

Rozdział 1

Wstęp

1.1 Cassandra

Apache Cassandra to darmowa, rozproszona baza danych przeznaczona do zarządzania dużymi ilościami ustrukturyzowanych danych. Projekt powstał, aby umożliwić realizację funkcjonalności przeszukiwania skrzynki odbiorczej użytkownika na portalu Facebook¹. Funkcjonalność ta nie była możliwa do zrealizowania w oparciu o tradycyjny, relacyjny model bazy danych. W 2008 Facebook udostępnił kod źródłowy Cassandra, która następnie (w 2010 roku) została wcielona do projektów pod opieką fundacji Apache². [1]

Apache Cassandra jest częścią ruchu NoSQL³. Chociaż brakuje oficjalnej, ustanowionej odgórnie definicji pojęcia NoSQL możemy wyróżnić zestaw cech wspólnych baz danych, które powszechnie zaliczane są do tej grupy:

- Wewnętrzna reprezentacja danych w bazach NoSQL nie opiera się na tabelach i relacjach.
- Brak pełnej implementacji języka SQL. Chociaż niektóre bazy NoSQL wykorzystują język zapytań o składni podobnej do SQL, żadna z nich nie implementuje jej w pełni.
- Bazy NoSQL posiadają elastyczny model danych. Nie wykorzystują sztywnego, narzuconego z góry schematu.
- Model danych w bazach NoSQL jest zazwyczaj ukierunkowany na łatwe klastrowanie, chociaż istnieją wyjątki od tej reguły (na przykład Neo4j⁴).
- Bazy NoSQL są przystosowane do wymagań nowoczesnych portali internetowych, które muszą obsługiwać ogromne zbiory danych z szybkim czasem dostępu. [3]

¹Popularny portal społecznościowy dostępny pod adresem <http://fb.com>.

²Fundacja wspierająca powstawanie oprogramowania o otwartym źródle.

³NoSQL - *Not only SQL* (ang. nie tylko SQL).

⁴Neo4j - baza danych oparta wewnętrznie o grafy.

Konsekwencją ostatniej cechy jest odejście od idei właściwości ACID. Zestaw cech ACID⁵ to minimalny zbiór cech, które gwarantują realizację transakcyjności w bazie danych. [2] Klastrowanie sprawia, że nie da się zachować spójności danych na wszystkich węzłach, stąd wymagane są inne mechanizmy zachowania poprawności danych.

1.2 Historia Cassandra

Placeholder.

1.3 Architektura

Placeholder.

⁵ACID - skrót od *Atomicity, Consistency, Isolation, Durability* (ang. atomowość, jednorodność, izolacja, trwałość)

Bibliography

- [1] DataStax. *Introduction to Apache Cassandra*. Apr. 14, 2014. URL: <http://www.datastax.com/docs/0.8/introduction/index>.
- [2] Jim Gray, ed. *The Transaction Concept: Virtues and Limitations*. Proceedings of Seventh International Conference on Very Large Databases. 19333 Vallco Parkway, Cupertino CA 95014: Tandem Computers Incorporated, 1981.
- [3] Martin Fowler Pramod J. Sadalage. *NoSQL Distilled*. Pearson Education, 2013. ISBN: 978-0-321-82662-6.