

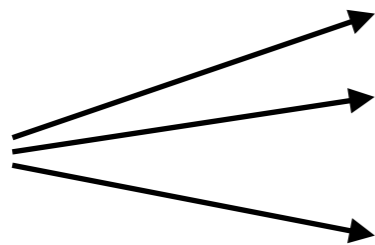
# Inputs



# Outputs

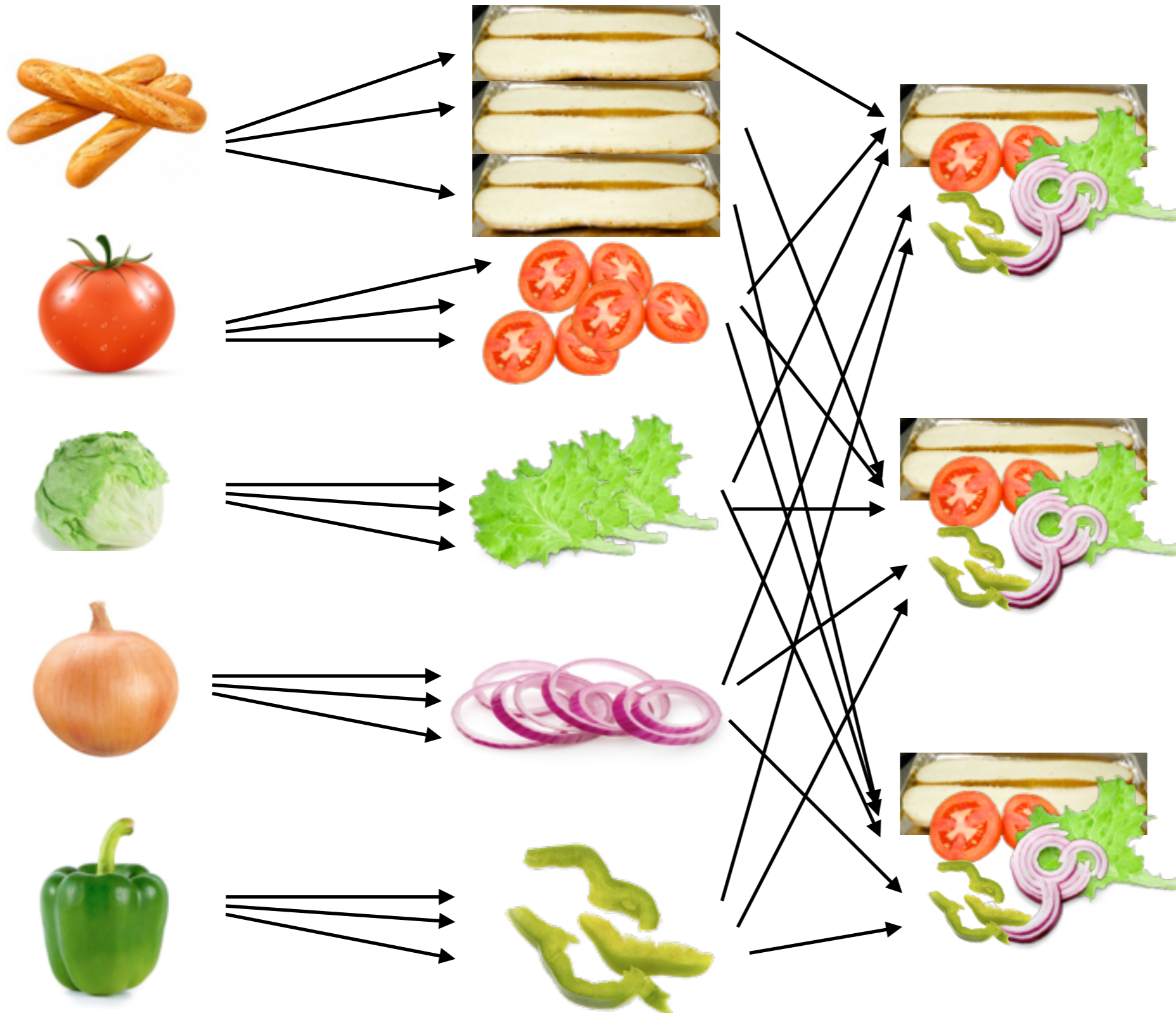


Map



Map

Shuffle

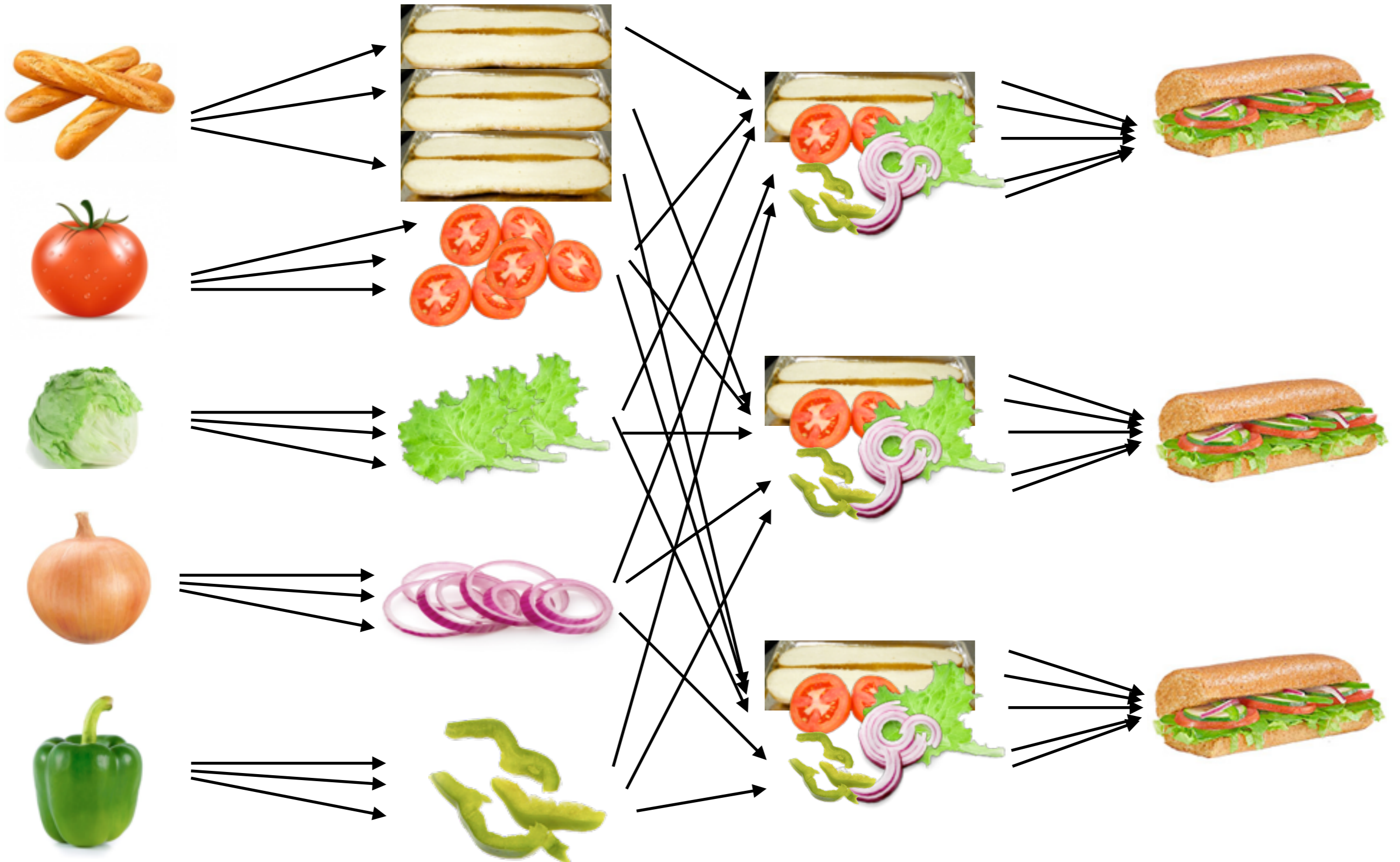




Map

Shuffle

Reduce



# Inputs

index.html

lecture-1.html

lab-3.html

midterms.html

# Outputs

CS110 index.html:1 lecture-1.html:1

filesystems lecture-1.html:1 midterms.html:1

processes lab-3.html:1 midterms.html:1

quicksort index.html:1

systems index.html:1

threads lab-3.html:1 midterms.html:1

unix lecture-1.html:1

Welcome index.html:1

# Map

index.html → Welcome index.html  
index.html → CS110 index.html  
index.html → systems index.html

