами KNDC являются сбор и передача данных с сейсмических и инфразвуковых станций, обработка поступающих данных, хранение и обмен данными с другими национальными и международными центрами, проведение различных научных исследований. Для решения этих и других задач на базе Центра создана сложная информационная система, включающая систему наблюдений, телекоммуникационное оборудование, каналы связи, компьютерное оборудование и программное обеспечение. Как любая другая система, информационная система Центра данных нуждается в постоянной модернизации в силу многих причин, таких, как применение новых версий программного обеспечения, возросшие требования к безопасности, увеличение объемов передаваемых и накопленных данных, дополнительные требования к качеству данных и др. В 2016 г. был разработан план модернизации информационной системы Центра данных, реализация которого была начата в 2017 г. и продолжается по настоящее время.

Система наблюдений и коммуникаций и обмен данными

На рисунке 1 показано расположение сейсмических и инфразвуковых станций на территории Казахстана, находящихся под оперативным управлением РГП «Институт геофизических исследований» [2], а на рисунке 2 – телекоммуникационная система передачи данных от сейсмических и инфразвуковых станций в KNDC и другие организации, сотрудничество с которыми ведется по ряду международных договоров и соглашений.

Из рисунка 2 видно, что, исходные сейсмические данные (в режиме реального времени) передаются в Международный центр данных в Вену, в Американский национальный центр данных во Флориду, в Центр международного консорциума IRIS. Ведется обмен результатами обработки данных. Так, автоматический сейсмологический бюллетень передается в Европейский средиземноморский центр (EMSC) в Париж (Франция), заключительный интерактивный сейсмологический бюллетень поступает в Международный сейсмологический центр (ISC) в Англию, сводки по сильным землетрясениям мира передаются в Центр геофизической службы России в г. Обнинск.

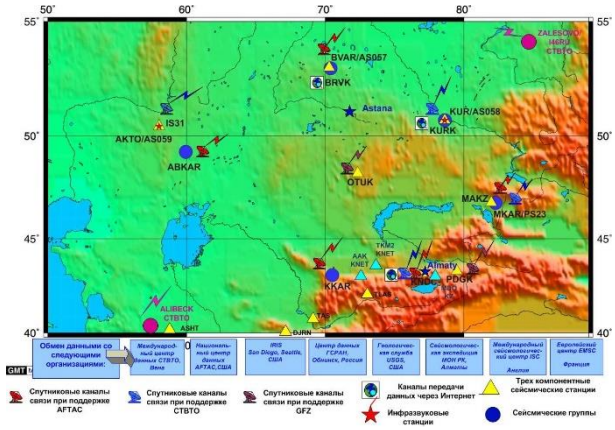


Рисунок 1. Расположение сейсмических и инфразвуковых станций, работающих под оперативным управление Института геофизических исследований

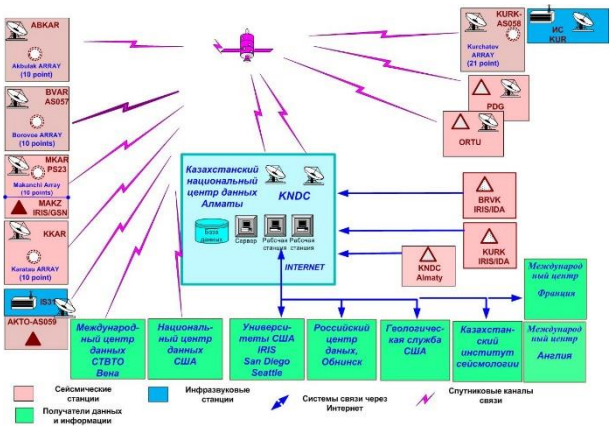


Рисунок 2. Коммуникационная схема передачи данных сети станций, работающих под оперативным управлением Института геофизических исследований

По запросу данными Центра пользуются специалисты разных стран. В то же время специалистам Центра данных в Казахстане доступна любая информация, имеющаяся в Международных центрах данных, как исходная, так и информация с результатами обработки.

