

УДК 004.02

## ГОРИЗОНТАЛЬНОЕ МАСШТАБИРОВАНИЕ ИСТОЧНИКОВ ДАННЫХ ПРИ АГРЕГАЦИИ ДАННЫХ В СИСТЕМЕ УМНОГО ГОРОДА

Волков А.А. (Университет ИТМО)

Научный руководитель – к.т.н, Тесля Н.Н.  
(Университет ИТМО)

В работе рассмотрен вопрос построения архитектуры системы агрегации данных для умного города решающей задачу простого горизонтального масштабирования источников данных. Приведены преимущества и недостатки централизованного подхода с единой базой данных и распределенного сервисного подхода с несколькими базами данных под каждый источник данных.

**Введение.** Любая архитектура информационной системы не может покрыть все требования к разработке и эксплуатации системы на протяжении всего жизненного цикла системы, так как требования меняются с течением времени. Из этого можно утверждать, что архитектура не может быть статичной, это постоянно изменяемая динамическая сущность. В данной работе будет описана задача выбора подходящей архитектуры для простого добавления новых источников данных в систему агрегации данных умного города.

**Основная часть.** Качество работы системы умного города напрямую коррелирует с количеством входных данных. Для умного города при большом количестве открытых источников данных возникает вопрос простоты добавления еще одного источника данных в систему. Эту задачу можно обозначить как горизонтальное масштабирование источников данных. На текущий момент можно выделить два способа хранения информации при её агрегации: централизованный и распределенный.

Централизованный способ является самым простым и представляет собой сохранение всех данных в одну общую базу данных. Обработка данных проводится также в рамках одной программы. Это самый простой способ реализации и подходит для небольших проектов. При добавлении новых источников сложность системы начнет расти нелинейно. Это приводит к повышению расходов на сопровождение системы, связанных с разнообразием вариантов описания данных и их обработки. Все эти варианты требуются описывать в контексте одной программы и одной базы данных. Во-вторых, этот способ требует универсального подхода для всех источников данных, что в некоторых случаях является неоптимальным решением.

Распределенный подход подразумевает разделение программ хранения и обработки данных на независимые части. При таком подходе упрощается параллельная разработка и подключение новых источников данных в систему. Однако при этом также растет сложность организации взаимодействия между независимыми частями, особенно в сравнении с централизованным подходом.

В работе рассмотрен один из вариантов реализации распределенного подхода. Для простого масштабирования каждый источник данных выделен в отдельный модуль вместе с обработчиком. В центре системы расположен модуль, который направлен на взаимодействие с пользователем. Это модуль можно обозначить ядром системы. Возможно подключать дополнительные модули с частными данными, которые можно расширить, используя открытые данные.

**Выводы.** В качестве заключения можно сказать, что распределенный подход в агрегации, хранении и обработке данных из гетерогенных источников наиболее предпочтителен для системы умного города. Предложенный подход был опробован в задаче визуализации на карте открытых данных по аренде велосипедов в Берлине. Актуальные данные представляли собой список геоточек, которые можно получить с помощью веб-запроса. После чего этот список визуализировался на карте. Данная задача была разбита на разработку трех сервисов:

пользовательский модуль, модуль получения геоточек и модуль визуализации. Это позволило четко разделить данные и упростить добавление новых источников. При дальнейшем развитии системы, если появятся новые источники геоданных, то это потребует только настройки взаимодействия с пользовательским модулем.