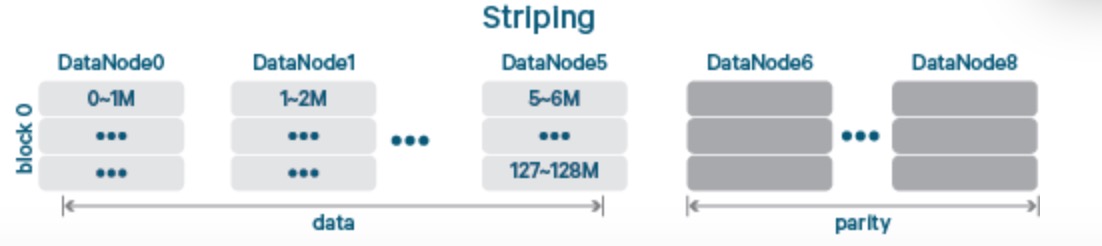
HDFS Erasure Code使用

Erasure Code的执行原理如下图所示：



1. 相关概念

* Code Policy，一个文件的校验块、数据块数目及cell大小，格式:RX-6-3-1024K
* Cell，校验计算时以cell为单位，默认4KB

1. EC相关配置：

|  |  |
| --- | --- |
| 参数 | 描述 |
| io.erasurecode.codec.rs.rawcoders | 默认：rs\_native,rs\_java |
| io.erasurecode.codec.rs-legacy.awcoders | 默认：rs-legacy\_java |
| io.erasurecode.codec.xor.rawcoders | 默认：xor\_native,xor\_java |

|  |  |
| --- | --- |
| 参数 | 描述 |
| dfs.namenode.ec.system.default.policy | 默认Code Policies，分别对应文件的数据块-校验块-cell大小，默认：RS-6-3-1024K，文件被分为6个数据块、3个校验快及Cell大小为1024K |
| dfs.namenode.ec.policies.max.cellsize | Cell的最大值，默认4194304，4MB |
| dfs.datanode.ec.reconstruction.stripedread.  buffer.size | 读数据的缓存大小，默认64KB |
| dfs.datanode.ec.reconstruction.threads | DN中用于EC Block的重建，默认8个线程 |
| dfs.datanode.ec.reconstruction.xmits.weight | 副本恢复和EC恢复的相对权重，默认05 |

1. 命令

*hdfs ec [generic options]*

*[-setPolicy -path <path> [-policy <policyName>] [-replicate]]*

*[-getPolicy -path <path>]*

*[-unsetPolicy -path <path>]*

*[-listPolicies]*

*[-addPolicies -policyFile <file>]*

*[-listCodecs]*

*[-enablePolicy -policy <policyName>]*

*[-disablePolicy -policy <policyName>]*

*[-help [cmd ...]]*

1. 测试(hadoop 3.2)

*$ hdfs ec -listCodecs //编解码器*

*Erasure Coding Codecs: Codec [Coder List]*

*RS [RS\_NATIVE, RS\_JAVA]*

*RS-LEGACY [RS-LEGACY\_JAVA]*

*XOR [XOR\_NATIVE, XOR\_JAVA]*

*$ hdfs ec -listPolicies*

*ErasureCodingPolicy=[Name=RS-10-4-1024k, Schema=[..., Id=5], State=DISABLED*

*ErasureCodingPolicy=[Name=RS-3-2-1024k, Schema=[..., Id=2], State=DISABLED*

*ErasureCodingPolicy=[Name=RS-6-3-1024k, Schema=[..., Id=1], State=ENABLED //默认*

*ErasureCodingPolicy=[Name=RS-LEGACY-6-3-1024k, Schema=[..., Id=3], State=DISABLED*

