

СУБД Cassandra: обзор архитектуры

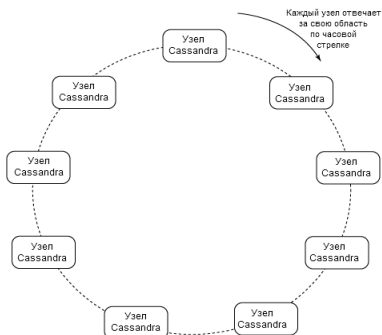
Добродей С.Г., Бородуля А.В.

Белорусский национальный технический университет

Apache Cassandra- это реализация семейства NoSQL Column, поддерживающая модель данных Big Table (большая таблица) с использованием некоторых архитектурных аспектов системы Amazon Dynamo. Некоторые из преимуществ Cassandra:

- высокая масштабируемость и надежность без элементов, отказ которых приводит к выходу из строя всей системы;
- реализация семейства NoSQL Column;
- SQL-подобный язык запросов (начиная с версии 0.8) и поддержка поиска посредством вторичных индексов;
- настраиваемая согласованность и поддержка репликации;
- гибкая схема.

Архитектура кластера Cassandra показана на рисунке Cassandra — распределенная система. Кластер состоит из нескольких узлов и распределяет данные между этими узлами (или секционировает их по терминологии баз данных).



Кластер Cassandra

Для распределения элементов данных по узлам Cassandra использует последовательное хэширование. Cassandra использует хэш-алгоритм для вычисления хэш-значений ключей каждого элемента данных, хранящегося в Cassandra. Диапазон хэш-значений или все возможные хэш-значения распределяется между узлами кластера Cassandra. Затем Cassandra назначает каждому элементу данных свой узел, и этот узел отвечает за хранение и управление этим элементом данных.