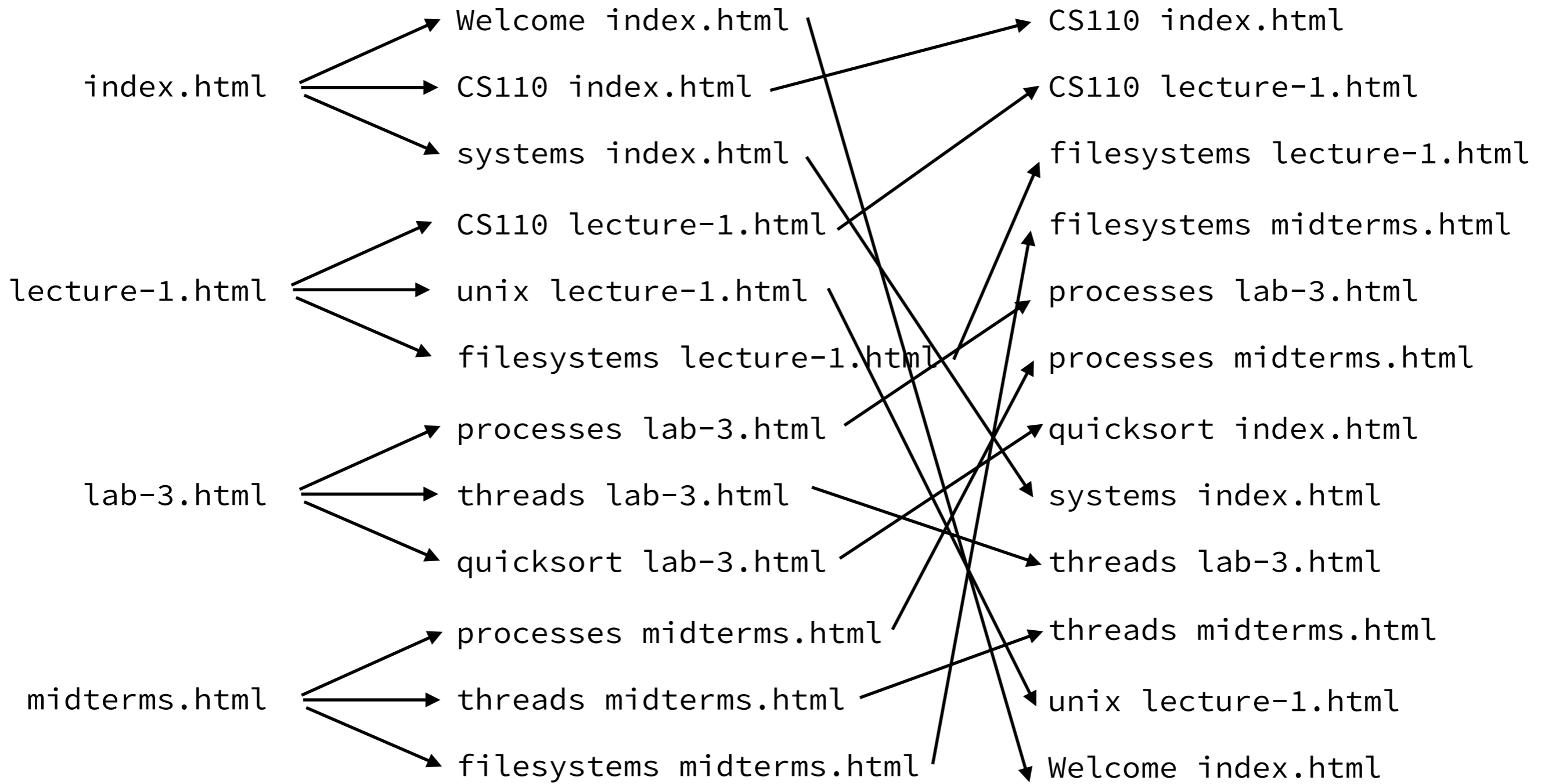
lecture-1.html → CS110 lecture-1.html  
lecture-1.html → unix lecture-1.html  
lecture-1.html → filesystems lecture-1.html

lab-3.html → processes lab-3.html  
lab-3.html → threads lab-3.html  
lab-3.html → quicksort lab-3.html

midterms.html → processes midterms.html  
midterms.html → threads midterms.html  
midterms.html → filesystems midterms.html

