

ABSTRACT

The diploma project on the topic "Software architecture of fault-tolerant big data processing systems" consists of an introduction, four chapters, conclusions, a list of reference sources, appendices.

The total volume of work is 102 pages. The list of reference sources consists of 37 items. The work includes 20 figures, 23 tables, and 2 appendixes.

The purpose of the research is to improve the quality of existing software architectures of a fault-tolerant big data processing system.

The object of research is the software of big data processing systems.

The subject of the research is methods, models and means of building the architecture of a fault-tolerant big data processing system.

In the article, an overview of the problems of fault-tolerance during big data processing was carried out. The available tools for big data processing and building such systems have been reviewed. As a result of the work, an improved architecture for fault-tolerant big data processing systems was proposed. With the help of the reviewed tools and techniques for increasing fault-tolerance, the architecture ensures high throughput and is resistant to faults with fast fault recovery. Also, a billing data processing system was developed according to the developed enhanced architecture of fault-tolerant big data processing systems.

Keywords: ARCHITECTURE, FAULT-TOLERANCE, BIG DATA, RECOVERY, KUBERNETES, SPARK, STRUCTURED STREAMING, KAFKA.