Paper Evaluation

Amazon Aurora: Design Considerations for High Throughput Cloud-Native Relational Databases

Michael Lanthier

Amazon Aurora is a relational datbase designed to operate in a distributed environment while providing consistency and high preformance. By de-coupling computation from storage, Aurora moves data storage into a quorom system accross multiple data centers allowing for data to survive and remain available even if a full data center goes down. Aurora is also able to reduce network traffic by only allowing redo logs to be transmitted over the network which allows the storage tier to generate the database in the background.

Questions/Comments

1. Not the best comment, but some hard evidence/ numbers would be nice. They talk about their own failure rates but never mention what they are. I suppose this is probably secrecy but some numbers for context especially for people like me who do not really know how often systems fail in an industrial setting would be nice. This also seems like a very good data point to have as quorum replication is rended useless if a majority of the nodes fail.

2. What is Aurora using S3 for? They say they are using it to backup data and restore when needed. Are they just storing a full extra copy of the db in S3 that is being updated in the background as transactions run? I really am not sure what it is four and if it is necessary at all. It does not seem to be imparative to normal operation as it is barely mentioned.

3. If the background construction of the database in the storage nodes get sufficiently far behind when constructing from the logs will there be a significant performance hit and if there is how do they mitigate it? It seems like they try to mitigate it with the LSN allocation limit (LAL) so the storage system does not get far behind but their 10 million number seems arbitrary and knowing why that would be chosen would be handy.