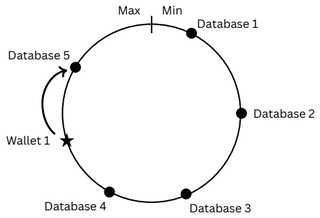
Consistent hashing

Consider a hash function and the minimum and maximum values it would produce. They are joined in a circle like shown below.



The ID of each wallet in the database is unique. Its hash value will first be calculated. The resulting hash value can lie anywhere on the circumference of the circle.

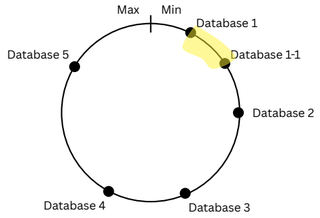
The hash value of the name of each database is also calculated and plotted on the above circle.

Data for the wallet will be stored in the database closest to it, moving clockwise on the circle above.

Why consistent hashing?

The are several methods to distribute independent sets of data over several database partitions. Consistent hashing is a method to ensure that only a small portion of the data needs to be migrated if a new database server is added or removed.

Consider the hashing ring below for which a new database (database 1-1) has been added. Only the data in the highlighted region of the ring needs to be moved from Database 2 to Database 1-1 when the new database was added.



This is a significant improvement compared to if a simple hash function like below was used to distribute the data. Data on all servers will need to be migrated when a new database is added if the scheme below was used.

Database to store data = Hash value % Number of databases