CSCI-B 649 Engineering Cloud Computing

Assignment 2: Implementing MapReduce

server.py

The server.py script implements the key-value store server. I have reused code from last assignment to implement this. The key-value store uses TCP socket to communicate with clients. It allows multiple client to perform get operation simultaneously, but any one client can perform write operation on the key-value store at a time. The key-value store server provides the following functionalities:

- set: Takes 3 arguments, the unique key that is to be used to store data, the length of data and the data which is to be stored on the key-value store.
- get: Takes 1 argument, the unique key that was provided to store data on the server. The same key is used to retrieve the data by clients using get method.
- append: Takes 3 argument, the key of data block to which new data is to be added, the length of new data and the new data block (only supported for objects of class list for now).

The key-value server uses a dictionary to store data, which is serialized using pickle, and stored on the disk for persistence.

The data sent to key-value store is serialized using pickle by clients which is then unserialized at the server. When a client fetches data from using get request, the data is serialized before sending over network which is then unserialized by the client. This mechanism ensures that clients can store and retrieve any valid python object that is serializable.

Note: The key-value store is not started by the main program and hence must be launched separately by running the server.py script.

config.json

This is the configuration file that stores the configuration used by the master and main programs for their working. The name of this file is passed as command line argument to the main program and it must be stored in the same directory as the main program. The configuration file stores the following information:

- main: The name of main file and address where it is executing
- config: The name of configuration file
- mapper: The name of file that contains code for map task
- reducer: The name of file that contains code for reduce task
- store: The name of file executing key-value store server. It also stores, the address where the server is executing and the port number which the server is listening to
- input_data: The pickle file that contains the input data and key that is to be used to store the input data on key-value store server

main.py

This is the driver program that launches master and consumes its services to complete a task using map-reduce algorithm. The driver program has following important tasks:

- It logs all the major events that happen during the execution
- It stores the configuration file and input data on key-value store that will later be utilized by master for its working.
- It sends the mapper.py and reducer.py scripts, implementing map and reduce functions respectively, along with master.py, implementing the master program, to the node executing master program, if required.
- It launches the master program on local machine or remote host and consumes its services using RPCs.
- It prints and logs the output of map-reduce

master.py

This is the master program that implements map-reduce as a library and exposes API to allow programs to interact with it and consume its services. The master program is launched by the main program and it executes till it is issued a command to shut down. The master creates a SimpleXMLRPCServer object to allow other programs to consume its services using RPCs. It is provided 3 command line arguments when it is launched by main program:

- The key that was used to store the configuration file by driver program
- The address where key-value server is running
- The port at which the key-value store is listening

The master program provides the following three API calls:

- init_cluster: This method is called by driver program for following tasks:
    - connecting to key-value store
    - retrieving configuration file from key-value store
    - provisioning gcloud instances for worker nodes, if the addresses where worker nodes are to be executed is different from that of node where master is executing
    - retrieving input data from key-value store, partitioning into almost equal parts according to number of workers, and storing each partition on key-value store which are later used by worker nodes as input
    - making sure that the key "intermediate_data" contains empty list which will later be populated with intermediate data by worker nodes
- runmapred: This is the heart of master and executes map-reduce tasks. This method has the following flow:
    - launch a new thread, for each worker, which will create a process to start the worker that will execute the map task specified in mapper.py. The worker executes till it either finishes or throws an error. If all the workers have successfully finished map tasks, the method moves to launch a worker for reduce task implemented in reducer.py. If not, it

calls the destroy method to stop the master and report the error to driver program. *Note: The use of multithreading is to ensure that the method won't launch a worker for reduce task until all the workers implementing map task have finished. Each thread creates a new process to launch the worker and hence this program can be extended to execute workers on remote machines.*

- destroy: This method is provided by master for cleanup operations. The method call performs the following cleanups:
  - close the connection of master to key-value store
  - destroy gcloud instances, if provisioned for worker nodes
  - shutdown the SimpleXMLRPCServer and close the master program

All the important events that happen at master are sent to driver program to be included in the log file.

