

2.1 Key/Value Stores Dynamo / Voldemort

23 November 2021 17:09

Dynamo

Amazon 2007

Za pristup podacima se samo koriste primarni ključevi

Decentralizovan i simetričan P2P

Skaliranje lako jer dodavanja i uklanjanja čvora lako

Dynamo kao baza za pravljenje Cassandra

Konzistentnost se namerno žrtvuje u korist dostupnosti (write operacija se uvek izvršava)

Pravi se kao prsten čvorova i svaki čvor ima random poziciju

Kad se doda ili ukloni čvor za to znaju samo dva suseda u čvoru

Virtuelni čvorovi - jedan fizički čvor dobija nekoliko pozicija u prstenu

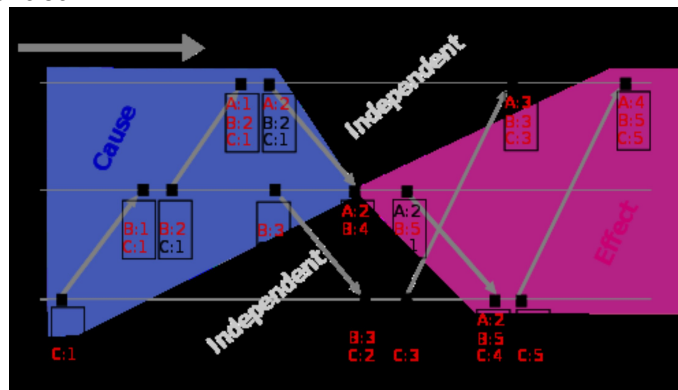
Replikacija

N kopija - min 3 čvora (kao master slave gde ako neko odkaze nešto ga zameni)

Scale out

Eventually consistent

Vector clock



Azuriranje kopija neće se desiti odmah već sa odlaganjem

Redosled operacija izvršiti na isti način za svaku repliku ali ne u isto vreme

Read / Write zahtev

Bira se čvor da bude kordinator operacije - najbliži ili najneopterećeni čvor

N - broj kopija podatka koji postoji u sistemu

W - kvorum - broj čvorova koji treba da potvrdi da je operacija uspešna da bi se priznala operacija, obično većina 50%+

Konzistentnost se podesava za read i write

Consistency level	Read / Write
ONE	1 replica
QUORUM	$N/2 + 1$
ALL	N

Write

Kordinator generiše verziju i vektor clock

Kordinator šalje zahtev u N čvorova koji sadrže repliku podatka, ukoliko W čvorova potvrdi operacija je uspešna

Read

Koordinator šalje zahtev u N čvorova koji sadrže repliku traženog podatka, ukoliko R čvorova odgovori operacija je uspešna

Ukoliko postoje različite verzije podatka, razrešava se konflikt, generiše se važeća verzija podatka i ona se upisuje nazad (read repair)

Hinted handoff

Upis

Ako cvor koji dobije podatke je nedostupan kordinator salje podatke drugom cvoru da ih cuva sve dok originalni cvor ne postane dostupan I onda se upisuju / pamte podaci

Citanje

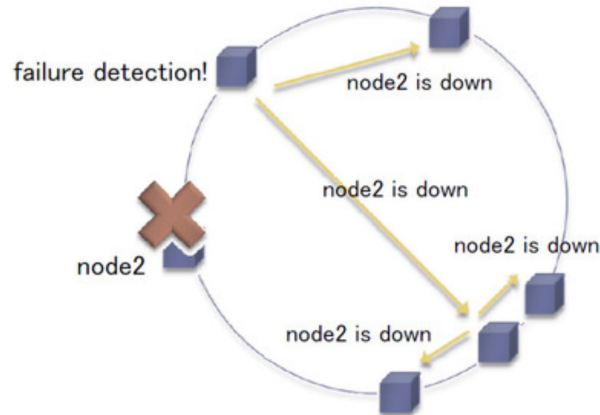
Ako nije dostupan cvor za citanje moze kvorum da izglasa da citanje nije uspelo

Ili da drugi cvor cuva podatke nedostupnog cvora dok ovaj ne postane dostupan I onda mu predaje podatke

Obrada otkaza

Gossip (epidemic) protocol

Kad sused cvora koji je otkazao primeti da je taj cvor otkazao on salje informacije drugim cvorovima koji salju drugima I kao epidemija se siri informacija o otkazu cvora



////////////////////////////////////

Voldemort

LinkedIn

Baziran na Dynamo resenju