

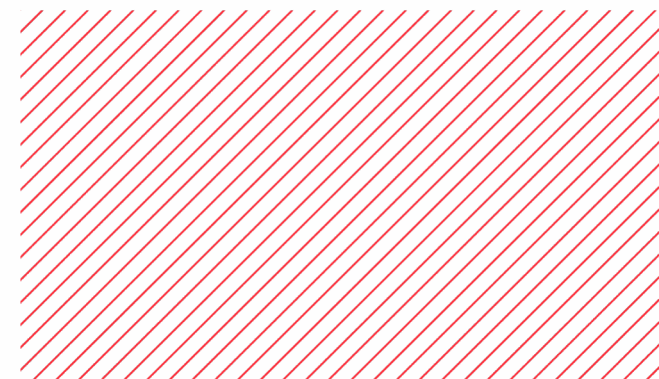
академия
больших
данных

mail.ru
group

made

Введение в большие данные

Дмитрий Музалевский 16.09.2020



Знакомство – Дмитрий Музалевский



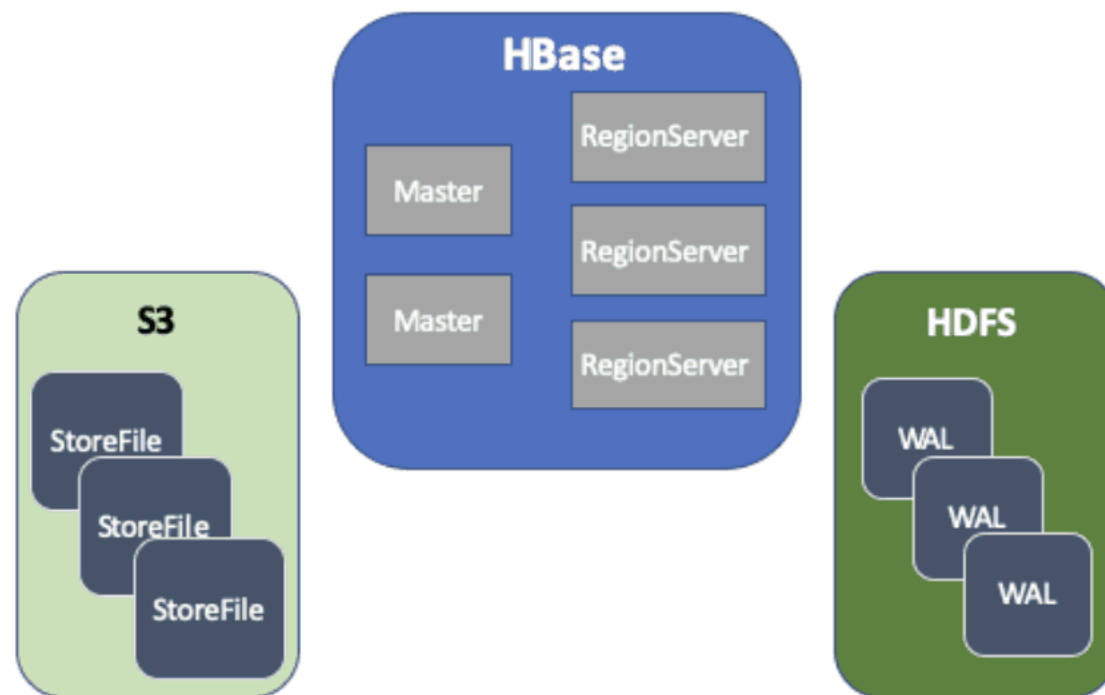
Lead Data Scientist, Берлин

Более 8 лет опыта в области анализа данных и машинного обучения. Опыт работы и проектов в различных индустриях: телеком, онлайн-ритейл, банковская сфера, финтех и медтех.