Система контроля качества данных и мониторинг информационных систем

Объем данных, передаваемых коммуникационной системой, по современным меркам, невелик и составляет порядка 800 Мб принятых и 600 Мб отправленных данных в сутки. Однако к большинству данных предъявляются высокие требования по качеству и своевременности получения. Наиболее высокими являются требования к качеству данных первичных станций мониторинга, входящих в МСМ, таким как сейсмическая группа PS23-Маканчи (MKAR) [3] и инфразвуковая станция IS31-Актюбинск (I31KZ) [4]. Согласно требованиям МСМ объем полученных и передаваемым данным по этим станциям должен составлять более 98 процентов от максимально возможного.

На рисунке 3 приведена типовая схема маршрутизации данных для отдельно взятого пункта наблюдения, из которого видно, что данные на пути от сейсмической станции до системы хранения проделывают путь, включающий кольцевые буферы, преобразование форматов, маршрутизацию по локальным сетям, пересылку по каналам связи, копирование, архивирование и, в конечном итоге, обработку и хранение.

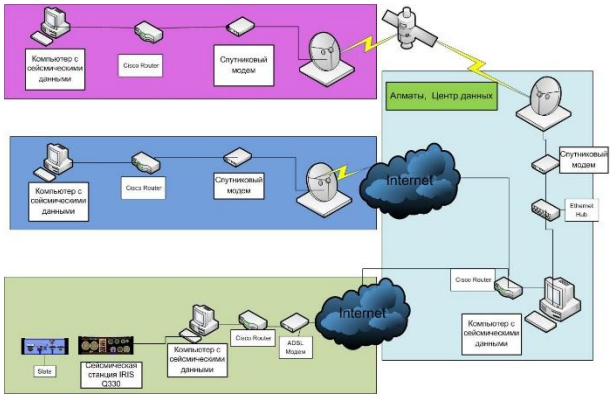


Рисунок 3. Типовая схема маршрутизации данных

На каждом из этих этапов возможны сбои, влекущие потери данных или задержку их поступления. Для контроля за поступлением данных создана автоматическая система, обеспечивающая расчет и статистический анализ качества и объемов поступающих данных [5]. Все результаты заносятся в специальную базу данных, доступ к которой возможен через вебсайт Центра данных. Однако система не дает исчерпывающей информации о типе, локализации и не имеет возможности сигнализировать о случившейся внештатной ситуации. В мировой практике, начиная со второй половины нулевых годов XXI столетия, для решения подобных проблем стали активно внедряться универсальные системы мониторинга, обеспечивающие контроль за состоянием объектов информационной инфраструктуры и обработку исключительных ситуаций [6]. С учетом этой практики в Центре данных на первом этапе проведено сравнение существующих в мире систем мониторинга и выбрана одна из наиболее перспективных и функциональных систем – ZABBIX [7] , которая распространятся по лицензии GNU GPL2. ZABBIX имеет богатое сообщество пользователей, хорошую поддержку и расширенный программный интерфейс (API), что делает возможным интеграцию данной системы с другими программными сервисами Центра данных и его действующим веб-сайтом. С учетом возможностей ZABBIX, применительно к задачам Центра данных, планируется осуществить мониторинг следующих типов элементов информационной инфраструктуры:

– мониторинг приложений, скриптов, сервисов;

– мониторинг почтового сервера, веб-сервера, сервера базы данных и др.;

– мониторинг журнальных файлов (файлов логирования и аудита);

– мониторинг сервера виртуализации, виртуальных машин и контейнеров;

– мониторинг элементов коммуникации (маршрутизаторы, коммутаторы);

– мониторинг периферийного оборудования (принтеры, сканеры, телефонные станции);

– мониторинг состояния элементов аппаратного обеспечения (жесткие диски, память, процессоры);

