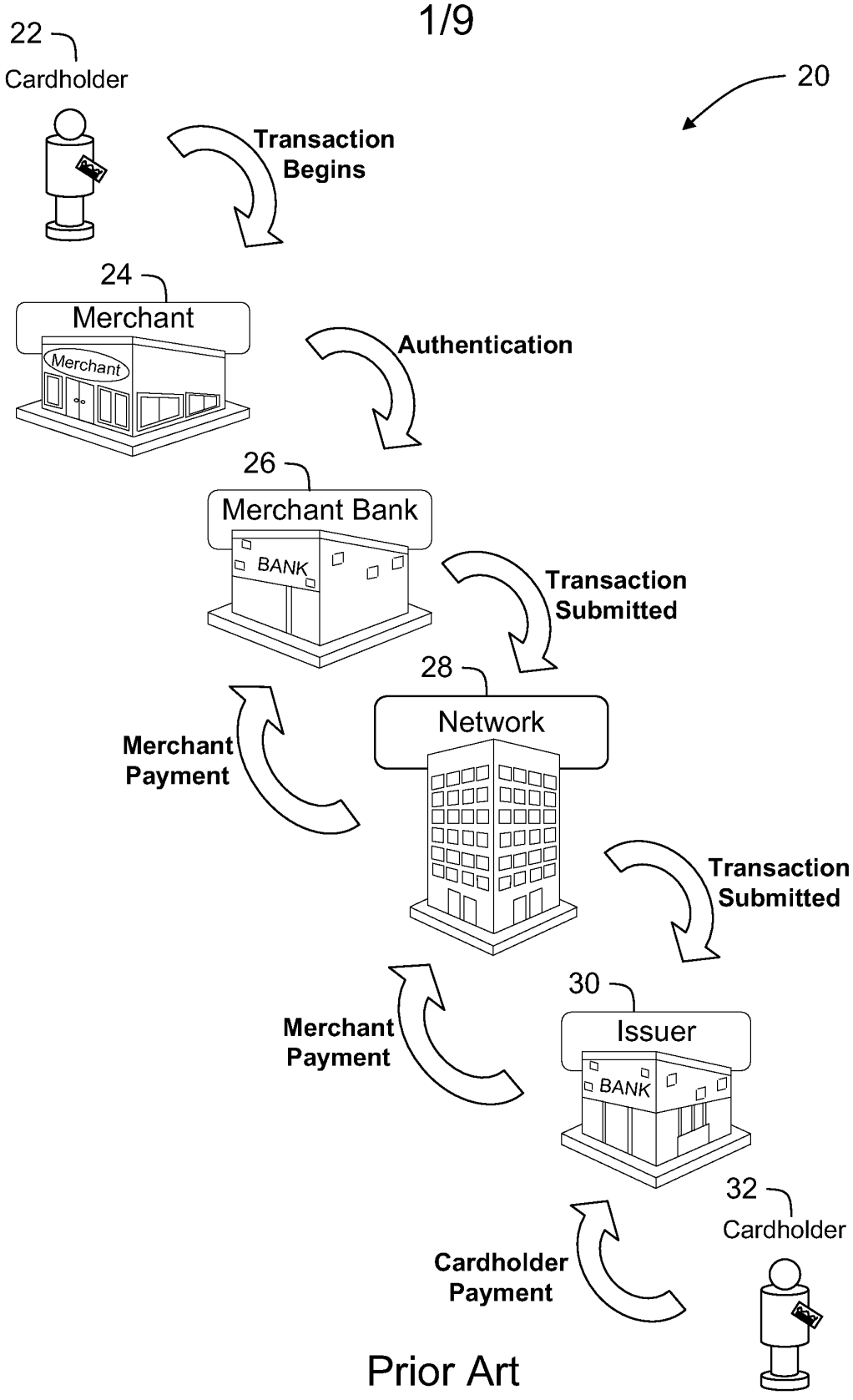


Prior Art

FIG. 1

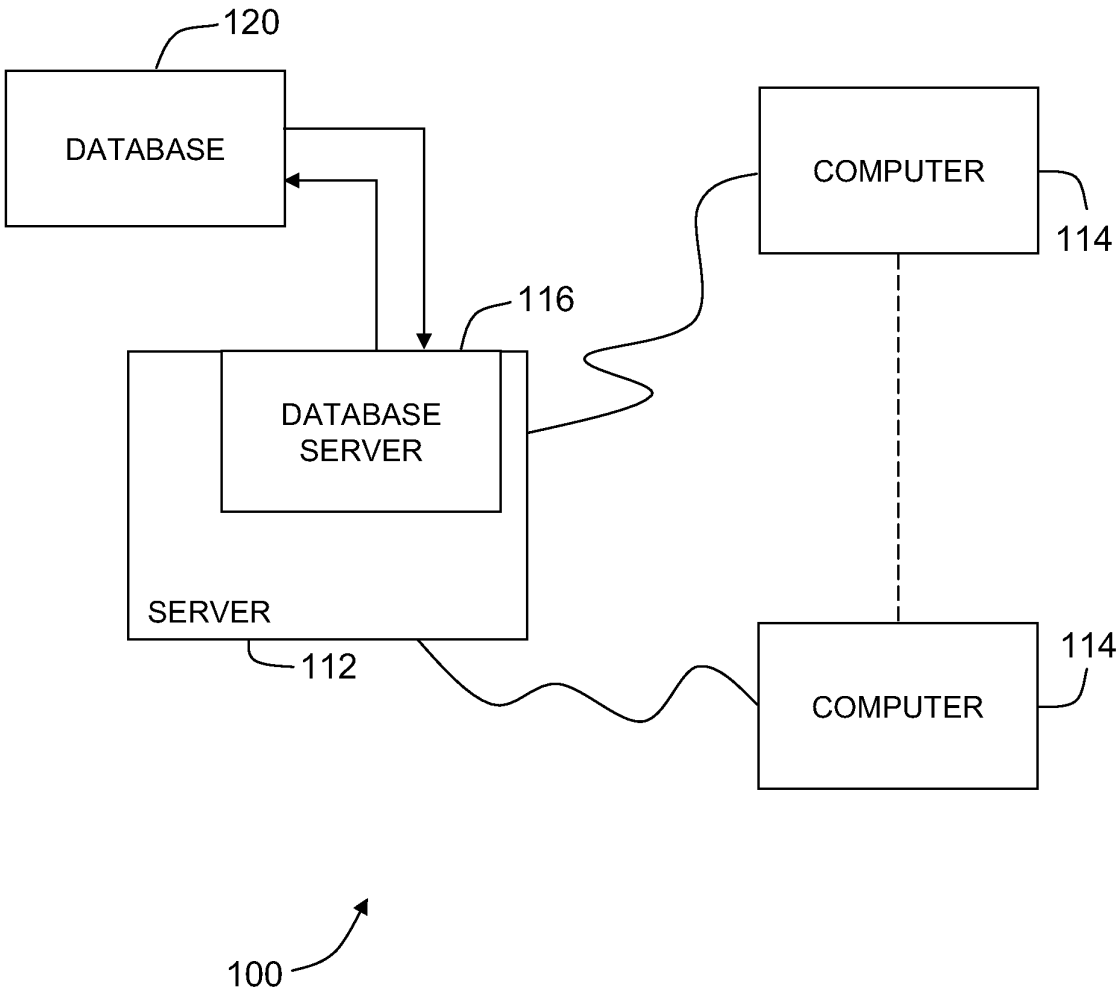

120
DATABASE
COMPUTER
114
116
DATABASE
SERVER
SERVER
112
114
COMPUTER
