МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ ИМЕНИ М.В.ЛОМОНОСОВА

Голозубов О.М., Колесникова В.М., Никитенко Д.А.

МЕТОДИЧЕСКОЕ ПОСОБИЕ

«Создание типового аграрно-почвенного дата-центра»

Электронная версия

РАЗРАБОТАНО К **ПРОГРАММЕ ДОПОЛНИТЕЛЬНОГО ОБРАЗОВАНИЯ «СОЗДАНИЕ ТИПОВОГО ПОЧВЕННОГО ДАТА-ЦЕНТРА»**
(открыта 21 мая 2020 г. на факультете почвоведения МГУ имени М.В.Ломоносова)

Москва
2020

УДК 631.4

МЕТОДИЧЕСКОЕ ПОСОБИЕ
«Создание типового аграрно-почвенного дата-центра»
Электронная версия. М. 2020 г. 76 с.

Голозубов Олег Модестович, МГУ имени М.В.Ломоносова, факультет почвоведения, кафедра географии почв, вед.н.с., к.б.н.

Колесникова Варвара Михайловна, МГУ имени М.В.Ломоносова, факультет почвоведения, кафедра географии почв, доцент, к.б.н.

Никитенко Дмитрий Александрович, МГУ имени М.В. Ломоносова, Научно-исследовательский вычислительный центр, Лаборатория параллельных информационных технологий, с.н.с., к.ф-м.н

Методическое пособие рекомендовано для лиц, имеющих среднее профессиональное и высшее образование, специалистов центров агрохимической службы Минсельхоза РФ, специалистов по аграрно-почвенному мониторингу, ГИС-специалистов в почвоведении и агрономии, заинтересованных в приобретении или совершенствовании навыков администрирования баз данных, настройки веб-приложений и использования ГИС совместно с СУБД. В пособии изложены основы построения моделей почвенных данных в виде реляционных структур; методы проектирования основных технологических циклов сбора, обработки и использования аграрно-почвенной информации; приемы администрирования баз данных и программных приложений в распределенной вычислительной среде. В пособии подробно излагаются современные методы стандартизации, гармонизации, и унификации объектно-атрибутивной модели почвенных данных, рассматривается концептуальная модель и международные стандарты аграрно-почвенного описания. Настоящее издание поможет слушателям курса «Создание типового почвенного дата-центра» приобрести навыки создания дата-центра на базе локальной сети и/или виртуальной машины; организовать технологические циклы и инфраструктуру обработки информации; создать модель данных (объектную модель, семантическую модель и т.д.) для регионального (или ведомственного) почвенного дата-центра (ПДЦ); разработать базовые стандарты для форм ввода данных. Рекомендуется также научным сотрудникам других учреждений, занимающимся исследованием почв и обработкой полученных данных, созданием собственных баз данных.

УДК 631.4

Оглавление

Введение

Особенностью России является большое разнообразие представленных на ее территории природных зон, что определяет огромную роль нашей страны в сохранении природного биоразнообразия и создании необходимых условий для его всестороннего изучения. Кроме того, большое значение приобретают региональные исследования. При этом особо актуальным становится вопрос унификации почвенно-географических данных, представляемых в информационные системы из разных региональных источников. Наиболее важным в свете решения задач рационального природопользования остается мониторинг плодородия почв. Помимо включения в информационные системы данных из научных источников, накопленных за весь период развития наук о почве, переведенных в цифровой формат, одним из путей привлечения информации является использование массива данных региональных аграрных центров, собранных для решения практических вопросов рационального сельскохозяйственного использования почв. Объединение разнородной информации о почвах требует гармонизации данных – согласования методов проводимых исследований, специфики отбора почвенных проб, методик проведения аналитических исследований. В последнее время немаловажную роль в расширении массива почвенных данных играют расчетные методы с использованием педотрансферных функций почв.

Россия участвует в программе глобального почвенного партнерства, являясь членом ФАО ООН, поэтому вопросы гармонизации аграрно-почвенных данных до международных стандартов являются первостепенной задачей для участия в международных программах и осуществления информационного обмена данными. Осуществление этих задач становится возможным благодаря созданию распределенной сети аграрно-почвенных дата-центров, объединенных единым стандартом хранения, представления и

обмена почвенными данными (рис. 1).

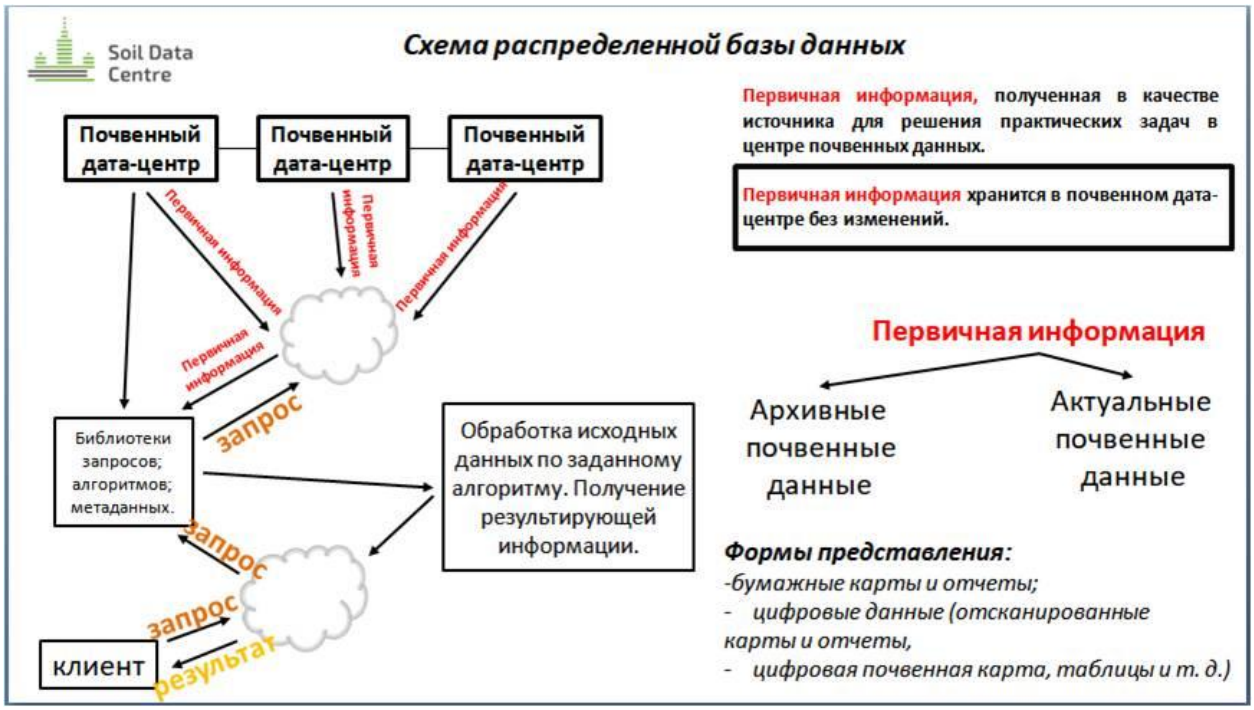

Рис.1 Функционирование сети аграрно-почвенных Дата-центров

1. Стандартизация в хранении, представлении и обмене почвенными данными

Задачи стандартизации, гармонизации, многоязычности и унификации объектно-атрибутивной модели почвенных данных. Концептуальная модель и международные стандарты, определение проектно-объектной модели. Региональные стандарты - аналоги стандартов серии SQ ISO 28258.

1.1. Информационный обмен атрибутивными почвенными данными на международном уровне: проблемы и перспективы

Проблема сопоставимости атрибутивных данных для информационного обмена актуальна для современного этапа развития национальных и региональных информационных систем. Россия участвует в программе глобального почвенного партнерства (FAO Global Soil Partnership (GSP), являясь членом ФАО ООН с 2006 года. Одной из главных задач программы является организация международной сети национальных институтов почвенной информации и осуществление информационного обмена сопоставимой почвенной информацией. В основу рабочей программы партнерства положено использование серии международных стандартов (ISO 28258), определение понятия почвенного исследования и описания иерархии почвенных индикаторов, представление пространственных природных объектов (ISO 11074, ISO 19156) и др., связанных также с такими международными проектами, как IUSS GlobalSoilMap, INSPIRE EU, OGC Soil Interoperability Experiment (Soil IE). Таким образом, главная задача при информационном обмене сводится к обеспечению сопоставимости (гармонизации) почвенных данных.

Для России в настоящее время актуальна задача обеспечения сопоставимости между различными подразделениями – генераторами почвенной информации (агрохимцентрами, НИИ, отдельными хозяйствами

и другими организациями и ведомствами) в первую очередь, а затем гармонизация этих данных до международных стандартов, принятых в других странах, для участия в различных программах. Значительная площадь территории нашей страны, представленность широкого спектра природных зон, со своей спецификой организации почвенного покрова и его использования, обуславливает определенную сложность в согласовании единых стандартов проведения аграрно-почвенных исследований. Кроме того, отсутствует определенный опыт практики следования международным стандартам по причине специфики российской научной школы.

Полнота представления почвенно-географических данных, предусмотренная в проекте «Почвенно-географическая база данных России» (ПГБД РФ) превосходит принятый в международных системах уровень (8). Разработанные с учетом национальной специфики программы этого проекта могут быть использованы для решения фундаментальных научных и педагогических задач и для решения прикладных задач. Вместе с тем актуальной задачей настоящего этапа развития ПГБД является поиск механизма обмена почвенной информацией международного уровня. При этом важнейшим условием корректного использования разнородной атрибутивной почвенной информации является гармонизация данных.

1.2. Разнородность почвенной информации, привлекаемой для решения практических задач

В ведомствах, координирующих агрохимические исследования разного уровня, зачастую отсутствовала единая схема описания и представления данных о свойствах почв. Многие годы сбор информации о почвах, необходимой для решения конкретных прикладных задач, ограничивался узким диапазоном исследований. Проблемы унификации почвенных данных связаны, прежде всего, с различной полнотой описания почв. Главной проблемой является использование различных методов определения химических свойств почв, отсутствие их корректного указания в ведомостях

агрохимических анализов, особенно это касается материалов прошлых лет. Отдельная задача связана с сопоставлением данных, которые приведены в различных единицах измерения. В архивных материалах использованы различные форматы представления данных о свойствах почв.

Для унификации данных о почвах и почвенном покрове необходима концептуальная модель привлечения информации, как атрибутивной, так и пространственной (рис. 2).

Рис.2 Концептуальные модели стандарта обмена почвенной информацией России

1.3. Источники почвенно-географической информации. Объекты исследований

Наиболее важным звеном в концептуальной модели стандарта представления почвенно-географической информации является выбор перечня основных объектов.

ГОСТ 27593-88 определяет следующие объекты, для которых может быть задана пространственная локализация:

· Элементарный почвенный ареал (почвенный контур);

· Почвенный профиль;

· Почвенный горизонт;

· Пробная площадка;

· Единичная проба - Почвенный образец;

· Объединенная проба – Смешанный образец:

В методиках Министерства сельского хозяйства добавлены понятия поля, тестового полигона, схемы формирования смешанного образца.

В структуре атрибутивного блока ПГБД России (8) и ЕГРПР (6) предложена следующая иерархическая схема (рис.3):

Разрез

Профиль

Горизонт Слой Морфон

Морфологический элемент

Образец

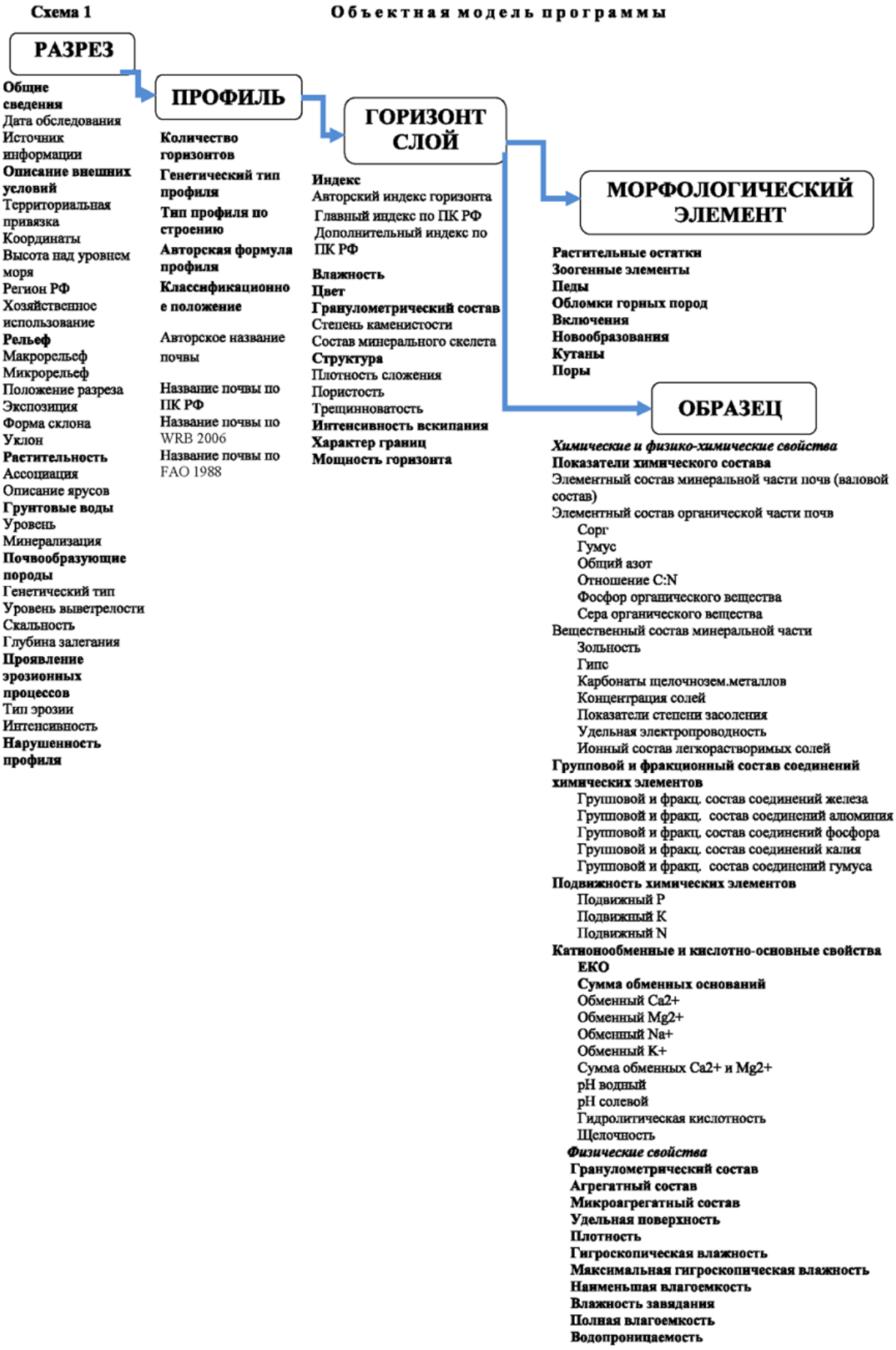

Рис.3 Схема атрибутивного блока ИС ПГБД РФ

Кроме того, современная цифровая почвенная картография (стандарт ISO 28258 и концептуальная модель проекта INSPIRE) такими формализованными понятиями как:

· Место обследования;

· Точка обследования;

· Элемент профиля (состоит из горизонта или слоя);

· Представительный профиль (т.е. усредненный для некоторой территории);

· Слой (в отличие от горизонта) как вертикальная координата.

Также вводятся новые понятия для рассчитанных (производных или выведенных), но также локализованных географически (возможно также и во времени) объектов, таких как:

· Почвенное тело и покрытие (аналог ареала и контурной карты);

· Грид или растровая карта.

