

Debate on data exploration

Material

- Hands – on Sharding with MongoDB

Description

Consider that you must develop a “vademecum” for helping decision making processes related to data sharding using MongoDB. Use the following questions to interact with the members of your group to develop a set of hints that would contain your vademecum.

- What is a sharding key? Is the choice of a sharding key directly dependent of the sharding strategy? Explain and give examples.
- Explain how does apparently mongodb choses the number of intervals used to shard a collection in the Interval oriented strategy? In which situations would such a strategy be well adapted for sharding a collection?
- In which situations would the hash based strategy be interesting for a collection to be sharded?
- Which of the strategies interval or sharding would lead to a more balanced distribution of data across shards, interval or hash?
- What are the advantages and disadvantages of allowing access to shards directly through their server and not only through the query router?
- Give an example of situation where tag based sharding would be an interesting option?
- What happens when a new shard is added to a cluster containing already other shards with data?
- How would you test whether a sharded collection was an interesting solution in comparison to a centralized one?

To Do

- Prepare a document / slides summarizing your examples and discussion results.
- Choose a spokesperson of the group who will present your conclusions to the group.
- Hand in your document / slides to genoveva.vargas@gmail.com