Prof. Jayant Harista

Title: Mutating Database Engines to be Bio-friendly

Abstract: Database systems are touted to be the panacea for the world's information processing problems. While this claim is justified with regard to industrial and commercial applications, there is comparatively poor support for scientific applications, such as computational genomics. As a result, scientists have been largely forced to store their data in flat files, resulting in long running times for their queries. This situation will become only worse in the future given the exponential increase in the size of scientific data repositories and the complexity of the queries posed on them.

In this talk, we will look at how database engines can be redesigned to efficiently support the processing of biological information, especially with regard to genomic sequences. We will make the case that a holistic solution requires reworking the core components of the engine, including the data storage mechanisms, index structures, and query optimization techniques, thereby providing rich research opportunity for algorithmic, structural, architectural and mathematical innovations.