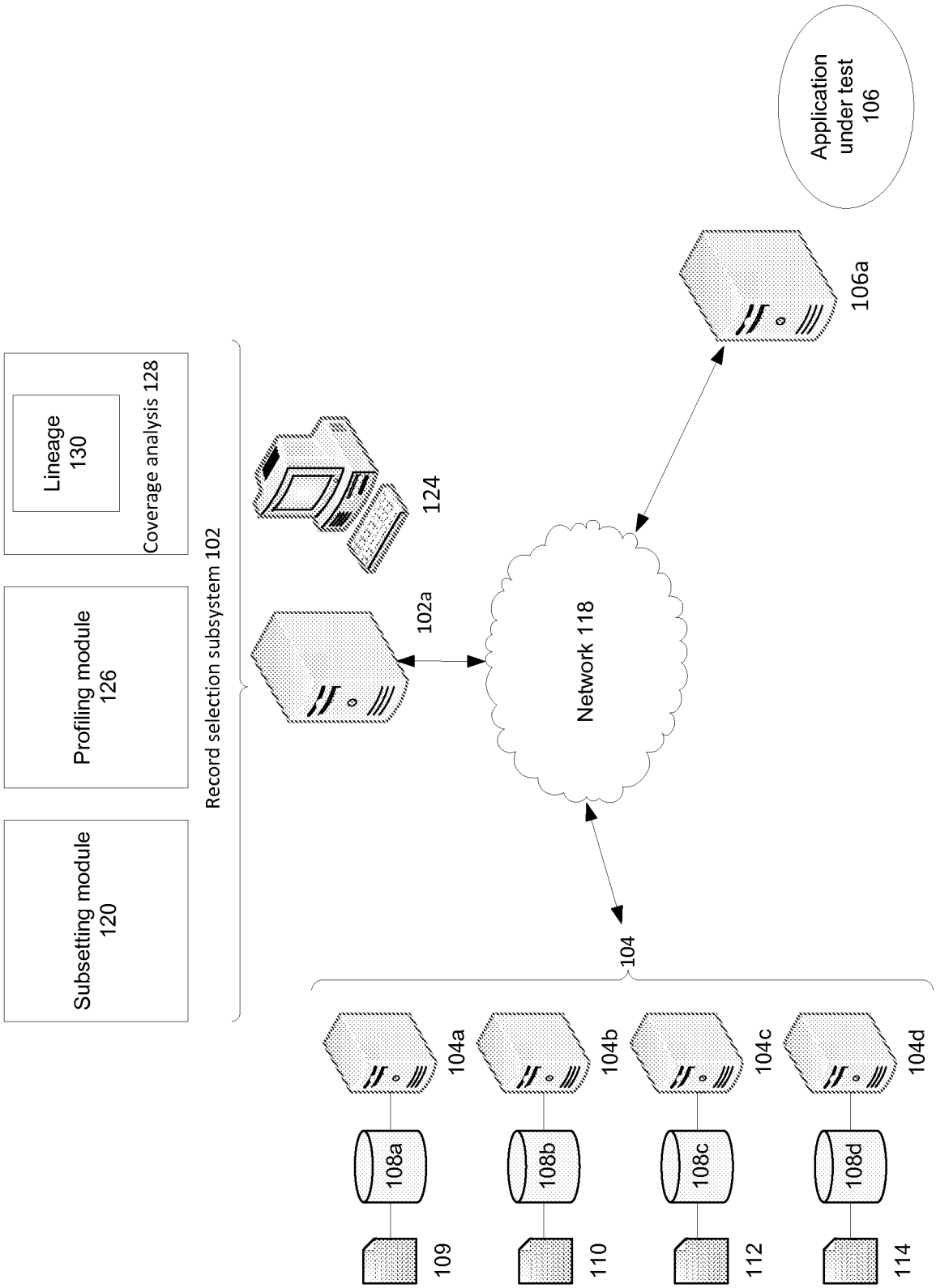

Application under test 106
106a
Network 118
102a
124
Record selection subsystem 102
Subsetting module 120
Profiling module 126
Lineage 130
Coverage analysis 128
104
104a
104b
104c
104d
108a
108b
108c
108d
109
110
112
114