Рис.4 Объектная модель

Все объекты (рис.4) характеризуются определенным набором свойств (показателей, индикаторов) наблюдаемых, измеренных или выведенных. Эти свойства (как в соответствии с ГОСТ, так и в соответствии с ISO) могут быть разделены на химические, физические, биологические или морфологические и т.п.

Выводимые или рассчитанные объекты могут содержать как наблюдаемые в исходных объектах свойства, так и новые. Например, расчет нового свойства запаса органического углерода в зависимости от глубины гумусового горизонта. При расчете свойств могут использоваться почвенные свойства, а также показатели условий почвообразования (рельеф, осадки и температура и т.п.).

В одних случаях (проект INSPIRE) стандарт требует полного и конечного перечня свойств, в других предполагается возможность его расширения, однако сложность современного представления модели почвенного объекта требует применения современных мощных средств

формализации и онтологических описаний (понятийно-сущностных в терминах реляционных БД).

1.4. Концепция гармонизации почвенных данных

Главным вопросом при решении проблемы гармонизации разнородной почвенной информации, участвующей в информационном обмене, является выбор стандартов представляемых данных.

Гармонизация понимается как взаимное согласование, унификация, координация, упорядочение, обеспечение взаимного соответствия, сопоставимости. Наиболее остро проблема гармонизации почвенных данных обозначилась в процессе становления цифровой почвенной картографии (ЦПК).

Стержневой идеей ЦПК является автоматический расчет почвенного показателя в заданном месте в заданное время в зависимости от почвенных и непочвенных показателей известных в другом месте (и в другое время) – так называемая scorpan-модель. То есть, фактически, формирование множества универсальных алгоритмов такого расчета над некоторым набором метаданных. Эта идея породила ряд международных проектов, направленных на получение глобальной и бесшовной почвенной карты мира, что потребовало выработать для почвенной информации ряд стандартов и регламентов.

Гармонизация стандарта — приведение его содержания в соответствие с другим стандартом для обеспечения взаимозаменяемости данных, взаимного понимания результатов испытаний и информации, содержащейся в стандартах. В такой же степени гармонизация может быть отнесена и к техническим регламентам. Длительная история формирования международных стандартов ISO, их преемственность, привычность практики, привели к тому, что во всех международных проектах стандарты, так или иначе, есть, и они весьма сходны. Для гармонизации необходимы два

стандарта и либо их сведение, либо обеспечение взаимопонимания (или формирования «ситуационной осведомлённости» (Situational Awareness) - определенность всех понятий, используемых при описании объекта).

Стандарт представления почвенной информации представляет собой информационную модель, которая обычно описывается с помощью таких инструментов, как UML – унифицированный язык моделирования, XML-XSD – схема и язык расширенной разметки, словарей метаданных и связей в СУБД, методах описания ресурсов RDF и т.д. В информационной модели в соответствии с концептуальной моделью должны быть определены сведение схем объектов, сведение методов, сведение показателей, сведение значений. Можно назвать несколько особенностей, которые нужно принимать во внимание при формировании информационной модели:

· Историчность и национальные стандарты.

· Данные, полученные за разные периоды времени, требуют особой осторожности при гармонизации. Они могут быть несопоставимы также по причине изменения технологий измерений, методов, классификаций и т.п.

· При гармонизации следует различать идентичность и инфицированность. Если некий показатель или объект не может быть сведен к другому, то от описания этого показателя требуется формализованное единообразие.

· Для архивных почвенных карт как объекта обычно производится дизагрегация легенд (атрибутивной информации), после чего каждое из свойств рассматривается и гармонизируется изолированно.

· Кроме детерминированной, возможна и вероятностная гармонизация, когда один показатель преобразуется в другой, например, в виде распределения с известной формой, и характеристиками.

· Иерархичность элементов модели или принадлежность.

· Временное усреднение (также как и пространственное) является методом, который должен быть указан в описании.

· Можно привести множество примеров определения педотрансферных функций.

Предполагается, что любой стандарт должен содержать в себе как возможность расширения – включения новых объектов, свойств, методов и значений, так и механизмы реализации этого расширения. Что, безусловно, является достаточно сложной задачей в условиях распределенности и асинхронности почвенных дата-центров.

1.5. Опыт создания многоязычного комплекса

На факультете почвоведения МГУ имени М.В.Ломоносова был разработан многоязычный программный комплекс «SOIL_ML_MultyL» для создания структурированного описания почв (11). Комплекс предназначен для использования как независимыми исследователями (для решения авторских задач), так и государственными структурами - агрохимцентрами, лабораториями, экспертными организациями (для целей сельскохозяйственного мониторинга). База метаданных создана на основе типовой объектной модели, содержит внушительный список показателей свойств почв, методов определений их свойств и единиц измерений. Возможности SOIL_ML_MultyL позволяют расширять перечень объектов и показателей, а также настраивать структуру описания почвенного объекта в соответствии со специализацией исследований. Главной особенностью является «многоязычность» программы (возможность работы на русском, английском, азербайджанском и румынском языках), которая позволяет объединить усилия специалистов разных стран для обеспечения гармонизации с международными стандартами в области почвоведения. Комплекс рассчитан на пользователей-почвоведов и других участников

обмена почвенной информацией в сети Интернет и входит в состав Информационной Системы Почвенно-Географической Базы Данных РФ (11).

1.6. Функциональные возможности программного комплекса SOIL_ML_MultyL

Комплекс может быть использован:

- в качестве средства первичной структуризации данных и объектной модели в рамках проектного подхода при развертывании почвенного дата-центра;

- для формирования и хранения структурированных почвенных описаний в формате XML аналогичных международным стандартам ISO28258 и SoilML;

- для обеспечения информационного обмена почвенными описаниями в распределенной сети почвенных дата-центров как между дата-центрами, так и между интернет-пользователями и дата-центрами.

Комплекс состоит из базы метаданных, тестовой базы почвенных описаний и многоцелевого пакета программ V8. База метаданных содержит:

1) Перечень показателей (indicators) почвенных свойств, включая перечень географически определенных объектов почвенного описания;

2) Перечень методов (methods) извлечения, определения, значений и единиц измерения почвенных свойств вместе с их кратким описанием;

3) Перечень типов данных (data types), с помощью которых могут быть описаны свойства почвенного объекта;

4) Перечень тематических разделов (Pages), объединяющих большие группы почвенных свойств, относящихся к определенной области знаний;

5) Перечень тематических подразделов (Partitions), группирующих сходные почвенные свойства внутри одного раздела;

6) Перечень наименований «Специализаций» пользователя, для которых можно задать специализированный сокращенный перечень показателей,

ориентированный на конкретную задачу;

7) Перечень наименований «Концепций» - способов структурирования (разделения) множества почвенных показателей на тематические разделы и подразделы. Указанные перечни представляют собой таблицы реляционной базы данных, связанные отношениями «один-к-многим» и «многие-к-многим» (через вспомогательные таблицы индексации) (рис.5).

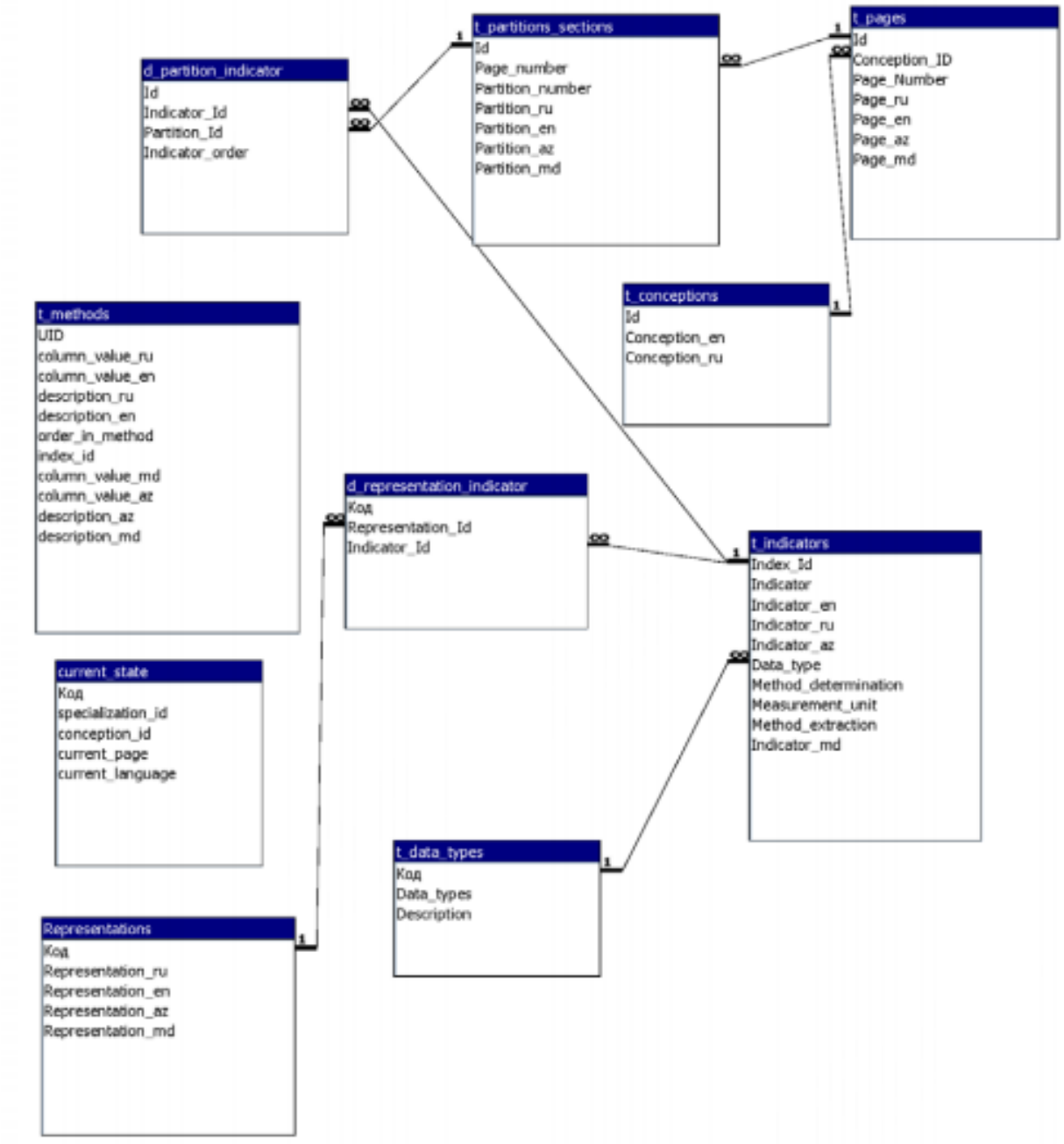

Рис.5 Схема базы мета-данных

База метаданных содержит дублирующиеся текстовые поля (наименования, описания и т.п.) для каждого из 4-х языков многоязычной версии – русский, английский, азербайджанский и румынский. Основное назначение базы метаданных – поддержание единого справочника, обеспечение соответствия терминов, определений и стандартов на международном уровне. База метаданных рассчитана в первую очередь для применения в республиках, объединенных общей советской школой почвоведения. База метаданных содержит объектную модель – как часть перечня показателей (Indicators). Сюда входят как обычные для международных стандартов объекты (профиль, горизонт, слой, образец и т.п.), так и принятые в сельскохозяйственном мониторинге объекты: смешанный образец, полигон и т.п. Назначением базы метаданных также является формирование иерархически структурированных схем описания почвенных данных (рис. 6), то есть, стандартов, в рамках которых может осуществляться обмен почвенными данными.

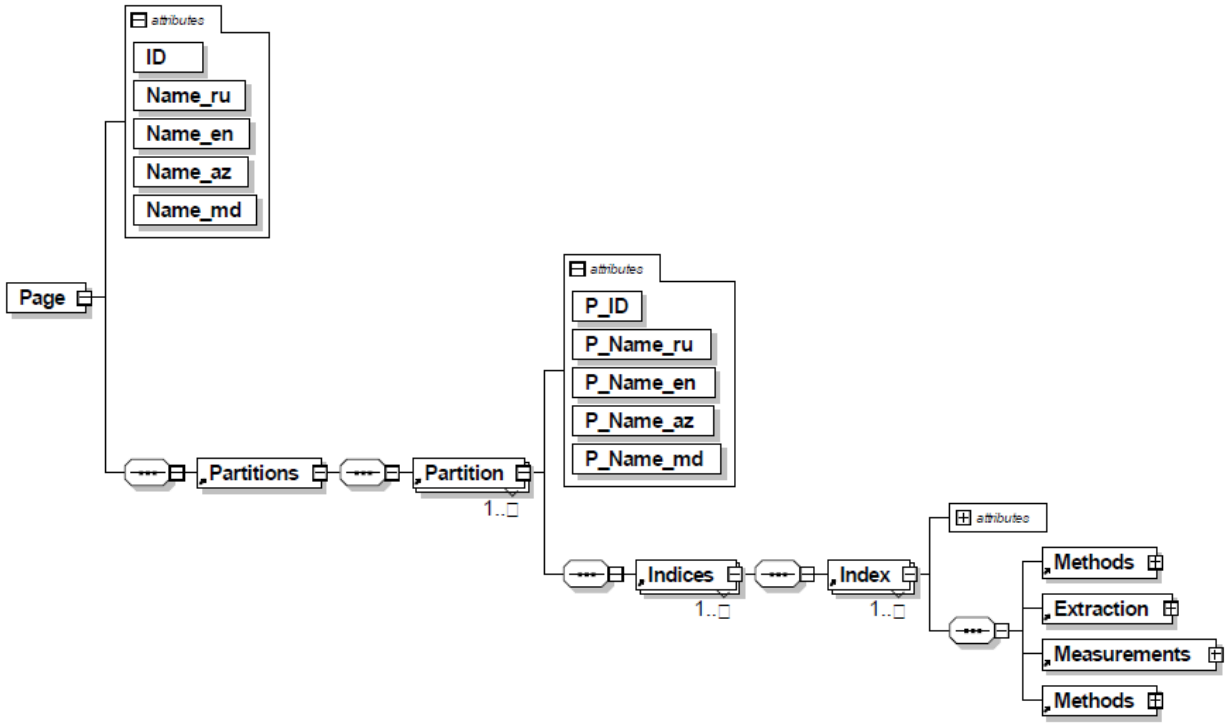

Рис.6 Дерево схемы описания почвенных данных

Тестовая база почвенных описаний предназначена для хранения и редактирования персональных почвенных описаний на локальном компьютере, объединения почвенных описаний в группы (например, несколько горизонтов в один профиль). Тестовая БД хранится в отдельном файле. Функциональные возможности пакета программ V8 и область применения Пакет программ V8 работает с базой метаданных и тестовой базой, расположенными в одном каталоге.

2. Инфраструктура и администрирование типового почвенного Дата-центра

Концепция клиент-серверной организации хранения почвенной информации. Состав системного и программного обеспечения (СПО) дата-центра. Веб-сервисы и стандарты обмена в сети Интернет: GeoRSS, WFS, WMS, GeoJSON, основанные на XML-стандартах представления почвенных данных.

2.1. Структурная и функциональная схема типового аграрно-почвенного Дата-центра

Аграрно-почвенный дата-центр (АПДЦ) представляет собой аппаратно-программный комплекс, обеспечивающий автоматизацию технологического цикла обработки природно-почвенной информации, включая планирование обследований, накопление и хранение пространственно-атрибутивной информации результатов обследований, формирование отчетов, паспортов и рекомендаций, а также включенный в состав ИС ПГБД РФ*. Схема организационной структуры типовых аграрно-почвенных дата-центров (АПДЦ) представлена на рисунках 7 и 8.

