

Week 6 – Managing large databases – Part I

Introduction

- Data volumes grows daily and/or number of users grows daily requiring the need to provide stronger and more powerful infrastructure.
- Data access may be critical for business 24x7x365 or maintenance windows may be allowed for hours or a day. For example, Chick Filet is a business that closes on Sundays while other restaurants are open 24x7.
- Data access may be consistent throughout the week, month, and year or may have spikes on certain days of the week, month, or year. For example, holidays for retailer.
- With globalization and the cloud based software services, the user base is distributed geographically necessitating the need for a better solution than the old model of a giant centralized computer system over a single massive centralized database.

Technologies and Terminology

- Database Replication – replicating data from one database to another in order to reduce load on a particular server by making other identical servers available for users, also provides for better fault tolerance. Replication maybe done with frequent snapshot copies or at the transaction level.
- Table Partitioning – storing the rows of a large table across multiple partitions.
- Redundancy – duplication of a critical component (which may be data) of a system for the sake of using it in the case of a failure of the primary component.
- Fail over – switching to a redundant standby component.
- Maintainability / Serviceability of a system – the ease and the speed by which a system can be maintained.

Availability

- Percent of time a system is available to serve its intended functions.
- A high availability system is a system that is tolerant for power outages, network outages, data center shutdown, and other environment failures.
- Typically measured with the “nines”. A 100% availability is 24 hours a day / 7 days a week / 365 days a year. A system with a 99.9 % availability (three nines) means it is down only 8.76 days a year. A system with 99.99 % availability (four nines) means that it is down 52.56 minutes out of the year.
- A Service Level Agreement (SLA) defines the allowed downtime per contract. It’s subject to interpretation.

Reliability

- Describes a system where the likelihood of failure –while operating the system- is low and therefore the data is available to the client without any knowledge of the system having to utilize the fail over in place.
- A more reliable system is a system that has strong Fail over measures, i.e. is resilient to failures with technologies in place to mask the failures from the business applications / business

analysts. For example, through the use of replication and clustering a system could be more fault tolerant as the processing is switched to the copies of the data when a failure occurs.

Difference between Availability and Reliability:

- Availability is the probability that a system is available when it's needed, i.e. it's not undergoing maintenance or in a failed state.
- Reliability is the probability that the system, when available, will produce the correct output. Reliability doesn't account for repair action that may be taking place, and is not related to the time it takes to get a failed component back up and running.
- So Availability is a function of Reliability and Maintainability.
- A system of reliable components that require a lot of maintenance is not a highly available system.

Scalability

- Scalable – system that can adjust to increased workload without having to be redesigned, e.g. another hardware server is added to adjust for the increased load during the holiday season.
- Scaling up by adding more CPU, memory, disk.
- Scaling out by adding more hardware servers.

Partitioning, Distributed Databases,....part 2