

Pencarian Anomali dengan Menggunakan Algoritma BLAST Menggunakan framework MapReduce pada Platform Big Data = Genomic Anomaly Searching with BLAST Algorithm using MapReduce Framework in Big Data Platform

Ramanti Dharayani, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=20504264&lokasi=lokal>

Abstrak

Berkembangnya teknologi pada bioteknologi berdampak pada banyaknya jumlah data yang harus diolah untuk melakukan analisis dalam pencarian anomali terutama pada data Genomic. Selain pertumbuhan data kebutuhan akan kompleksitas dalam melakukan proses olah data akan mengikuti seiring berkembangnya data genomic. Dengan berkembangnya data dan kompleksitas olah data, Big Data merupakan salah satu solusi untuk mengatasi perkembangan data dan olah data yang cukup kompleks. Dalam penelitian ini membahas bagaimana Big Data dapat melakukan pencarian anomali pada data genomic dengan menggunakan MapReduce dan algoritma BLAST pada platform Big Data. Hasil dari penelitian ini adalah pencarian anomali pada data genomic dapat dilakukan dengan menggunakan mapreduce pada platform big data dengan waktu yang lebih efisien.

.....Biofarma Corp should adopt big data on vaccine and serum development by analyze genomic sequencing using searching any anomaly. As the root problem, it the anomaly searching requires about 1.62 Terabytes data transient as primary data and 301 Gigabytes as secondary data to get analysis from genomic variance. Moreover Biofarma Corp spent 16 hours for one anomaly searching from 3 Terabytes vaccines. This study proposed big data implementation to handle anomaly searching processes by prioritize less time complexity and less spending storage. It was signaled by a research question, "How big data technology is applied in searching anomalies on genomic data". It aimed to implement big data system to facilitate large volume and complex data in order to fulfill business process on Biofarma Corp. It adopted framework architecture as brought by Demchenko, Ngo, and Membrey. This study has designed data flow from FASTA and FATQ as sources for anomalies searching processes. This data flow is facilitated in big data system as designed in this research. As main contribution, this research adopted MapReduce framework to run BLAST algorithm with less spending time. As comparison, MapReduce framework can handle 21, 33, and 55 K-Mer in four minutes respectively while 50 minutes were spent without MapReduce.