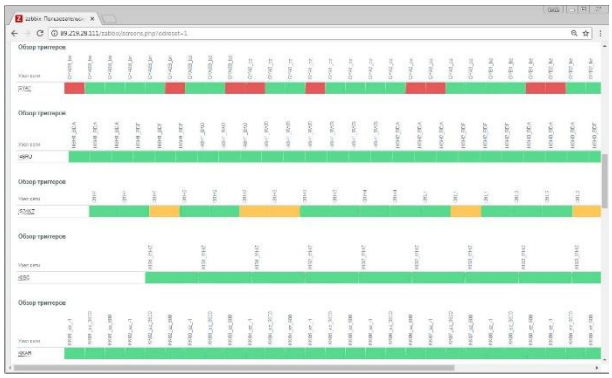
– мониторинг систем питания.

Процесс мониторинга основывается на том, что создается некий набор элементов, соответствующий набору наблюдаемых параметров. Каждый элемент, в свою очередь, связан с набором триггеров, настроенных на различные уровни входных параметров. Данные триггерам передают, так называемые агенты, получающие информацию от определенного источника. Получая и интерпретируя информацию о состоянии триггеров, ZABBIX выполняет предписанные ему действия.

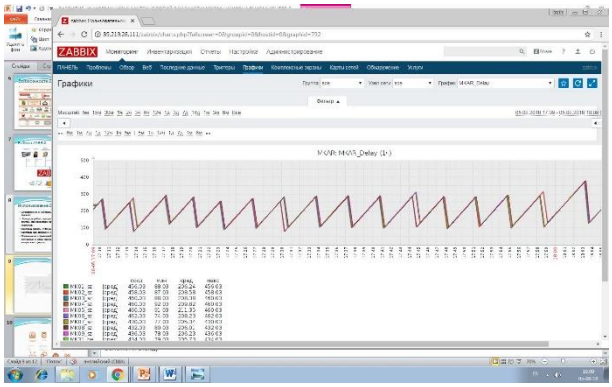
Из перечисленных возможностей ZABBIX в Центре данных реализован механизм, позволяющий системе мониторинга получать информацию о состоянии процессов поступления данных, об их задержках и статусе. Вся информация, проанализированная системой мониторинга, доступна в веб-интерфейсе ZABBIX, где можно выбрать наиболее подходящее ее представление: в зависимости от состояния триггеров (рисунок 4-а), в виде текущих значений, поступающих от агентов (рисунок 4-б).

Описанные сущности (элементы, агенты, триггеры и др.) системы мониторинга могут создаваться и настраиваться либо через веб-интерфейс, либо автоматически с помощью приложений, взаимодействующих через API ZABBIX. Ниже показан набор триг-

геров с различными пороговыми значениями (рисунок 5) и набор элементов (рисунок 6), где каждому элементу соответствует определённая сейсмическая станция, для каждой из которых создан набор триггеров, срабатывающих при определенных задержках поступления данных. В приводимом примере описывается ситуация, при которой задержка поступления данных менее 10 минут считается нормальной, но заслуживающей внимания; задержка от 10 минут до часа считается важной, требующей внимания, а задержка более часа считается чрезвычайной, при которой необходимо принимать меры.



а) представление информации о состоянии триггеров



б) представление информации в виде текущих значений наблюдаемой величины

Рисунок 4. Представление наблюдаемой информации в Веб интерфейсе системы ZABBIX

Важность	Имя ▲	Выражение
Чрезвычайная	AAK_be	{AAK:db.odbc.select[AAK_be,Delay].last()}=-1
Предупреждение	AAK_be	{AAK:db.odbc.select[AAK_be,Delay].last()}>600
Средняя	AAK_be	{AAK:db.odbc.select[AAK_be,Delay].last()}>3600
Чрезвычайная	AAK_bn	{AAK:db.odbc.select[AAK_bn,Delay].last()}=-1
Предупреждение	AAK_bn	{AAK:db.odbc.select[AAK_bn,Delay].last()}>600
Средняя	AAK_bn	{AAK:db.odbc.select[AAK_bn,Delay].last()}>3600
Чрезвычайная	AAK_bz	{AAK:db.odbc.select[AAK_bz,Delay].last()}=-1
Предупреждение	AAK_bz	{AAK:db.odbc.select[AAK_bz,Delay].last()}>600
Средняя	AAK_bz	{AAK:db.odbc.select[AAK_bz,Delay].last()}>3600