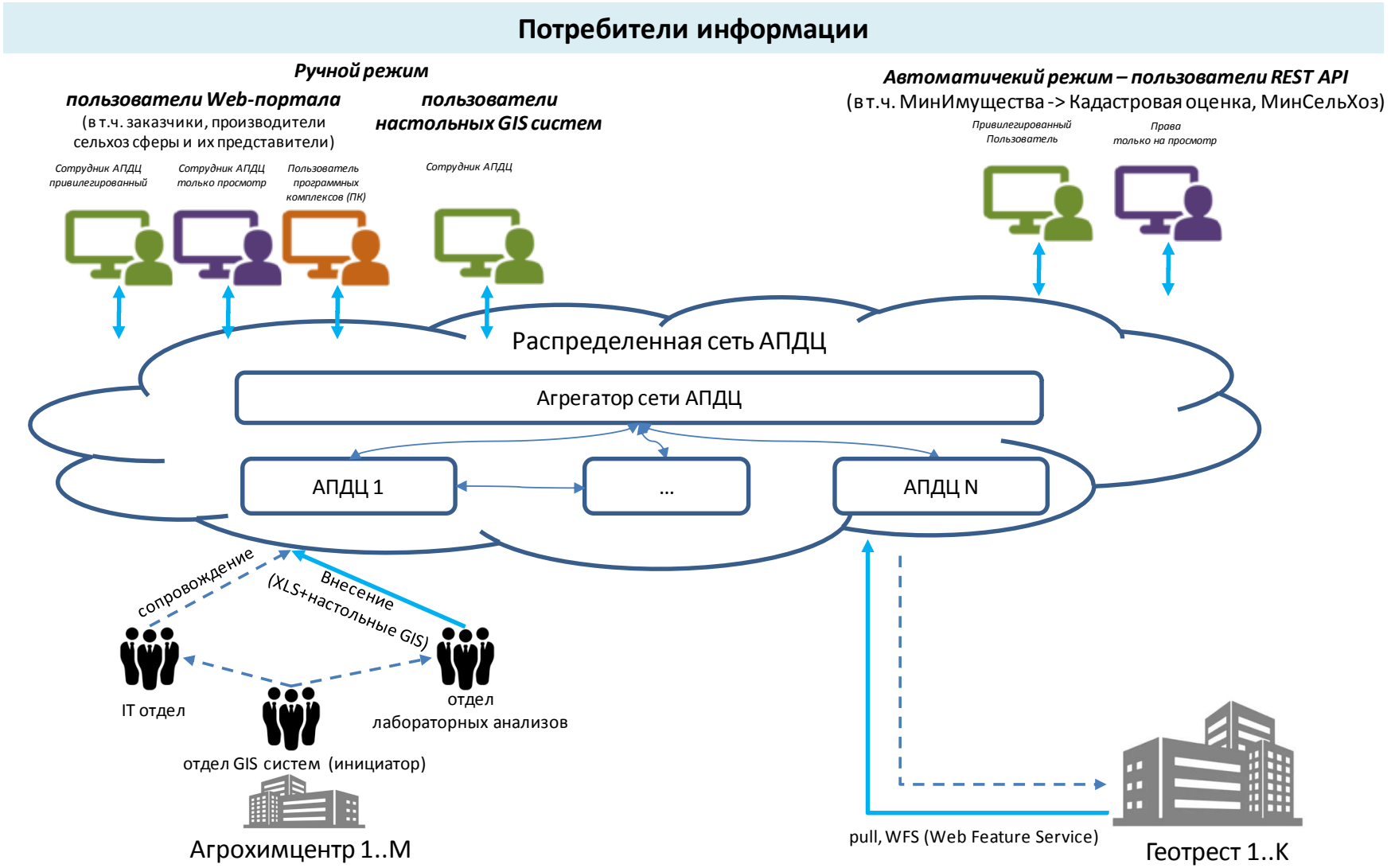

Рис.7. Схема организационной структуры аграрно-почвенных дата-центров (АПДЦ). Поставщики и потребители информаци

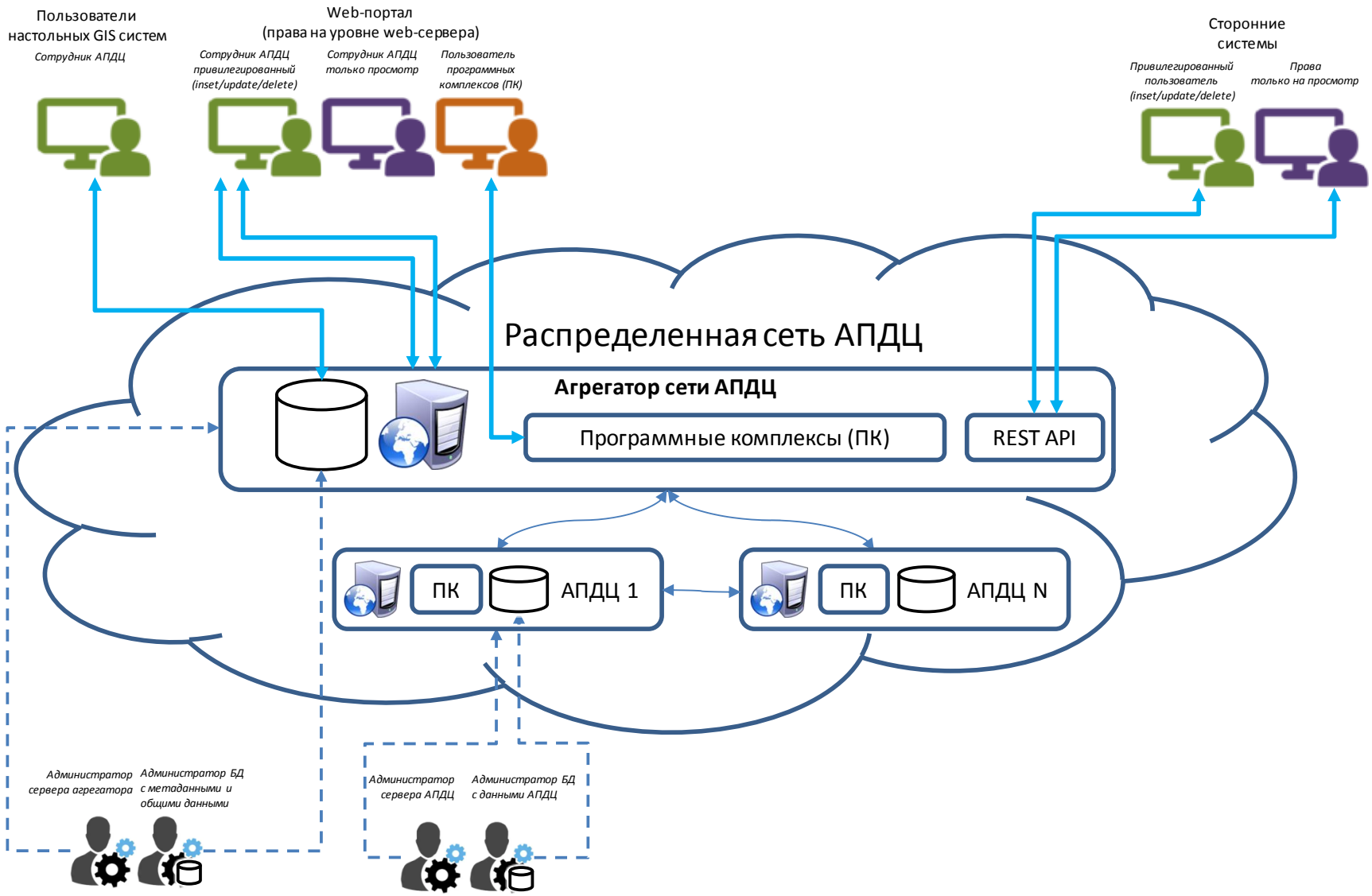

Рис. 8. Схема организационной структуры аграрно-почвенных дата-центров (АПДЦ). Категории пользователей.

Схема структурная комплекса программно-технических средств (структура программных средств)

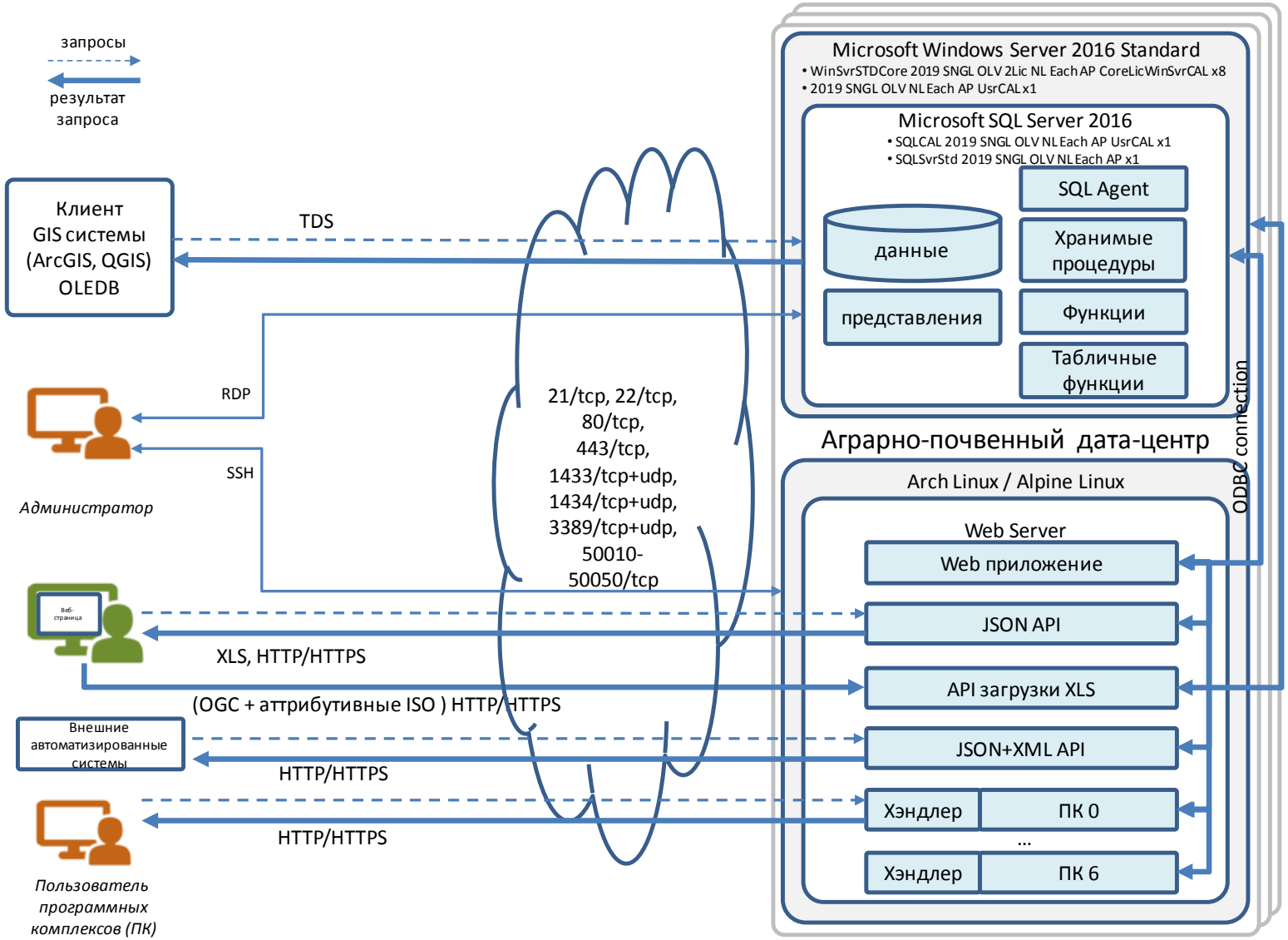

Рис.9. Схема структурная комплекса программно-технических средств (структура программных средств).

Схема структурная комплекса программно-технических средств (структура технических средств)

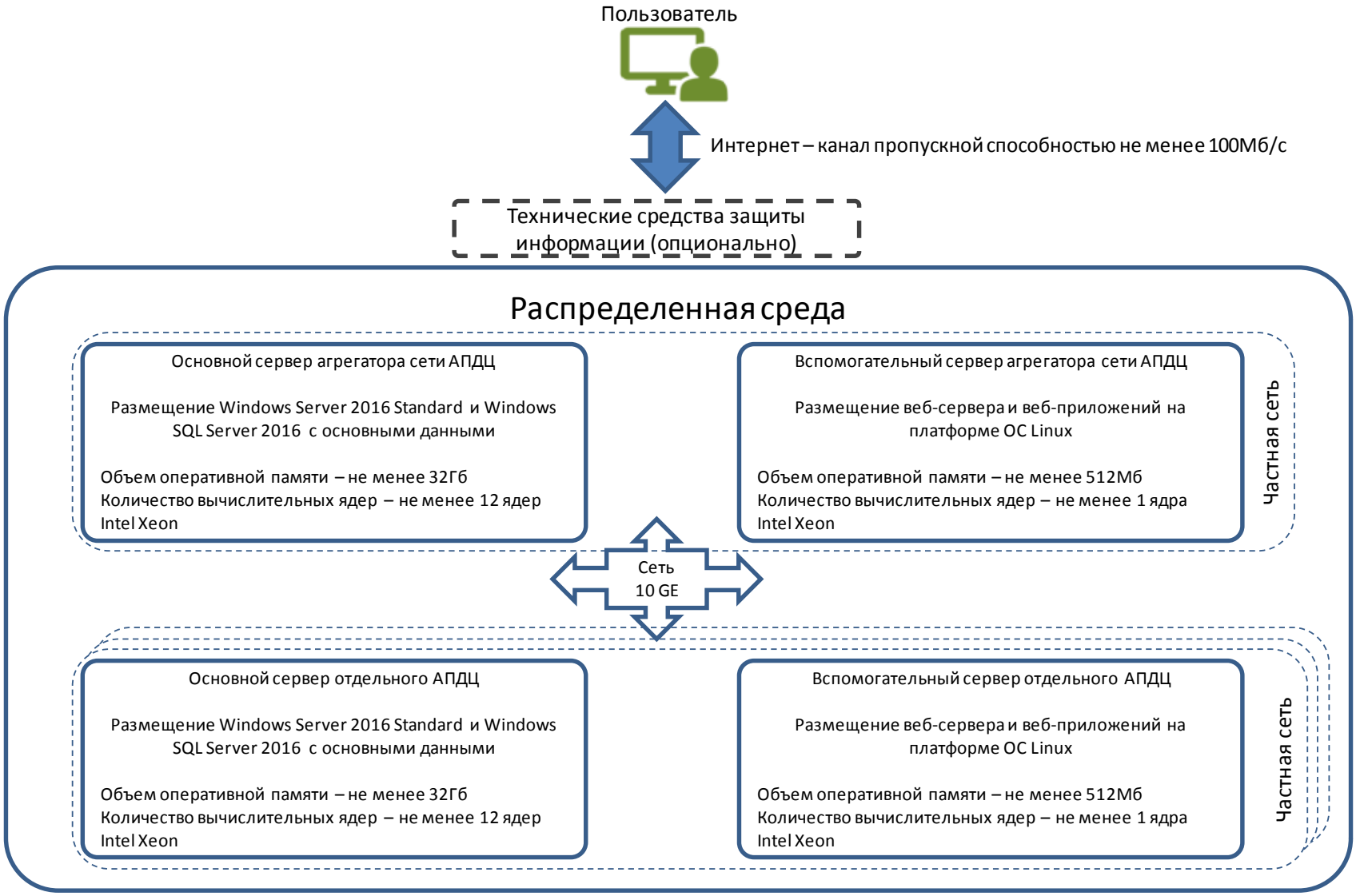

Рис.10. Схема структурная комплекса программно-технических средств (структура технических средств).

