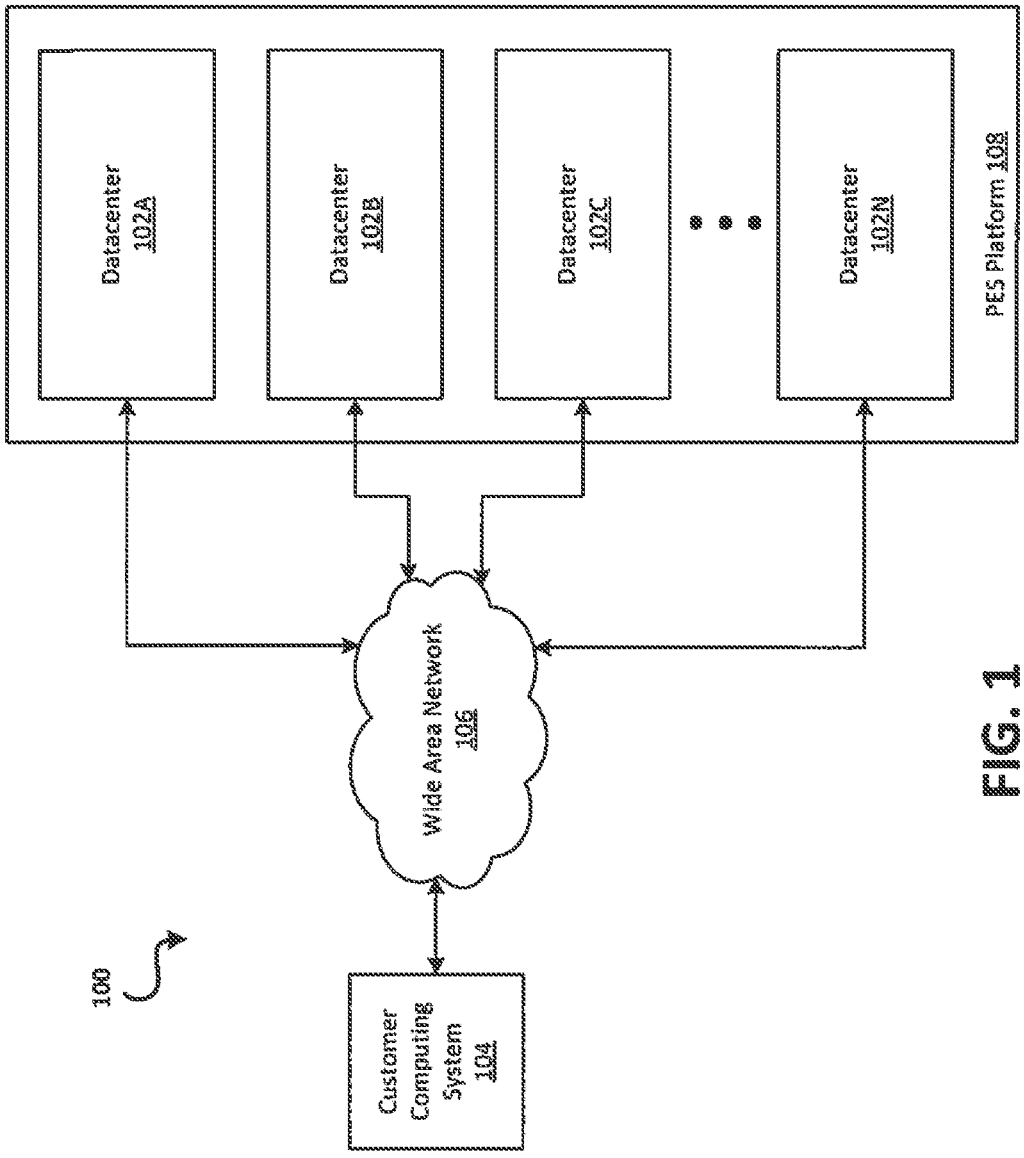
Datacenter 102A
Datacenter 102B
Datacenter 102C
Datacenter 102N
PES Platform 108
Wide Area Network 106
Customer Computing System 104
100