Note: I have not implemented fault tolerance right now and all the worker nodes that are to be used for map-reduce task must be specified in the configuration file beforehand.

mapper.py

The mapper script contains the code to execute the map task. It receives the following input, as command line arguments, from master when the master launches a worker node to execute map tasks:

- The address where key-value server is running and the port at which the server is listening
- The key of chunk that is to be used for map task
- The key which is to be used to store the intermediate data at key-value store

The mapper program fetches a chunk of input data, assigned to the worker implementing map task, from key-value server, performs map operation on it to create a list of key, value pairs and stores this list on the key-value server. It exits silently on successful completion otherwise it will throw an exception that is caught by master to realize that the worker failed to execute map task.

reducer.py

The mapper script contains the code to execute the reduce task. It receives the following input, as command line arguments, from master when the master launches a worker node to execute the reduce task:

- The address where key-value server is running and the port at which the server is listening
- The key which is to be used to store the intermediate data at key-value store
- The key that is to be used to store the result of reduce task at key-value server

The reduce program fetches the intermediate data, that is populated by all the workers implementing map tasks, from key-value server, performs reduce operation on it to create a dictionary of output key-value pairs and stores this dictionary on the key-value server. It exits silently on successful completion otherwise it will throw an exception that is caught by master to realize that the worker failed to execute reduce task.

Note: I have not implemented group by and sort operations. I am also using a single worker to implement the reduce task instead of using multiple workers.

Note: I ran the entire thing for the tasks of word count and inverted index. The logs and final state of key-value store are provided in folders indicating the tasks

Task: Word Count

Input: Emma (novel) by Jane Austen from nltk Gutenberg corpus as raw string. The following cleanup operations were performed:

- Removing of punctuations except "'" (apostrophe)
- Lowering all the words

Intermediate data: A list of words and their counts in individual chunks

Final output: A dictionary of words as keys and their total count in the entire input text as each words corresponding value

Task: Inverted Index

Input: Emma (novel), Persuasion (novel) and Sense and Sensibility (novel) by Jane Austen from NLTK Gutenberg corpus stored as dictionary with shortened name of novel as key and the content as value. The following cleanup operations were performed:

- Removing of punctuations except "'" (apostrophe)
- Lowering all the words

Intermediate data: A list of words with the name of novel in which it appears and frequency of the word in that novel's text.

Final output: A dictionary of words and a list of (novel, frequency in novel) tuples.