FIG. 1

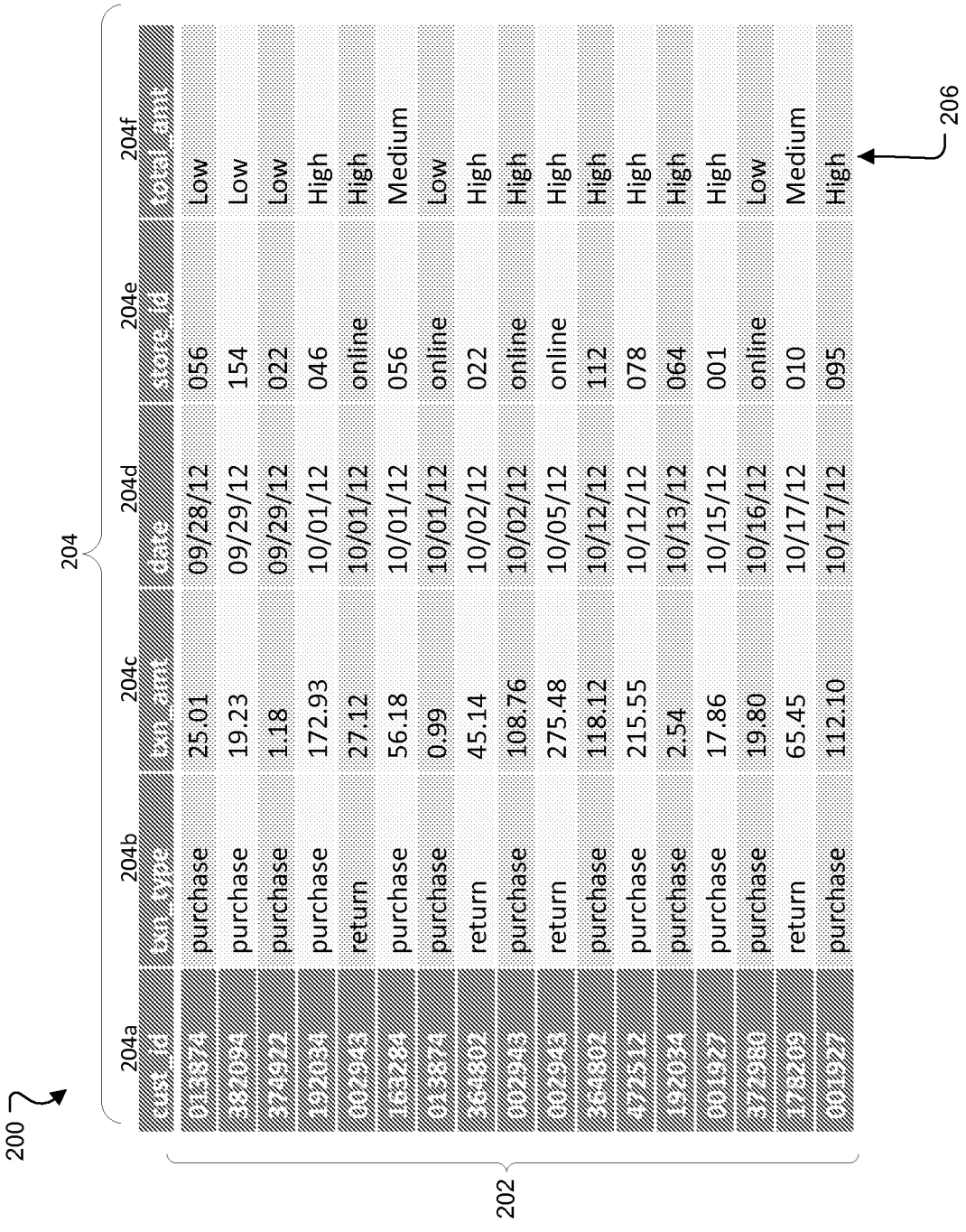
