нительно высокими характеристиками, отсутствием необходимости настройки, малым числом внешних элементов, лёгкостью управления и осуществления модуляции различных видов. Однако малое количество типов микросхем прямого цифрового синтеза с тактовыми частотами выше 300 МГц не позволяет им конкурировать с синтезаторами на основе ФАПЧ на частотах выше 200 МГц. Значительно уступают они пока синтезаторам на основе ФАПЧ и в спектральной чистоте выходного сигнала

КИСЕЛЁВ А.А., КАЛАШНИКОВ В. И., канд. техн. наук, проф.

ИССЛЕДОВАНИЕ ВНУТРЕННЕЙ СТРУКТУРЫ APACHE HADOOP. HDFS (HADOOP DISTRIBUTED FILE SYSTEM)

В условиях современного (информационного) общества своевременная и надёжная обработка информации выходит на первый план среди, без преувеличения, всех отраслей деятельности человека. Естественно, что с каждым днём объёмы обрабатываемых данных возрастают, а их важность – увеличивается. Это касается как, к примеру, важных производственных процессов (управление промышленным оборудованием и т.п.) или научно-исследовательской деятельности (моделирование сложных физических процессов, какие-либо громоздкие статистические расчёты, др.), так и в отрасли отдыха и развлечений (рендеринг сцен фильмов и локаций компьютерных/консольных игр и т.д.). Не стоит также забывать, что львиную долю информации современный человек получает из Интернета, а число его пользователей с каждым днём увеличивается, и, чтобы удовлетворять запросы столь огромного количества людей и сделать их работу более удобной, необходимо постоянно совершенствовать не только аппаратные, но и программные средства обработки данных. Естественно, что программное обеспечение ЦоД (Центров обработки данных, также распространено название «дата-центры»), выполняющих задачи, подобные вышеописанным, спроектировано и реализовано для выполнения именно параллельных и распределённых вычислений. Это позволяет:

- максимально возможно утилизировать все имеющиеся вычислительные мощности и объём памяти;
- максимально ускорить собственно обработку различных (особенно, крайне больших) объёмов данных;
- используя репликацию данных между узлами, достигнуть большой надёжности их хранения.

Как известно, информация в таких системах не может быть одновременно абсолютно своевременной, надёжной а также расположенной распределённо, но к этим показателям необходимо стремиться.

В качестве объекта изучения выбрана система Apache Hadoop по следующим причинам:

•система написана на языке Java - динамично развивающимся, широко распространённым, со множеством сопутствующих технологий;

•имеет открытый код, что облегчает её изучение;

•«вдохновлён» технологиями Google (компания, прославившаяся внедрением инновационных технологий программирования).

•используется такими компаниями, как: Yahoo! (более чем 100000 ЦП на более чем 25000 компьютерах. Наибольший кластер: 4000 узлов), Adobe (30 узлов HDFS), Amazon, AOL, др.

Основным компонентом, предназначенным для обмена данными, в Apache Hadoop является Hadoop Distributed File System (HDFS). Архитектура представлена на рис. 1. HDFS преследует ряд целей, а также обладает множеством особенностей:

• устойчивость к аппаратным сбоям;

• распределение доступа к данным;

• поддержка больших объёмов данных;

• «перенос вычислений «накладен» менее переноса данных»;

• высокая портируемость между программными и аппаратными платформами.

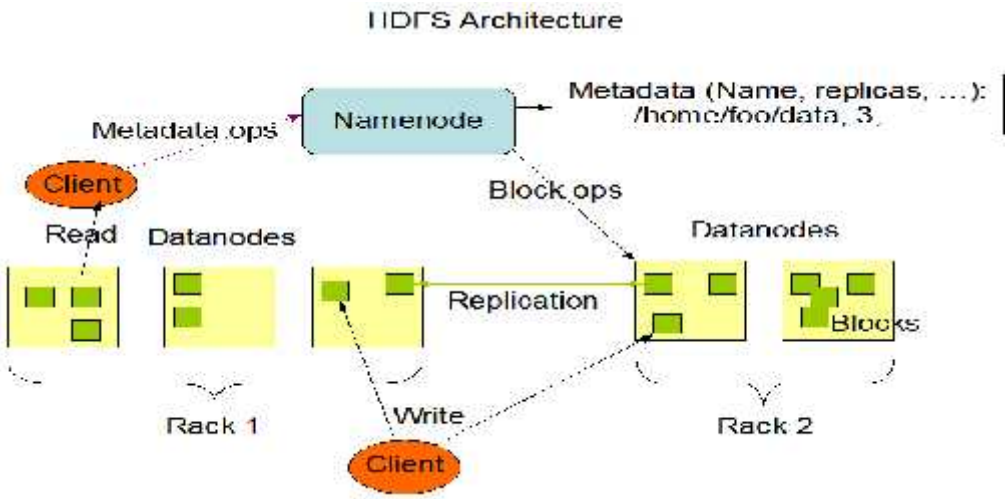

Рис. 1. Архитектура HDFS.

В докладе кратко рассмотрена структура и взаимосвязь Java-классов, реализующих HDFS, что может быть интересно Java-разработчикам, а также просто желающим «заглянуть» в реализацию HDFS.