100

FIG. 2

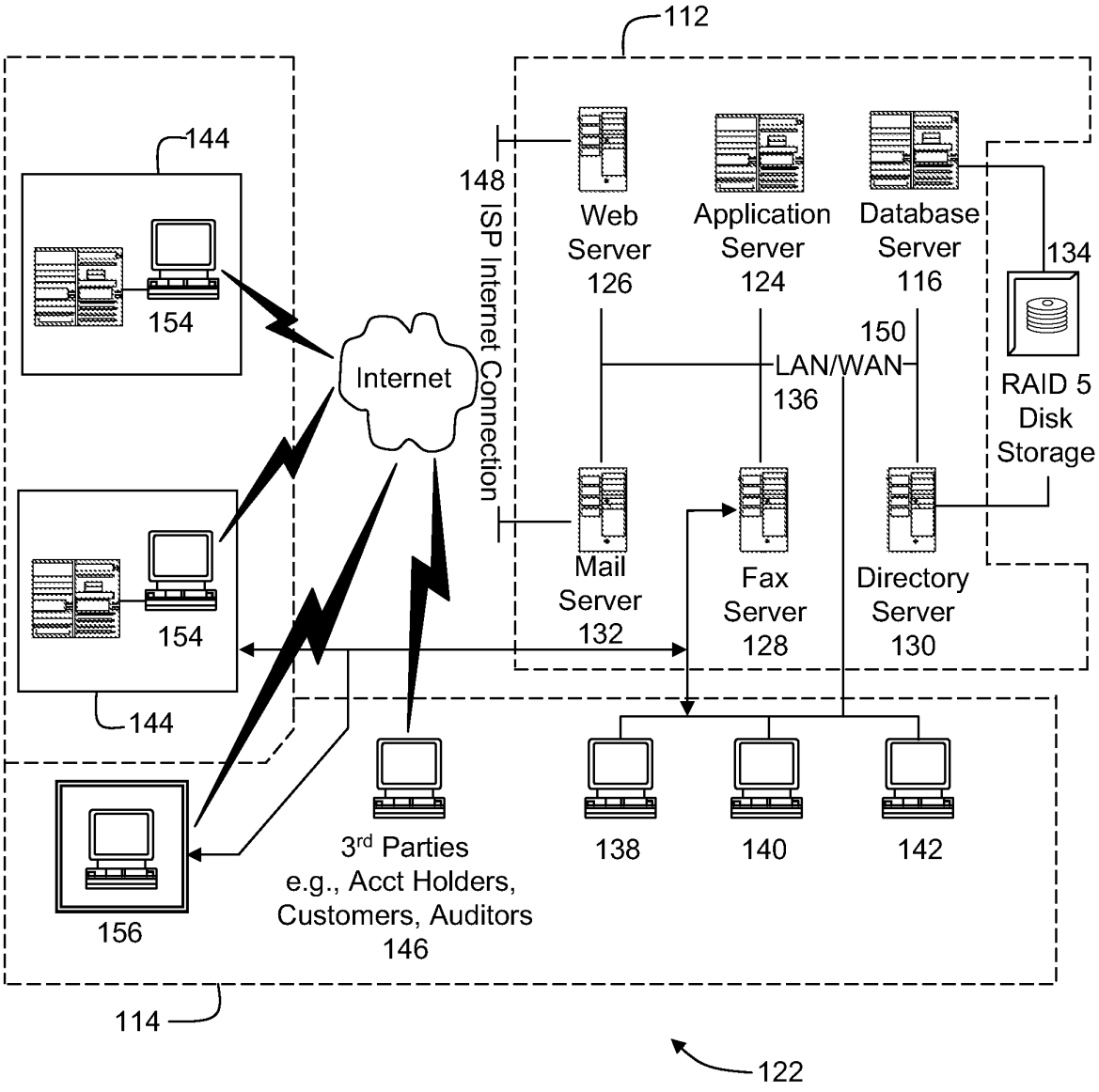

112
144
154
148
ISP Internet Connection
Web Server 126
Application Server 124
Database Server 116
134
RAID 5 Disk Storage
150
LAN/WAN 136
Internet
154
Mail Server 132
Fax Server 128
Directory Server 130
144
156
3rd Parties e.g., Acct Holders, Customers, Auditors 146
138
140
142
114
122