Схема структурная комплекса программно-технических средств (структура системных средств)

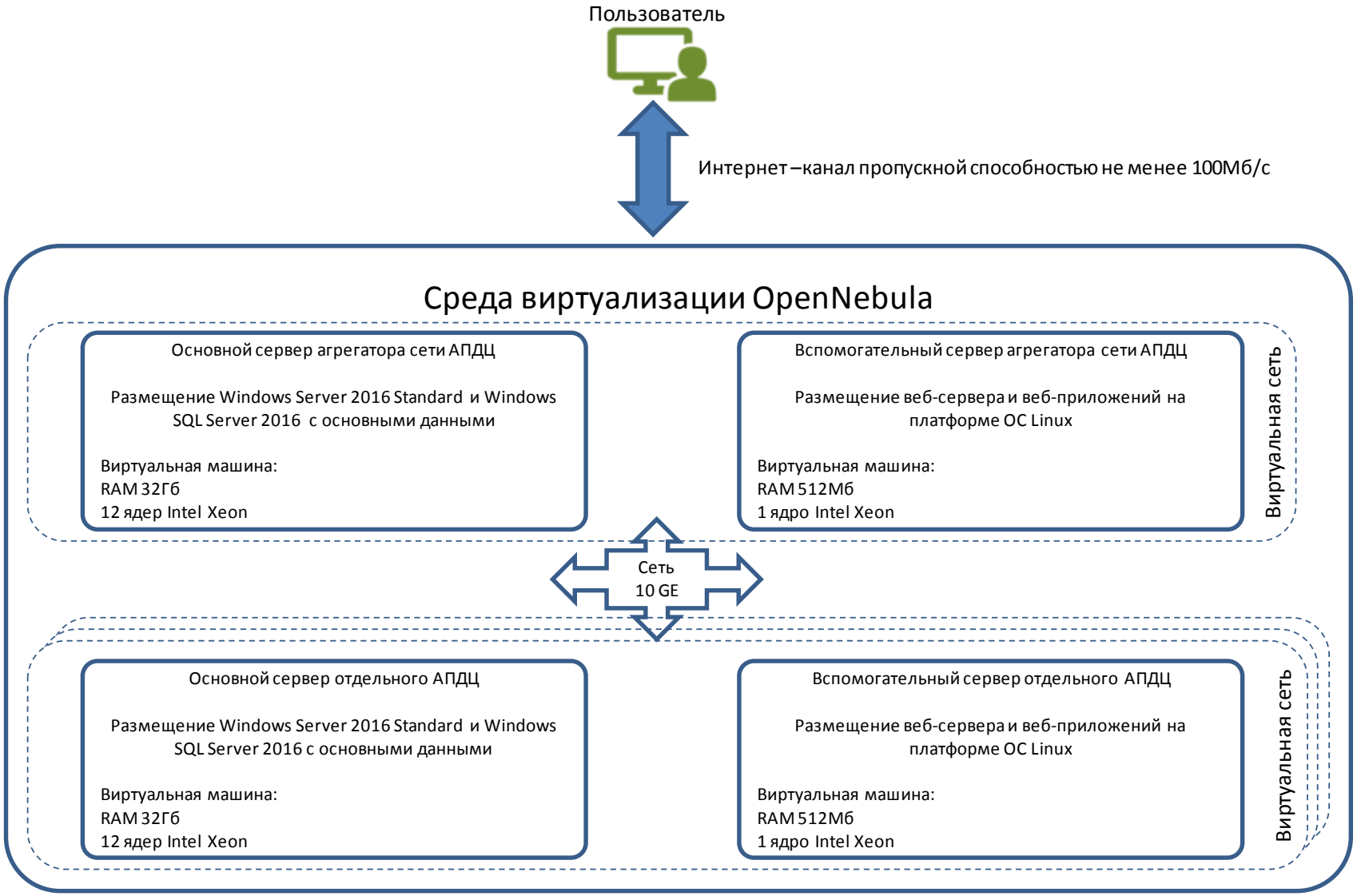

Рис. 11. Схема структурная комплекса программно-технических средств (структура системных средств).

2.2. Структура аграрно-почвенного Дата-центра

Аграрно-почвенный Дата-центр представляет собой многоуровневую (multi-tier) систему с открытым кодом (opensource), включающую следующие подсистемы:

• Программы автономного ввода, накопления и модификации почвенно-картографических данных (SoilMLMultiL – V8);

• Реляционную БД в формате MSSQLServer, содержащую более 200 связанных таблиц с пространственно-атрибутивной информацией, и справочниками, представления (View) и хранимые процедуры (SP);

• Набор обработчиков (хэндлеров), осуществляющих взаимодействие между автономными программами и серверами БД, между различными серверами БД в распределенной сети, между серверами БД и веб-порталом;

• Постоянно пополняемый набор хранимых процедур, осуществляющих выполнение алгоритмов решения прикладных задач (расчет запасов органического углерода, пригодности с/х земель и их оценки, доз удобрений и т.п.);

• Средства доступа к пространственно-атрибутивной информации непосредственно из типовых ГИС (ArcGIS, QGIS);

• Реализованные на языке XML и его подмножествах (GML, KML) описания метаданных для информационного обмена (перечисленных выше взаимодействий);

• Реализованные в виде XSD-схем и адаптированные к задачам ИС ПГБД РФ существующие международные стандарты: OGC 14-013r1 – ServiceIntegration; ISO 28258 – SoilQualityInformationExchange; ISO/TC 190/SC – SoilQualityFieldDescription; также должны быть учтены проекты разрабатываемых стандартов –ISO/FDIS 28258:2013(E), ISO TC 190/SC 1/WG 3–Soilquality – Digitalexchangeofsoil-relateddata;

• Программные средства верификации и валидации почвенных данных в соответствии с изменениями в объектной и концептуальной моделях указанных стандартов (современные стандарты публикуются и применяются в виде UML/XML текстов).

Структурная схема комплекса программно-технических средств (структура технических средств) типового аграрно-почвенного центра приведена на рисунках 9 и 10.

Типовой почвенный Дата-центр включает следующие подсистемы:

1. Подсистема авторизации и защиты информации;
2. Подсистема контроля авторских прав и интеллектуальной собственности;
3. Подсистема ведения дистанционного обучения и поддержки тестирования;
4. Подсистема виртуализации веб-сервисов как основу масштабируемости и тиражируемости распределенных систем;
5. Средства интерактивного отображения данных на картографических основах Yandex, Google, ESRI, обеспечивающие предоставление сведений посредством следующих webсервисов по открытым спецификациям OGC:WMS (WebMapService) – картографические сервисы, предоставляющие клиентам изображения;WFS (WebFeatureService) – сервисы, предоставляющие клиентам координатное описание объектов; WCS (WebCoverageService);
6. Подсистема ведения каталога веб-сервисов в соответствии с рекомендациями OGC.

3. Пример аппаратно-программной реализации типового аграрно-почвенного дата-центра (АПДЦ)

3.1. Анализ компонент в почвенно-географической распределенной системе АПДЦ (Информационной системы “Почвенно-географическая база данных России”) и требований к ее администрированию

Почвенно-географическая распределенная система АПДЦ состоит из нескольких центров, клиентов и источников данных. Клиентами являются как внутренние, так и внешние пользователи, а в качестве средств доступа они используют либо web-браузер, либо настольные GIS-системы. Так как специального клиентского ПО не предусмотрено, администрирование на клиентской стороне возлагается на пользователя-клиента. Со стороны АПДЦ необходимо обеспечить аутентификацию пользователей и защиту передаваемых ими данных.

Источники данных в данном проекте являются пассивными, то есть программное обеспечение АПДЦ получает от них данные с помощью запросов. ПО источников в данном проекте на разрабатывается и не контролируется, поэтому вопрос об администрировании в этом случае не стоит.

Состав АПДЦ. В него входит два сервера — на базе Microsoft Windows Server Standard (основной), и на базе Linux (вспомогательный). Первый сервер обеспечивает работу Microsoft SQL Server и доступ к его данным, в то время как второй обеспечивает работу Web-сервера и программных комплексов. Web-сервер обслуживает как клиентов, работающих по JSON+XML API, так и предоставляет доступ к web-страницам проекта (web-портал). Программные комплексы также пользуются web-сервером для обмена данными с подключёнными клиентами.

Таким образом, ключевыми компонентами одного АПДЦ являются:

ОС Windows Server Standard на основном сервере,

ОС Linux на вспомогательном сервере,

Microsoft SQL Server,

Web-сервер.

Стоит обратить внимание на то, что программные комплексы хотя и не общаются с клиентами напрямую, но имеют высокие полномочия в общем программном комплексе, а значит, к ним также должны быть сформулированы требования безопасности. Это же относится ко всем системным сервисам ОС.

Каждый вышеперечисленный компонент должен быть защищён стандартными средствами с учётом требований безопасности:

обеспечение идентификации и аутентификации пользователей,

защита от несанкционированного доступа,

использование средств шифрования интернет-соединения,

использование средств управления безопасностью.

3.2. Разработка подхода к администрированию почвенно-географической распределенной системы АПДЦ с учетом различных уровней администрирования

Администрирование почвенно-географической распределенной системы АПДЦ имеет два аспекта:

- администрирование уровня ОС и системного ПО (компоненты ОС, Microsoft SQL Server, Web-сервер, серверы удалённого доступа RDP И SSH и т. п.),
- администрирование прикладного уровня АПДЦ (управление пользователями, правами доступа, данными АПДЦ и т. п.).

На уровне ОС и системного ПО администрирование должно быть разграничено ролями по функциям, выполняемым в системе:

- суперпользователь (обновление и конфигурация системного ПО, обновление ПО АПДЦ, резервное копирование),
- администратор баз данных (управление данными в Microsoft SQL Server, резервное копирование базы данных),
- пользователь для обеспечения мониторинга и сбора статистики.

На уровне прикладного администрирования АПДЦ должны быть обеспечены две роли:

- привилегированный пользователь АПДЦ (управление пользователями и правами АПДЦ, выполнение служебных процедур),
- непривилегированный пользователь АПДЦ (мониторинг, статистика).

Кроме вышеобозначенных пользователей в системе должны присутствовать и не административные пользователи, их роли мы здесь не рассматриваютя.

3.3. Организация администрирования и сопровождения инфраструктуры, используемой почвенно-географической распределенной системы АПДЦ

На уровне администрирования системного ПО необходимо создать административных пользователей указанных в предыдущем пункте, а также обеспечить им удалённых доступ на серверы АПДЦ. Из соображений безопасности удалённый доступ суперпользователя рекомендуется отключить и обеспечить возможность выполнять действия от его имени для выделенного пользователя.

В случае ОС Windows это должен быть пользователь с правами администратора, например, «Администратор». Этот доступ должен предоставляться только администратору данного АПДЦ. В обязанности администратора ОС Windows входит обновление и настройка параметров работы системы Windows Server.

Для удалённого доступа в случае ОС Windows целесообразно использовать доступ по протоколу RDP. Для защиты соединения необходимо использовать либо соединение через VPN (можно использовать SSH-туннелирование порта), либо включённое шифрование в самом протоколе RDP (желательно стандартом FIPS или более стойким). Для упрощения

решения предлагается использовать второй вариант – включённое шифрование RDP с сертификатом сервера.

Для удалённого доступа в случае ОС Linux целесообразно использовать доступ по протоколу SSH. На сервере потребуется установить пакет openssh-server. При настройке параметров сервера необходимо отключить вход пользователя root по паролю, а также включить доступ для пользователя root и других пользователей только по открытому ключу (доступ по паролю отключить). Это позволит избежать атак на перебор паролей. Необходимо проверить,что для входа разрешаются ключи формата RSA/ECDSA с длиной ключа 2048 бит или более.

Пользователи смогут входить удалённо используя клиентские программы ssh из пакета openssh-client на системах под управлением ОС Linux или MacOS, а также putty, xshell, SecureCRT или любой другой ssh-клиент. Для передачи файлов можно использовать под Linux и MacOS команды scp, sftp, встроенную поддержку протокола sftp в пакете mc, а также поддержку протокола sftp в файловых менеджерах Nautilus, Konquer и других. В случае ОС Windows рекомендуется использовать программу WinSCP, а также плагин sftp для файлового менеджера FAR.

Рекомендуется полностью закрыть доступ ко всем сетевым портам, не используемым внешними пользователями и администраторами с помощью файрволла.

Доступ всех других административных пользователей также должен осуществляться вышеописанными способами. Администратор БД Microsoft SQL Server должен иметь права только на подключение к серверу баз данных и не должен иметь прав администратора. Даже в случае, если эти роли возложены на одного и того же человека, необходимо создать двух разных пользователей с разными правами для выполнения различных обязанностей.

Предлагается использовать пользователя Microsoft SQL Server с именем admin и правами администратора в сервере баз данных. Доступ осуществляется только со сложным паролем, копия которого должна храниться на удалённом носителе.

В обязанности администратора баз данных в рамках одного АПДЦ входит:

создание резервной копии базы данных,

восстановление резервной копии базы данных,

ручная выгрузка резервных копий базы данных,

ручная загрузка резервных копий базы данных,

ручная корректировка (удаление/обновление/чистка) данных в БД,

ручное изменение логики работы хранимых процедур.

Пользователь для осуществления мониторинга в ОС Linux также не должен иметь прав администратора (возможности выполнения любых команд через sudo от имени пользователя root). В случае, если такому пользователю необходимо предоставить права для выполнения какого-то действия в рамках своей роли с привилегиями администратора (например, перезапуск сервиса мониторинга), то в настройках sudo необходимо прописать право на выполнение именно этой команды данным пользователем.

В обязанности администратора ОС Linux входит:

обновление системы Alpine Linux,

обновление образов, из которых делаются контейнеры с веб-приложениями,

ручная перезагрузка серверов

3.4. Организация доступа по сети к инфраструктуре ЦХАБД. Определение протоколов и требуемых методов доступа

Структура доступа по сети к инфраструктуре ЦХАБД

Инфраструктура ЦХАБД МГУ имени М.В. Ломоносова установлена на площадке суперкомпьютера «Ломоносов-2» суперкомпьютерного комплекса МГУ имени М.В. Ломоносова. Оборудование установлено в выделенных шкафах. Инфраструктурное обеспечение (электропитание, поддержание климатических условий, сетевая связность) обеспечивается компонентами инфраструктуры суперкомпьютера «Ломоносов-2».

Сетевая часть инфраструктуры ЦХАБД построена на основе технологии Ethernet со скоростью 10 Гбит/с. При этом каждый сервер из состава инфраструктуры ЦХАБД подключен с коммутаторам сети ЦХАБД двумя каналами с пропускной способностью 10 Гбит/с каждый для отказоустойчивости и повышения пропускной способности.

Сеть ЦХАБД представляет собой изолированный фрагмент сети. Внешняя связность (с сетью Интернет) обеспечивается при помощи подключения я ядру сети суперкомпьютера «Ломоносов-2». Это подключение выполнено двумя каналами с пропускной способностью 10 Гбит/с каждый. По этому подключению обеспечивается связь инфраструктуры ЦХАБД с внешними сетями (сетью Интернет).

Ядро сети суперкомпьютера «Ломоносов-2» подключено к сети МГУ имени М.В. Ломоносова двумя каналами с пропускной способностью 1 Гбит/с каждый. Через это подключение обеспечивается доступ пользователей инфраструктуры ЦХАБД и суперкомпьютера «Ломоносов-2» к соответствующим ресурсам.

Сеть МГУ присоединена к внешним сетям (сети Интернет) каналом с пропускной способностью 1 Гбит/с. Скорее всего, пользователи той части сети АПДЦ, которая развернута на мощностях инфраструктуры ЦХАБД, будут находиться внести МГУ. Поэтому пропускная способность и загруженность именно этого канала, по всей видимости, окажется решающей

при обеспечении доступа к части сети АПДЦ, расположенной в МГУ. Нам это ограничение представляется малозначимым, так как, с большой вероятностью, на пути от пользователя АПДЦ будут встречаться каналы с меньшей пропускной способностью. Способы работы с АПДЦ проектируются с расчетом пропускной способности канала от пользователя до АПДЦ в 100 Мбит/с. С этой точки зрения, имеющиеся каналы связи инфраструктуры ЦХАБД с внешними сетями не являются ограничением для развертывания сети АПДЦ.