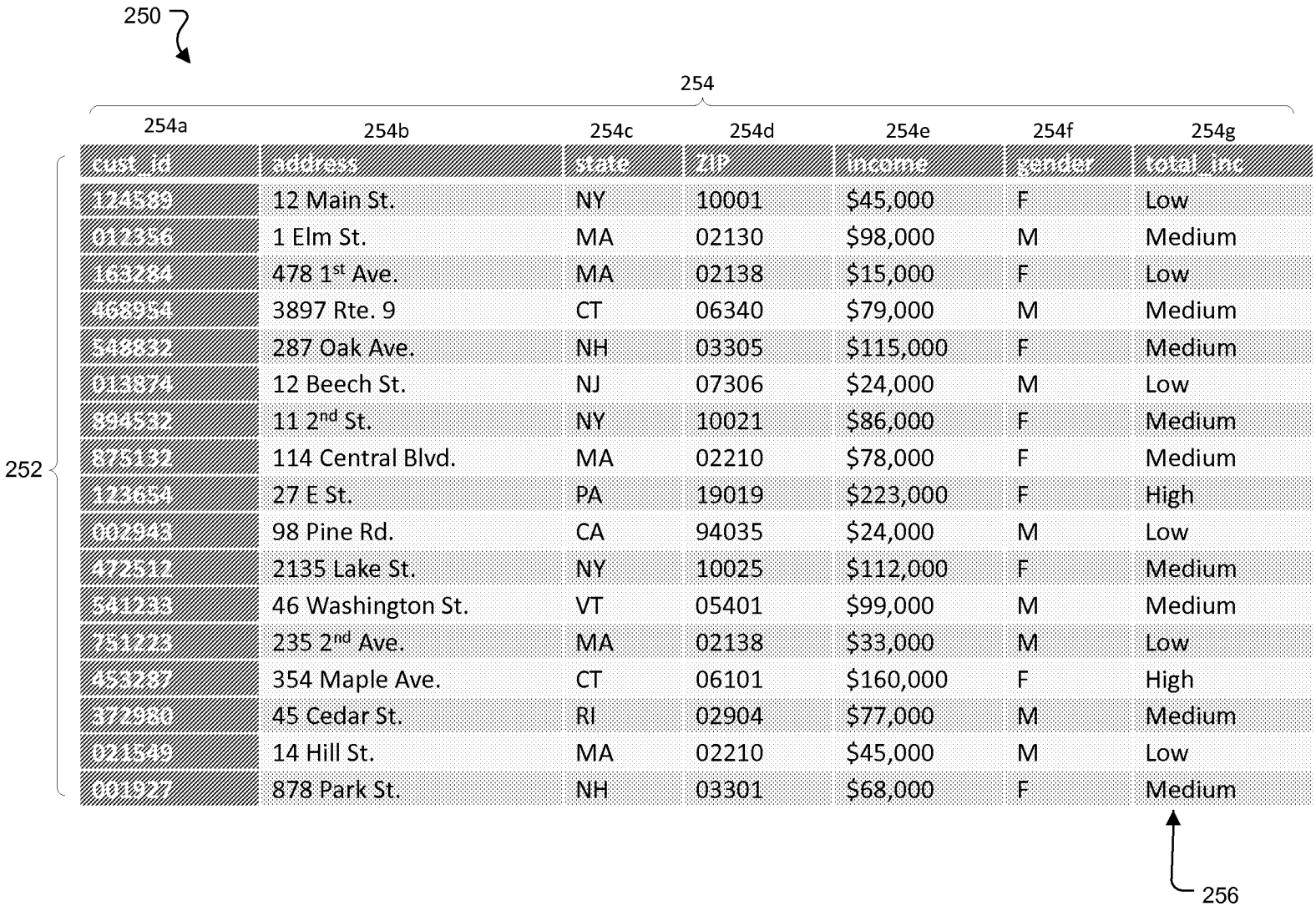

