

БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
Факультет радиофизики и компьютерных технологий
Кафедра интеллектуальных систем

Аннотация к дипломной работе

**Обработка больших объемов данных на основании модели
асинхронных потоков**

Сайков Павел Викторович

Научный руководитель: старший преподаватель А.В. Курочкин

Минск, 2021

РЕФЕРАТ

Дипломная работа: 62 страницы, 18 рисунков, 13 таблиц, 26 источников.

СТРУКТУРЫ ДАННЫХ, БАЗЫ ДАННЫХ, ВЫСОКОНАГРУЖЕННАЯ СИСТЕМА, МАСШТАБИРОВАНИЕ, WEB CLICK DATA, СОБЫТИЙНАЯ МОДЕЛЬ, ПОТОКОВАЯ ОБРАБОТКА ДАННЫХ.

Цель работы - на базе существующих современных программных решений в области потоковой обработки данных, разработать инструмент для сбора статистической информации, используя механизм потокового чтения из некоторого источника.

Методология проведения работы – компьютерное программирование

Данная работа направлена на рассмотрение современных подходов к хранению данных, а также способов взаимодействия и обработки.

В частности, рассматривается вопрос оперативного и долговременного хранения данных, что включает в рассмотрение структур данных, систем управления баз данных и индексирования.

Также рассмотрению подвергается вопрос высоконагруженных систем, а именно какими характеристиками обладает высоконагруженная система, какую роль в таких системах играет масштабирование, разновидности таких систем и их особенности.

Далее в работе рассматривается платформа «Tarantool», на основании которой можно спроектировать высоконагруженную систему, а именно рассмотрению подвергаются ее архитектурные особенности и задачи, который данный инструмент позволяет решить.

Также рассмотрена концепция потоковой обработки данных, на основании которой было разработано программное решение, позволяющее агрегировать статистическую информацию на основании структурированных данных.

РЭФЕРАТ

Дыпломная праца: 62 старонкі, 18 малюнкаў, 13 табліц, 26 крыніц.

СТРУКТУРЫ ДАДЗЕННЫХ, БАЗЫ ДАДЗЕННЫХ,
ЫСОКАНАГРУЖАНАЯ СІСТЭМА, МАШТАБАВАННЕ, WEB CLICK
DATA, ПАДЗЕЙНАЯ МАДЭЛЬ, СТУРМНЕЎВАЯ АПРЦОЎКА
ДАДЗЕННЫХ.

Мэта працы - на базе існуючых сучасных праграмных рашэнняў у вобласці струменевай апрацоўкі дадзеных, распрацаваць інструмент для збору статыстычнай інфармацыі, выкарыстоўваючы механізм струменевага чытання з некаторай крыніцы.

Метадалогія правядзення працы - камп'ютэрнае праграмаванне

Дадзеная праца накіравана на разгляд сучасных падыходаў да захоўвання дадзеных, а таксама спосабаў ўзаемадзеяння і апрацоўкі.

У прыватнасці, разглядаецца пытанне аператыўнага і доўгачасовага захоўвання дадзеных, што ўключае ў разгляд структур дадзеных, сістэм кіравання баз дадзеных і індэксавання.

Таксама разгляду падвяргаецца пытанне высоканагружаных сістэм, а менавіта якімі характарыстыкамі валодае высоканагружаная сістэма, якую ролю ў такіх сістэмах гуляе маштабаванне, разнавіднасці такіх сістэм і іх асаблівасці.

Далей у рабоце разглядаецца платформа “Tarantool”, на падставе якой можна спраектаваць высоканагружаную сістэму, а менавіта разгляду падвяргаюцца яе архітэктурныя асаблівасці і задачы, які дадзены інструмент дазваляе вырашыць.

Таксама разгледжана канцэпцыя струменевай апрацоўкі дадзеных, на падставе якой было распрацавана праграмнае рашэнне, якое дазваляе аграгаваць статыстычную інфармацыю на падставе структураваных дадзеных.

ABSTRACT

Thesis: 62 pages, 18 figures, 13 tables, 26 sources.

DATA STRUCTURES, DATA BASES, HIGHLOAD SYSTEMS, SKEALING, WEB CLICK DATA, EVENT-DRIVEN MODEL, STREAM DATA PROCESSING.

Objectives – to develop a tool for collecting statistical information using the mechanism of streaming reading from some source on the basis of existing modern software solutions in the field of streaming data processing.

Methods – computer programming.

This work is aimed at considering modern approaches to data storage, as well as ways of interaction and processing.

In particular, the issue of operational and long-term data storage is considered, which includes consideration of data structures, database management systems and indexing.

The issue of high-loaded systems is also considered, namely, what characteristics does a highly-loaded system have, what role scaling plays in such systems, the types of such systems and their features.

Further, the work considers the «Tarantool» platform, on the basis of which it is possible to design a high-load system, namely, its architectural features and the tasks that this tool allows to solve are examined.

The concept of streaming data processing was also considered, on the basis of which a software solution was developed that allows aggregating statistical information based on structured data.