FIG. 3

High Level Ensemble Prediction Process Overview

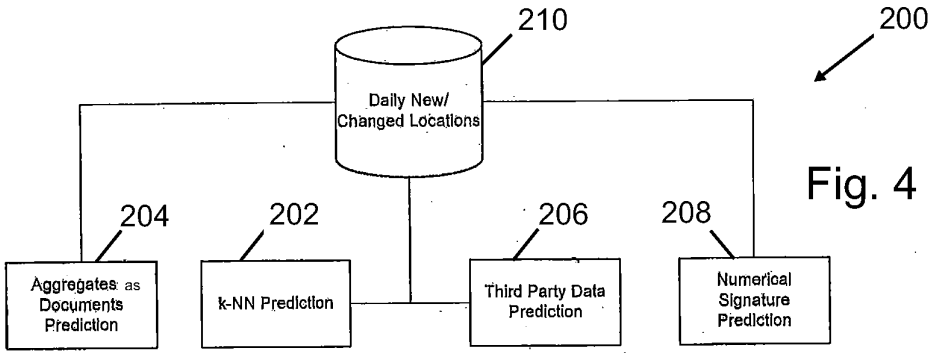


Fig. 4

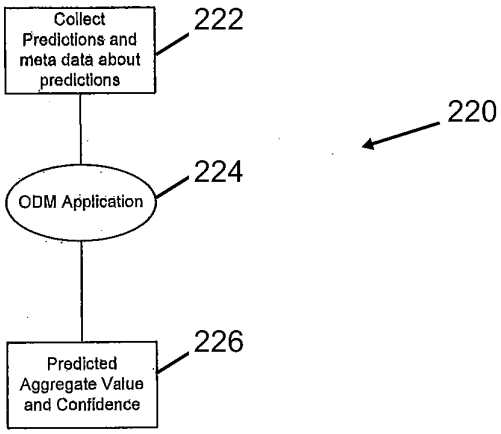


Fig. 5

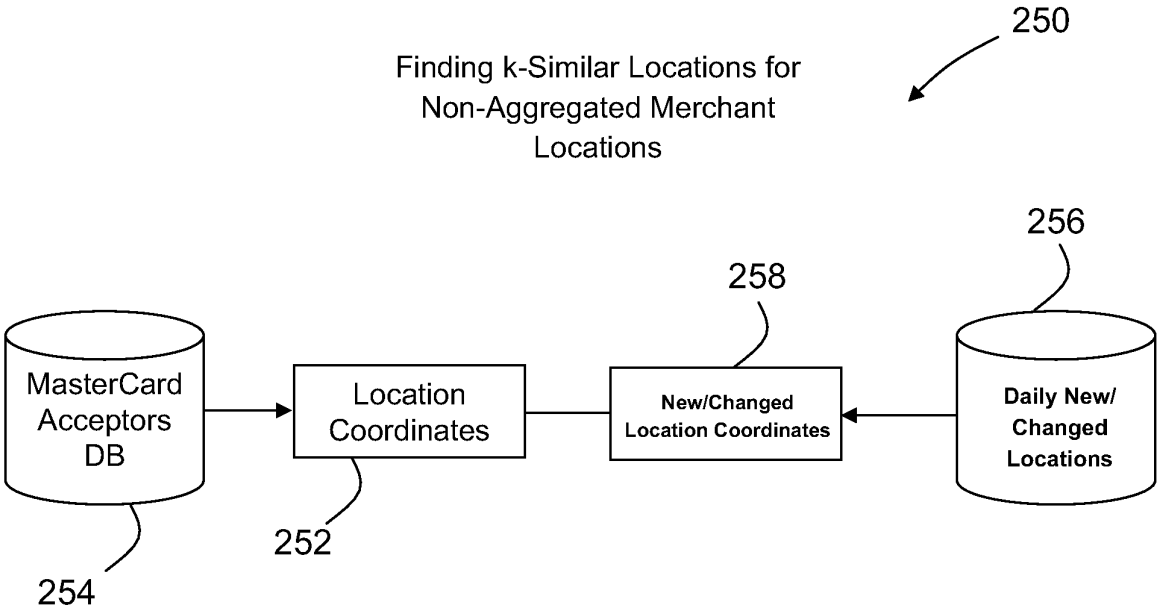

250
Finding k-Similar Locations for Non-Aggregated Merchant Locations
256
258
MasterCard Acceptors DB
Location Coordinates
New/Changed Location Coordinates
Daily New/ Changed Locations
252
254

