



# **АКТУАЛЬНЫЕ ПРОБЛЕМЫ ТЕОРИИ, МЕТОДОЛОГИИ И ПРАКТИКИ НАУЧНОЙ ДЕЯТЕЛЬНОСТИ**

**Сборник статей  
Международной научно-практической конференции  
25 июня 2020 г.**

**НАУЧНО-ИЗДАТЕЛЬСКИЙ ЦЕНТР «АЭТЕРНА»  
Омск, 2020**

УДК 00(082) + 001.18 + 001.89  
ББК 94.3 + 72.4: 72.5  
А 437

А 437

**АКТУАЛЬНЫЕ ПРОБЛЕМЫ ТЕОРИИ, МЕТОДОЛОГИИ И ПРАКТИКИ НАУЧНОЙ ДЕЯТЕЛЬНОСТИ:** сборник статей Международной научно-практической конференции (25 июня 2020 г., г. Омск). - Уфа: Аэтерна, 2020. – 318 с.

ISBN 978-5-00177-008-4

**Настоящий сборник составлен по итогам Международной научно-практической конференции «АКТУАЛЬНЫЕ ПРОБЛЕМЫ ТЕОРИИ, МЕТОДОЛОГИИ И ПРАКТИКИ НАУЧНОЙ ДЕЯТЕЛЬНОСТИ», состоявшейся 25 июня 2020 г. в г. Омск. В сборнике статей рассматриваются современные вопросы науки, образования и практики применения результатов научных исследований.**

Все материалы сгруппированы по разделам, соответствующим номенклатуре научных специальностей.

Сборник предназначен для широкого круга читателей, интересующихся научными исследованиями и разработками, научных и педагогических работников, преподавателей, докторантов, аспирантов, магистрантов и студентов с целью использования в научной и педагогической работе и учебной деятельности.

Все статьи проходят рецензирование (экспертную оценку). **Точка зрения редакции не всегда совпадает с точкой зрения авторов публикуемых статей.** Статьи представлены в авторской редакции. Ответственность за точность цитат, имен, названий и иных сведений, а так же за соблюдение законов об интеллектуальной собственности несут авторы публикуемых материалов.

При использовании опубликованных материалов в контексте других документов или их перепечатке ссылка на сборник статей научно-практической конференции обязательна.

**Полнотекстовая электронная версия сборника размещена в свободном доступе на сайте <https://aeterna-ufa.ru/arh-conf/>**

Сборник статей постатейно размещён в научной электронной библиотеке elibrary.ru по договору № 242 - 02 / 2014К от 7 февраля 2014 г.

ISBN 978-5-00177-008-4

УДК 00(082) + 001.18 + 001.89  
ББК 94.3 + 72.4: 72.5

© ООО «АЭТЕРНА», 2020  
© Коллектив авторов, 2020

литься под давлением являются высокая стоимость пресс - формы, сложность и длительность ее изготовления.

### **Список использованной литературы**

1. Учебное пособие по дисциплине «Обработка металлов резанием, станки, инструмент» для обучающихся дневного отделения специальности 15.02.12 «Монтаж, техническое обслуживание и ремонт промышленного оборудования (по отраслям)» 2020г.

© И.В. Воронова, М.И. Назарова, 2020

**УДК 004**

**Галенина А.А.**

студент Курского Государственного Университета, г. Курск

**Galenina Alina Alexandrovna**

student of Kursk State University, Kursk

## **ОБЗОР АРХИТЕКТУРЫ APACHE HADOOP**

## **SURVEY OF THE ARCHITECTURE APACHE HADOOP**

### ***Аннотация***

Когда - то Hadoop был именем маленького желтого слоненка, но сейчас за этим словом стоит целая платформа, которая позволяет хранить большие данные в распределенной среде и обеспечивать быструю скорость доступа и обработки этих данных. Данная статья рассказывает об истории возникновения фреймворка. Приводится четкое описание ядра современной платформы Hadoop. Также рассматривается архитектура основных компонентов ядра.

### ***Ключевые слова***

Hadoop, MapReduce, архитектура Hadoop, фреймворк, распределенные вычисления, Apache, HDFS, YARN.

### ***Abstract***

Once Hadoop was the name of a small yellow elephant but today it is a large framework that allows you to store big data in a distributed environment and provide fast access and processing for this data. This article tells about the history of the framework. A clear description of the core of the modern Hadoop framework is provided. The architecture of the core key components is also considered.

### ***Keywords***

Hadoop, MapReduce, Hadoop architecture, framework, distributed computing, Apache, HDFS, YARN.

История возникновения Hadoop началась с двух имен – Майк Кафарелла (Mike Cafarella) и Дуг Каттинг (Doug Cutting). Они работали над проектом с открытым исходным кодом

Nutch, целью которого была индексация веб - страниц и использование уже проиндексированных страниц для поиска.