FIG. 1

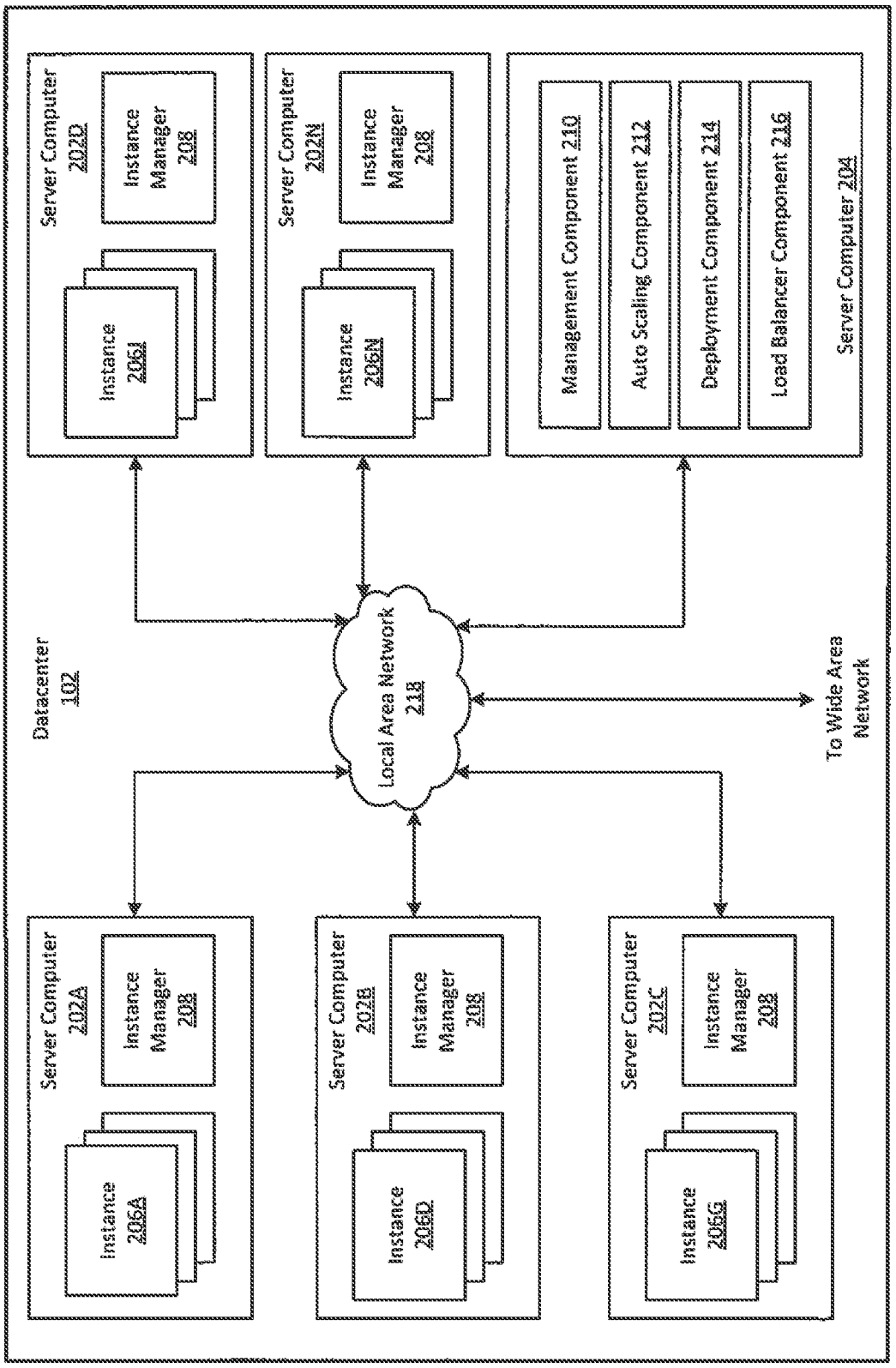
Datacenter 102
Server Computer 202D
Instance Manager 208
Instance 206J
Server Computer 202N
Instance Manager 208
Instance 206N
Management Component 210
Auto Scaling Component 212
Deployment Component 214
Load Balancer Component 216
Server Computer 204
Local Area Network 218
To Wide Area Network
Server Computer 202A
Instance Manager 208
Instance 206A
Server Computer 202B
Instance Manager 208
Instance 206D
Server Computer 202C
Instance Manager 208
Instance 206G

