

**Анализ возможностей облачных технологий для
распределённого хранения и обработки данных
дистанционных наблюдений с целью мониторинга состояния
окружающей среды**

Сорокин А.А. (1), Прошин А.А. (2), Королев С.П. (1), Бурцев М.А. (2), Мальковский С.И. (1), Лупян Е.А. (2), Смагин С.И. (1)

(1) Вычислительный центр ДВО РАН, Хабаровск, Россия
(2) Институт космических исследований РАН, Москва, Россия

Цели проекта

- ▶ Исследования и разработка современных методов и алгоритмов хранения, обработки и анализа сверхбольших объемов данных дистанционного зондирования с использованием возможностей облачных вычислительных сред;



Направления исследований

- ▶ Создание программных инструментов, позволяющих формировать и поддерживать распределенные архивы данных дистанционных наблюдений и результатов их обработки в таких средах, а также предоставлять пользователям (в том числе различных специализированных информационных систем) сервисы для работы с ними;
- ▶ Создание прототипа системы распределённого хранения данных ДЗЗ с применением современных высокопроизводительных решений.



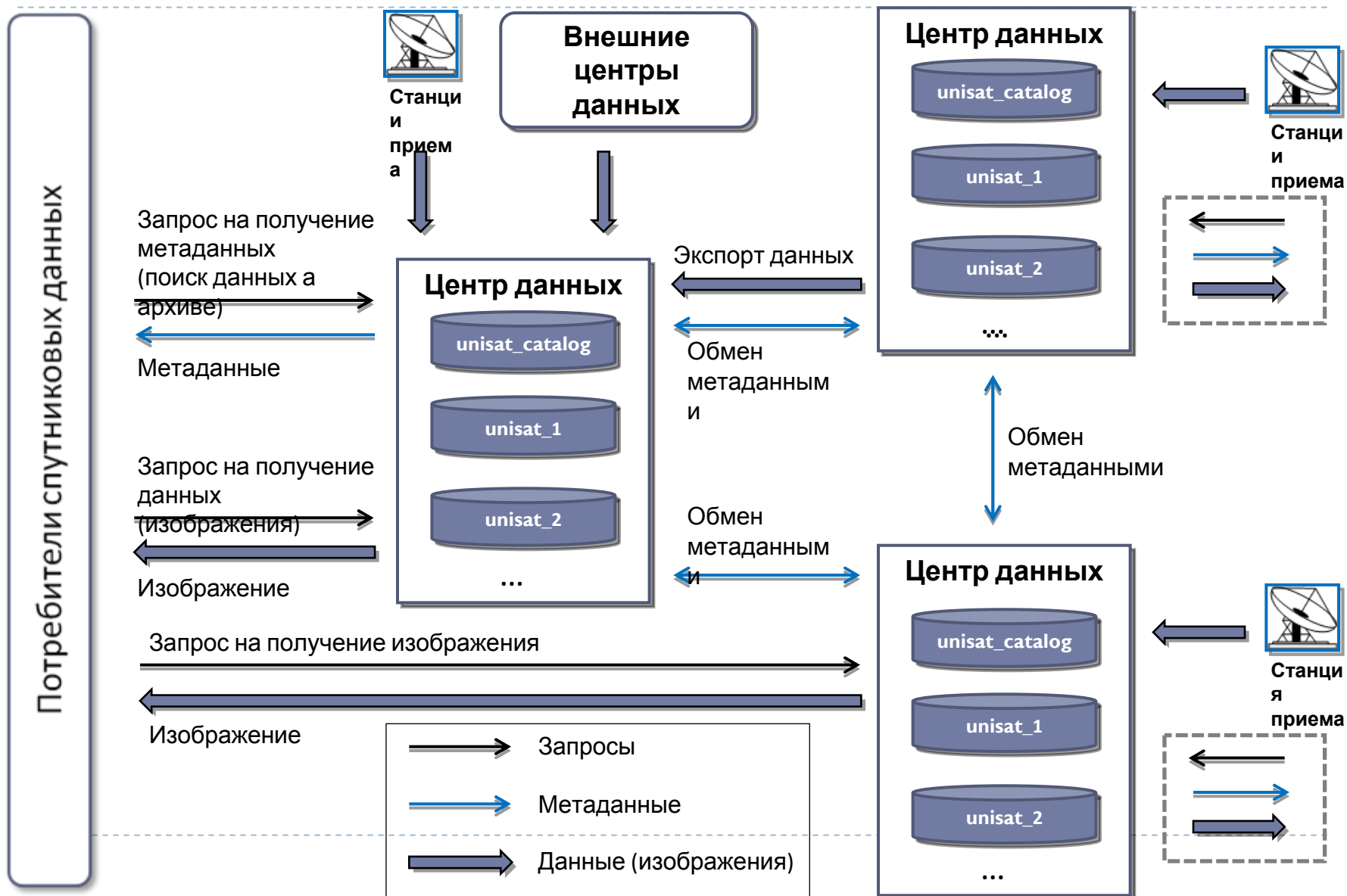
Технология UNISAT - описание

Технология UNISAT обеспечивает:

- ▶ Единообразное распределённое хранение разнотипных данных и каталогов развёрнутых описаний к ним;
- ▶ Предоставление пользователю достаточно сложных инструментов для работы с данными и их анализа, которые до недавнего времени были доступны только в рамках специализированного программного обеспечения;
- ▶ Предоставление доступа к «виртуальным» продуктам, т.е. продуктам, которые динамически формируются на момент запроса их пользователями на основе данных, имеющихся в архивах. Такой подход позволяет обеспечить доступ пользователей к интересующим их новым информационным продуктам, не прибегая при этом к переобработке исходных данных и не увеличивая размер имеющихся архивов.



Технология UNISAT - схема



Рассматривавшиеся технологии

- ▶ Amazon S3;
- ▶ Google Storage;
- ▶ Открытые облачные решения (Openstack SWIFT);
- ▶ Кластерные решения (LustreFS и др.);



Openstack SWIFT – плюсы и минусы

- ▶ Распределённость;
- ▶ Масштабируемость;
- ▶ Устойчивость к отказам аппаратуры;
- ▶ Возможность доступа к данным через RestFUL-сервисы;
- ▶ POSIX-несовместимо.



Проблемы POSIX-несовместимых хранилищ

- ▶ Возможность доступа только к файлу целиком;
- ▶ Проблематичность полноценной реализации механизма «виртуальных» продуктов и честной выборки по зоне интереса.



Интеграция технологий Unisat и Openstack SWIFT

1. Модернизация архивов данных – перевод их в формат, не подразумевающий прямого доступа к отдельной части файла (тайловый архив).
 2. Использование надстроек для Openstack SWIFT, обеспечивающих работу с ним, как с файловым хранилищем прямого доступа. Решение необходимо для данных, требующих возможности доступа к отдельным участкам файла данных, в том числе для построения «виртуальных продуктов» и отображения произвольной выбранной территории в заданном масштабе.
 3. Использование кэширующих серверов для предоставления данных.
-



S3QL - Обеспечение прямого файлового доступа к Openstack

S3QL (<https://bitbucket.org/nikratio/s3ql/>) - файловая система, использующая FUSE, для работы с удалёнными облачными хранилищами, в т.ч. Amazon S3, Google Storage и Openstack SWIFT.

Система позволяет использовать ресурсы таких хранилищ в виде стандартизированной UNIX файловой системы, поддерживая стандартные функции чтения и записи, а также сжатия данных, шифрования, дедупликации и многие другие.



S3QL – плюсы и минусы возможной реализации

- ▶ Возможность прямого доступа к хранимым файлам;
- ▶ Прозрачность с т.з. системы доступа – хранилище работает, как локальная файловая система;
- ▶ Необходимость работы через выделенный шлюз с метаданными;
- ▶ Необходимость обеспечения устойчивости шлюза с метаданными;



Условия и результаты испытаний

- ▶ Хранилище на базе Openstack SWIFT – ВЦ ДВО РАН, Хабаровск;
- ▶ Узел доступа на базе S3QL – ИКИ РАН, Москва;
- ▶ Канал связи – L2, 200 Мбит;
- ▶ Тестовые наборы данных – группы файлов с характерными для архивируемых сцен размерами, т.е. мегабайты, десятки мегабайт, сотни мегабайт, гигабайты;
- ▶ Устойчивые скорости записи данных - ~2,5 МБ/с;
- ▶ Устойчивые скорости чтения данных - ~7 МБ/с;



Предварительные выводы

- ▶ Система Openstack SWIFT оптимальна для систем доступа к файловым архивам (как, например, архив данных Sentinel-2 на Amazon S3, <http://sentinel-pds.s3-website.eu-central-1.amazonaws.com/>) или для систем, представляющих данные в виде фиксированных тайлов;
- ▶ Для полноценного использования всех возможностей технологии UNISAT и Openstack SWIFT требуется доработка промежуточного ПО сопряжения.



Спасибо за внимание!

Работы проводятся при поддержке проектов Российского фонда фундаментальных исследований (№15-29-07953) и Программы фундаментальных исследований ДВО РАН «Дальний Восток» (№15-1-4-072).