Word Count

```
amitsyadav@DESKTOP-H1P4CCM:~$ python3 main.py config.json
main - INFO - Output: {'stockings': 1, 'jane': 271, 'flesh': 1, 'opposed': 4, 'perturbation': 4, 'cavil': 1, 'subordination': 1, 'acceptable': 8, 'restrained': 2, 'across': 7, 'matchi': 1, 'newsi': 1, 'itvery':
2, 'refreshments': 1, 'possession': 5, 'woodhousefor': 1, 'obligation': 5, 'regards': 2, 'reprehensible': 1, 'improvement': 9, 'hoped': 42, 'among': 51, 'distinctly': 7, "o'clock": 7, 'seeming': 10, 'discern': 3
, 'uncertainty': 4, 'bench': 1, 'younodding': 1, 'pronounce': 2, 'fare': 1, 'force': 6, 'company': 32, 'stars': 1, 'attraction': 8, "ballroom's": 1, 'amends': 4, 'occurrence': 2, 'traits': 1, 'road': 17, 'happen
s': 1, 'law': 5, 'augusta"': 1, 'talentsi': 1, 'mercantile': 1, 'weddingpresentso': 1, 'defence': 2, 'miserably': 3, 'downstairs': 4, 'perceiving': 3, 'frightened': 12, 'gratitude': 29, 'ballroom': 5, 'safest':
3, 'either': 62, 'lately': 20, 'probably': 33, 'unequal': 7, 'correspond': 1, 'dancingnot': 2, 'creep': 1, "housekeeper's": 1, 'advantage': 27, 'uncertaintyat': 1, 'measure': 7, 'narrower': 1, 'complaint': 7, 'm
anners': 52, 'enoughto': 1, 'cake': 3, 'opposite': 7, 'bloom': 4, 'belovedwhich': 1, 'prevails': 1, 'pique': 1, 'unconcern': 1, 'employment': 10, 'interfered': 1, 'aired': 1, 'unwholesome': 5, 'served': 1, 'sure
': 200, 'cousin': 2, 'improves': 1, 'lasted': 3, 'speculation': 5, 'astonishingly': 1, 'rides': 1, 'delicacy': 16, 'accomplished': 11, 'moments': 11, 'famous': 3, 'alland': 3, 'notice': 24, 'intent': 3, 'emma's"
: 77, 'respectable': 20, 'recommend': 20, "body's": 29, 'entertainment': 6, 'repugnance': 1, 'reproof': 3, 'all': 831, 'addressing': 9, 'bravado': 1, 'knightleywhat': 1, 'just': 162, 'sensible': 24, 'spirits': 6
4, 'rude': 5, 'but': 1251, 'rush': 2, 'clever': 26, 'introductionof': 1, 'engross': 1, 'society': 47, 'call': 42, 'neighbourhood': 13, 'creating': 6, 'imprudence': 3, "brother's": 6, 'shopor': 1, 'itthere': 2, '
retaining': 1, 'begged': 10, 'inquiry': 12, 'strictly': 2, 'minute': 15, 'whatever': 18, 'intentions': 9, 'instrumental': 1, 'chamber': 1, 'reluctant': 1, 'homewards': 1, 'precedent': 1, 'put': 62, 'feeling': 54
, 'sounds': 3, 'tooknowing': 1, 'treatment': 6, 'courtship': 2, 'decline': 2, 'turn': 28, 'sense': 55, 'sluggish': 1, 'befriend': 1, 'wasa': 1, 'choseand': 1, 'hernothing': 1, 'ceased': 6, 'weekly': 1, 'admirati
on': 24, 'towards': 66, 'returning': 16, 'since': 64, 'changeable': 1, 'impertinently': 1, 'constant': 6, 'courage': 6, 'stopping': 3, 'feelingit': 1, 'hair': 12, 'properlythere': 1, 'younger': 2, 'quaint': 1, '
discreet': 3, 'different': 39, 'ample': 6, 'thingit': 2, 'expenses': 1, 'events': 3, 'trade': 4, 'linendraper': 2, 'voted': 1, 'elegant': 25, 'impolite': 1, 'outwardly': 1, 'confidence': 17, 'appreciate': 1, 'of
ten': 91, 'new': 35, 'suddenness': 2, 'female': 7, 'briskmoving': 1, 'notbut': 2, 'carriagewho': 1, 'permit': 1, 'lay': 11, 'try': 22, 'object': 36, 'guests': 7, 'welllooking': 2, 'partner': 10, 'youso': 1, 'sid
ebut': 1, 'ways': 15, 'dowe': 1, 'enscombe': 35, 'occupied': 15, 'preferable': 2, 'sir"': 1, 'separation': 4, 'vainby': 1, 'closet': 1, 'foolish': 9, 'playyes': 1, 'sick': 9, 'persevered': 2, 'impressed': 2, 'ex
ercise': 17, 'propensity': 1, 'liberty': 9, 'everbut': 1, 'xviii': 3, 'wellbehaved': 1, 'mere': 16, 'recollected': 3, 'arrange': 3, 'restored': 5, 'performer': 1, 'is': 1230, 'sposo': 4, 'allies': 1, 'behaviour'