FIG. 6

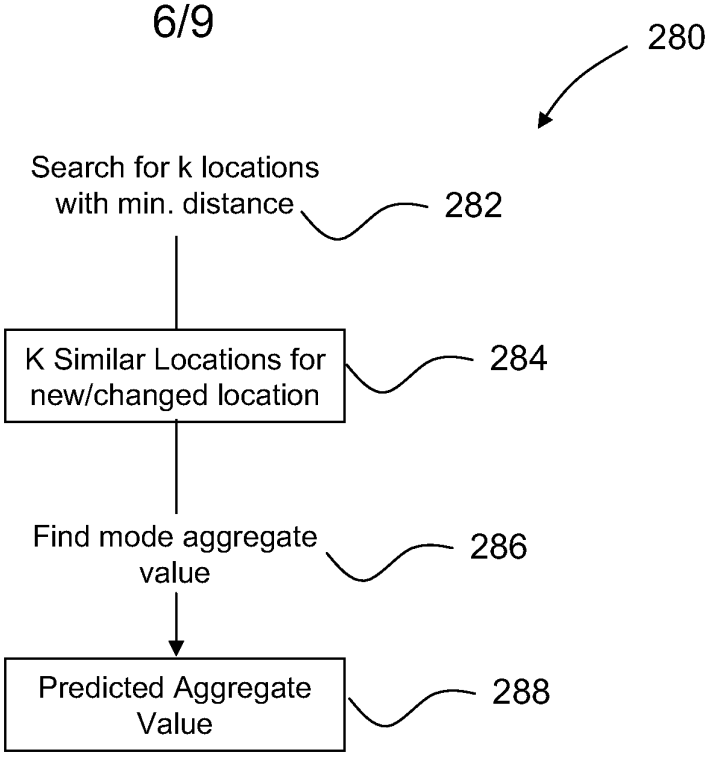

280
Search for k locations with min. distance
282
K Similar Locations for new/changed location
284
Find mode aggregate value
286
Predicted Aggregate Value
288

