

АНАЛИЗ ПОВЕДЕНИЯ НА ОСНОВЕ ЛОГ-ФАЙЛОВ С ИСПОЛЬЗОВАНИЕМ СРЕДСТВ ДЛЯ РАБОТЫ С BIG DATA

В работе рассматривается задача определения, на основе анализа лог-файла сетевой компьютерной игры, «потенциальных клиентов» среди множества игроков. Под потенциальными клиентами будем понимать тех, которые, при воздействии на них рекламы, вкладывали бы в игру денежные средства. Рекламу, конечно, можно отправлять всем игрокам, но это не эффективно и затратно.

Для решения подобных задач традиционно применяется когортный анализ [1]. Особенность применения когортного анализа в контексте данной задачи заключается в выделении групп пользователей, игровое поведение которых схоже с поведением игроков, которые вложили в игру денежные средства. В дальнейшем на них должно быть направлено целевое воздействие средствами рекламы или стимулированием их игрового интереса.

В компьютерных играх для социальных сетей количество игроков и данных растёт с очень большой скоростью. Обработку лог-файлов, размер которых за месяц активной игры достигает сотен гигабайт, эффективно осуществлять с использованием фреймворка Apache Spark. Для эффективного проведения анализа необходимо использовать значительные вычислительные мощности, что достигается горизонтальным масштабированием системы. Apache Spark обеспечи-

вает также большую скорость обработки данных и имеет простой функционал. Разработку возможно проводить на таких языках как Python, Java, Scala, R. Хранить данные предполагается в такой системе хранения данных как HIVE.

В заключении нужно отметить, что эффективная реализация анализатора лог-файлов может принести немалую прибыль игровым компаниям. Подобным образом, при незначительных изменениях, можно детектировать проблемы в работе систем, программ и сервисов, а также применять анализатор в других приложениях, связанных с Big Data.

Литература

1 Маккинли, У. Python и анализ данных / У. Маккинли. – М. : ДМК Пресс, 2015. – 482 с.