: 23, 'deserved': 10, 'above': 12, 'progress': 9, 'awayparticularly': 1, 'protection': 2, 'saffron': 1, 'account': 55, 'dulness': 3, 'counter': 1, 'master': 13, 'grew': 10, 'efforts': 1, 'white': 7, 'amazing': 3
, 'stories': 1, 'ballwhy': 1, 'eltonjane': 1, 'aunt': 51, 'effect': 17, 'action': 6, 'dangers': 2, 'feelingthere': 1, 'absentshe': 1, 'intimate': 8, 'instrumentthough': 1, 'many': 137, 'applications': 3, 'espali
er': 1, 'christmas': 10, 'arm': 12, 'herno': 1, 'hands': 22, 'voices': 4, 'charity': 6, 'cutting': 5, 'itcertainly': 1, 'provoke': 1, 'unexpectedly': 1, 'richmond': 13, 'retract': 3, 'table': 23, 'apart': 3, 'in
sensibility': 2, 'matchmaking': 6, 'now': 300, 'ran': 9, 'particulars': 12, 'popularity': 4, 'testifies': 1, 'knows': 30, 'adieus': 1, 'patient': 3, 'visibly': 2, 'glibly': 1, 'sanguinely': 1, 'young': 190, 'the
re': 518, 'remaining': 3, 'some': 262, 'himand': 3, 'attacked': 2, 'forgotten': 18, 'crisis': 4, 'once': 80, 'secret': 22, 'answer': 63, 'starting': 1, 'request': 4, 'shamefulof': 1, 'appledumpling': 1, 'drew':
6, 'becoming': 12, 'shook': 8, 'restraints': 2, 'extreme': 11, 'word': 87, 'lamp': 2, 'flowed': 1, 'pleasantly': 8, 'extremelythought': 1, 'advised': 4, 'silenced': 3, 'took': 47, 'harshly': 1, 'cards': 4, 'himt
o': 1, 'practice': 3, 'mediocre': 1, 'deceive': 3, 'contemplate': 1, 'floated': 1, 'restraint': 3, 'nearly': 14, 'aware': 25, 'trifling': 8, 'specimen': 1, 'avoid': 13, 'hundred': 14, 'suffered': 11, 'weymouth':
17, 'barnes': 1, 'worn': 2, 'use': 37, 'views': 17, 'halfcircle': 1, 'tomorrow': 21, 'crowded': 2, 'by': 567, 'evermrs': 1, 'excessively': 10, 'convenient': 5, 'likeliest': 1, 'sightoh': 1, "cole's": 12, 'artle
ss': 4, 'checking': 3, 'sign': 3, 'pacing': 1, 'considers': 4, 'eminent': 1, 'soliloquy': 1, 'catching': 4, 'perform': 2, 'drawing': 13, 'itat': 1, 'churchill': 188, 'unhappiness': 2, 'stairs': 4, 'second': 29,
'allnot': 1, 'happy': 120, 'upstarts': 1, 'shower': 1, 'amid': 1, 'nephew': 12, 'somebody': 15, 'accrue': 1, 'advantages': 12, 'entrapped': 1, 'proprieties': 1, 'arthurhow': 1, 'saidbut': 1, 'friendsaid': 1, 'em
ma': 750, 'purchased': 1, 'consideration': 23, 'disposed': 19, 'hearing': 31, 'twice': 6, 'disappoint': 3, 'stuck': 1, 'merchant': 1, 'prevailed': 6, 'polite': 2, 'reasons': 2, 'speeches': 4, 'hindquarter': 2, '
lookout': 2, 'sweetness': 6, 'uncivil': 1, 'birmingham': 2, 'sincere': 5, 'bottom': 3, 'rationality': 1, 'triumph': 5, 'sensations': 13, 'consciously': 1, 'viii': 3, 'prefer': 6, 'awkwardness': 5, 'dialogues': 1
, 'reigning': 1, 'event': 22, 'finger': 3, 'bearers': 1, 'water': 6, 'dixonvery': 1, 'carriages': 11, 'attract': 1, 'whose': 39, 'youno': 1, 'didbut': 1, 'press': 3, 'shoulder': 1, 'does': 127, 'likelihood': 2,
'fainted': 2, 'change': 60, 'magnified': 1, 'married': 45, 'soundfor': 1, 'extravagant': 1, 'highburyto': 1, 'composedly': 2, 'pains': 10, 'palpably': 3, 'sensibilities': 1, 'westonswho': 1, 'doubted': 8, 'stric
t': 3, 'begging': 1, 'hawkins': 14, 'itelegance': 1, 'completely': 26, 'long': 142, 'occasionally': 5, 'judged': 11, 'stoppingit': 1, 'insuperable': 1, 'that': 1759, 'unpersuadable': 3, 'days': 47, 'largesized':
1, 'west': 1, 'caro': 4, 'settled': 39, 'barouchelandau': 7, 'today': 23, 'proofs': 4, 'stablesi': 1, 'roundly': 1, 'july': 2, 'catch': 14, 'entire': 8, 'consequent': 1, 'principles': 6, 'saying': 49, 'remained
': 16, 'became': 23, 'stronger': 8, 'fondly': 4, 'tomorrowand': 1, 'taste': 34, 'italian': 1, 'excessively"': 1, 'hastily': 6, 'plenty': 5, "elton'sprobably": 1, 'reverie': 2, 'greater': 30, 'standard': 1, 'allo