FIG. 7

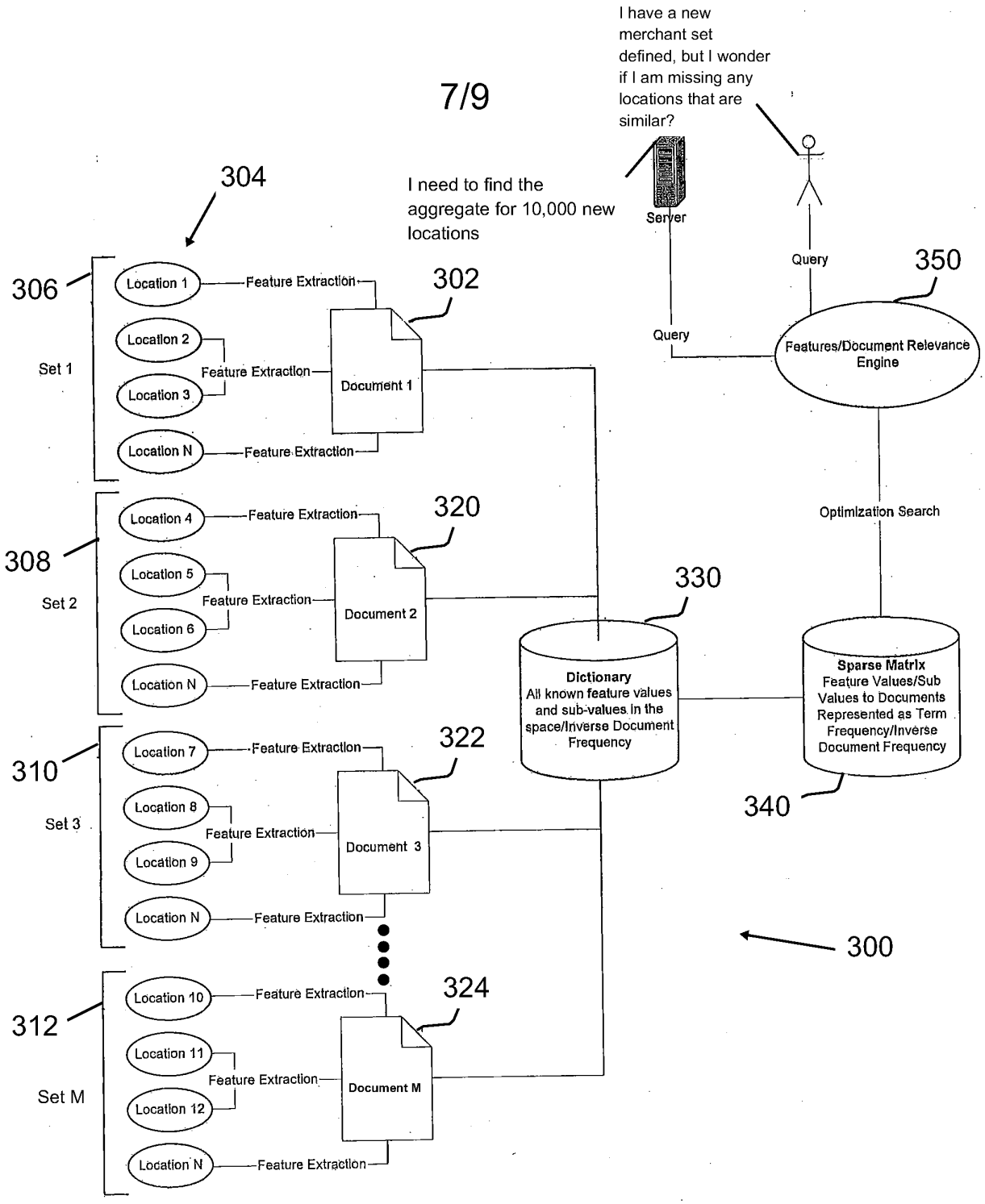

I have a new merchant set defined, but I wonder if I am missing any locations that are similar?
I need to find the aggregate for 10,000 new locations
Server
Query
Query
350
Features/Document Relevance Engine
304
306
Set 1
Location 1
Location 2
Location 3
Location N
Feature Extraction
302
Document 1
308
Set 2
Location 4
Location 5
Location 6
Location N
320
Document 2
330
Dictionary
All known feature values and sub-values in the space/Inverse Document Frequency
Optimization Search
Sparse Matrix
Feature Values/Sub Values to Documents Represented as Term Frequency/Inverse Document Frequency
340
310
Set 3
Location 7
Location 8
Location 9
Location N
322
Document 3
300
312
Set M
Location 10
Location 11
Location 12
Location N
324
Document M