В результате исследования Майк Кафарелла и Дуг Каттинг пришли к выводу, что разрабатываемая ими система будет стоить около 500000\$ на оборудовании с ежемесячной эксплуатационной стоимостью в 30000\$ [6; с.34]. Данная система была достаточно дорогостоящей. Но вскоре они поняли, что архитектура данной системы не способна проиндексировать миллиард страниц.

Некоторым временем позже Майк Кафарелла и Дуг Каттинг наткнулись на статью, опубликованную в 2003 году, в которой описывалась архитектура распределенной файловой системы Google, называемой GFS (Google File System). Эта статья оказалась тем, что они так долго искали. Данная статья помогла решить проблемы хранения очень больших файлов, которые создаются как часть процесса сканирования и индексации в Интернете.

Позже в 2004 году Google опубликовал еще одну статью, которая представила миру метод MapReduce. Наконец, эти две статьи привели к созданию платформы под названием «Hadoop».

Hadoop [3; с.23] – это платформа с открытым исходным кодом для написания и запуска распределенных приложений, которые обрабатывают большие объемы данных. Hadoop написан на языке программирования Java и входит в число проектов Apache самого высокого уровня.

Ядро современной платформы Hadoop [2, с. 53] представляет распределенная файловая система Hadoop (HDFS), Hadoop YARN, Hadoop MapReduce и Hadoop Common. HDFS и YARN – это два основополагающих компонента, которые необходимы для настройки Hadoop.

HDFS (Hadoop Distributed File System) расшифровывается как распределенная файловая система Hadoop. Это файловая система с блочной структурой, в которой каждый файл разделен на блоки заранее определенного размера. Эти блоки хранятся в кластере одной или нескольких машин.

Архитектура Apache Hadoop HDFS (рис. 1) состоит из одного NameNode (главного узла), а все остальные узлы являются DataNode (подчиненными узлами).

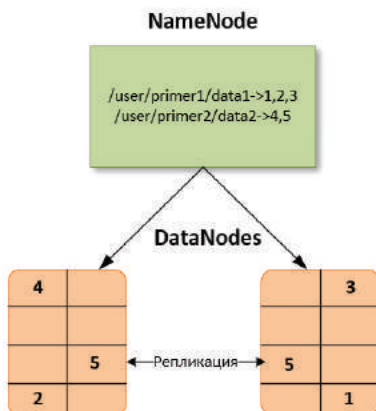


Рисунок 1. Архитектура HDFS

NameNode [1; с. 128] – это главный узел в архитектуре HDFS Apache Hadoop, который поддерживает и управляет блоками, присутствующими в узлах DataNode (подчиненных узлах). NameNode – это сервер с высокой степенью доступности, который управляет пространством имен файловой системы и контролирует доступ клиентов к файлам. Архитектура HDFS построена таким образом, что пользовательские данные никогда не находятся в NameNode. Данные находятся только на узлах DataNode.

Функции NameNode:

- поддерживает и управляет DataNode (подчиненными узлами);
- записывает метаданные всех файлов, хранящихся в кластере, (местоположение сохраненных блоков, размер файлов);
- записывает каждое изменение в метаданных файловой системы. (если файл удален в HDFS, NameNode немедленно запишет это в EditLog);
- получение отчетов от узлов DataNode с целью проверки работоспособности узлов;
- учет всех блоков в HDFS и в каких узлах эти блоки находятся.

В отличие от NameNode, DataNode [4; с.154] – это стандартное аппаратное обеспечение, то есть недорогая система, которая не отличается высоким качеством или высокой доступностью. DataNode – это блочный сервер, который хранит данные в локальном файле.

Функции DataNode:

- хранит фактические данные;
- выполняет низкоуровневые запросы чтения и записи от клиентов файловой системы;
- периодически посылают запрос в NameNode, чтобы сообщить об общем состоянии HDFS, по умолчанию эта частота установлена в 3 секунды.

Помимо этого, существует третий компонент, называемый Secondary NameNode. Secondary NameNode работает одновременно с основным NameNode как вспомогательный. Многие считают, что Secondary NameNode является резервным для NameNode, но это не так.

Функции вторичного NameNode:

- считывание всех файловых систем и метаданных из ОЗУ NameNode и запись их на жесткий диск или в файловую систему;
- объединение EditLogs с FsImage из NameNode;
- загрузка EditLogs из NameNode через равные промежутки времени и применение этих данных к FsImage. Новый FsImage копируется обратно в NameNode, который используется всякий раз, когда NameNode запускается в следующий раз.

YARN [5; с. 58]– система планирования заданий и управления кластером (Yet Another Resource Negotiator), которую также называют MapReduce 2.0 (MRv2).