Программное обеспечение АПДЦ развертывается на инфраструктуре ЦХАБД в виде комплектов виртуальных машин (2 виртуальных машины на один дата-центр). Этим машинам создается виртуальная локальная сеть, изолированная от других сетей. По этой виртуальной сети осуществляется взаимодействие между виртуальными машинами одного АПДЦ.

Доступ пользователей и администраторов к сервисам АПДЦ

Для предоставления сервисов АПДЦ внешним пользователям одной из виртуальных машин АПДЦ назначается глобально маршрутизируемый IP-адрес (используется протокол IPv4). На этом адресе запускается программное обеспечение (ПО) веб-сервера для обеспечения доступа к АПДЦ через браузер. Также на этом адресе доступны и другие сетевые сервисы, необходимые пользователям и администраторам АПДЦ.

Основным способом доступа к сервисам АПДЦ является веб-интерфейс (протоколы HTTP и HTTPS, порты 80/tcp и 443/tcp). Также поверх HTTP и HTTPS функционируют API для внешних автоматизированных систем и программных комплексов в рамках ПО АПДЦ.

Дополнительно клиентам GIS-систем предоставляется прямо доступ к серверу баз данных MS SQL по протоколу Tabular Data Stream (TDS). Для него используются порты 1433/tcp, 1434/tcp и 1434/udp.

Доступ администратора осуществляется по протоколу RDP (порт 3389/tcp) и SSH (22/tcp, возможен перенос на нестандартный порт для уменьшения потока обращений в поисках уязвимостей).

Для передачи больших файлов дополнительно может использоваться протокол FTP (порт 21/tcp) или SFTP (порт 22/tcp).

3.5. Анализ структуры взаимодействия компонент в почвенно-географической распределенной системе АПДЦ

Структура программных средств АПДЦ

На рис. 9 представлена структура программных средств АПДЦ.

Каждый аграрно-почвенный дата-центр (АПДЦ) (его серверная часть, предоставляющая сервисы для клиентов) состоит из двух виртуальных машин. В одной виртуальной машине работает основной сервер отдельного АПДЦ, в другой – вспомогательный сервер отдельного АПДЦ.

Основная функция основного сервера отдельного АПДЦ – работа системы управления баз данных MS SQL Server.

Основная функция вспомогательного сервера отдельного АПДЦ – предоставление доступа к сервисам АПДЦ на основе данных, хранящимся на основном сервере отдельного АПДЦ.

Кроме отдельных АПДЦ для объединения сервисов предусмотрен агрегатор сети АПДЦ. Устройство агрегатора сети АПДЦ аналогично устройству отдельного АПДЦ, он состоит из двух виртуальных машин. В одной виртуальной машине работает основной сервер агрегатора сети АПДЦ, в другой – вспомогательный сервер агрегатора сети АПДЦ.

Основной сервер отдельного АПДЦ

Основная функция основного сервера отдельного АПДЦ – сервер баз данных. На нем хранятся и обрабатываются данные АПДЦ.

Технические требования, предъявляемые к виртуальной машине основного сервера отдельного АПДЦ

К виртуальной машине, на которой работает основной сервер отдельного АПДЦ, предъявляются следующие технические требования:

объем оперативной памяти: не менее 32 ГБайт;

количество ядер процессора: не менее 12;

тип ядер процессора: Intel Xeon.

Программное обеспечение основного сервера отдельного АПДЦ

Операционная система основного сервера отдельного АПДЦ имеет следующие характеристики:

а) тип операционной системы: Windows Server;

б) версия операционной системы: 2016;

в) редакция операционной системы: Standard.

Необходимые лицензии операционной системы:

а) WinSvrSTDCore 2019 SNGL OLV 2Lic NL Each AP CoreLicWinSvrCAL x8

б) 2019 SNGL OLV NL Each AP UsrCAL x1

Администрирование операционной системы осуществляется удаленно при помощи протокола RDP.

Система управления баз данных (СУБД) основного сервера отдельного АПДЦ имеет следующие характеристики:

а) Тип СУБД: Microsoft SQL Server

б) Версия СУБД: 2016

Необходимые лицензии СУБД:

а) SQLCAL 2019 SNGL OLV NL Each AP UsrCAL x1

б) SQLSvrStd 2019 SNGL OLV NL Each AP x1

В СУБД данные хранятся в виде более 200 связанных таблиц с пространственно-атрибутивной информацией и справочниками. Данные организованы в соответствии с реляционной моделью данных.

Для удобства обращения к данным и разделения прав доступа организованы представления данных (Views).

В рамках СУБД функционирует SQL Agent. Эта служба в составе СУБД Microsoft SQL Server позволяет запускать по расписанию задания, связанные с действиями с данными, хранящимися в СУБД.

Для переноса части алгоритмов обработки данных в СУБД организованы хранимые процедуры (Stored Procedures, SP). Хранимые процедуры позволяют выполнять некоторые действия по заранее заданному

алгоритму. При этом хранимые процедуры могут использоваться и для управления правами доступа к разным объектам, хранящимся в СУБД.

Функции в рамках СУБД – заранее запрограммированный способ вычисления по заданной формуле.

Табличные функции, в отличие от обычных функций, могут возвращать не одно значение (один кортеж), а несколько кортежей, то есть таблицу.

Вспомогательный сервер отдельного АПДЦ

Основная функция вспомогательного сервера отдельного АПДЦ – предоставление доступа к сервисам АПДЦ на основе данных, хранящимся на основном сервере отдельного АПДЦ.

Технические требования, предъявляемые к виртуальной машине вспомогательного сервера отдельного АПДЦ

К виртуальной машине, на которой работает вспомогательный сервер отдельного АПДЦ, предъявляются следующие технические требования:

а) объем оперативной памяти: не менее 512 МБайт;

количество ядер процессора: не менее 1;

тип ядер процессора: Intel Xeon.

Программное обеспечение вспомогательного сервера отдельного АПДЦ

Операционная система вспомогательного сервера отдельного АПДЦ имеет следующие характеристики:

а) тип операционной системы: Arch Linux или Alpine Linux.

Необходимые лицензии операционной системы: операционная система является свободно распространяемой под лицензией GNU General Public License (GNU GPL).

Администрирование операционной системы осуществляется удаленно при помощи протокола SSH.

Основным программным обеспечением, выполняющимся на вспомогательном сервере отдельного АПДЦ, является веб-сервер Apache HTTPD.

При помощи веб-сервера реализованы программные компоненты, отвечающие за предоставление сервисов АПДЦ пользователям:

а) web –приложение;

б) JSON API;

в) API загрузки XLS;

г) JSON+XML API;

д) Специализированные программные комплексы (ПК).

Web-приложение является основным способом доступа к сервисам АПДЦ для клиентов, получающих доступ при помощи веб-браузера через веб-страницу. Доступ к web-приложению осуществляется по протоколу HTTP (HTTPS). Помимо отображения в виде веб-страницы web-приложение может выдавать табличную информацию в формате Microsoft Excel (XLS).

JSON API является дополнительным способом получения тех же возможностей, которые предоставляет web-приложение, только в виде стандартизованного API, с которым могут взаимодействовать программные системы. Доступ к нему также предоставляется по протоколу HTTP (HTTPS).

API загрузки XLS предназначено для загрузки больших объемов табличной информации, представленной в формате Microsoft Excel (XLS). Основным протоколом доступа является HTTP (HTTPS), однако для упрощения операций с файлами большого объема могут использоваться протоколы FTP и/или SFTP. Загружаемый файл сопровождается атрибутами, оформленными в соответствии со стандартами OGC 14-013r1 – ServiceIntegration; ISO 28258 – SoilQualityInformationExchange; ISO/TC 190/SC – SoilQualityFieldDescription.

JSON+XML API предназначено для получения доступа к сервисам АПДЦ из внешних автоматизированных систем. В рамках этого компонента реализовано REST API. Пользователями внешних автоматизированных систем могут быть органы власти Российской Федерации, такие как МинСельХоз или МинИмущества.

Программные комплексы (ПК) – специализированные компоненты для решения конкретных задач. Каждый программный комплекс функционирует в рамках веб-сервера и состоит из программного обеспечения собственно программного комплекса и хэндлера. Доступ к хэндлеру осуществляется по протоколу HTTP (HTTPS). Хэндлер реализует интерфейс и осуществляет взаимодействие с пользователем программного комплекса.

Способы взаимодействия между основным и вспомогательным серверами отдельного АПДЦ

Для реализации сервисов для клиентов АПДЦ вспомогательный сервер отдельного АПДЦ взаимодействует с основным сервером отдельного АПДЦ. Способы взаимодействия для различных компонентов, работающих на вспомогательном сервере отдельного АПДЦ, перечислены ниже.

Web-приложение для своей работы обращается к СУБД, работающей на основном сервере отдельного АПДЦ. Связь обеспечивается драйвером ODBC, и осуществляется по протоколу TDS.

JSON API для своей работы обращается к СУБД, работающей на основном сервере отдельного АПДЦ. Связь обеспечивается драйвером ODBC, и осуществляется по протоколу TDS.

API загрузки XLS для своей работы обращается к СУБД, работающей на основном сервере отдельного АПДЦ. Связь обеспечивается драйвером ODBC, и осуществляется по протоколу TDS. Кроме этого, файл XLS, полученный от пользователя, сохраняется в каталог, который доступен на основном сервере отдельного АПДЦ по протоколу NFS.

JSON+XML API для своей работы обращается к СУБД, работающей на основном сервере отдельного АПДЦ. Связь обеспечивается драйвером ODBC, и осуществляется по протоколу TDS.

Программные комплексы для своей работы обращается к СУБД, работающей на основном сервере отдельного АПДЦ. Связь обеспечивается драйвером ODBC, и осуществляется по протоколу TDS.

Способы взаимодействия между пользователями и основным и вспомогательным серверами отдельного АПДЦ

В зависимости от категории пользователей применяются описанные ниже типы взаимодействия.

Основная часть пользователей АПДЦ получается доступ к сервисам АПДЦ через web-приложение. Некоторая часть дополнительной функциональности реализуется JSON API. Для загрузки данных в АПДЦ применяется API загрузки XLS. Взаимодействие пользователя с этими компонентами происходит при помощи протокола HTTP (HTTPS). Загрузка больших файлов может осуществляться при помощи протокола FTP (SFTP).

Внешние автоматизированные системы для взаимодействия с АПДЦ используют сервис, предоставляемый JSON+XML API. Доступ к этому сервису осуществляется по протоколу HTTP (HTTPS).

Пользователи программных комплексов получаются к ним доступ по протоколу HTTP (HTTPS). Хэндлер внутри программного комплекса отвечает за обработку запросов от пользователя, остальная часть программного комплекса – за реализацию прочей логики работы.

Клиенты GIS-систем, таких как ARCGIS, QGIS, получают прямой доступ к СУБД, работающей на основном сервере отдельного АПДЦ. Доступ к СУБД осуществляется по протоколу TDS при помощи драйвера OLEDB.

Администрирование основного и вспомогательного серверов отдельного АПДЦ

Администрирование основного сервера отдельного АПДЦ осуществляется при помощи встроенных средств установленной на нем операционной системы Microsoft Windows Server – по протоколу Remote Desktop Protocol (RDP).

Администрирование вспомогательного сервера отдельного АПДЦ осуществляется при помощи сервера OpenSSH, входящего в состав ПО операционной системы. Доступ к серверу осуществляется по протоколу SSH.

Основной сервер агрегатора сети АПДЦ

Основная функция основного сервера агрегатора сети АПДЦ – сервер баз данных. На нем хранятся и обрабатываются данные АПДЦ.

Технические требования, предъявляемые к виртуальной машине основного сервера агрегатора сети АПДЦ

К виртуальной машине, на которой работает основной сервер агрегатора сети АПДЦ, предъявляются следующие технические требования:

а) объем оперативной памяти: не менее 32 ГБайт;

б) количество ядер процессора: не менее 12;

в) тип ядер процессора: Intel Xeon.

Программное обеспечение основного сервера агрегатора сети АПДЦ

Операционная система основного сервера агрегатора сети АПДЦ имеет следующие характеристики:

а) тип операционной системы: Windows Server;

б) версия операционной системы: 2016;

в) редакция операционной системы: Standard.

Необходимые лицензии операционной системы:

а) WinSvrSTDCore 2019 SNGL OLV 2Lic NL Each AP CoreLicWinSvrCAL x8

б) 2019 SNGL OLV NL Each AP UsrCAL x1

Администрирование операционной системы осуществляется удаленно при помощи протокола RDP.

Система управления баз данных (СУБД) основного сервера агрегатора сети АПДЦ имеет следующие характеристики:

а) Тип СУБД: Microsoft SQL Server

б) Версия СУБД: 2016

Необходимые лицензии СУБД:

а) SQLCAL 2019 SNGL OLV NL Each AP UsrCAL x1

б) SQLSvrStd 2019 SNGL OLV NL Each AP x1

В СУБД данные хранятся в виде более 200 связанных таблиц с пространственно-атрибутивной информацией и справочниками. Данные организованы в соответствии с реляционной моделью данных.

Для удобства обращения к данным и разделения прав доступа организованы представления данных (Views).

В рамках СУБД функционирует SQL Agent. Эта служба в составе СУБД Microsoft SQL Server позволяет запускать по расписанию задания, связанные с действиями с данными, хранящимися в СУБД.

Для переноса части алгоритмов обработки данных в СУБД организованы хранимые процедуры (Stored Procedures, SP). Хранимые процедуры позволяют выполнять некоторые действия по заранее заданному алгоритму. При этом хранимые процедуры могут использоваться и для управления правами доступа к разным объектам, хранящимся в СУБД.

Функции в рамках СУБД – заранее запрограммированный способ вычисления по заданной формуле.

Табличные функции, в отличие от обычных функций, могут возвращать не одно значение (один кортеж), а несколько кортежей, то есть таблицу.

Вспомогательный сервер агрегатора сети АПДЦ

Основная функция вспомогательного сервера агрегатора сети АПДЦ – предоставление доступа к сервисам сети АПДЦ на основе данных, хранящимся на основном сервере агрегатора сети АПДЦ.

Технические требования, предъявляемые к виртуальной машине вспомогательного сервера агрегатора сети АПДЦ

К виртуальной машине, на которой работает вспомогательный сервер агрегатора сети АПДЦ, предъявляются следующие технические требования:

а) объем оперативной памяти: не менее 512 МБайт;

б) количество ядер процессора: не менее 1;

в) тип ядер процессора: Intel Xeon.

Программное обеспечение вспомогательного сервера агрегатора сети АПДЦ

Операционная система вспомогательного сервера агрегатора сети АПДЦ имеет следующие характеристики:

а) тип операционной системы: Arch Linux или Alpine Linux.

Необходимые лицензии операционной системы: операционная система является свободно распространяемой под лицензией GNU General Public License (GNU GPL).

Администрирование операционной системы осуществляется удаленно при помощи протокола SSH.

Основным программным обеспечением, выполняющимся на вспомогательном сервере агрегатора сети АПДЦ, является веб-сервер Apache HTTPD.

При помощи веб-сервера реализованы программные компоненты, отвечающие за предоставление сервисов сети АПДЦ пользователям:

а) средства интерактивного отображения данных;

б) подсистема ведения каталога веб-сервисов в соответствии с рекомендациями OGC;

в) специализированные программные комплексы (ПК).