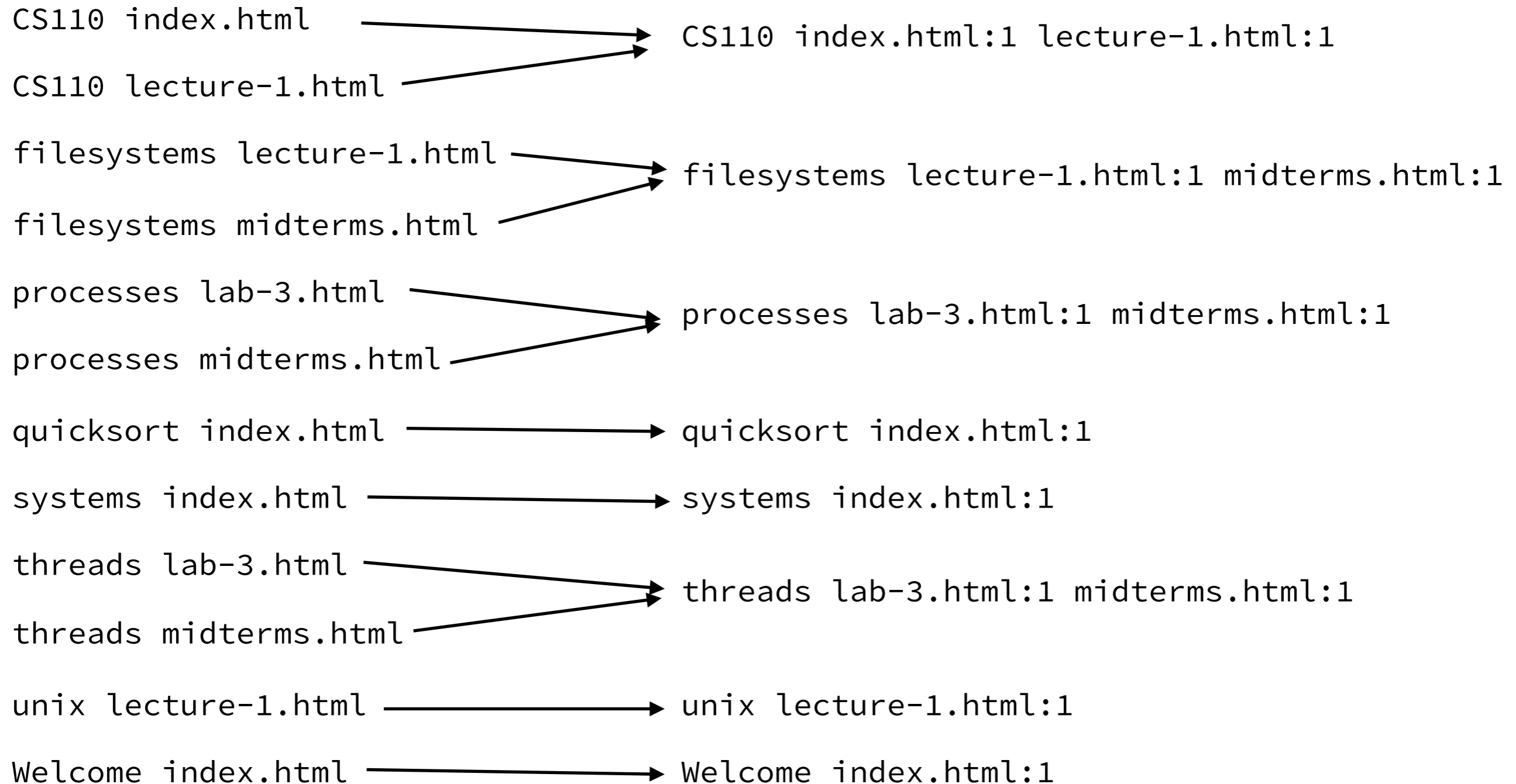
# Map

# Shuffle



# Reduce

(shuffle output)







# More MapReduce applications

- **Log analysis: which IP addresses are suspicious?**
  - Mapper: log line  $\rightarrow$  (IP : content accessed)
  - Reducer: (IP : all content accessed)  $\rightarrow$  probability of being malicious
- **Nearby gas stations**
  - Mapper: gas station  $\rightarrow$  (POI : distance to station)
  - Reducer: (POI : distance to all nearby stations)  $\rightarrow$  (POI : closest station)