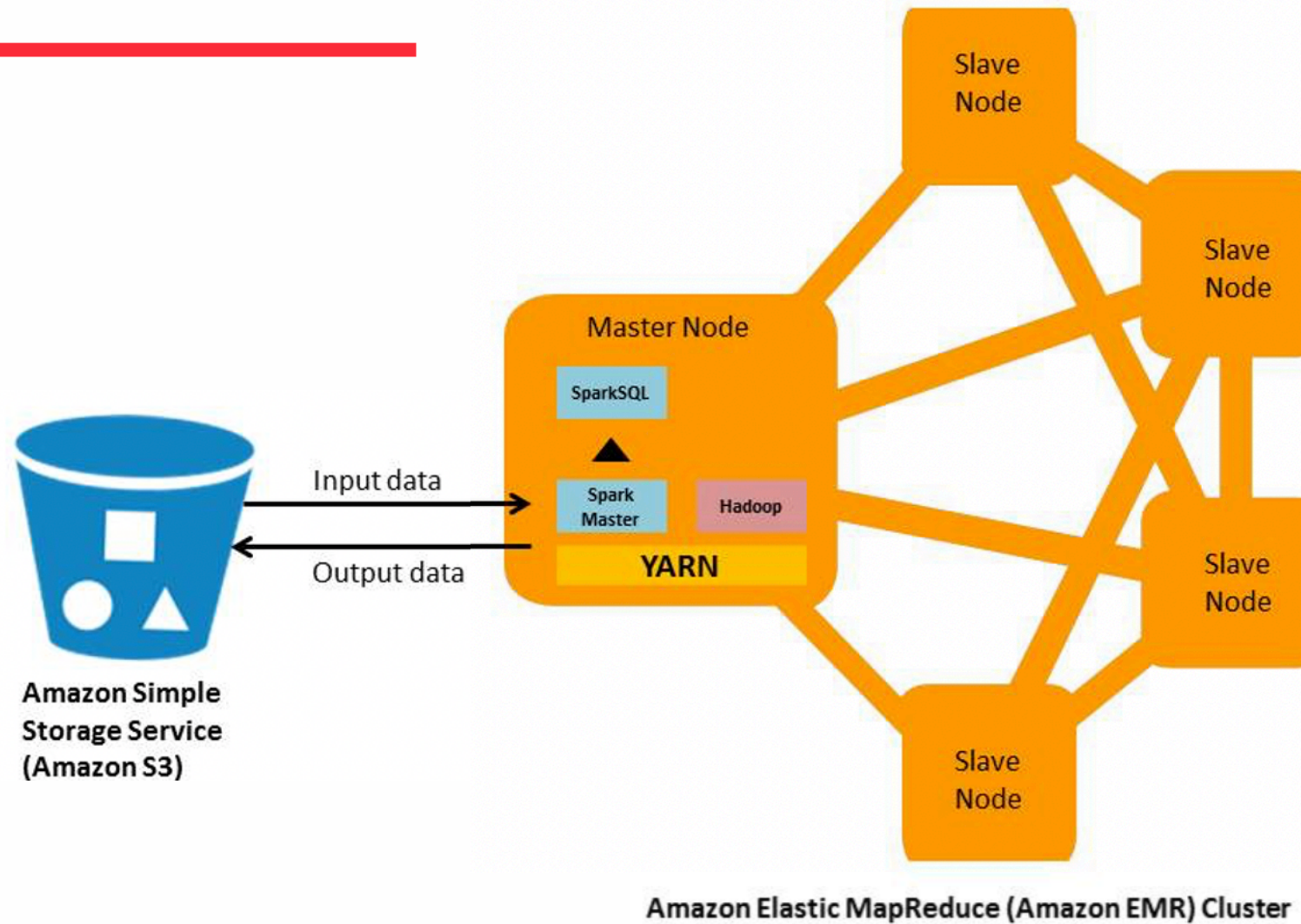
В настоящий момент работаю в медицинской сфере – обработка звуковых сигналов и улучшение слуховых аппаратов. Занимаю позицию Lead Data Scientist и веду работу команды, занимающейся аналитикой больших объемов данных и машинным обучением полного цикла.

Образование: бакалавриат МГТУ им. Баумана "Компьютерный анализ и интерпретация данных"; магистратура НИУ МАИ "Математические методы в экономике".

S3-Hbase-HDFS



Elastic Map Reduce (EMR)





-ls и -du

-ls <path>

Показывает содержимое директории, показывает имена, разрешения, владельца, размер и дату модификации для каждой записи.

-lsr <path>

Ведет себя как -ls, но отображает записи в поддиректориях пути рекурсивно.

-du <path>

Показывает использование disk storage

-dus <path>

Ведет себя как -du, только выводя суммарную информацию



-mv, -cp и -put

-mv <src><dest>

Переносит файл или директорию от source до destination в HDFS

-cp <src><dest>

Копирует файл или директорию от source до destination в HDFS

-distcp <src><dest>

Распределенно копирует файл или директорию от source до destination в HDFS

-put <localSrc> <dest>

Копирует файл или директорию из локальной системы от source до destination в DFS



-get, -getmerge и -cat

-get [-crc] <src> <localDest>

Копирует файл или директорию HDFS из source в локальную систему

-getmerge <src> <localDest>

Извлекает все файлы по пути source в HDFS и копирует их в один объединённый файл в локальной системе

-cat <filename>

Показывает содержимое указанного файла



-mkdir, -rm и -rmr

-mkdir <path>

Создает директорию по указанному пути в HDFS

-rm <path>

Удаляет файл или пустую директорию по указанному пути

-rmr <path>

Удаляет файл или директорию по указанному пути