FIG. 2

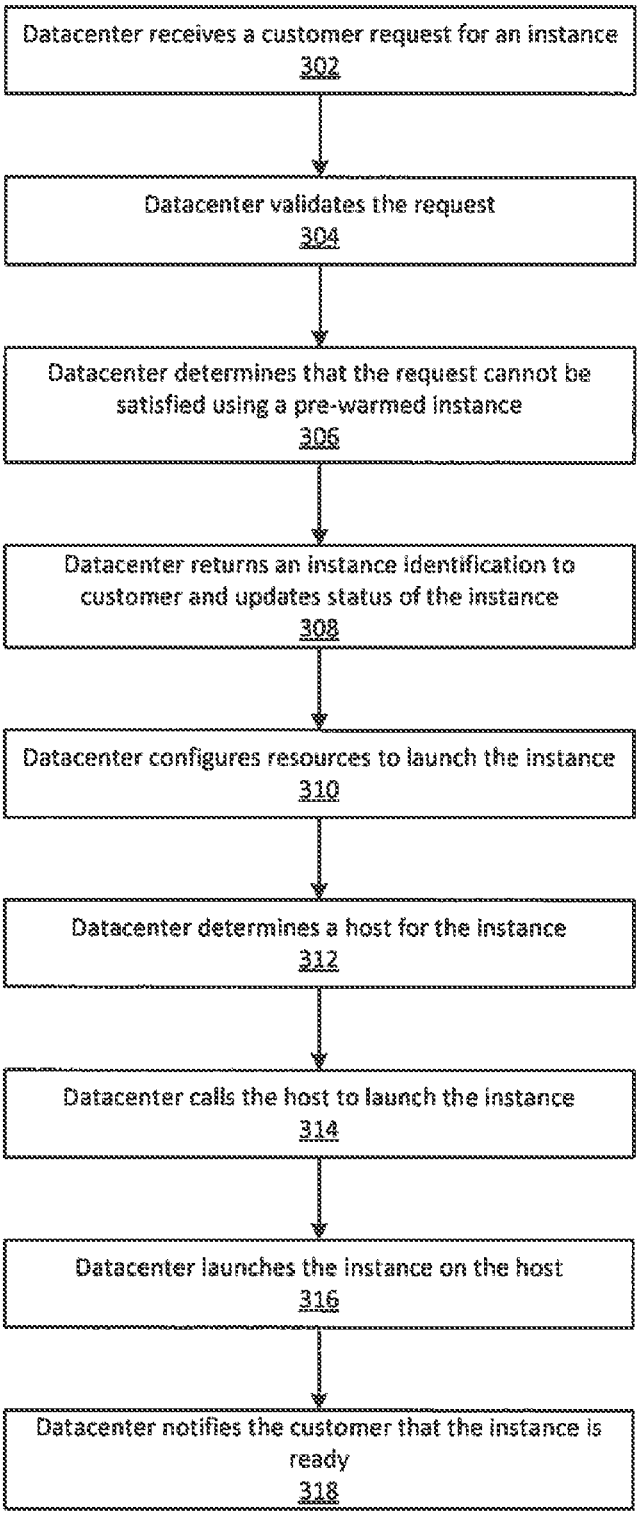
Datacenter receives a customer request for an instance
302
Datacenter validates the request
304
Datacenter determines that the request cannot be satisfied using a pre-warmed instance
306
Datacenter returns an instance identification to customer and updates status of the instance
308
Datacenter configures resources to launch the instance
310
Datacenter determines a host for the instance
312
Datacenter calls the host to launch the instance
314
Datacenter launches the instance on the host
316
Datacenter notifies the customer that the instance is ready
318