Рисунок 5. Набор триггеров системы мониторинга Zabbix для контроля поступающих данных

Имя ▲	Группы элементов данных	Элементы данных	Триггеры	Графики
AAK	Группы элементов данных 1	Элементы данных 3	Триггеры 9	Графики 1
ABKAR	Группы элементов данных 1	Элементы данных 12	Триггеры 36	Графики 1
AKTO	Группы элементов данных 1	Элементы данных 3	Триггеры 9	Графики 1
ASHT	Группы элементов данных 1	Элементы данных 6	Триггеры 18	Графики 1
BRVK	Группы элементов данных 1	Элементы данных 3	Триггеры 9	Графики 1
BVAR	Группы элементов данных 1	Элементы данных 12	Триггеры 36	Графики 1
DJRN	Группы элементов данных 1	Элементы данных 3	Триггеры 9	Графики 1
DZET	Группы элементов данных 1	Элементы данных 3	Триггеры 9	Графики 1
GYAT	Группы элементов данных 1	Элементы данных 12	Триггеры 36	Графики 1

Рисунок 6. Набор подконтрольных элементов, соответствующий списку станций

Представление информации и способ визуализации может изменяться в зависимости от задачи, возможна также группировка отдельных графиков и представлений в одну область просмотра.

На рисунке 7 показан сводный лист, где отображается состояние подконтрольных элементов и значений задержек поступающих данных по всей сети сейсмического мониторинга.

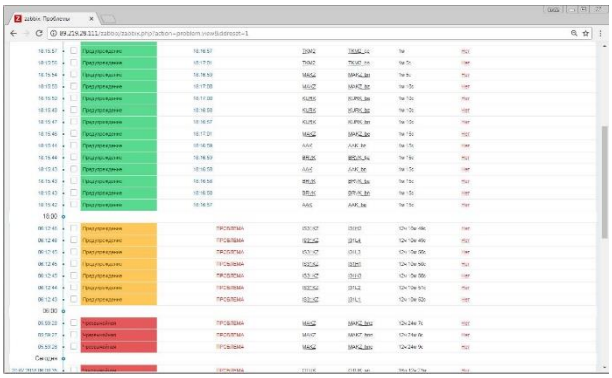


Рисунок 7. Сводный лист с информацией о задержках поступающих данных для всей сети сейсмического мониторинга

Кроме визуализации состояния элементов мониторинга через веб-интерфейс ZABBIX возможны и другие способы получения информации. Например, ZABBIX может отправлять сообщение по электронной почте, используя gsm-модем, возможна отправка sms-сообщения или же сообщения через электронный мессенджер, но крайне важной способностью ZABBIX является то, что он может сам предпринять попытки к устранению возникших проблем. Например, в некоторых случаях может быть перезагружен компьютер, перезапущена определенная программа или сервис, выполнена очистка жесткого диска и др.

СИСТЕМА ВИРТУАЛИЗАЦИИ

Следующим направлением развития информационных систем в Центре данных является внедрение и использование системы виртуализации. Актуальность данной задачи обусловлена несколькими причинами. Во-первых, начиная со времени создания Центра данных для задач сейсмической обработки основным парком компьютерной техники являлись рабочие станции SUN, работающие на платформе SPARC и операционной системе Unix, которые в настоящее время морально и физически устарели. В настоящее время происходит плановый переход на новое программное обеспечение, работающее на платформе x86 под операционной системой Linux. Следуя мировой тенденции, использование систем виртуализации позволяет значительно сократить экономические издержки на приобретение нового оборудования, когда вместо одного-двух десятков физических серверов используется один корпоративный сервер начального уровня. Во-вторых, благодаря множеству механизмов и технологий, применяемых в системах виртуализации как на программном, так и аппаратном уровне, повышается общая отказоустойчивость системы. В-третьих, система мониторинга ZABBIX имеет встроенные механизмы взаимодействия с системами виртуализации, что облегчает внедрение обоих систем.

