

# Технологии параллельной обработки данных в задачах анализа больших объемов СПУТНИКОВЫХ СНИМКОВ

Кихтенко В.А., Чубаров Д.Л.

Институт вычислительных технологий СО РАН

# Системы доступа к данным ДЗЗ

## I. Библиотеки — *Reverb, Mercury, MODIS Subsets*

- поиск по метаданным, загрузка снимков, обработка локально

## II. Геопорталы — *Earthdata, Giovanni*

- доступ к данным онлайн, средства обработки определены создателем геопортала

## III. Роботы — *Google Earth Engine*

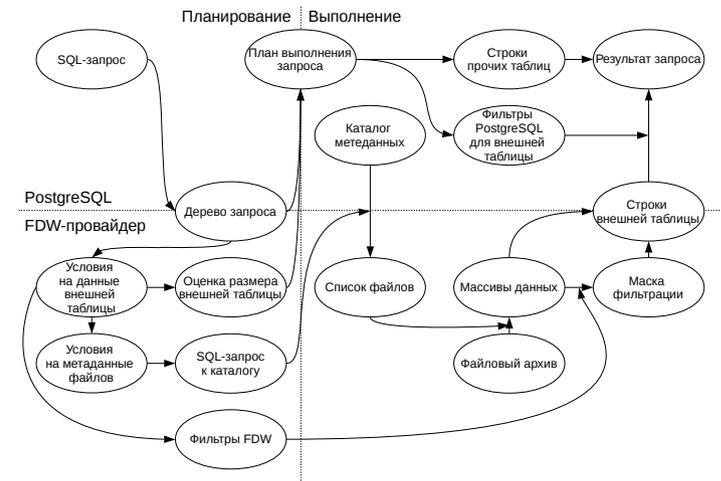
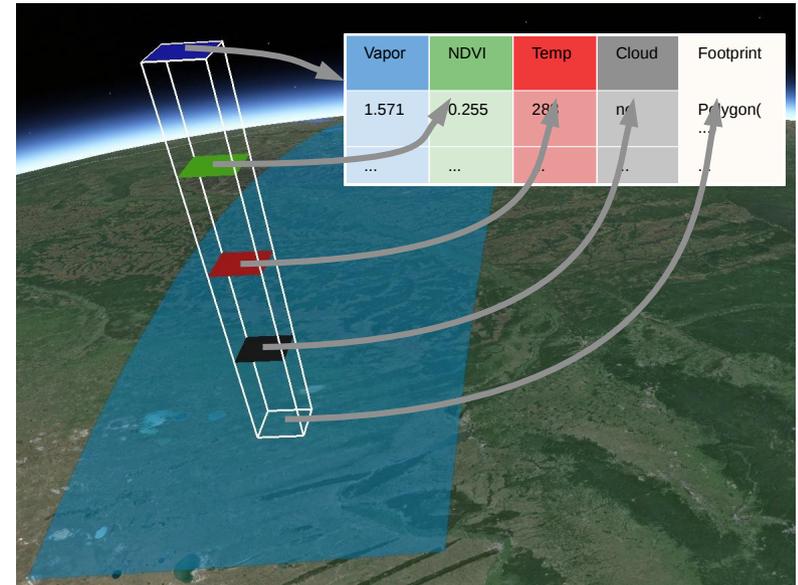
- загружается код программы обработки, робот возвращает результат

# Мотивация

- Объемы данных не позволяют работать с ними локально
- Воспроизводимость и публикация результатов исследований
- Система типа III может использоваться для создания систем типа II.

# Система hVault

- Отображение спутниковых снимков в виртуальные таблицы СУБД
- Обработка данных задается SQL-запросом



# Уроки hVault

## ✓ Положительные:

- Абстракция от источника, формата и места хранения данных
- Хранение данных в исходном формате
- Стандартные язык запросов и протокол доступа

## ✗ Отрицательные:

- Производительность
- Модель пространственных данных (таблицы)

# Системы параллельной обработки данных

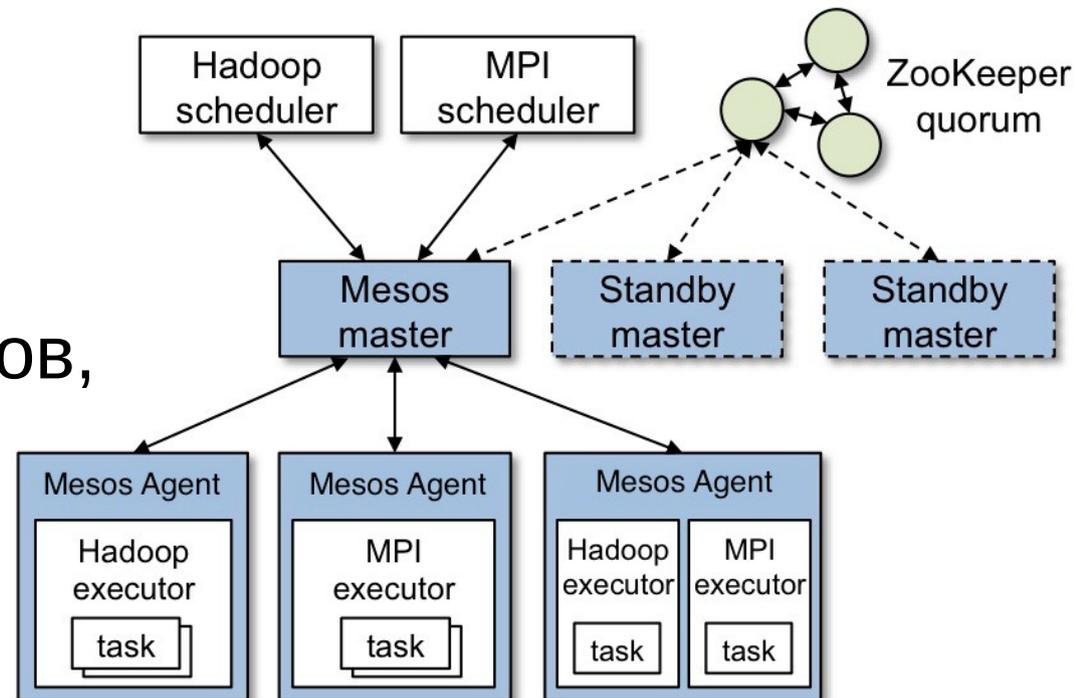
- Обеспечивают автоматическое распараллеливание обработки
- Требуют описания алгоритма в специальном виде:
  - Map-reduce: Apache Hadoop, Spark
  - Многомерные массивы: SciDB
- Модель данных не учитывает специфику пространственных данных