wable': 4, 'ofoh': 1, 'windowsi': 1, 'finery': 5, 'separated': 3, 'might': 325, 'pleasantdid': 1, 'interview': 2, 'grandpapa': 3, 'ladies': 44, 'hasty': 5, 'professional': 1, 'ofthe': 2, 'discussion': 5, 'liked'
: 15, 'boast': 2, 'carefulness': 1, 'refusal': 5, 'beavers': 1, 'honest': 5, 'active': 7, 'kindly': 11, 'taking': 26, 'pass': 31, 'ourselvesoh': 1, 'sends': 3, 'slighted': 6, 'wine': 6, 'admitted': 12, 'praise':
31, 'moderation': 2, 'coachman': 4, 'exceedingly': 11, 'six': 11, 'eltonshere': 1, 'entitle': 1, 'blunder': 8, 'glancing': 5, 'resolution': 31, 'bore': 7, 'drove': 3, 'hours': 23, 'tones': 1, 'fatigue': 1, 'adv
isable': 2, 'holding': 1, 'worlds': 2, "wife's": 2, 'elton': 317, 'quickness': 8, 'dissipation': 1, 'blessing': 12, 'invited': 15, 'my': 718, 'hitherthe': 1, 'tolerable': 8, 'wallis': 3, 'dinners': 2, 'townmrs':
1, 'tears': 8, 'fancying': 13, 'shut': 9, 'boy': 12, 'dies': 1, 'although': 1, 'ladya': 1, 'disposedhad': 1, 'achievement': 1, 'bring': 38, 'truewell': 1, 'decorated': 1, 'intention': 7, 'interruption': 2, 'int
ending': 4, 'lip': 2, 'notions': 6, 'pictures': 2, 'flower': 1, 'talking': 52, 'easy': 27, 'whist': 2, 'lightened': 2, 'beforeevery': 1, 'wise': 6, 'likely': 45, 'popular': 1, 'gracious': 3, 'contrast': 5, 'degr
ee': 55, 'native': 5, 'wins': 1, 'patroness': 2, 'exaggeration': 1, 'divined': 1, 'forerunner': 1, 'overtrimmedquite': 1, 'during': 17, 'successively': 1, 'girls': 12, 'leave': 55, 'witness': 4, 'negligent': 5,
'indolent': 1, 'remarks': 1, 'nowhere': 2, 'glow': 5, 'prodigy': 2, 'deprived': 1, 'whom': 72, 'handsomeshe': 1, 'commission': 5, 'reproached': 1, 'one': 428, 'single': 18, 'three': 65, 'debased': 1, 'wishedbut'
: 1, 'forlorn': 3, 'wretchedest': 1, 'irritate': 1, 'wants': 10, 'donwell': 44, 'illbred': 1, 'umbrella': 1, 'fireside': 4, 'will': 555, 'mailsit': 1, 'breathe': 3, 'comfortable': 34, 'retreated': 2, 'several':
8, 'comparative': 2, 'unexpected': 4, 'firstshe': 1, 'honourably': 1, 'enchanted': 1, 'clock': 1, 'inconvenience': 10, 'draughts': 2, 'compliance': 2, 'secondary': 2, 'fellowmourners': 1, 'milesnay': 1, 'wrap':
1, 'dear': 231, 'examining': 4, 'misfortune': 5, 'behindhand': 3, 'prejudices': 1, 'space': 4, 'blue': 4, 'disorder': 3, 'it': 2400, 'form': 18, 'waverings': 1, 'attachmentwhether': 1, 'liquor': 1, 'beneath': 3,
'alas': 4, 'resisting': 1, 'nothingmakes': 1, 'pleasure': 114, 'issued': 1, 'objectif': 1, 'barely': 5, 'momentbut': 2, 'understanding': 26, 'parcels': 3, 'whi': 1, 'thorough': 12, 'oppress': 1, 'pardon': 9, 'd
espair': 4, 'commend': 1, 'listen': 18, 'himself': 136, 'established': 5, 'colesworthy': 1, 'noticeshort': 1, 'wornout': 1, 'awayhe': 1, 'heron': 1, 'coldness': 7, "friend's": 11, 'seen': 73, 'ingratitude': 3, '
putting': 9, 'deari': 2, 'consumption': 2, 'flirtation': 2, 'wounded': 1, 'knew': 88, 'wreath': 1, 'suchand': 1, 'height': 6, 'says': 29, 'driven': 2, 'themthe': 2, 'amongst': 3, 'travel': 5, 'reflecting': 1, 'h
azard': 4, 'cause': 17, 'escape': 14, 'disclaiming': 1, 'impressedso': 1, 'woeful': 1, 'serious': 31, 'fix': 9, 'whoever': 3, 'merest': 1, 'forthi': 1, "herself'but": 1, 'explanations': 3, 'recur': 3, 'represent
ing': 3, 'looked': 91, 'persuasion': 11, 'coldfetching': 1, 'refuse': 15, 'todayfor': 1, 'distance': 24, 'smiling': 28, 'calmness': 6, 'yourselfto': 1, 'jealous': 3, 'friendliness': 3, 'guinea': 1, 'original': 5
```

