Sharding – What is it and why is it important?

Sharding is a way of partitioning your datastore to benefit from the computing power of more than one server.  For instance many web-facing databases get sharded on user\_id, the unique serial number your application assigns to each user on the website.

Sharding can bring you the advantages of horizontal scalability by dividing up data into multiple backend databases.  This can bring tremendous speedups and performance improvements.

Sharding, however has a number of important costs.

* reduced availability
* higher administrative complexity
* greater application complexity

High Availability is a goal of most web applications as they aim for always-on or 24×7 by 365 availability.  By introducing more servers, you have more components that have to work flawlessly.  If the expected downtime of any one backend database is 1/2 hour per month and you shard across five servers, your downtime has now increased by a factor of five to 2.5 hours per month.

Administrative complexity is an important consideration as well.  More databases means more servers to backup, more complex recovery, more complex testing, more complex replication and more complex data integrity checking.

Since Sharding keeps a chunk of your data on various different servers, your application must accept the burden of deciding where the data is, and fetching it there.  In some cases the application must make alternate decisions if it cannot find the data where it expects.  All of this increases application complexity and is important to keep in mind.

[Sean Hull asks on Quora – What is Sharding and why is it important?](https://www.quora.com/What-is-sharding-and-why-is-it-important)

Related posts:

1. [High Availability – What is it and why is it important?](https://www.iheavy.com/2011/06/13/high-availability-what-is-it-and-why-is-it-important/)
2. [Point-in-time Recovery – What is it and why is it important?](https://www.iheavy.com/2011/07/15/point-in-time-recovery-what-is-it-and-why-is-it-important/)
3. [How To Build Highly Scalable Web Applications For The Cloud](https://www.iheavy.com/2010/12/30/how-to-build-highly-scalable-web-applications-for-the-cloud/)
4. [Zero Downtime – What is it and why is it important?](https://www.iheavy.com/2011/06/23/zero-downtime-what-is-it-and-why-is-it-important/)
5. [Capacity Planning – What is it and why is it important?](https://www.iheavy.com/2011/06/26/capacity-planning-what-is-it-and-why-is-it-important/)