Средства интерактивного отображения

Каждый программный комплекс функционирует в рамках веб-сервера и состоит из программного обеспечения собственно программного комплекса и хэндлера. Доступ к хэндлеру осуществляется по протоколу HTTP (HTTPS). Хэндлер реализует интерфейс и осуществляет взаимодействие с пользователем программного комплекса.

Структура технических средств сети АПДЦ

На рис. 10 приведена структура технических средств сети АПДЦ.

Сеть АПДЦ состоит из агрегатора сети АПДЦ и отдельных АПДЦ. Они работают в распределенной среде, связанной каналами связи с пропускной способностью не менее 10 Гбит/с.

Агрегатор сети АПДЦ состоит из основного сервера агрегатора сети АПДЦ и вспомогательного сервера агрегатора сети АПДЦ. Это пара серверов связана между собой изолированной частной сетью для внутреннего обмена данными.

Каждый отдельный АПДЦ состоит из основного сервера отдельного АПДЦ и вспомогательного сервера отдельного АПДЦ. Это пара серверов связана между собой изолированной частной сетью для внутреннего обмена данными.

Клиенты сети АПДЦ могут получать доступ к сервисам отдельного АПДЦ или доступ к сервисам сети АПДЦ через агрегатор сети АПДЦ. Подключение к точке предоставления сервиса производится через публичную сеть (сеть Интернет). Канал, используемый для подключения, должен иметь пропускную способность не менее 100 Мбит/с.

В случае если в сети АПДЦ обрабатываются данные, к которым предъявляются специальные требования по их защите, между распределенной средой, в которой функционирует сеть АПДЦ, и пользователем, могут устанавливаться технические средства защиты информации. Состав таких средств определяется требованиями, предъявляемыми к системам, в которых может обрабатываться информация с соответствующими характеристиками.

Структура системных средств АПДЦ

На рис. 11 приведена структура системных средств сети АПДЦ. Сеть АПДЦ состоит из агрегатора сети АПДЦ и отдельных АПДЦ.

Распределенная среда, в которой функционирует сеть АПДЦ, реализована при помощи среды виртуализации OpenNebula, работающей на оборудовании инфраструктуры ЦХАБД МГУ. Каждый сервер (основной сервер агрегатора сети АПДЦ, вспомогательный сервер агрегатора сети АПДЦ, основной сервер отдельного АПДЦ, вспомогательный сервер отдельного АПДЦ) реализован в виде виртуальной машины, работающей под управлением гипервизора QEMU-KVM. Для каждой пары основной сервер –

вспомогательный сервер создается отдельная изолированная виртуальная сеть, недоступная другим виртуальным машинам.

Внутренняя сеть инфраструктуры ЦХАБД МГУ предоставляет среду связи между виртуальными машинами с пропускной способностью 20 Гбит/с.

Связь инфраструктуры ЦХАБД МГУ с внешними сетями (сетью Интернет) обеспечивается каналом связи с пропускной способностью 1 Гбит/с. Этот канал не ограничивается в использовании виртуальными машинами, выполняющимися на оборудовании инфраструктуры ЦХАБД МГУ.

4. Реализация функциональных требований к АПДЦ в рамках сети АПДЦ

4.1. Подсистема авторизации и защиты информации

Подсистема авторизации и защиты информации должна реализовываться в каждом компоненте сети АПДЦ, предоставляющим сервисы пользователям, и в других компонентах, где это необходимо по соображениям безопасности.

Минимальный уровень реализации авторизации и защиты информации предполагает использование средств, встроенных в использованное для построения сети АПДЦ программное обеспечение.

Так, операционные системы имеют средства контроля прав доступа к файлам и другим объектам. СУБД MS SQL Server имеет развитые средства назначения прав доступа пользователям СУБД на уровне отдельных таблиц, представлений, хранимых процедур и т.п.

Веб-сервер Apache HTTPD имеет средства авторизации, позволяющие разграничить доступ к разным областям веб-приложений для разных пользователей. Пользователь при этом может аутентифицироваться как по сетевому адресу, так и при помощи более сложных механизмов.

Все упомянутые выше в данном пункте программные компоненты реализуют защиту информации путем разграничения доступа к ней.

4.2. Подсистема контроля авторских прав и интеллектуальной собственности

Для контроля авторских прав и интеллектуальной собственности вся хранящаяся в АПДЦ информация маркируется в метаданных владельцем этой информации (организации, от имени которой эта информация внесена).

Для контроля использования информации проводится контроль того, какие из пользователей запрашивали какую информацию. Для простейшей реализации такого контроля можно записывать файлы журналов веб-сервера, в которые помимо стандартной информации заносятся сведения о владельце информации.

4.3. Подсистема ведения дистанционного обучения и поддержки тестирования

Подсистема ведения дистанционного обучения и поддержки тестирования предназначена для поддержки процесса обучения. В рамках этой подсистемы могут создаваться выделенные АПДЦ для обучения, в базе данных которых содержатся учебные или копия реальных данных. Суть такого подхода заключается в том, что при обучении информация может быть модифицирована и стать ошибочной. В то же время для полноценного обучения необходимо иметь возможность проводить полный цикл работы с информацией, включающий ее изменение и удаление.

4.4. Подсистема виртуализации веб-сервисов как основа масштабируемости и тиражируемости распределенных систем

Под подсистемой виртуализации веб-сервисов подразумевается использование технологий виртуализации при создании основного и вспомогательного серверов агрегатора сети АПДЦ и отдельного АПДЦ. Использование технологий виртуализации имеет следующие преимущества.

Облегчается тиражирование создаваемых решений. Разворачивание копий виртуальных машин с заранее созданного шаблона – быстрая и легкая операция.

Облегчается поддержка создаваемых решений. Использование виртуализации позволяет не закладываться на конкретные аппаратные решения.

Легкое копирование виртуальных машин также означает легкость масштабирования созданных средств. Для создания еще одного АПДЦ. Фактически, нужно создать дополнительные виртуальные машины и сеть между ними.

Облегчается адаптация разработанных решений под повышающуюся нагрузку. В отличие от физических серверов, изменение конфигурации виртуальных машин гораздо проще, и в некоторых случаях может быть выполнено даже без остановки работы.

4.5. Средства интерактивного отображения данных на картографических основах

Средства интерактивного отображения данных на картографических основах Yandex, Google, ESRI, обеспечивающие предоставление сведений посредством следующих web-сервисов по открытым спецификациям OGC: WMS (WebMapService) – картографические сервисы, предоставляющие клиентам изображения; WFS (WebFeatureService) – сервисы, предоставляющие клиентам координатное описание объектов; WCS (WebCoverageService)

Картографические сервисы Yandex, Google, ESRI имеют картографическую основу, на которой можно отобразить свои данные. Для этого у названных сервисов существует API, доступное по протоколу HTTPS. Для его работы необходимо загрузить файл со сведениями, которые нужно отобразить, и выбрать параметры отображения.

В рамках создаваемых средств предусматривается возможность преобразования данных в форматах OGC в формат, пригодный для картографических сервисов, и сопряжение двух типов форматов.

4.6. Подсистема ведения каталога веб-сервисов в соответствии с рекомендациями OGC

Стандарт OGC Catalogue Services описывает модель для унифицированного описания сервисов и наборов данных с пространственной информацией. Также стандартизован способ предоставления доступа к этой информации по протоколу HTTP (HTTPS).

В рамках создаваемых средств создается реализация этого стандарта для описания тех сервисов, которые предоставляются пользователям в рамках сети АПДЦ.

5. Разработка подхода к защите доступа и обеспечению информационной безопасности

Разработка подхода к разграничению доступа на базе общедоступного инструментария, реализующего различные уровни доступа и защиты доступа по каналам связи к объектам почвенно-географической распределенной системе АПДЦ

5.1. Разграничение доступа по сети средствами операционных систем

В состав операционных систем, применяемых на основном и вспомогательном серверах отдельного АПДЦ, а также на основном и вспомогательном серверах агрегатора сети АПДЦ имеются средства фильтрации сетевого трафика (брандмауэры).

В ОС Windows Server, используемой на основном сервере отдельного АПДЦ и основном сервере агрегатора сети АДЦ, это Windows Firewall.

В ОС семейства Linux, используемых на вспомогательном сервере отдельного АПДЦ и вспомогательном сервере агрегатора сети АДЦ, это пакет iptables.

Оба средства имеют развитые средства по разграничению доступа на основе сетевых адресов клиентов и тех сервисов, к которым эти клиент пытаются подключаться. Поддерживаемые типы информации, на основе которой может приниматься решение и применении ограничений к трафику:

IP-адрес подключающегося клиента;

а) диапазон IP-адресов, которому принадлежит адрес подключающегося клиента;

транспортный протокол, используемый подключающимся клиентом (TCP,UDP, …);

порт транспортного протокола, с которого устанавливается соединение;

IP-адрес сервера, к которому осуществляется подключение (при наличии у сервера нескольких адресов, можно ограничивать подключения на части из них или устанавливать разные правила для разных адресов);

порт транспортного протокола, к которому осуществляется подключение.

Также доступна и другая информация о трафике, которая используется реже.

На основе правил, построенных с использованием указанных сведений о трафике, могут быть приняты следующие решения:

а) разрешить соединение;

запретить соединение;

ограничить соединение (например, разрешить установление не более заданного количества соединений с одинаковыми параметрами или ограничить полосу пропускания для заданного соединения);

Двумя возможными стратегиями построения правил фильтрации трафика являются «разрешено все, что не запрещено» (черные списки) и «запрещено все, что не разрешено» (белые списки).

Стратегия использования белых списков хорошо подходит в ситуации, когда у определенного сервиса имеется ограниченный набор клиентов, которые имеют фиксированные сетевые адреса. В этом случае можно составить белый список таких адресов (диапазонов адресов), запретив подключения со всех остальных адресов. Такое использование в значительной степени затрудняет организацию несанкционированного доступа к сервисы, особенно если сети, откуда получают доступ клиенты,

сами по себе в достаточной мере защищены от несанкционированного доступа.

Примером сервиса АПДЦ, который целесообразно защищать с применением стратегии белого списка – доступ клиентов GIS-систем к СУБД, работающей на основном сервере отдельного АПДЦ. Это привилегированный доступ, который должен предоставляться только небольшому количеству сотрудников АПДЦ. Поэтому можно перечислить сетевые адреса рабочих мест этих сотрудников, запретив доступ всем остальным.

Стратегия использования черных списков применяется там, где заранее невозможно определить сетевые адреса клиентов, которым надо получать доступ к сервису. В этом случае доступ разрешается всем, кроме некоторого списка адресов, откуда, например, проводились попытки получения несанкционированного доступа или другой вид атаки.

Примером сервиса, к которому практически невозможно составить белый список – доступ к веб-страницам, особенно если этот доступ необходим клиентам с мобильных устройств.

Промежуточной стратегией ограничения доступа, которая может быть применена в сети АПДЧ, может быть ограничение по сетевым адресам, принадлежащим отдельным странам. Списки IP адресов с привязкой по странам могут быть получены или из баз географической привязки IP-адресов (GeoIP), или из сведений региональных интернет-регистраторов (RIR): ARIN, RIPE. NCC, APNIC, LACNIC, AfrNIC.

Разграничение доступа средствами веб-сервера Apache HTTPD

Веб-сервер Apache HTTPD имеет развитые средства для разграничения доступа. Однако основная функциональность этих средств ориентирована на аутентификацию средствами протокола HTTP и разграничение доступа н основе этой аутентификации. Однако такой способ аутентификации не применяется широко, и не знаком значительной части пользователей. Реализация его в рамках сети АПДЦ не представляется перспективной.

Кроме разграничения доступа на основе имени аутентифицированного пользователя, Apache HTTPD предоставляет средства разграничения доступа на основе IP-адреса клиента (модуль mod_authz_host). В отличие от разграничения доступа по IP-адресу средствами ОС, разграничение доступа по IP-адресу средствами веб-сервера позволяет регулировать доступ к отдельным областям сайта (веб-приложения).

Определение области веб-сайта, к которому ограничен доступ, может производиться на основании следующих данных:

а) URL в запросе HTTP;
б) соответствия URL заданному регулярному выражению (regexp);
в) имени запрашиваемого файла в файловой системе;
г) соответствия имени запрашиваемого файла маске оболочки (wild-card, shell glob);
д) соответствия имени запрашиваемого файла регулярному выражению (regexp);
е) имени каталога файловой системы, в который транслируется обрабатываемый запрос;
ж) соответствия имени каталога файловой системы, в который транслируется обрабатываемый запрос, маске оболочки (wild-card, shell glob);
з) соответствия имени каталога файловой системы, в который транслируется обрабатываемый запрос, регулярному выражению;
и) методу HTTP в обрабатываемом запросе;
к) заголовкам HTTP в обрабатываемом запросе;

Определение сущности клиента может производиться на основании следующей информации:

а) IPv4-адрес клиента;
б) соответствие IPv4-адреса клиента паре адрес сети/маска сети;
в) соответствие IPv4-адреса клиента паре адрес сети/длина адреса сети (CIDR notation);

г) IPv6-адрес клиента;

д) соответствие IPv6-адреса клиента паре адрес сети/длина адреса сети (CIDR notation);

е) символическое имя, соответствующее IP-адресу клиента;

ж) соответствие символического имени, соответствующего ШЗ-адресу клиента, маске оболочки;

з) соответствие IP-адреса клиента IP-адресам, получаемым из разрешения заданного символического имени.

Решение о возможности доступа принимается на основе любых сочетаний вышеуказанных признаков. Это дает возможность применять гибкие подходы к разграничению доступа средствами веб-сервера Apache HTTPD.

Разграничение доступа средствами СУБД Microsoft SQL Server

Применяемая в сети АПДЦ СУБД Microsoft SQL Server имеет развитые средства управления доступом. Все соединения к СУБД требуют аутентификации. После аутентификации соединение ассоциировано с каким-то именем пользователя, и все права доступа относятся к этому пользователю.

Microsoft SQL Server поддерживает следующие способы аутентификации:

а) аутентификация Windows (для аутентификации пользователя в SQL Server необходимо пройти аутентификацию в ОС Windows);

б) собственная аутентификация SQL Server.

Аутентификация Windows удобна в том случае, когда все пользователи, которым необходимо иметь доступ к СУБД, имеют учетные записи в операционной системе. Однако это же приводит к необходимости более тщательной настройки разрешений для пользователей в рамках операционной системы, чтобы не предоставить тем пользователям, которым нужен доступ исключительно к СУБД, разрешений на доступ к другим объектам ОС.

Аутентификация средствами SQL Server позволяет разделить пользователей ОС и пользователей СУБД, однако в этом случае пользователям, имеющим учетные записи и в ОС, и в СУБД, требуется проходить аутентификацию дважды.

Разрешения могут быть выданы:

а) конкретному пользователю;

б) группе пользователей;

в) роли.

Разрешения могут применяться:

а) к базе данных целиком;

б) к таблице;

в) к колонке;

г) к представлению (view);

д) к функции;

е) к хранимой процедуре;

ж) другим объектам СУБД.

Помимо разрешения на выполнение действия с объектом, также может быть предоставлено разрешение на предоставление этого же разрешения другим пользователям (группам, ролям).

Отметим, что в рамках сети АПДЦ доступ к СУБД имеют две категории приложений: приложения, выполняемые на вспомогательных серверах агрегатора сети АПДЦ и отдельного АПДЦ, и клиенты GIS-систем. Учетные записи, от имени которых будут осуществлять доступ эти приложения, требуют особо тщательного отношения в настройке разрешений, так как в обоих случаях обе этих категории могут стать потенциальным вектором для получения несанкционированного доступа как доступные удаленно.

5.2. Разграничение доступа средствами прикладных приложений

Самым гибким вариантом реализации разграничения доступа является реализация такого разграничения в прикладном программном обеспечении. В

случае АПДЦ это программное обеспечение, которое работает в рамках веб-сервера вспомогательных серверов агрегатора сети АПДЦ и отдельного АПДЦ. Именно на этом уровне известны все объекты, к которым нужно разграничивать доступ, и все характеристики субъектов доступа, на основании которых необходимо принимать решение о предоставлении или отказе в доступе.