Inverted Index

amitsyadav@DESKTOP-H1P4CCM: ~   – ☐ X

amitsyadav@DESKTOP-H1P4CCM:~$ python3 main.py config.json
main - INFO - Output: {'ascent': [('persuasion', 2)], 'selfishness': [('persuasion', 5), ('sense', 6), ('emma', 3)], 'lasting': [('persuasion', 3), ('sense', 5), ('emma', 5)], 'crime': [('persuasion', 1), ('sens
e', 3)], 'requested': [('persuasion', 2), ('sense', 3), ('emma', 1)], 'drawingrooms': [('persuasion', 4)], 'symptom': [('persuasion', 1), ('sense', 5)], 'syllable': [('persuasion', 1), ('sense', 7), ('emma', 10)
], 'transgressions': [('persuasion', 1)], 'september': [('persuasion', 4), ('sense', 2), ('emma', 5)], 'occur': [('persuasion', 5), ('sense', 7), ('emma', 10)], 'instantaneously': [('persuasion', 1), ('sense', 1
)], 'conceal': [('persuasion', 2), ('sense', 8), ('emma', 9)], 'asserted': [('persuasion', 2), ('sense', 1), ('emma', 2)], 'considering': [('persuasion', 11), ('sense', 7), ('emma', 20)], 'acquainted': [('persua
sion', 27), ('sense', 26), ('emma', 32)], 'passed': [('persuasion', 40), ('sense', 49), ('emma', 53)], 'step': [('persuasion', 3), ('sense', 2), ('emma', 11)], 'publicity': [('persuasion', 1), ('emma', 1)], 'dan
gerous': [('persuasion', 9), ('sense', 3), ('emma', 8)], 'justified': [('persuasion', 2), ('sense', 5), ('emma', 5)], 'page': [('persuasion', 2), ('sense', 1), ('emma', 2)], 'lighted': [('persuasion', 1), ('emma
', 1)], 'foremost': [('persuasion', 1)], 'bending': [('persuasion', 1), ('emma', 1)], 'disposition': [('persuasion', 7), ('sense', 21), ('emma', 23)], 'dropping': [('persuasion', 1)], "elizabeth's": [('persuasio
n', 7)], 'gig': [('persuasion', 2)], 'dinners': [('persuasion', 3), ('sense', 2), ('emma', 2)], 'kinds': [('persuasion', 1)], 'witness': [('persuasion', 1), ('sense', 5), ('emma', 4)], 'oddly': [('persuasion', 1
)], 'tenacious': [('persuasion', 1)], 'projecting': [('persuasion', 1), ('sense', 2)], 'resigned': [('persuasion', 1), ('sense', 2), ('emma', 2)], 'corpselike': [('persuasion', 1)], 'dropt': [('persuasion', 1),
('sense', 4), ('emma', 3)], 'thoroughfare': [('persuasion', 1)], 'fare': [('persuasion', 1), ('emma', 1)], 'not': [('persuasion', 933), ('sense', 1236), ('emma', 2120)], 'unknown': [('persuasion', 2), ('sense',
5), ('emma', 3)], 'explicit': [('persuasion', 1)], "farmer's": [('persuasion', 1), ('emma', 1)], 'dash': [('persuasion', 1)], 'concert': [('persuasion', 20), ('emma', 1)], 'adopt': [('persuasion', 1), ('sense',
6), ('emma', 4)], 'me': [('persuasion', 187), ('sense', 423), ('emma', 545)], 'prey': [('persuasion', 1)], 'ah': [('persuasion', 9), ('sense', 8), ('emma', 41)], 'viewed': [('persuasion', 1), ('sense', 1), ('emm
a', 2)], 'interchange': [('persuasion', 4), ('emma', 1)], 'attachments': [('persuasion', 2), ('sense', 3), ('emma', 1)], 'historian': [('persuasion', 1)], 'inaction': [('persuasion', 1)], 'rose': [('persuasion',
1), ('sense', 14), ('emma', 8)], 'midst': [('persuasion', 3), ('sense', 2), ('emma', 7)], 'holds': [('persuasion', 3), ('sense', 1), ('emma', 5)], 'urged': [('persuasion', 1), ('sense', 10), ('emma', 1)], 'guid
e': [('persuasion', 2), ('sense', 1), ('emma', 2)], 'oppression': [('persuasion', 1), ('sense', 3), ('emma', 1)], "edward's": [('persuasion', 2), ('sense', 43)], 'age': [('persuasion', 11), ('sense', 19), ('emma
', 21)], 'due': [('persuasion', 13), ('sense', 13), ('emma', 22)], 'dwell': [('persuasion', 3), ('sense', 1), ('emma', 4)], "misunderstood'": [('persuasion', 1)], 'reported': [('persuasion', 1)], 'extinguish': [
