

Rozproszone systemy operacyjne

Projekt – RSO Sharding

Autorzy:

* Tomasz Adamiec
* Piotr Cebulski
* Marek Kowalski
* Mateusz Rosiewicz
* Paweł Sokołowski
* Marcin Wnuk

Warszawa, 2013

* 1. Czym jest Sharding?

Jest to rodzaj podziału bazy danych polegający na rozbiciu dużych fragmentów na mniejsze, szybsze oraz łatwiejsze w zażądaniu części nazywane shardami.

* 1. Struktura

Sharding odbywa się wewnątrz klastra shardów, który składa się z następujących elementów:

* Shardów. Shard jest kontenerem przechowującym podzbiór danych. Aplikacje nie odwołują się bezpośrednio do shardów ale do instancji mongo.
* Serwerów konfiguracyjnych. Każdy serwer konfiguracyjny przechowuje metadane opisujące klaster.
* Instancji mongo. Instancje mongo kierują polecenia odczytu i zapisu z aplikacji do shardów.

Shardy składają się na jedną bazę danych. Mogą być przechowywane na różnych serwerach i nie zawierają części wspólnych.

* 1. Idea

Idea shardingu wynika z obserwacji pracy bazy danych. Przy liniowym wzroście rozmiaru bazy i ilości transakcji na jednostkę czasu, czas odpowiedzi rośnie wykładniczo.

Dodatkowo tworzenie i utrzymywanie dużej bazy danych w jednym miejscu może wzrastać nieliniowo ze względu na potrzebę zastosowania kosztownego sprzętu. Utrzymywanie mniejszych fragmentów na tańszych jednostkach jest nieproporcjonalnie tańsze.

Częstym przypadkiem jest podział wynikający z położenia geograficznego. Przykładowo klienci z konkretnego regionu przypisani są do odpowiedniego serwera (shardu).

* 1. Cel Projektu

Sharding w mongodb pozwala na automatyczny podział danych oraz obciążenia na serwery. Naszym celem będzie poprawa algorytmów za to odpowiedzialnych. Planujemy poprawić wydajność zapisu danych poprzez optymalny rozkład obciążenia na poszczególne serwery.