

Обзор высокомасштабируемых СУБД

Добродей С.Г., Бородуля А.В.

Белорусский национальный технический университет

На протяжении многих лет реляционные базы данных были наиболее популярным решением для хранения и обработки данных. Однако, современный мир постоянного обмена данными принес с собой существенные изменения в том, как профессионалы должны управлять данными и использовать их для того, чтобы добиться максимальной отдачи от бизнеса. . Переход от мэйнфреймов к клиент-серверной архитектуре, а после к миру онлайн сервисов и социальных сетей привёл к появлению бесчисленного количества приложений и постоянного растущего объема данных, которые трудно хранить и обрабатывать в классических реляционных базах данных. На смену им пришли высокомасштабируемые СУБД. В данной работе рассматривается MySQL Cluster, как продукт основанный на принципах реляционных баз данных, а так БД Apache Cassandra как представитель активно развивающегося направления NoSQL баз данных.

MySQL Cluster состоит из набора компьютеров, каждый выполняет ряд процессов, включая сервер MySQL, узлы хранения для NDB, сервер управления и (возможно) специализированные программы доступа к данным. Все эти программы работают вместе, чтобы формировать MySQL Cluster. Когда данные сохраняются в кластере, то данные таблицы хранятся в оперативной памяти. Данные, сохраненные в узлах памяти MySQL Cluster могут быть зеркалированы, таким образом кластер может обрабатывать свои индивидуальных узлов памяти, теряя только тот ряд транзакций, которые прерваны из-за потери состояния транзакции. Так как в прикладных программах имеется возможность обрабатывать свои транзакции, эти прерванные транзакции не являются источником проблем.

Apache Cassandra – масштабируемая NoSQL БД с открытым исходным кодом, отлично подходящая для задач управления большим количеством неструктурированных, полуструктурированных и структурированных наборов данных в нескольких датацентрах или в облаке. Cassandra обеспечивает непрерывную доступность, линейную масштабируемость. Она может использоваться на нескольких серверах без единой точки отказа.

В данной статье были рассмотрены основные концепции и механизмы работы СУБД Cassandra и MySQL Cluster.