('persuasion', 1)], 'merit': [('persuasion', 4), ('sense', 13), ('emma', 13)], 'vacancies': [('persuasion', 1), ('emma', 1)], 'injuries': [('persuasion', 1), ('sense', 2), ('emma', 1)], 'prevail': [('persuasion'
, 1), ('sense', 5), ('emma', 4)], 'derived': [('persuasion', 1), ('sense', 2), ('emma', 1)], 'brightness': [('persuasion', 1), ('sense', 2)], 'sound': [('persuasion', 11), ('sense', 5), ('emma', 10)], 'approacha
ble': [('persuasion', 1)], 'solitariness': [('persuasion', 1)], 'of': [('persuasion', 2568), ('sense', 3569), ('emma', 4274)], 'shook': [('persuasion', 3), ('sense', 4), ('emma', 8)], 'illa': [('persuasion', 1)]
, 'season': [('persuasion', 4), ('sense', 5), ('emma', 6)], 'sake': [('persuasion', 13), ('sense', 32), ('emma', 23)], 'powerful': [('persuasion', 4), ('sense', 1), ('emma', 5)], 'advantage': [('persuasion', 22)
, ('sense', 25), ('emma', 27)], 'nothingness': [('persuasion', 2)], 'continue': [('persuasion', 1), ('sense', 4), ('emma', 11)], 'to': [('persuasion', 2782), ('sense', 4086), ('emma', 5149)], 'escaped': [('persu
asion', 4), ('sense', 7), ('emma', 7)], 'nursed': [('persuasion', 1), ('emma', 3)], 'next': [('persuasion', 42), ('sense', 51), ('emma', 62)], 'speedy': [('persuasion', 1), ('sense', 5), ('emma', 1)], 'smile': [
('persuasion', 13), ('sense', 23), ('emma', 42)], 'lose': [('persuasion', 7), ('sense', 4), ('emma', 19)], 'noisy': [('persuasion', 1), ('sense', 5), ('emma', 2)], 'prospects': [('persuasion', 3), ('sense', 4),
('emma', 8)], 'needed': [('persuasion', 1), ('sense', 6), ('emma', 2)], 'behaving': [('persuasion', 2), ('emma', 3)], 'privations': [('persuasion', 1), ('emma', 4)], 'depending': [('persuasion', 1), ('sense', 2)
, ('emma', 1)], 'isle': [('persuasion', 1)], 'test': [('persuasion', 1), ('sense', 1)], 'illjudged': [('persuasion', 1), ('sense', 3), ('emma', 2)], 'feelings': [('persuasion', 75), ('sense', 72), ('emma', 99)],
'canvass': [('persuasion', 1)], 'a': [('persuasion', 1591), ('sense', 2047), ('emma', 3073)], 'modesty': [('persuasion', 2), ('sense', 3), ('emma', 1)], 'rekindled': [('persuasion', 1)], 'custom': [('persuasion
', 2), ('emma', 3)], 'learning': [('persuasion', 3), ('sense', 2)], 'brightened': [('persuasion', 1), ('sense', 1), ('emma', 1)], 'smokes': [('persuasion', 1), ('sense', 1)], 'letter': [('persuasion', 33), ('sen
se', 74), ('emma', 102)], 'consulting': [('persuasion', 1), ('emma', 4)], 'autumnal': [('persuasion', 1)], 'penance': [('persuasion', 3), ('emma', 6)], 'compelled': [('persuasion', 1)], 'surname': [('persuasion'
, 1)], 'convenience': [('persuasion', 5), ('sense', 5), ('emma', 6)], 'proof': [('persuasion', 15), ('sense', 14), ('emma', 23)], "drop's": [('persuasion', 1)], 'higher': [('persuasion', 6), ('sense', 2), ('emma
', 10)], 'frequent': [('persuasion', 3), ('sense', 9), ('emma', 8)], 'improvement': [('persuasion', 7), ('sense', 8), ('emma', 9)], 'reason': [('persuasion', 21), ('sense', 53), ('emma', 40)], 'staring': [('pers
uasion', 2)], 'regret': [('persuasion', 13), ('sense', 22), ('emma', 26)], 'week': [('persuasion', 25), ('sense', 33), ('emma', 28)], 'stockings': [('persuasion', 1), ('sense', 1), ('emma', 1)], 'byron': [('pers
uasion', 2)], 'cottagers': [('persuasion', 1)], 'trusting': [('persuasion', 1), ('sense', 2), ('emma', 3)], "owners'": [('persuasion', 1)], 'navy': [('persuasion', 13), ('sense', 1)], 'street': [('persuasion', 2