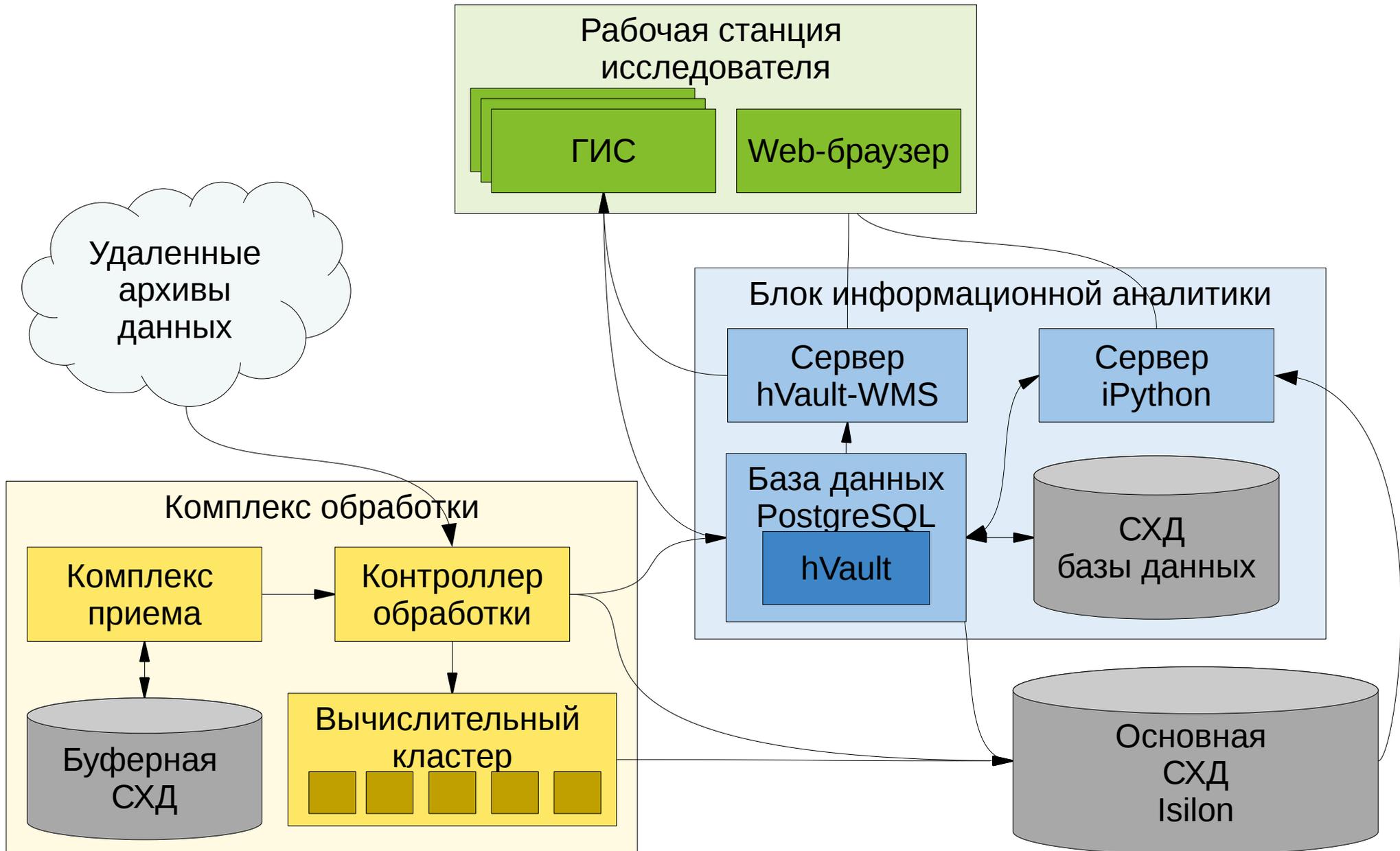
# Разделение ресурсов

- Архив ДДЗ обычно поддерживает множество различных сервисов и ресурсов.
- Сервисы могут использовать аппаратные ресурсы совместно:

- Виртуальные машины и контейнеры
- Планировщики ресурсов, например Mesos
- Все вычисления и доступ к данным через систему типа III



# Архитектура архива ДДЗ



# Выводы

- Система hVault демонстрирует перспективность предложенного подхода
- Эффективная реализация возможна на основе систем параллельной обработки данных общего назначения
- Требуется разработка API для удаленного доступа и программной обработки пространственных данных

Спасибо за внимание!