Как и в случае выбора системы мониторинга, на первом этапе был проведен анализ существующих систем виртуализации, изучены существующие типы и способы их построения и выработано технически обоснованное решение. В качестве аппаратной платформы для построения системы виртуализации выбран сервер Supermicro CSE-825TQC. Благодаря использованию двух процессоров Xeon E5 и жестких дисков, подключенных через внешний RAID контроллер интерфейсом SAS2, сервер обладает высокими показателями производительности. Наличие двух блоков питания с функцией горячей замены обеспечивает отличные показатели отказоустойчивости. Сервер может использоваться для широкого спектра задач на предприятиях среднего масштаба. В качестве системы виртуализации выбрана система Proxmox Virtual Environment (Proxmox VE) − система виртуализации с открытым исходным кодом, основанная на Debian GNU/Linux [8]. Proxmox в качестве гипервизоров использует KVM (Kernel-based Virtual Machine) и LXC (Linux Containers) и способна выполнять любые поддерживаемые KVM операционные системы (Linux, BSD, Windows и другие) с минимальными потерями производительности. KVM— программное решение, обеспечивающее виртуализацию в среде Linux на платформе x86, которая поддерживает аппаратную виртуализацию, а LXC – система виртуализации на уровне операционной системы для запуска нескольких изолированных экземпляров операционной системы Linux на одном узле. LXC не использует виртуальные машины, а создаёт виртуальное окружение с собственным пространством процессов и сетевым стеком. Все экземпляры LXC используют один экземпляр ядра операционной системы.

На рисунке 8 показаны различия в подходе использования двух видов виртуализации – виртуальных машин и контейнеров.

Система виртуализации Proxmox обладает функциями и возможностями, основными из которых являются:

– простое управление через веб-интерфейс;
– мониторинг нагрузки в реальном времени;
– статистика и информативные графики нагрузки сервера виртуализации и каждой виртуальной машины в отдельности – по оперативной памяти, cpu, hdd, сети в разрезе последний час/день/неделя/месяц/год;
– библиотека установочных образов (в локальном или удаленном хранилище);
– подключение к «физической» консоли гостевых систем непосредственно из браузера;
– объединение серверов в кластер с возможностью живой миграции виртуальных машин (без остановки гостевой системы);
– быстрое развертывание гостевых систем из шаблонов;

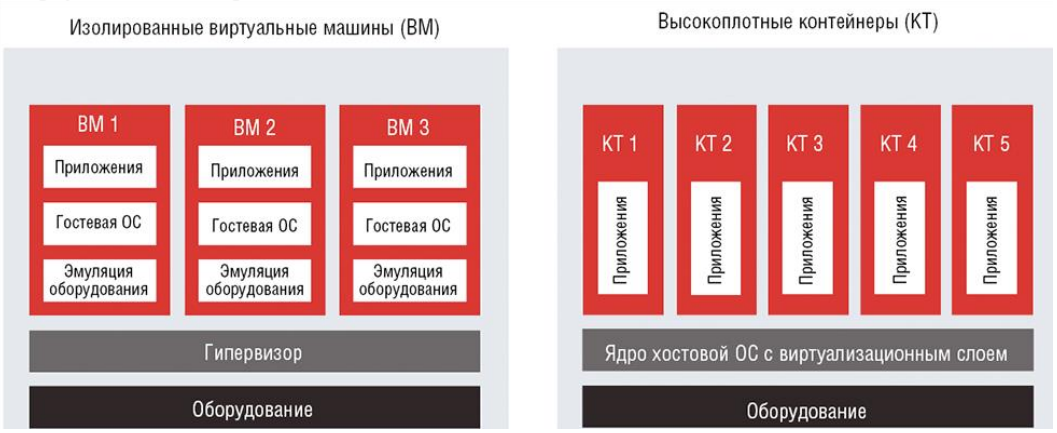


Рисунок 8. Принцип организации двух систем виртуализации – виртуальных машин и контейнеров

– сохранение образа состояния виртуальной машины (snapshot), формирование дерева состояний и возможность отката на любую из точек;

– автоматическое резервное копирование виртуальных машин.