В Hadoop версии 1.0, которая также называется MRv1 (MapReduce версии 1), MapReduce выполнял функции обработки и управления ресурсами. Он представлял собой один процесс JobTracker и произвольное количество процессов TaskTracker. Job Tracker распределяет ресурсы, выполняет планирование и контролировал обработку заданий. Трекеры периодически сообщали о своем прогрессе в Job Tracker.

Такая конструкция привела к узким местам в рамках масштабируемости из-за одного Job Tracker. IBM упоминает, что согласно Yahoo!, практические пределы такой конструкции достигаются с кластером из 5000 узлов и 40 000 задач, выполняемых одновременно.

Чтобы преодолеть все эти проблемы, YARN был представлен в версии 2.0 Hadoop в 2012 году компаниями Yahoo и Hortonworks. Основная идея, стоящая за YARN, состоит в том,

чтобы облегчить MapReduce, взяв на себя ответственность за управление ресурсами и планирование работ.

YARN состоит из следующих основных компонентов:

1. ResourceManager запускается на NameNode и управляет распределением ресурсов в кластере.
2. NodeManager работает на DataNode и отвечает за выполнение задачи на каждом узле данных.
3. Application Master управляет жизненным циклом пользовательских заданий и потребностями в ресурсах отдельных приложений.
4. Container – пакет ресурсов, включая оперативную память, процессор, сеть, жесткий диск, используемых для одного узла DataNode.

Таким образом, Apache Hadoop – это выдающаяся технология, которая питает современную ИТ - индустрию. Данная платформа помогает решить проблему хранения больших и разнородных данных. Hadoop ориентирован на горизонтальное масштабирование. Решения, реализованные в данной платформе, увеличивают скорость доступа и обработки данных. Помимо этого, Hadoop имеет открытый исходный код. Поэтому данная платформа так стремительно набрала популярность.

### **СПИСОК ИСПОЛЬЗОВАННОЙ ЛИТЕРАТУРЫ**

1. Aven, J. Sams Teach Yourself Hadoop in 24 Hours / J. Aven. – М.: Sams Publishing, 2017. – 488 p. – Текст : непосредственный.
2. DeRoos, D. Hadoop For Dummies / D. DeRoos. – М.: John Wiley & Sons, Inc., 2014. – 411 p. – Текст : непосредственный.
3. Lam, C. Hadoop in Action / C. Lam. – М.: Manning Publications, 2010. – 325 p. – Текст: непосредственный.
4. Owens, J Hadoop Real World Solutions Cookbook / J. Owens, B. Femiano, J. Lentz; М.: Packt Publishing, 2013. – 318 p. – Текст : непосредственный.
5. Tannir. K Optimizing Hadoop for MapReduce / K. Tannir. М.: Packt Publishing, 2014. – 120 p. – Текст : непосредственный.
6. White, T. Hadoop: The Definitive Guide / T. White. – М.: Yahoo Press, 2012. – 688 p. – Текст : непосредственный.

© А.А. Галенина, 2020

**УДК 004**

**Е.Ю. Егоров**

адъюнкт ВИ ФСИН РФ, г. Воронеж, РФ

**Л.В. Россихина**

док. тех. наук, проф. ВИ ФСИН РФ, г. Воронеж, РФ

глав. науч. сотр. Академии управления МВД РФ, г. Москва, РФ

### **ПОСТАНОВКА ЗАДАЧИ ОПТИМИЗАЦИИ РЕМОНТА ЭЛЕМЕНТОВ ИНТЕГРИРОВАННОЙ СИСТЕМЫ БЕЗОПАСНОСТИ**

Аннотация

В статье рассматривается задача оптимизации ремонта элементов интегрированной системы безопасности. Предложен алгоритм обслуживания наибольшего числа заявок на

**Научное издание**

**АКТУАЛЬНЫЕ ПРОБЛЕМЫ ТЕОРИИ,  
МЕТОДОЛОГИИ И ПРАКТИКИ  
НАУЧНОЙ ДЕЯТЕЛЬНОСТИ**

**Сборник статей  
Международной научно-практической конференции  
25 июня 2020 г.**

**В авторской редакции**

Издательство не несет ответственности за опубликованные материалы.

Все материалы отображают персональную позицию авторов.

Мнение Издательства может не совпадать с мнением авторов

Подписано в печать 29.06.2020 г. Формат 60х84/16.

Печать: цифровая. Гарнитура: Times New Roman

Усл. печ. л. 15,1. Тираж 500. Заказ 1243.



**Отпечатано в редакционно-издательском отделе  
НАУЧНО-ИЗДАТЕЛЬСКОГО ЦЕНТРА «АЭТЕРНА»**

**450076, г. Уфа, ул. М. Гафури 27/2**

**<https://aeterna-ufa.ru>**

**[info@aeterna-ufa.ru](mailto:info@aeterna-ufa.ru)**

**+7 (347) 266 60 68**