6), ('sense', 47), ('emma', 7)], 'independence': [('persuasion', 6), ('sense', 7), ('emma', 10)], 'pacify': [('persuasion', 1)], 'upper': [('persuasion', 3), ('emma', 3)], 'fastidious': [('persuasion', 2), ('sen
se', 2), ('emma', 1)], 'accession': [('persuasion', 1), ('emma', 1)], 'rents': [('persuasion', 1)], 'violence': [('persuasion', 1), ('sense', 3), ('emma', 1)], 'objects': [('persuasion', 3), ('sense', 11), ('emm
a', 7)], 'labourers': [('persuasion', 1), ('emma', 1)], 'contemptible': [('persuasion', 1), ('sense', 1), ('emma', 3)], 'consulted': [('persuasion', 4), ('sense', 2), ('emma', 4)], "sixtytwo'": [('persuasion', 1
)], 'richer': [('persuasion', 3), ('sense', 2)], 'precipitate': [('persuasion', 1)], 'along': [('persuasion', 13), ('sense', 10), ('emma', 4)], 'repulsively': [('persuasion', 1)], 'words': [('persuasion', 23), (
'sense', 28), ('emma', 47)], 'sensible': [('persuasion', 27), ('sense', 13), ('emma', 24)], 'among': [('persuasion', 39), ('sense', 21), ('emma', 51)], 'grey': [('persuasion', 1), ('sense', 8), ('emma', 1)], 'du
tiful': [('persuasion', 2), ('emma', 1)], 'monkford': [('persuasion', 8)], 'probable': [('persuasion', 6), ('sense', 11), ('emma', 5)], 'enter': [('persuasion', 6), ('sense', 12), ('emma', 8)], 'intervals': [('p
ersuasion', 5), ('sense', 3), ('emma', 1)], 'reconcile': [('persuasion', 3), ('sense', 1), ('emma', 2)], 'well': [('persuasion', 145), ('sense', 208), ('emma', 351)], 'completion': [('persuasion', 1), ('sense',
1), ('emma', 1)], 'only': [('persuasion', 218), ('sense', 286), ('emma', 329)], 'christmas': [('persuasion', 6), ('sense', 2), ('emma', 10)], 'guided': [('persuasion', 3), ('sense', 5), ('emma', 6)], 'pitch': [(
'persuasion', 1), ('sense', 1), ('emma', 2)], 'dignity': [('persuasion', 11), ('sense', 4), ('emma', 5)], 'satisfactory': [('persuasion', 1), ('sense', 7), ('emma', 6)], 'privileges': [('persuasion', 2), ('emma'
, 1)], 'amongst': [('persuasion', 3), ('sense', 8), ('emma', 3)], 'enjoyed': [('persuasion', 3), ('sense', 4), ('emma', 6)], 'husbands': [('persuasion', 2), ('sense', 4), ('emma', 4)], 'dawdling': [('persuasion'
, 1), ('sense', 1), ('emma', 1)], 'mackenzie': [('persuasion', 1)], 'concluding': [('persuasion', 2), ('sense', 3), ('emma', 4)], 'firstrate': [('persuasion', 2), ('emma', 4)], 'cordially': [('persuasion', 1), (
'sense', 2), ('emma', 4)], 'incapable': [('persuasion', 1), ('sense', 4), ('emma', 2)], 'happier': [('persuasion', 5), ('sense', 7), ('emma', 10)], 'cheer': [('persuasion', 1), ('emma', 4)], 'relative': [('persu
asion', 3), ('sense', 4), ('emma', 5)], 'space': [('persuasion', 6), ('sense', 2), ('emma', 4)], 'passerby': [('persuasion', 1)], 'discussion': [('persuasion', 3), ('sense', 4), ('emma', 5)], 'dealings': [('pers
uasion', 1), ('emma', 2)], 'save': [('persuasion', 6), ('sense', 1), ('emma', 6)], 'resembling': [('persuasion', 1), ('sense', 1)], 'black': [('persuasion', 5), ('sense', 1), ('emma', 5)], 'contrast': [('persuas
ion', 1), ('sense', 3), ('emma', 5)], 'affection': [('persuasion', 9), ('sense', 78), ('emma', 49)], 'irish': [('persuasion', 2), ('emma', 4)], 'moderatesized': [('persuasion', 1), ('emma', 1)], 'simple': [('per
suasion', 2), ('sense', 17), ('emma', 17)], 'mistaken': [('persuasion', 9), ('sense', 20), ('emma', 21)], 'gift': [('persuasion', 1), ('sense', 4)], 'park': [('persuasion', 4), ('sense', 51), ('emma', 2)], 'sea'
: [('persuasion', 18), ('emma', 11)], 'seen': [('persuasion', 58), ('sense', 56), ('emma', 73)], 'forgetfulness': [('persuasion', 1)], 'connected': [('persuasion', 5), ('sense', 5), ('emma', 11)], 'boots': [('pe