

Файловая система HDFS (Hadoop Distributed File System)

Файловая система HDFS предназначена для хранения файлов больших размеров, поблочно распределенных между узлами вычислительного кластера.

Все блоки в HDFS (кроме последнего блока файла) имеют одинаковый размер, и каждый блок может быть размещен на нескольких узлах. Размер блока и коэффициент репликации (количество узлов, на которых должен быть размещен каждый блок) определяются в настройках на уровне файла. Благодаря репликации обеспечивается устойчивость распределенной системы к отказам отдельных узлов.

Файлы в HDFS могут быть записаны лишь однажды (модификация не поддерживается), а запись в файл в одно время может вести только один процесс. Организация файлов в пространстве имен — традиционная иерархическая: есть корневой каталог, поддерживается вложение каталогов, в одном каталоге могут располагаться и файлы, и другие каталоги.

Развертывание экземпляра HDFS предусматривает наличие центрального узла имен (англ. name node), хранящего метаданные файловой системы и метаинформацию о распределении блоков, и серии узлов данных (англ. data node), непосредственно хранящих блоки файлов.

Узел имен отвечает за обработку операций уровня файлов и каталогов — открытие и закрытие файлов, манипуляция с каталогами. Узлы данных непосредственно обрабатывают операции по записи и чтению данных. Узел имен и узлы данных снабжаются веб-серверами, отображающими текущий статус узлов и позволяющими просматривать содержимое файловой системы. Административные функции доступны из интерфейса командной строки.