Управление виртуальными машинами и администрирование самого сервера Proxmox проводятся через веб-интерфейс, либо через стандартный интерфейс командной строки Linux (рисунок 9). Для создаваемых виртуальных машин доступно множество опций: используемый гипервизор, тип хранилища (файл образа или LVM), тип эмулируемой дисковой подсистемы (IDE, SCSI или VirtIO), тип эмулируемой сетевой карты, количество доступных процессоров и другие. С сайта разработчиков можно загрузить готовые шаблоны (как дистрибутивы общего назначения, так и настроенные под конкретную задачу, например MySQL сервер, веб сервер, сервер контроля версионности Git и др.). Также можно создавать и собственные шаблоны, либо конвертировать настроенные операционные системы в шаблон.

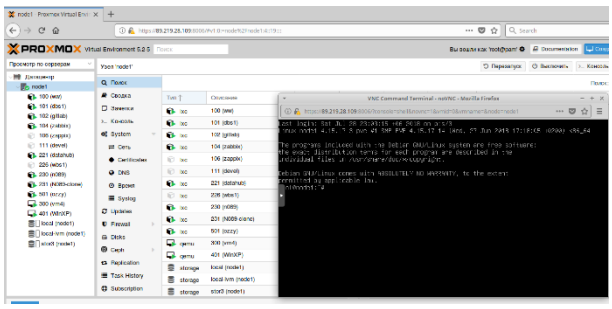


Рисунок 9. Веб интерфейс системы виртуализации Proxmox

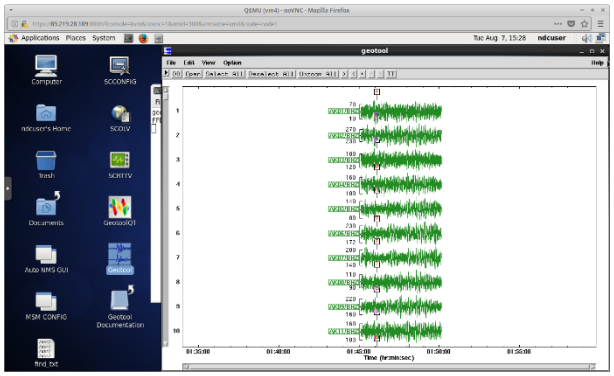


Рисунок 10. Пример работающего программного обеспечения NDC-in-a-Box на виртуальной машине с доступом через веб-браузер

Построенная система уже включена в повседневную эксплуатацию и на ее платформе работает несколько серверов и рабочих станций под управлением Linux и Windows. Так, например, созданы виртуальные машины с установленными pop3/smtp/imap4, dns, proxy и др. серверами. В среде виртуальных контейнеров работает web сервер Центра данных, сервер баз данных, сервер мониторинга ZABBIX, программное обеспечение для получения данных от отдельных сейсмических станций. На рисунке 10 показано, как на виртуальной машине Linux функционирует пакет программного обеспечения для получения и анализа сейсмических, гидроакустических, инфразвуковых и радионуклидных данных – NDC-in-a-Box, предоставляемый национальным центрам данных для работ в поддержку ДВЗЯИ [9].

Для повышения надежности организовано автоматическое резервное копирование снимков операционных систем по расписанию. Существует также резервный сервер Proxmox, на котором при необходимости можно развернуть любой из созданных архивов. В перспективе планируется приобретение и установка еще нескольких серверов для системы виртуализации Proxmox, что позволит создать кластер серверов, обеспечивающий более высокую надежность и производительность.