FIG. 3

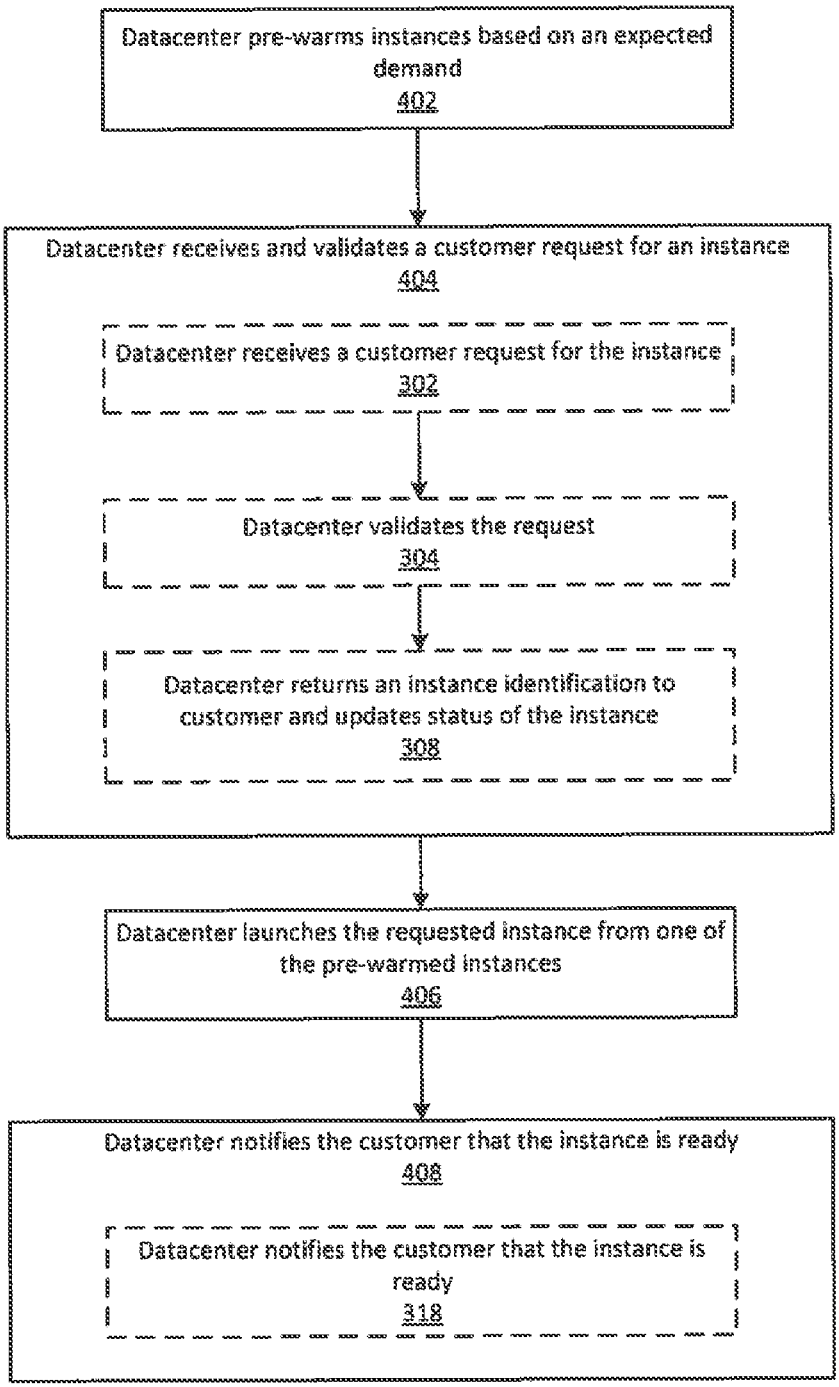
Datacenter pre-warms instances based on an expected demand
402
Datacenter receives and validates a customer request for an instance
404
Datacenter receives a customer request for the instance
302
Datacenter validates the request
304
Datacenter returns an instance identification to customer and updates status of the instance
308
Datacenter launches the requested instance from one of the pre-warmed instances
406
Datacenter notifies the customer that the instance is ready
408
Datacenter notifies the customer that the instance is ready
318

FIG. 4

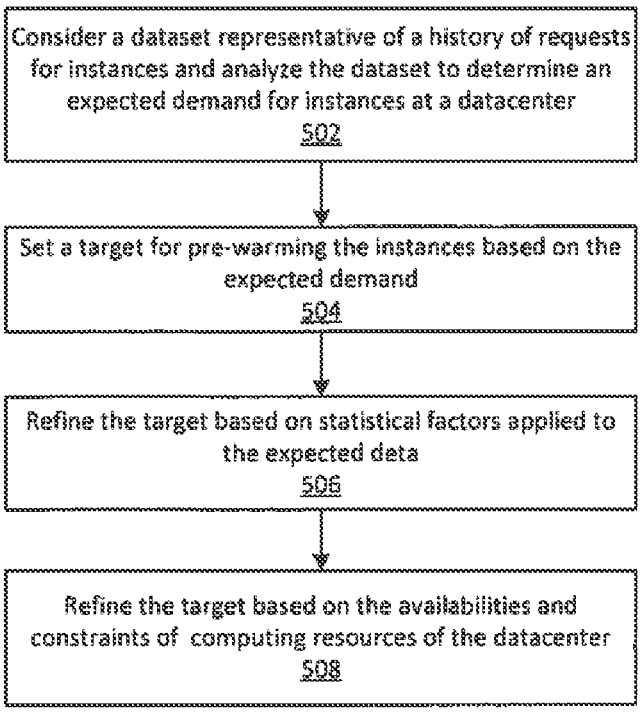
Consider a dataset representative of a history of requests for instances and analyze the dataset to determine an expected demand for instances at a datacenter
502
Set a target for pre-warming the instances based on the expected demand
504
Refine the target based on statistical factors applied to the expected deta
506
Refine the target based on the availabilities and constraints of computing resources of the datacenter
508