Разграничение доступа между пользователями операционной системы

При создании сети АПДЦ не предполагается, что пользователи отдельных АПДЦ или агрегатора сети АПДЦ будут иметь доступ к сервисам операционной системе, предоставляемым аутентифицированным пользователям. Такой уровень доступа будут иметь администраторы АПДЦ. Вопросы администрирования АПДЦ выходят за рамки данного документа.

5.3. Защита каналов связи

Все протоколы, используемые для коммуникации с внешними для отдельного АПДЦ или агрегатора сети АПДЦ, субъектами, имеют возможность построения безопасного соединения. Для протокола HTTP – это перенаправление на использование HTTPS.Протокол TDS имеет возможность организации защищенного канал при помощи технологии SSL. Используемый при администрировании ОС Windows протокол RDP также имеет возможность использования SSL. Используемый при администрировании Linux протокол SSH изначально разработан для безопасного использования и также имеет все необходимые средства.

Связь между основным и вспомогательным серверами отдельного АПДЦ или между основным и вспомогательным серверами агрегатора сети АПДЦ изолируется средствами используемой системы виртуализации. Такая изоляция считается достаточной, и использовать дополнительные средства защиты для коммуникаций по этому каналу нецелесообразно.

Протоколы, которые будут применяться в коммуникациях между агрегатором сети АПДЦ и отдельными АПДЦ, на текущем этапе разработки не специфицированы. Может быть удобно применить протоколы, не

поддерживающие организацию защищенного канала, и защиту коммуникаций осуществлять при помощи технологий организации виртуальных частных сетей (VPN). И даже если использованные протоколы поддерживают организацию защиты, применение VPN может оказаться целесообразным для того, чтобы вынести нагрузку по шифрованию трафика (основной вид вычислительной нагрузки при организации защищенных каналов) на отдельные устройства, физические или виртуальные.

5.4. Особые случаи, требующие дополнительных инструментов защиты

В некоторых случаях обработка данных может иметь нормативные требования по обеспечению безопасности. К таким случаям относится обработка персональных данных, обработка сведений, составляющих государственную тайну, и другие. В некоторых случаях повышенные требования по обеспечению безопасности могут быть обусловлены документами, принятыми негосударственными организациями (например, стандарты PCI DSS).

В части требования к обеспечению безопасности обработки информации могут быть предъявлены следующие требования:

а) организационное сопровождение (обеспечение процессов обработки информации в соответствии с требованиями);
б) использование программного обеспечения, имеющего заданные сертификаты;
в) использование сертифицированных средств криптографической защиты информации (СКЗИ);
г) использование средств обнаружения и предотвращения вторжений;
д) другие требования.

В настоящее время в рамках сети АПДЦ не планируется обработка информации с особенными требованиями по ее защите. Однако если в будущем такая обработка потребуется, необходимо будет учесть соответствующие требования. В некоторых случаях это может потребовать

изменения используемого программного обеспечения вплоть до смены используемых операционных систем

5.5. Предложение реализации разработанного подхода, основанного на возможностях ЦХАБД МГУ

Возможности ЦХАБД МГУ предоставляются пользователям инфраструктуры ЦХАБД МГУ в виде создания виртуальных машин (одна или несколько на проект), создания частных сетей (одна или несколько на проект), обеспечение связи с внешними сетями (сетью Интернет), предоставление пространства для хранения данных.

Все вышеперечисленные сервисы предоставляются при помощи среды виртуализации OpenNebula и системы хранения на основе Ceph.

Предоставление виртуальных машин для создания агрегатора сети АПДЦ или создания отдельного АПДЦ

Для создания отдельного АПДЦ или создания агрегатора сети АПДЦ выделяется две отдельных виртуальных машины. Характеристики создаваемых виртуальных машин приведены в соответствующем разделе.

На выделенные виртуальные машины производится установка базового варианта операционной системы.

На основной сервер отдельного АПДЦ и основной сервер агрегатора сети АПДЦ устанавливается ОС Microsoft Windows Server. При этом не производится активация с вводом лицензионного ключа. Ввод ключа производится администратором агрегатора сети АПДЦ или отдельного АПДЦ после передачи ему контроля. Таким образом, ответственность за приобретение лицензий лежит на лице, эксплуатирующем АПДЦ.

На вспомогательный сервер отдельного АПДЦ и вспомогательный сервер агрегатора сети АПДЦ устанавливается ОС Arch Linux или Alpine Linux. Так как эти ОС распространяются под свободной лицензией GPL, вопрос ввода лицензионных ключей и т.п. подтверждения прав на Ос не стоит.

Создание частной сети

Средствами среды виртуализации OpenNebula создается частная сеть для пары основной сервер – вспомогательный сервер (агрегатора сети АДЦ или отдельного АПДЦ). Эта сеть доступна только двум виртуальным машинам, для которых она выделена. Ее изоляция обеспечивается средствами среды виртуализации.

Обеспечение связи с внешними сетями

Для обеспечения возможности связи с внешними сетями (сетью Интернет) вспомогательному серверу агрегатора сети АПДЦ или отдельного АПДЦ выделяется глобально маршрутизируемый IP-адрес. Средствами среды виртуализации обеспечивается возможность устанавливать соединения на этот адрес или с этого адреса.

Этот адрес указывается как адрес подключения для клиентов отдельного АПДЦ или агрегатора сети АПДЦ.

Предоставление пространства для хранения данных

Пространство для хранения данных выделяется в виде блочного устройства RBD Ceph. Выделенное блочное устройство системой виртуализации OpenNebula передается в виртуальную машину как обычное блочное устройство (диск). Создание файловой системы на этом блочном устройстве и создание структур, необходимых для хранения данных, осуществляется средствами операционных систем.

5.6. Анализ сценариев взаимодействия разных групп пользователей с подсистемами АПДЦ

Выделение сценариев взаимодействия пользователей с системой, или ролей пользователей, может осуществлять с различных точек зрения, таких как методы аутентификации и права доступа к данным или функциональная роль (администрирование, пользователь, внешний запрос и т.п.).

С точки зрения организационной структуры (рис.7) можно выделить две большие группы пользователей: поставщики информации и потребители

информации, имеющие в результате проведенного анализа следующую структуру относительно автоматизированности действий:

— ручной режим

- пользователи Web-портала
  - ♦ привилегированный сотрудник АПДЦ
  - ♦ сотрудники АПДЦ с правами просмотра
  - ♦ пользователь программных комплексов
- пользователи настольных GIS систем
  - ♦ сотрудник АПДЦ

— автоматический режим – сторонние пользователи REST API

- привилегированный пользователь
- пользователь с правами только на просмотр

С точки зрения доступа к системным настройкам в рамках каждого АПДЦ можно выделить минимум два уровня администрирования, что отдельно рассмотрено в других разделах:

— администрирование виртуальной машины или сервера, на котором развернуто ПО;

— администрирование БД.

С технологической точки зрения сценарии взаимодействия пользователей с системой делятся на следующие группы, рассмотренные в соответствующих разделах:

— доступ через Web-сервер (в т.ч. через web-портал);

— доступ через специализированные API;

— доступ через RDP и SSH для администрирования серверов;

— доступ к базе данных и ее компонентам через TDS.

5.7. Требования безопасности к рассматриваемым АПДЦ

Электронные средства идентификации и аутентификации пользователей

Каждый из авторизованных пользователей АПДЦ имеет свою учетную запись и пароль. Доступ к системе без пароля запрещён. Все пароли должны

состоять из набора букв в разных регистрах, цифр и дополнительных символов. Пароли должны быть достаточной длины, что исключает возможность их подбора за разумное время.

Каждый пользователь однозначно идентифицируется на основании своей учётной записи, которая определяет в дальнейшем его права и уровень доступа. На всех пользователей налагаются требования следить за сохранностью и конфиденциальностью выданных им реквизитов доступа.

Для усиления алгоритма авторизации может применяться двухфакторная авторизация с помощью подтверждения личности пользователя SMS сообщением, специальным аппаратным ключом, сканером отпечатка пальца и т.д.

Защита от несанкционированного доступа к данным

После авторизации пользователей в системе, разграничение прав доступа контролируются встроенными в операционную систему Windows Server 2016 и в базу данных MS SQLServer средствами, а также программными продуктами Заказчика.

Ограничение прав пользователя должно базироваться на инклюзивном принципе: все действия по умолчанию запрещены, а списки действий в рамках используемого программного обеспечения отдельно разрешены для пользователей определенных категорий. Для этого должны быть определены группы пользователей в каждой категории и список разрешений для каждой групп. Количество таких групп и разрешенные действия определяются логикой работы АПДЦ и стоящими перед пользователями задачами. Заказчик следит за тем, чтоб каждая группа пользователей получала доступ только к той части информации, которая необходима ей для работы.

Доступ к полной копии данных и управлению программами должен быть только у ограниченной группы администраторов АПДЦ. Обычные пользователи не имеют доступ к исходным данным, а только к результатам обработки их запросов. Действия пользователей должны логироваться и

анализироваться администраторами АПДЦ на предмет неправомочных действий.

Заказчик следит за тем, что все обычные действия на серверах АПДЦ выполнялись под непривилегированной учетной записью, а административный доступ с расширенными правами осуществлялся только в крайних случаях и только для выделенной группы администраторов АПДЦ. Действия администраторов должны быть регламентированы и протоколироваться.

Построенная система безопасности охватывает инфраструктуру ЦХАБД, виртуальные сервера АПДЦ, каналы связи между ними и до рабочего компьютера конечного пользователя. При существующей архитектуре нет возможности контролировать программное окружение на рабочих компьютерах конечных пользователей, поэтому запросы с них должны считаться потенциально не безопасными.

При необходимости усиления контроля за действиями пользователей АПДЦ, гибкости и удобством управления большим количеством клиентов, Заказчиком может быть развернута доменная структура на основе Active Directory (AD) для управления групповыми политиками.

Несанкционированный доступ к данным может произойти при ошибках программного обеспечения, выбранных политик разграничения доступа, нарушении безопасности каналов связи или при компрометации учетной записи пользователя.

Все программные компоненты АПДЦ проходят предварительное тестирование, описанное в пункте 2.12.

Для защиты подключений к виртуальным серверам АПДЦ применяются меры, писанные в пункте 9.5, описывающем технологию защищённых узлов доступа к АПДЦ.

Для выявления случаев компрометации учётных записей пользователей, администраторами АПДЦ производится анализ логов действий пользователей. В программном обеспечении могут быть настроены

оповещения (алерты) случаев запросов на превышения необходимых привилегий для более пристального анализа действий пользователей.

Данные, находящие на виртуальных серверах АПДЦ в облаке ЦХАБД и копий бэкапов локализованы в охраняемом помещении, к которому нет доступа посторонних лиц. Обмен данными между компонентами АПДЦ, находящихся на площадке ЦХАБД, происходит в выделенной VLAN, что исключает доступ к данным посторонних программ и сервисов.

Средства мониторинга и фильтрации содержимого web-ресурсов

После развертывания АПДЦ или смены версий программных компонент для проверки комплекса на отсутствие известных ошибок может быть проведён процесс сканирования открытых портов и протоколов сканерами безопасности (Vulnerability Scanner). Список подобных сканеров можно посмотреть здесь: https://owasp.org/www-community/Vulnerability_Scanning_Tools

После любого обновления версий программного обеспечения или ввода в эксплуатацию новых программных модулей по протоколу, описанному в пункте 2.12, сканирование должно быть произведено заново.

При необходимости дополнительного контроля доступа и сохранности конфиденциальных данных могут разворачиваться программно-аппаратные комплексы IDS (Intrusion Detection System) (рис.4).

При необходимости может быть инициирован процесс соответствия системы АПДЦ международному сертификату по информационной безопасности ISO/IEC 27001.

Средства управления безопасностью информации

Риски перерывов в работе АПДЦ могут быть связаны с нарушений работоспособности инфраструктуры комплекса, утратой или повреждением данных, непреднамеренными действиями авторизованного персонала или с действиями злоумышленников.

1) К программно-техническим факторам риска относятся:

a) События локализованные в пределах ЦХАБД: сбои аппаратуры датацентра, выход из строя отдельных компьютерных компонент, нарушение сетевой связанности элементов комплекса, аварии питания или климатического контроля;
b) Потери данных из-за выхода из строя носителя данных;
c) События вне пределов датацентра, которые не контролируются персоналом ЦХАБД (события непреодолимой силы): глобальные отключение питания, повреждение внешних магистральных каналов связи, временные технические проблемы провайдеров доступа в Интернет, технические и регламентные работы с критическим оборудованием для работы ЦХАБД, проводимые в рамках поддержания и развития инфраструктуры МГУ ;
d) Программный сбой при сочетании определенных факторов (баг) на оборудовании комплекса или в программном обеспечении АПДЦ;

2) К непреднамеренным действиям авторизованного персонала (человеческий фактор) относятся:
   a) Действия оператора, которые могут привести к потере или порче данных;
   b) Нарушение логической целостности из-за ошибки входных данных или действий оператора;
3) Доступ к работе комплекса предоставляется только авторизированным пользователям. Получения доступа к конфиденциальным данным или несанкционированные изменения в работе комплекса должны рассматриваться как злоумышленные действия.

Для предотвращения проблем с функционированием комплекса, описанные в пункте 1a, все элементы системы АПДЦ размещаются на высоконадёжном оборудовании серверного класса в облаке ЦХАБД:

— Все критические компоненты инфраструктуры продублированы – каждый из серверов оснащён 2 блоками питания, запитанных по независимым фидерам.

— Компьютерная сеть организована на дублирующих коммутаторах включенных параллельно таким образом, что выход из строя одного из них не приводит к остановке работы облака ЦХАБД.
— Всё оборудование расположено в специальном помещении, в которое нет доступа для посторонних лиц. Система климат-контроля поддерживает в помещении оптимальную температуру и не допускает перегрев оборудования. Помещение оборудовано системой аварийного пожаротушения.
— Оборудование облака ЦХАБД обеспечено мощным источником бесперебойного питания с батареями достаточной ёмкости, который позволяет корректно заглушить рабочие сервисы при авариях питающего напряжения.
— Работоспособность и нагрузка на всё компьютерные элементы постоянно мониторится, что позволяет оперативно локализовать вышедшее из строя оборудование и планировать нагрузку в облаке ЦХАБД
— Элементы АПДЦ расположены на виртуальных серверах в облаке ЦХАБД. Даже при наличии проблем в работе физического сервера, на котором в данный момент времени работает виртуальная машина АПДЦ, система управления виртуализацией автоматически переносит виртуальный сервер АПДЦ на другой исправный физический сервер в облаке ЦХАБД. Таким образом, достигается независимость работы комплекса от аппаратных сбоев оборудования.

Для исключения потери данных из-за выходя из строя физического носителя (пункт 1b) в качестве хранилища выбрано решение на базе Software Defined Storage (SDS) Ceph. Выбор был обусловлен комплексом следующих уникальных особенностей:

— В работе хранилища упор сделан на надежность сохранения данных:
    - все дисковые операции транзакционны;

- в конфигурации по-умолчанию данные хранятся одновременно в 3 экземплярах (replica=3);
- в конфигурации хранилища можно выбрать условия, что данные распределены по дискам и серверам таким образом, что выход из строя любого диска или даже целого сервера не приведёт к остановке работы хранилища и доступности данных;
- при выходе из строя каких-то компонент, система управления автоматически перестроит конфигурацию для распределения данных по рабочим дискам и серверам без перерывов в работе хранилища (ребалансинг);
- в случае одновременного выхода из строя большого количество компонентов хранилища, данные перейдут в режим только для чтения, что предотвратит их порчу;

— У SDS Ceph модель разработки с открытым исходным кодом, что гарантирует отсутствие недокументированных возможностей по не санкционированному доступу к данным (back door). Такая модель разработки так же предполагает возможность использования программных и аппаратных компонент без привязки только к одному к определенному производителю (Vendor Lock).