РАЗРАБОТКА И ВНЕДРЕНИЕ НОВОГО ПРОГРАММНОГО ОБЕСПЕЧЕНИЯ ДЛЯ ЗАДАЧ СЕЙСМИЧЕСКОГО МОНИТОРИНГА

Для создания целостной рабочей системы и в рамках перехода на операционные системы Linux необходимо написание большого количества собственных программ или адаптация существующих. Первой из приоритетных задач является установка и настройка новой версии программы для автоматического детектирования и локализации сейсмических событий DP/EP (Data Processor/Evaluation Plan), предоставленной на безвозмездной основе Норвежским национальным центром данных – NORSAR [10]. Новое программное обеспечение позволяет повысить точность локализации сейсмических событий для ряда районов Казахстана и Центральной Азии в автоматическом режиме, а также увеличить количество детектируемых событий. Второй приоритетной задачей является создание комплекса программ для загрузки и хранения данных в формате CSS3.0 (Center for Seismic Studies v.3.0) в базах данных MySQL, а также манипуляций с ними. Разработка выполнена на языке Python3. Для решения некоторых задач использовался дополнительный фреймворк ObsPy - проект с открытым исходным кодом, предназначенный для обработки сейсмических данных. Созданы и продолжают разрабатываться программы для расчета азимутов сейсмических событий, расчета магнитуд и энергетического класса, преобразования форматов, копирования и верификации данных и др.

Для повышения эффективности программного обеспечения изучен и внедрен еще один программный продукт – система контроля версий Git [11] – набор консольных утилит, отслеживающих изменения в файлах (исходных файлах программ, но не ограниченных ими). Использование Git позволяет вести корпоративную разработку, фиксировать и отслеживать изменения проектов и файлов, анализировать и проводить тестирование различных версий. Git является распределенным, не зависит от одного централь-

ного сервера, на котором хранятся файлы и работает полностью локально, сохраняя данные в папках на жестком диске - репозитории. Тем не менее, существует возможность сохранения репозитория онлайн, что сильно облегчает совместную работу над проектом команды разработчиков. Для этой цели в виртуальном пространстве Центра данных создан веб-сервер Git – GitLab (рисунок 11).

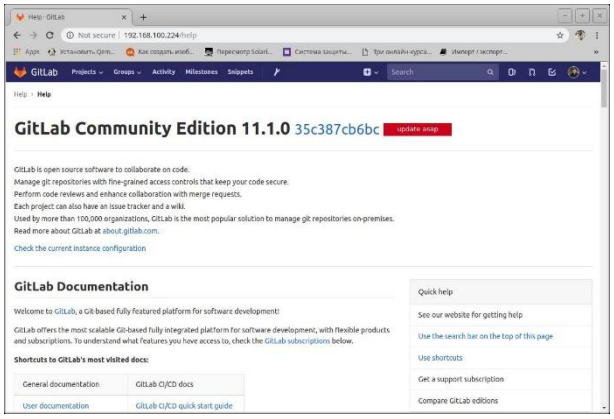


Рисунок 11. Веб-интерфейс виртуального Git-сервера GitLab

Git значительно облегчает работу на каждом из этапов разработки программного обеспечения (написание программного кода, тестирование, отладка, ввод в эксплуатацию), благодаря встроенным механизмам ветвления, фиксации изменений, контроля версионности, сервер GitLab позволяет обмениваться исходным кодом, вести независимое выполнение некоторых этапов, а встроенная в GitLab база знаний wiki позволяет хранить справочную информацию и вести документацию по каждому из проектов.

Заключение

Основной задачей модернизации, проводимой в Казахстанском национальном центре данных, является создание новой функциональной, масштабируемой и высокотехнологичной информационной системы с использованием современных программно-технических решений, основанных, главным образом, на свободно распространяемом и открытом программном обеспечении. Модернизация затрагивает многие значимые элементы информационной системы, построенной в Казахстанском национальном центре данных, и непосредственно влияет на качество решения задач, возложенных на Центр данных в рамках выполнения международных договоров и соглашений.

