

数据集格式

block数据

文件: block\_part1.csv ( 逗号分隔的csv文件 · 数据包含表头 )

csv格式：

名字	数据类型	说明
blockID	int	区块编号
hash	varchar(64)	该区块的hash值
block_timestamp	long int	unix时间戳 · 在此次实验 · 用该值代表该区块上的所有交易的时间

csv示例数据：

blockID	hash	block_timestamp
92094	00000000000ab2bab31eebdc15927c12917b1ce7a4cef80b0e81ecd22844bcac	1289874960

transaction数据

文件：tx\_data\_part1.csv ( 逗号分隔的csv文件 · 数据包含表头 )

csv格式：

名字	数据类型	说明
tx_id	long int	交易编号
blockID	int	该交易所属的区块id · 区块的更多信息查看block数据
from	varchar(34)	交易的转出方 · 其中有些账号的长度不足34 · 最少为31位 · 最大34位
amout	double	交易的金额 · 最小金额为0.0001
to	varchar(34)	交易的转出方 · 其中有些账号的长度不足34 · 最少为33位 · 最大34位

csv示例数据：

tx_id	blockID	from	amout	to
639408	129047	1KBfnLThzVLM32pURdBZHgx3xU33UNb3bR	1.07	1A9yC99DzbjvEN2LWU35yn4e7AriMJ46HZ

补充说明

- 1. tx\_data\_part1\_v1.csv把原来的tx\_data\_part1.csv的tx\_id重新编号 · 从0开始
- 2. tx\_data\_part2.csv是任务4-1的测试数据集 · 格式与tx\_data\_part1.csv保持一致 · 新增数据集一共有30万条交易记录 · txid从1,000,000开始编号 · 不产生新的区块&账号
- 3. 在处理所有的csv文件时 · 注意每一行末尾的"\n"或者"\r\n"