200
202
204
204a
cust_id
204b
activity
204c
tx_amt
204d
date
204e
store_nr
204f
rewards_level
013374 purchase 25.01 09/28/12 056 Low
364084 purchase 19.23 09/29/12 154 Low
224477 purchase 1.18 09/29/12 022 Low
122161 purchase 172.93 10/01/12 046 High
032913 return 27.12 10/01/12 online High
163284 purchase 56.18 10/01/12 056 Medium
013874 purchase 0.99 10/01/12 online Low
764301 return 45.14 10/02/12 022 High
002443 purchase 108.76 10/02/12 online High
017463 return 275.48 10/05/12 online High
364301 purchase 118.12 10/12/12 112 High
172512 purchase 215.55 10/12/12 078 High
122084 purchase 2.54 10/13/12 064 High
011327 purchase 17.86 10/15/12 001 High
374364 purchase 19.80 10/16/12 online Low
178403 return 65.45 10/17/12 010 Medium
001927 purchase 112.10 10/17/12 095 High
206

FIG. 2A



cust_id	address	state	ZIP	income	gender	total_inc
124589	12 Main St.	NY	10001	$45,000	F	Low
012356	1 Elm St.	MA	02130	$98,000	M	Medium
163284	478 1st Ave.	MA	02138	$15,000	F	Low
469954	3897 Rte. 9	CT	06340	$79,000	M	Medium
548832	287 Oak Ave.	NH	03305	$115,000	F	Medium
013874	12 Beech St.	NJ	07306	$24,000	M	Low
894532	11 2nd St.	NY	10021	$86,000	F	Medium
875132	114 Central Blvd.	MA	02210	$78,000	F	Medium
123654	27 E St.	PA	19019	$223,000	F	High
007943	98 Pine Rd.	CA	94035	$24,000	M	Low
472512	2135 Lake St.	NY	10025	$112,000	F	Medium
541238	46 Washington St.	VT	05401	$99,000	M	Medium
751723	235 2nd Ave.	MA	02138	$33,000	M	Low
453287	354 Maple Ave.	CT	06101	$160,000	F	High
372980	45 Cedar St.	RI	02904	$77,000	M	Medium
021549	14 Hill St.	MA	02210	$45,000	M	Low
001927	878 Park St.	NH	03301	$68,000	F	Medium

FIG. 2B

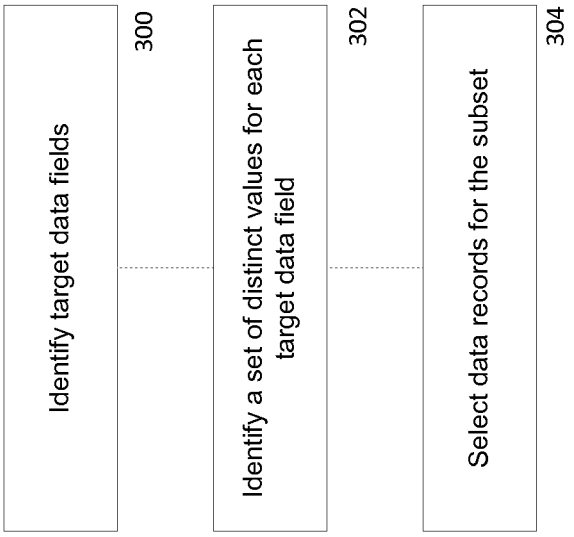

Identify target data fields
300
Identify a set of distinct values for each target data field
302
Select data records for the subset
304

