

Принципы построения масштабируемых систем управления и контроля данных на основе программных компонент с открытым кодом

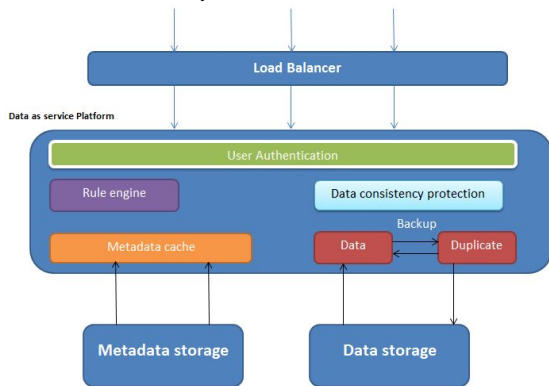
Тимофеев Б.М.

Санкт-Петербургский государственный университет

Цель: данная работа посвящена разработке системы хранения и управления данными на основе компонент с открытым исходным кодом. Среди требований данной системы следует также отдельно выделить отказоустойчивость.

Актуальность: В современном мире, в связи со стремительным ростом объемов хранимой информации, становятся актуальными новые подходы к ее хранению, при которых время поиска существенно сокращается по сравнению с классическими методами. Системы для управления данными становятся все более и более востребованными, также многим их пользователям необходима возможность использования метаданных. В контексте моей работы одним из таких подходов является раздельное хранение данных и метаданных.

Общая архитектура системы: Архитектура системы, которая разрабатывается мной, может быть представлена с помощью следующей схемы:



Таким образом, система включает в себя следующие компоненты:

- Платформа для хранения данных
- Платформа для хранения метаданных
- Модуль авторизации пользователей
- Балансировщик нагрузки на систему

- Система резервного копирования данных
- Модуль управления политиками и правилами (с помощью него пользователи могут создавать правила, по которым данные будут храниться и использоваться)
- Распределенный кэш метаданных (для ускорения доступа к метаданным)

Реализованная функциональность: В настоящее время мной реализованы следующие компоненты:

1. **Система для хранения данных.** В качестве компонента, находящегося в открытом доступе была использована система хранения данных Ceph, то есть требование масштабируемости соблюдено (в силу особенностей Ceph).
2. **Система для хранения метаданных.** В настоящее время реализована при помощи реляционной базы данных HSQLDB, в силу удобства ее эксплуатации как встроенной БД с языком Java. В дальнейшем планируется разработка решения с использованием NoSQL базы данных, например, MongoDB, в силу того, что метаданные хранятся в формате "ключ"-значение
3. **Система авторизации пользователей.** Доступ к данным и метаданным осуществляется после проверки прав пользователей на требуемое действие, ACL (Access Control List, список контроля доступа) в настоящее время хранится в системе для хранения метаданных, для добавления новых пользователей планируется использование авторизации Google или Facebook (то есть использование существующего аккаунта).



Все вышеописанные компоненты взаимодействуют между собой при помощи веб-сервера, разработанного мной при помощи языка Java (в силу его большой распространенности) с использованием технологий Spring MVC и Hibernate.

Выводы: Построение конкретной системы управления и контроля данными во многом зависит от предъявляемых к ней требований. В контексте данной работы были рассмотрены компоненты, находящиеся в открытом доступе, с целью удешевления разработки, также могут быть рассмотрены коммерческие решения (таким образом, надежность системы повысится). Как показало исследование, количество компонентов со свободной лицензией достаточно обширно, соответственно необходимо проводить тщательный анализ с целью выбрать наиболее достойного "кандидата". Список компонентов системы, указанный выше, может также меняться в зависимости от требований, главное, чтобы было сохранено главное свойство подобного рода систем — универсальность, т.е. независимость от предметной области.