— Данное программное обеспечение находится в стадии активной разработки, постоянно выходят новые версии и идет работа над исправление ошибок в существующих релизах.

— Совместимость с популярным стандартом S3, наличие POSIX совместимой файловой системы CephFS, возможность организации подключения по протоколам NFS, iSCSI и т.д.

— Нативная поддержка работы с системами виртуализации OpenStack, OpenNebula и др.

Описанные особенности выбранного продукта позволяют гарантировать отсутствие перерывов в работе хранилища при единичных случаях выхода из строя оборудования и обеспечить максимальную

сохранность данных при максимальной гибкости вариантов доступа к ним. Хранилище Ceph позволяет наращивать ёмкость без перерывов в работе (известны практические примеры эксплуатации хранилища с 10000 дисками), что позволяет, при необходимости, достижения любой разумной ёмкости данных для текущих и будущих применений. Гибкость вариантов подключения выбранного продукта гарантирует доступность данных при развитии функционала АПДЦ и создании новых вариантов использования накопленной информации.

События, описанные в пункте 1c, относятся к обстоятельствам непреодолимой силы и не контролируются сотрудниками ЦХАБД. Персонал находится в постоянном контакте с администрацией МГУ и городскими службами и будет заблаговременно информировать Заказчика о всех планируемых событиях.

В случае необходимости может быть рассмотрен вариант, при котором строится аналогичная инфраструктура АПДЦ в резервном датацентре, которая вступает в строй в случае глобальных катаклизмов, связанных с недоступностью или невозможностью функционирования основного варианта АПДЦ. В этом случае разрабатывается протокол переключения работы датацентров «основной»-«резервный» и вырабатываются требования к консистентности данных в разных экземплярах АПДЦ.

Вся инфраструктура ЦХАБД расположена в госучереждении на территории Российской Федерации, что исключает политические риски, связанные с ограничением доступа к зарубежным или коммерческим ресурсам.

Персонал ЦХАБД постоянно мониторит бюллетени безопасности имеющие отношения к оборудованию ЦХАБД и используемому программному обеспечению для устранения рисков, описанных в пункте 1d. Заказчик, со своей стороны, следит за тем, чтоб на виртуальных серверах АПДЦ происходила регулярная установка обновлений безопасности операционной системы и сигнатур антивирусных программ.

Для предотвращения потери данных, описанных в пункте 2a, виртуальные сервера АПДЦ регулярно проходят процедуру бэкапа, что позволяет восстановить работоспособность комплекса на дату последнего бэкапа. Все бэкапы производятся автоматически без перерывов в работе сервисов по согласованному с Заказчиком графику.

Контроль за вводом новых данных и сохранение логической связанности информации (пункт 2b) осуществляется программным обеспечением Заказчика.

Все оборудование ЦХАБД находится в помещении, к которому нет доступа у посторонних лиц, поэтому физический контакт с оборудованием возможен только для сотрудников ЦХАБД и представителей Заказчика.

Технологии защищенных узлов подключения к Интернет и открытым сетям

Транспортное шифрование в подключениях к Сети

Вся коммуникация между виртуальными серверами АПДЦ в пределах ЦХАБД происходит в выделенной VLAN. Для предотвращения получения доступа неавторизованных лиц к конфиденциальным данным или несанкционированного изменения в работе комплекса все критические потоки информации между территориально-распространёнными компонентами АПДЦ осуществляются только по зашифрованным каналам связи, что исключает возможность прослушивания (снифинг). Такие каналы связи могут быть организованы либо внешними средствами (VPN), либо с помощью канального шифрования, если такая возможность поддерживается используемым программным обеспечением. Использование открытых каналов (например, порт 80 HTTP) допускается только для дальнейшего перенаправления пользователя на закрытые каналы (порт 443 HTTPS).

На виртуальных серверах АПДЦ должен быть включён брандмауэр (файрволл), в котором закрыты все порты кроме портов из списка разрешенных (см. рис.9).

Все используемые протоколы взаимодействия с компонентами АПДЦ должны быть сконфигурированы с поддержкой максимального уровня безопасности:

— Протокол RDP (TCP порт 3389) для удалённого подключения к виртуальными серверам под управлением операционной системы Microsoft Windows Server 2016 должен быть сконфигурирован с уровнем безопасности SSL TLS и выбран высокий уровень шифрования клиентских подключений. Для соединения должны быть использованы FIPS-совместимые алгоритмы шифрования. Для дополнительной безопасности подобных подключений доступ по этому протоколу может быть ограничен только со списком определенных IP адресов, используемых администраторами АПДЦ. Возможно так же организация дополнительного прокси-сервера RDP подключений, когда сотрудники из группы администраторов АПДЦ авторизуются вначале на прокси-сервере с определенным IP, а с него уже подключаются к виртуальным серверам АПДЦ. В этом случае на брандмауэре (firewall) виртуальных серверов АПДЦ должен быть настроен для фильтрации RDP подключений только со списка IP прокси серверов. Для уменьшения входного трафика, сканирующего известные порты, имеет смысл перенести порт RDP подключений на нестандартный номер порта.

— Для управления серверами под управлением операционной системы Unix (Linux) должен использоваться протокол SSH (TCP порт 22) версии не ниже 2.0. Подключение к виртуальным серверам АПДЦ должно происходить с помощью предварительно сгенерированных сертификатов, доступ по паролю запрещен. Для дополнительной безопасности доступ по этому протоколу может быть разрешен только с перечня определенных IP адресов. Такое подключение должно быть доступно только пользователями из специальной группы

администраторов АПДЦ, к которым предъявляются требования следить за сохранностью сертификатов и пароли к ним.

— Информация, передаваемая по портам TCP 1433 и 1434 для подключения к Microsoft MS SQLServer, должна передаваться только по защищенным каналам связи. Для дополнительного усиления безопасности подключения по порту TCP 1434, используемого для административного соединения с базой данных, могут быть разрешены подключения только с заранее определенного списка IP адресов, используемых администраторами АПДЦ. Для усиления безопасности можно рассмотреть вариант с уменьшением количества открытых портов, когда подключения по порту TCP 1434 для административного управления MS SQLServer должны быть разрешены только с самого виртуального сервера (localhost). В этом случае при возникновении подобных задач сотрудники из группы администраторов АПДЦ сначала подключается по RDP к виртуальному серверу, а с него уже решает задачи по управлению MS SQLServer.

— Доступ по протоколам, в которых информация передается в открытом виде (например, FTP, порт 20, 21) по открытым каналам связи запрещена. При необходимости подключения по такому протоколу должен быть предварительно организован шифрованный канал связи средствами VPN.

Использование конфиденциальной информации определенного уровня может так же потребовать шифров ГОСТа для каналов обмена информацией с соответствующими государственными учреждениями.

В случае использования информации составляющей Государственную тайну, необходима организация выделенных каналов связи для обмена информацией между компонентами АПДЦ без контакта с сетью Интернет, либо со специальными авторизованными гейтами во внешние сети.

5.8. Разработка подхода на базе общедоступного инструментария, реализующего защиту от копирования информации почвенно-географической распределенной системы БД

Под защитой от копирования понимается не санкционированный доступ к информации, превышающий требования политик разграничения доступ выбранных групп пользователей. Такой доступ может быть получен в результате ошибок используемого программного обеспечения, подключения к потокам информации между компонентами АПДЦ или между АПДЦ и авторизованным пользователем во время легальных операций (прослушивание каналов связи, снифинг).

Все программные компоненты АПДЦ тестируются перед использованием в рабочем окружении с помощью шаблонов возможного применения, разработанных Заказчиком.

После ввода в эксплуатацию текущий набор программных компонентов проходит аудит сканерами безопасности (Vulnerability Scanner). Система должна периодически проходить подобный аудит по мере обновления шаблонов сканеров безопасности и при обнаружении новых типов методик вторжений (эксплойтов).

Кроме общеизвестных шаблонов нарушений информационной безопасности, разработчики программного обеспечения АПДЦ должны постоянно проводить аудит кода на предмет возможных сценариев не санкционированного доступа к данным.

Для предотвращения доступа к данным во время передачи информации между компонентами АПДЦ и легальными пользователями, все каналы связи для передачи конфиденциальной информации должны быть зашифрованы.

Физический доступ посторонних лиц к элементам АПДЦ, расположенном в датацентре ЦХАБД невозможен, т.к. все оборудование находится в охраняемом помещении, доступ к которому есть только у сотрудников ЦХАБД.

5.9. Разработка подхода к регулярному тестированию на проникновение всех сервисов АПДЦ и выработка предложений по назначению ответственных

Разработчиками программного обеспечения подготавливается инструкция по использованию АПДЦ и возможным видам ошибок в процессе работы.

Администраторы АПДЦ следят за событиями на виртуальных серверах с помощью логов операционной системы, используемого программного обеспечения, а также логов действий пользователей. В случае детектирования подозрительных действий сотрудники из группы администраторов АПДЦ разбираются с инцидентов и, при необходимости, получают консультацию разработчиков по реакции программного обеспечения на обнаруженные действия.

В программном обеспечении могут быть настроены оповещения (алерты) случаев запросов на превышения необходимых привилегий для более пристального анализа действий пользователей.

Администраторами АПДЦ планируются действия по периодическому аудиту сканерами безопасности (Vulnerability Scanner) по мере обнаружения новых типов методик вторжений (эксплойтов) и обновления шаблонов сканеров безопасности.

В случае обнаружения фактов проникновения в систему АПДЦ или получения несанкционированного доступа к данным, должен быть выработан регламент действий в подобной ситуации. Возможны различные сценарии в зависимости от серьёзности последствий такого проникновения:

— Полное отключения внешнего доступа к АПДЦ.

— Ограничение внешнего доступа для простых пользователей на время восстановительных работ сотрудниками из группы администраторов АПДЦ.

— Сохранение внешнего доступа, но изменение функциональности предоставляемых сервисов на время восстановительных работ отдельных компонент АПДЦ.
— Ограничение доступа только для ограниченной группы пользователей (например, из-за компрометации учетных данных).

В случае обнаружение проблем с функционированием виртуальных серверов, администраторы АПДЦ обращаются к сотрудникам ЦХАБД для совместного определения причин нестандартного поведения программно-аппаратного комплекса и решения возникших проблем.

5.10. Разработка подхода к обслуживанию и тестированию систем АПДЦ

Поддержание работоспособности АПДЦ заключается в повседневных задачах администрирования работы комплекса, описанных в пункте 2.11. Стабильность работы АПДЦ должно осуществляться на всех уровнях функционирования комплекса:

— Облачной инфраструктура ЦХАБД по поддержки работоспособности виртуальных серверов.
— Поддержка работы операционной системы Microsoft Windows Server 2016 и основных программных компонент – базы данных MSSQL и web-server IIS – установка патчей безопасностей и обновление антивирусных сигнатур.
— Поддержка работоспособности уникальных программных компонент АПДЦ

Создание новых программных модулей должно следовать требованиям безопасности на используемом языке программирования и подвергаться аудиту.

Список рабочих программ на комплексе АПДЦ определяется решаемыми задачами. Не допускается запуск сторонних программ или изменении конфигурации рабочего окружения программ комплекса на виртуальных серверах АПДЦ без их предварительного тестирования на

испытательном полигоне с программной и аппаратной конфигурацией аналогичной рабочей конфигурации АПДЦ. При возникновении необходимости изменении (обновление) версии программного обеспечения, оно должно пройти процесс тестирования на типичные рабочие шаблоны, список которых определяется Заказчиком. По результатам тестирования составляется акт, в котором фиксируются версии тестируемых программ и прописывается процедура замены старых версий на новые в рабочем окружении. Дата и время предполагаемого изменения конфигурации определяется заранее и доносится до всех пользователей АПДЦ и работников ЦХАБД.

На финальной стадии обновления версии программного обеспечения администраторами АПДЦ проводится аудит безопасности новой версии.

ЛИТЕРАТУРА

1. Белобров В. П., Мазиков В. М. Использование аэрофотоснимков при определении контрастности почвенного покрова по засолению // Известия АН СССР, Серия географическая М.: Наука, 1979. №2. С. 121-129.
2. Белоусова Н. И., Мешалкина Ю. Л. Методические аспекты создания почвенно-атрибутивной базы данных //Бюллетень Почвенного института им. ВВ Докучаева. – 2009. – № 64. – С. 23-33.
3. ГОСТ 27593-88 Почвы. Термины и определения. М.: Стандартинформ, 2006. - 11 с.
4. ДеМерс М. Н. Географические информационные системы. М.: «Дата+» 1999 – 490 с.
5. Егоров В. В. и др. Классификация и диагностика почв СССР. – Колос, 1977.
6. Единый государственный реестр почвенных ресурсов России. Версия 1.0. Коллективная монография. – М.: Почвенный ин-т им. В.В. Докучаева Россельхозакадемии, 2014. – 768 с. ISBN 978-5-8125-1960-5
7. Капралов Е. Г., Кошкарев А.В., Тикунов В. С. Геоинформатика: в 2 кн. Изд-во «Академия», 2008 – 384 с.
8. Колесникова В.М., Алябина И.О., Воробьёва Л.А., Молчанов Э.М., Шоба С.А., Рожков В.А. Почвенная атрибутивная база данных России // Почвоведение. –2010. – №8. – С 899-908.
9. Крыщенко В. С., Голозубов О.М., Колесов В.В. // Математическое моделирование в почвоведении, ЮФУ 2012 г.
10. Мешалкина Ю.Л., Самсонова В.П. Презентация «Модель пространственной вариабельности почвенных свойств».
11. Многоязычная база метаданных, объектная модель и программный комплекс для стандартизации и обмена почвенной информацией Soil_ML_MultyL. Голозубов О.М., Исмаилов А.И., Колесникова В.М., Морозов И.В., Розлога Ю.Г., Филипчук В.Ф., Чернова О.В. 2019.

12. Общесоюзная инструкция по почвенным обследованиям и составлению крупномасштабных почвенных карт землепользований (ред. Ищенко Т.А.) Издательство: Колос, 1973 г. - 95 с.
13. Приказ Министерства сельского хозяйства Российской Федерации от 4 мая 2010 г. N 150 «Об утверждении Порядка государственного учета показателей состояния плодородия земель сельскохозяйственного назначения».
14. Рожков В.А., Рожкова С.В. Почвенная информатика / В.А. Рожков, С.В. Рожкова. – М.: Изд-во Мос. ун-та, 1993.190 с.
15. Руководство пользователя ArcGIS Desktop https://desktop.arcgis.com/ru/arcmap/10.3/main/get-started/arcgis-tutorials.htm
16. Рухович Д.И., Вагнер В.Б., Вильчевская Е.В., Калинина Н.В., Королева В.П. Проблемы использования тематических карт на территорию СССР при создании ГИС «Почвы России» // Почвоведение.– 2011. – №9. –С. 1043-1055.
17. Рухович Д.И., Королева П.В., Калинина Н.В., Вильчевская Е.В., Симакова М.С., Долинина Е.А., Рухович С.А. Государственная почвенная карта – версия ARCINFO // Почвоведение. – 2013. –№3. – С. 251-267.
18. Самсонова В.П. Пространственная изменчивость почвенных свойств: На примере дерново-подзолистых почв / В.П. Самсонова. – М.: Издательство ЛКИ, 2008. — 160 с.
19. Шоба С.А., Колесникова В.М., Голозубов О.М. //Область применения и основные приемы работы с программой локального ввода данных V7_7_TM (МЕТОДИЧЕСКОЕ ПОСОБИЕ). М. 2018 г. 92 с.