*ErasureCodingPolicy=[Name=XOR-2-1-1024k, Schema=[...., Id=4], State=DISABLE*

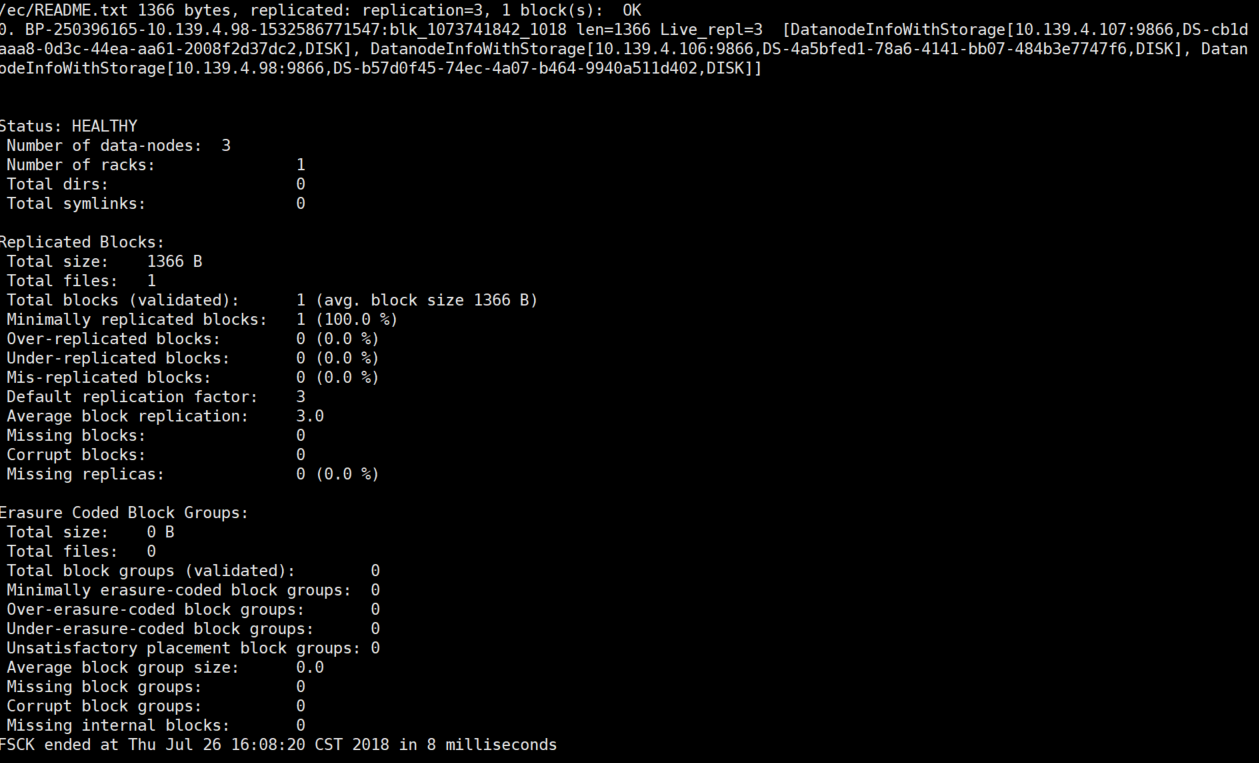
*$ hdfs ec -enablePolicy -policy XOR-2-1-1024k*

......

*ErasureCodingPolicy=[Name=XOR-2-1-1024k, Schema=[...., Id=4], State=ENABLED //已启用*

* *查看待处理文件的block信息*

*hdfs fsck /ec/README.txt -files -blocks -locations*

**

* 设置文件为EC Policy，[只能设置目录]

*hdfs ec -setPolicy -path /ec -policy XOR-2-1-1024k*

查看目录Policy如下：

*$ hdfs ec -getPolicy -path /ec*

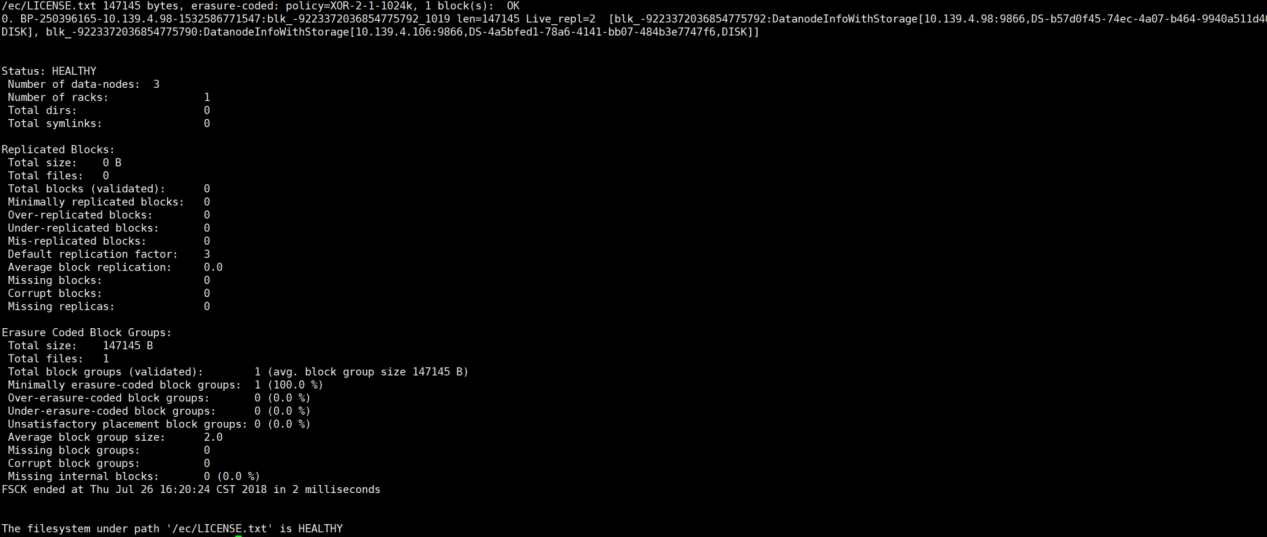
*XOR-2-1-1024k*

查看ec中的文件，文件block没有发生变化，EC仅对新加入的文件进行操作

* 文件操作

*hdfs dfs -put LICENSE.txt /ec*

查看文件block信息



文件存储为ec block，文件存储成功。