FIG. 3

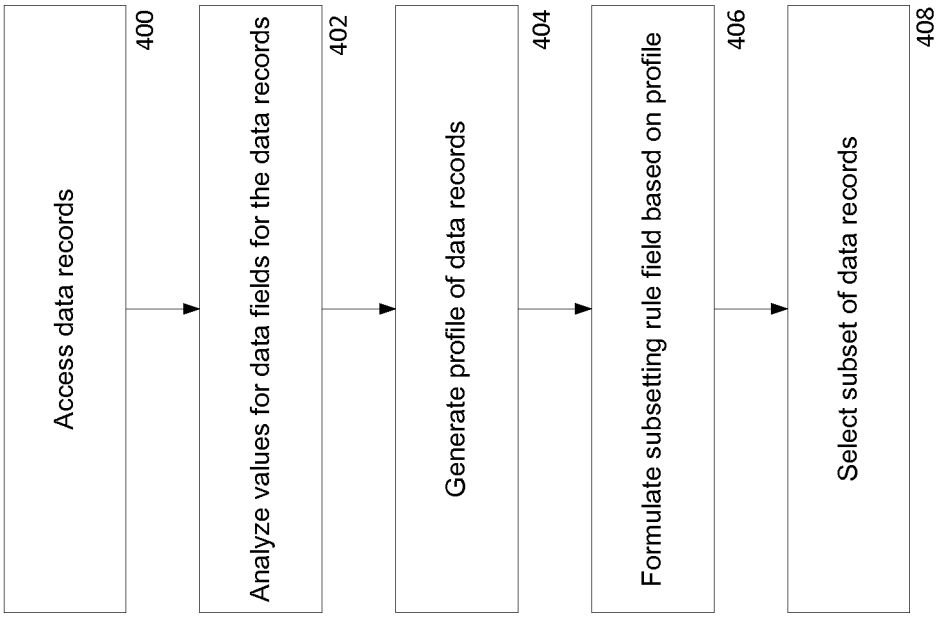

Access data records
400
Analyze values for data fields for the data records
402
Generate profile of data records
404
Formulate subsetting rule field based on profile
406
Select subset of data records
408

FIG. 4

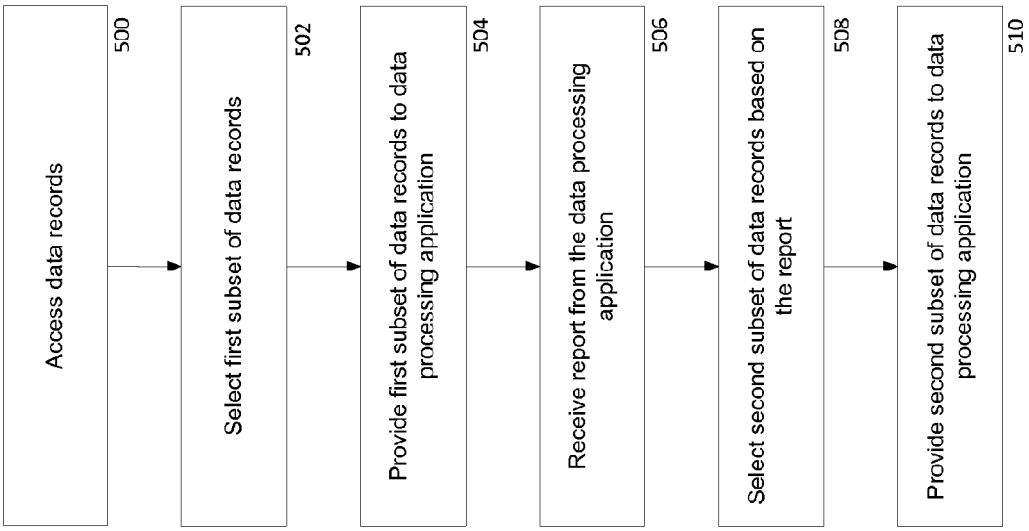

Access data records
500
Select first subset of data records
502
Provide first subset of data records to data processing application
504
Receive report from the data processing application
506
Select second subset of data records based on the report
508
Provide second subset of data records to data processing application
510

FIG. 5