Fig. 8

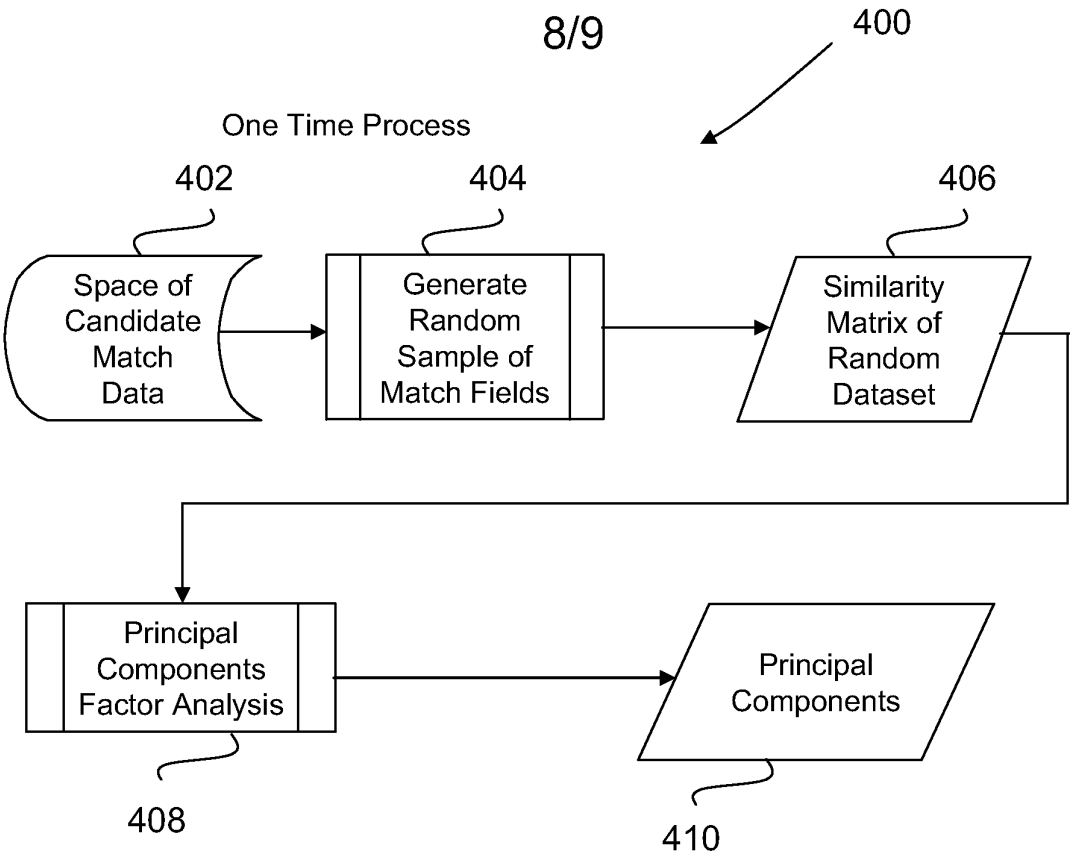
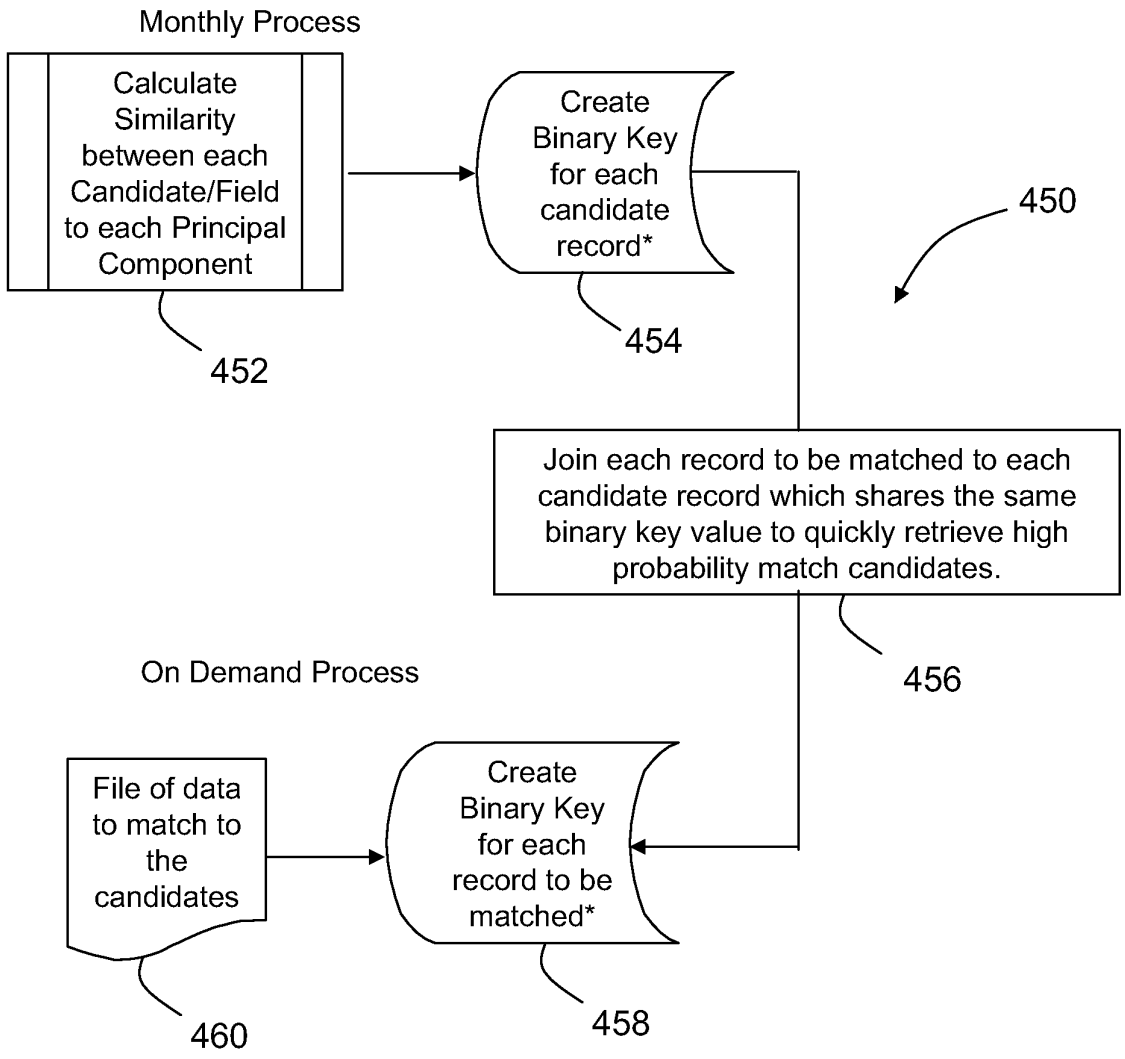

400
One Time Process
402
404
406
Space of Candidate Match Data
Generate Random Sample of Match Fields
Similarity Matrix of Random Dataset
Principal Components Factor Analysis
Principal Components
408
410

FIG. 9



*Each position of the binary key represents the given record/field's similarity to the corresponding principal component. If the similarity is above a predefined threshold, the given position of the key for that record will have a value of 1, otherwise it will have a value of 0

FIG. 10