Литература

1. Тухватулин, Ш.Т. Система геофизического мониторинга, созданная в Национальном ядерном центре Республики Казахстан, и ее возможности / Ш.Т. Тухватулин, Л.Н Тихомиров, Н.Н. Беляшова, Н.Н. Михайлова, В.Н Демин, В.Г. Марченко, И.И. Комаров // Вестник НЯЦ РК. − 2002. − Вып. 2. – С. 5–8.
2. Михайлова, Н.Н. Казахстанский центр сбора и обработки специальной сейсмической информации: функции, задачи, система телекоммуникаций, базы данных / Н.Н. Михайлова, И.И. Комаров, З.И. Синева, И.Н Соколова // Вестник НЯЦ РК. − 2001. − Вып. 2. – С. 21–26.
3. Михайлова, Н.Н. Оценка эффективности сейсмической группы PS23-Маканчи при регистрации региональных и телесейсмических событий / Н.Н. Михайлова, З.И. Синева // Вестник НЯЦ РК. − 2004. − Вып. 2. – С. 13–19.
4. Демин, В.Н. Новая инфразвуковая станция международной системы мониторинга в Казахстане IS31 «Актюбинск» / В.Н. Демин, В.Г. Кунаков, А.А. Смирнов // Вестник НЯЦ РК. − 2002. − Вып. 2. – С. 14–18.
5. Гордиенко, Д.Д. Система контроля качества данных в Центре сбора и обработки специальной сейсмической информации / Д.Д. Гордиенко, Н.А. Сейнасинов // Вестник НЯЦ РК. − 2009. − Вып. 2. – С. 113–118.
6. [Электронный ресурс]: https://haydenjames.io/20-top-server-monitoring-application-performance-monitoring-apm-solutions.
7. Olups R. Zabbix: Enterprise Network Monitoring Made Easy / A.Vacche, P. Uytterhoeven // Packt Publishing Ltd, 2017
8. [Электронный ресурс]: https://www.proxmox.com/en/proxmox-ve
9. [Электронный ресурс]: https://www.ctbto.org/fileadmin/user_upload/legal/treaty_text_Russian.pdf
10. Fyen, J. Event processor program package. In: NORSAR Semiannual Technical Summary. 1 Oct 1988 – 31 Mar 1989. Scientific Report 2-88/89, Kjeller, Norway.
11. [Электронный ресурс]: https://git-scm.com/

ҰЛТТЫҚ ҚАЗАҚСТАНДЫҚ ДЕРЕКТЕР ОРТАЛЫҒЫНДА (KNDC) АҚПАРАТТЫҚ ТЕХНОЛОГИЯЛАРДЫ ДАМЫТУ

Д.Д. Гордиенко, И.И. Комаров, Ю.А. Смирнов

Геофизикалық зерттеулер институты, Курчатов, Қазақстан

Ақпараттық технологияларды дамытудың заманауи талаптары мен үрдістеріне сәйкес деректерді жинаудың, өңдеудің және сақтаудың сапасын жоғарлату мақсатымен Ұлттық қазақстандық деректер орталығында жаңа бағдарламалық-техникалық құралдарды жетілдіру, жаңғырту және енгізу бойынша жоспарды іске асыру мәселесі қарастырылды.

DEVELOPMENT OF INFORMATION TECHNOLOGIES AT THE KAZAKHSTAN NATIONAL DATA CENTER (KNDC)

D.D. Gordiyenko, I.I. Komarov, U.A. Smirnov

Institute of Geophysical Research Kurchatov, Kazakhstan

The paper presents an implementation of a plan on improvement, upgrade and introduction of new software and technical means in Kazakhstan National Data Center for the purposes of improving the quality of data acquisition, processing and storage in accordance with modern requirements and trends of information technologies' development.