FINAL: QUESTION 9

Imagine an electronic medical record database designed to hold the medical records of every individual in the United States. Because each person has more than 16MB of medical history and records, it's not feasible to have a single document for every patient. Instead, there is a *patient* collection that contains basic information on each person and maps the person to a patient_id, and a*record* collection that contains one document for each test or procedure. One patient may have dozens or even hundreds of documents in the *record* collection.

We need to decide on a shard key to shard the *record* collection. What's the best shard key for the *record* collection, provided that we are willing to run inefficient scatter-gather operations to do infrequent research and run studies on various diseases and cohorts? That is, think mostly about the operational aspects of such a system. And by operational, we mean, think about what the most common operations that this systems needs to perform day in and day out.

- patient_id
- _id
- Primary care physician (your principal doctor that handles everyday problems)
- Date and time when medical record was created
- Patient first name
- Patient last name

SUBMIT