ПРИНЦИПЫ РЕАЛИЗАЦИИ ПРОГРАММНОГО МОДУЛЯ ДЛЯ СОЗДАНИЯ СПЕЦИАЛИЗИРОВАННЫХ ВЕБ-СЕРВИСОВ ОБРАБОТКИ И АНАЛИЗА ДАННЫХ

Афанасьев А.П., Волков С.Ю., Гринберг Я.Р., Джорджевич Т.С., Корх А.В., Смирнов С.А., Рубцов А.О., Сухорослов О.В.
Институт проблем передачи информации РАН

Целью проводимых исследований является разработка экспериментального образца программного комплекса (ЭО ПК) для создания масштабируемых облачных сервисов сбора и обработки данных (ССОД) на базе платформы Apache Hadoop. Актуальность данного направления исследований обусловлена растущей потребностью в организации сбора и обработки больших объемов данных с использованием технологий распределенных вычислений, как в науке, так и бизнесе.

Центральным компонентом разработанного ЭО ПК является модуль создания ССОД (МСС), предназначенный для поддержки создания специализированных веб-сервисов для сбора и обработки данных. Данный модуль использует в своей работе функциональность всех других модулей ЭО ПК. МСС представляет собой набор готовых компонент для реализации ССОД:
- адаптер для реализации веб-сервисов обработки данных (АВСОД) на базе модуля обработки данных;
- адаптер для реализации веб-сервисов выборки и анализа данных (АВСВАД) на базе модуля выборки и анализа данных.

В качестве базовой платформы для организации удаленного доступа к сервисам используются протоколы и технологии Web, основанные на архитектурном стиле REST. Благодаря унифицированному интерфейсу доступа к ресурсам, использованию открытых стандартов (HTTP, URI) и наличию множества проверенных временем реализаций (Web-серверы, библиотеки для работы с HTTP для всех современных языков программирования), REST обеспечивает максимальную свободу для независимой разработки Web-сервисов и соответствующих клиентских приложений. Немаловажным обстоятельством является простота разработки REST-сервисов в сравнении с другими рассмотренными технологиями. Все это позволяет максимально расширить как круг потенциальных разработчиков, так и пользователей ССОД, обеспечив при этом совместимость различных реализаций и широкое повторное использование сервисов.

При разработке адаптеров АВСОД и АВСВАД использована программная платформа MathCloud. Данная платформа включает среду выполнения проблемно-ориентированных вычислительных веб-сервисов, а также набор адаптеров для быстрого преобразования в сервисы существующих приложений и интеграции с вычислительными инфраструктурами. В рамках платформы MathCloud разработан унифицированный REST-интерфейс, который реализуют все сервисы, создаваемые с помощью платформы. Контейнер Everest представляет собой универсальную среду выполнения сервисов, реализующую унифицированный REST-интерфейс сервисов MathCloud. Универсальность контейнера заключается в том, что он позволяет легко, во многих случаях без написания программного кода, преобразовать в сервис широкий спектр существующих приложений. Кроме того, реализованный в контейнере механизм адаптеров позволяет подключать произвольные реализации обработчиков запросов. В частности, путем преобразования запроса к сервису в вычислительное задание, на уровне адаптера может быть реализован доступ к вычислительной инфраструктуре.

В рамках предлагаемого подхода, основанного на использовании платформы MathCloud, компоненты АВСОД и АВСВАД реализованы в виде адаптеров к контейнеру Everest. Адаптер АВСОД позволяет преобразовывать в сервис существующее MapReduce-приложение. А адаптер АВСВАД позволяет создавать сервисы по запуску произвольных Pig-программ на данных пользователя ССОД.

На базе компонента АВСОД реализован сервис общего назначения, предназначенный для запуска произвольных MapReduce-программ на Hadoop-кластере, доступном через ЭО ПК. При вызове сервису передаются следующие входные параметры: jar-файл с кодом MapReduce-программы на языке программирования Java; имя класса программы; дополнительные аргументы запуска программы и параметры MapReduce-задания. Сервис производит запуск указанной программы на кластере с помощью модуля обработки данных и предоставляет пользователю информацию о статусе выполнения задания. По окончания выполнения задания сервис выводит пользователю статистику выполнения задания.

Работа выполнена в рамках ФЦП «Исследования и разработки по приоритетным направлениям развития научно-технологического комплекса России на 2007-2013 годы» при финансовой поддержке Минобрнауки, государственный контракт № 14.514.11.4021.