



Design Document : Worker

Class

Tuple

- Key, value로 이루어진 tuple을 표현하는 클래스
- Key, value 값과 다른 타입으로의 변환해주는 메소드를 가진다.

Members

- key: Int
- value: List[Byte]
- fromBytes(List[Byte]): Tuple
 - Tuple의 companion object에 정의된 construct method이다.
- toStr: Str
 - 튜플을 string 형식으로 보기 좋게 바꿔주는 함수이다. 디버깅 목적으로 사용한다.
- <(other: Tuple)
 - Tuple의 key를 통해 대소를 비교하는 함수이다.

Block

- 각 worker에 저장된 블록(파일)을 표현하는 클래스
- 파일이름, 디렉토리, 튜플 개수 등의 정보를 포함한다.

Members

- val dir: String
- val temp_dir: String
 - Block을 sample, sort, divide 한 이후에 생기는 임시 파일들을 저장할 디렉토리이다.

- val filename: String
- val key_range: (Int, Int)
 - Partition에 따라 나눈 이후 블록이 포함하고 있는 key 값의 범위를 나타낸다.
- val file_size: Int
- val num_tuples: Int
- def toStream: Stream[Tuple]
- def toList: List[Tuple]
- def sorted: List[Tuple]
- def sample(sample_size: Int): List[Tuple]
- def divideByPartition(partition: List[Int]) : List[Block]
 - 주의: 반환된 list가 key range에 따라 정렬되어있을 것이라는 보장이 없다.
- def sortThenSaveTo(dst: Block): Unit

Worker Function

Initialization

- def initializeBlocks(file_name_list: List[String]): List[Block]

Sample

- def sampleFromBlocks(block_list: List[Block], sample_size: Int): List[Tuple]
- def sampleFromBlock(src: Block, sample_size: Int): List[Tuple]

Sort & Partitioning

- def sortBlocks(block_list: List[Block]): List[Block]
- def divideBlocksByPartition(blocks: List[Block], partition: List[Int]): List[Block]
 - 주의: 반환된 list가 key range에 따라 정렬되어있을 것이라는 보장이 없다.

Block Transfer

- def sendBlockTo(src: Block, ip: String): Unit
- def recvBlockFrom(ip: String): Block

Merge

- `def merge(input_blocks: List[Block]) : List[Block]`