

# **ЗАДАЧА ОПТИМАЛЬНОГО ВЫБОРА ТЕХНОЛОГИЙ РАСПРЕДЕЛЁННЫХ ВЫЧИСЛЕНИЙ НА КЛАСТЕРАХ ДЛЯ ФИНАНСОВОГО СЕКТОРА**

**А. А. Тащилин, А. К. Кузьмин**

*Саратовский государственный технический университет, Россия*  
E-mail: a.tashchilin@bk.ru, kuz\_alex\_konst@mail.ru

В статье рассматривается актуальность проблемы оптимального выбора технологий для организации распределённых вычислений в финансовой сфере. Ставится задача разработки подхода и соответствующего программного обеспечения, которые позволят получать рекомендации по выбору технологий с учетом различных параметров задачи и ограничений.

## **THE PROBLEM OF OPTIMAL SELECTION OF DISTRIBUTED COMPUTING TECHNOLOGIES ON CLUSTERS FOR THE FINANCIAL SECTOR**

**A. A. Tashchilin, A. K. Kuzmin**

The article discusses the relevance of the problem of optimal choice of technologies for organizing distributed computing in the financial sector. The task is to develop an approach and corresponding software that will make it possible to obtain recommendations on the choice of technologies taking into account various task parameters and restrictions.

Финтех, или финансовые технологии, охватывают широкий спектр инновационных решений, направленных на улучшение и оптимизацию финансовых операций и услуг. Это может включать в себя разработку новых платежных систем, цифровых валют, кредитных продуктов, управления рисками, а также внедрение технологий блокчейн и искусственного интеллекта в финансовые процессы.

Для решения сложных задач в области финансовых технологий, таких как скоринг потенциальных заёмщиков, прогнозирование рынка, оптимизация портфеля инвестиций и др., часто требуются большие вычислительные мощности. В таких случаях применяются распределенные вычисления на кластерах серверов, где задача делится на несколько подзадач, которые выполняются параллельно на различных узлах сети.

Это позволяет сократить время выполнения задачи и увеличить эффективность работы системы. Для организации распределенных вычислений используются специализированные программные и аппаратные средства, такие как Apache Spark [1], Apache Flink [2], Apache Storm [3] и многие другие.

Один из классических методов решения ресурсозатратных задач – это подход MapReduce. Эта парадигма была предложена компанией Google в 2004 году в статье "MapReduce: Simplified Data Processing on Large Clusters" [4].

В то время как статья содержала описание концепции, реализация не была

предоставлена, и несколько программистов из Yahoo разработали свою собственную реализацию в рамках работы над веб-краулером Nutch. Более подробно о развитии Hadoop можно узнать из статьи "The history of Hadoop: From 4 nodes to the future of data" [5].

Технология MapReduce позволяет разбить задачу на более простые подзадачи, которые могут быть выполнены параллельно на нескольких узлах вычислительного кластера. Затем из локальных решений, полученных на узлах кластера, собирается общее решение.

В финансовой сфере применение MapReduce позволяет эффективно обрабатывать и анализировать огромные массивы финансовых данных, включая информацию о торговых операциях, клиентах, рыночных трендах, рисках и прочее. Это позволяет финансовым учреждениям быстро и точно принимать решения, оптимизировать портфель инвестиций, управлять рисками и реагировать на изменения на рынке.

Преимущества использования технологии MapReduce в финтехе включают в себя:

- Увеличение скорости обработки данных: параллельная обработка данных на нескольких узлах кластера позволяет значительно ускорить процесс обработки данных.

- Снижение нагрузки на вычислительные ресурсы: распределенная обработка данных позволяет эффективно использовать ресурсы кластера, сокращая время выполнения задач и снижая нагрузку на сервера.

- Улучшение горизонтальной масштабируемости: технология MapReduce позволяет легко масштабировать вычислительные ресурсы в зависимости от объема данных, что позволяет более эффективно работать с растущими объемами данных.

- Повышение точности анализа данных: более быстрая и эффективная обработка данных с помощью MapReduce позволяет проводить более точный анализ финансовых данных, что помогает улучшить прогнозирование рыночных тенденций и принимать более обоснованные инвестиционные решения.

Решение об использовании подхода MapReduce сопряжено с выбором языка и технологий реализации, а также специализированного фреймворка для распределенных вычислений. Однако выбор может быть нетривиальным.

Основными критериями при выборе языка и технологий в данном случае являются профессиональные навыки команды разработчиков, специфика задачи и требования заказчика, а также популярность и поддержка технологий на рынке. На практике, часто критерии ограничиваются уже имеющимися знаниями и опытом команды, поэтому выбор может быть не всегда оптимальным.

Для улучшения процесса выбора технологий и языков программирования при разработке распределённых систем предлагается разработать подход, основанный на анализе и учете следующих факторов:

1. Специфика задачи: определение основных требований к системе, её функций и особенностей.

2. Требования заказчика: учёт пожеланий и предпочтений клиента относительно используемых технологий и инструментов.

3. Профессиональные навыки команды разработчиков: оценка опыта и знаний членов команды, их способность освоить новые технологии.

4. Популярность и поддержка технологий на рынке: изучение существующих решений и их применимости для конкретной задачи.

5. Характеристики данных: определение объёма и структуры данных, которые будут обрабатываться системой.

6. Доступные ресурсы: оценка времени, бюджета, технических возможностей, которые могут повлиять на выбор технологий.

Такой подход позволит компаниям эффективно использовать потенциал распределённых вычислений MapReduce, выбирать наиболее подходящие технологии и учитывать особенности каждой задачи. В результате повысится качество и эффективность разработки и реализации распределённых систем обработки данных. Актуальной видится разработка интеллектуальной системы поддержки принятия решений, которая накопит в себе знания, необходимые для выбора технологий разработки и параметров кластера исходя из имеющихся реалий и ограничений организации.

#### СПИСОК ЛИТЕРАТУРЫ

1. Оптимизация заданий Apache Spark. Часть 1. [Электронный ресурс]. URL: <https://datareview.info/article/optimizatsiya-zadaniy-apache-spark-chast-1/> (дата обращения: 09.10.2024).

2. Monitoring the Wikipedia Edit Stream. [Электронный ресурс]. URL: [https://nightlies.apache.org/flink/flink-docs-release-1.1/quickstart/run\\_example\\_quickstart.html](https://nightlies.apache.org/flink/flink-docs-release-1.1/quickstart/run_example_quickstart.html) (дата обращения: 09.10.2024).

3. Apache Storm. [Электронный ресурс]. URL: [https://www.tadviser.ru/index.php/%D0%9F%D1%80%D0%BE%D0%B4%D1%83%D0%BA%D1%82:Apache\\_Storm](https://www.tadviser.ru/index.php/%D0%9F%D1%80%D0%BE%D0%B4%D1%83%D0%BA%D1%82:Apache_Storm) (дата обращения: 09.10.2024).

4. MapReduce: Simplified Data Processing on Large Clusters. [Электронный ресурс]. URL: <https://static.googleusercontent.com/media/research.google.com/ru//archive/mapreduce-osdi04.pdf> (дата обращения: 09.10.2024).

5. A survey on evolution of big data with hadoop. [Электронный ресурс]. URL: [https://www.researchgate.net/publication/322070950\\_a\\_survey\\_on\\_evolution\\_of\\_big\\_data\\_with\\_hadoop](https://www.researchgate.net/publication/322070950_a_survey_on_evolution_of_big_data_with_hadoop) (дата обращения: 09.10.2024).