FIG. 5

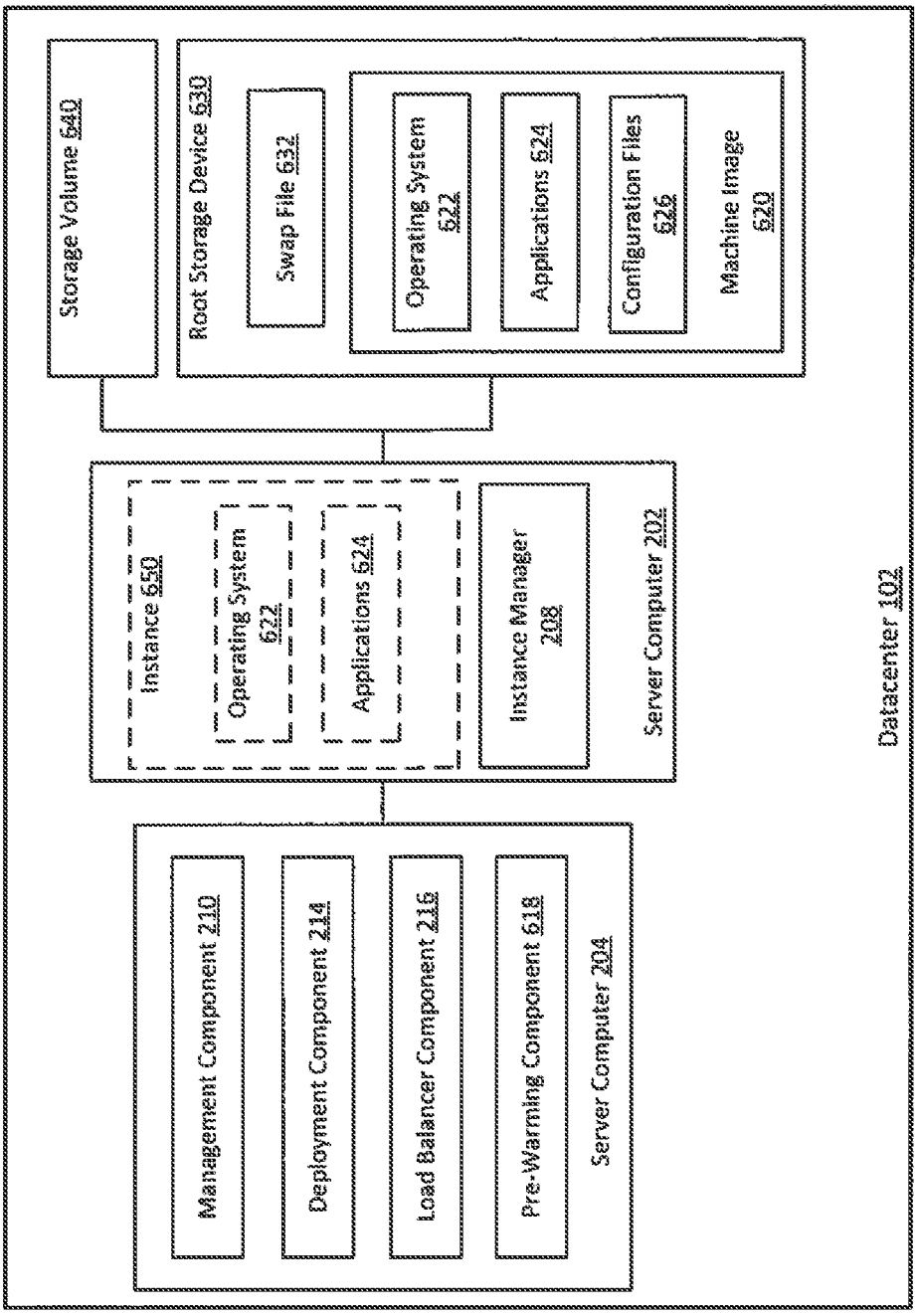
Datacenter 102
Storage Volume 640
Root Storage Device 630
Swap File 632
Operating System 622
Applications 624
Configuration Files 626
Machine Image 620
Instance 650
Operating System 622
Applications 624
Instance Manager 208
Server Computer 202
Management Component 210
Deployment Component 214
Load Balancer Component 216
Pre-Warming Component 618
Server Computer 204

FIG. 6

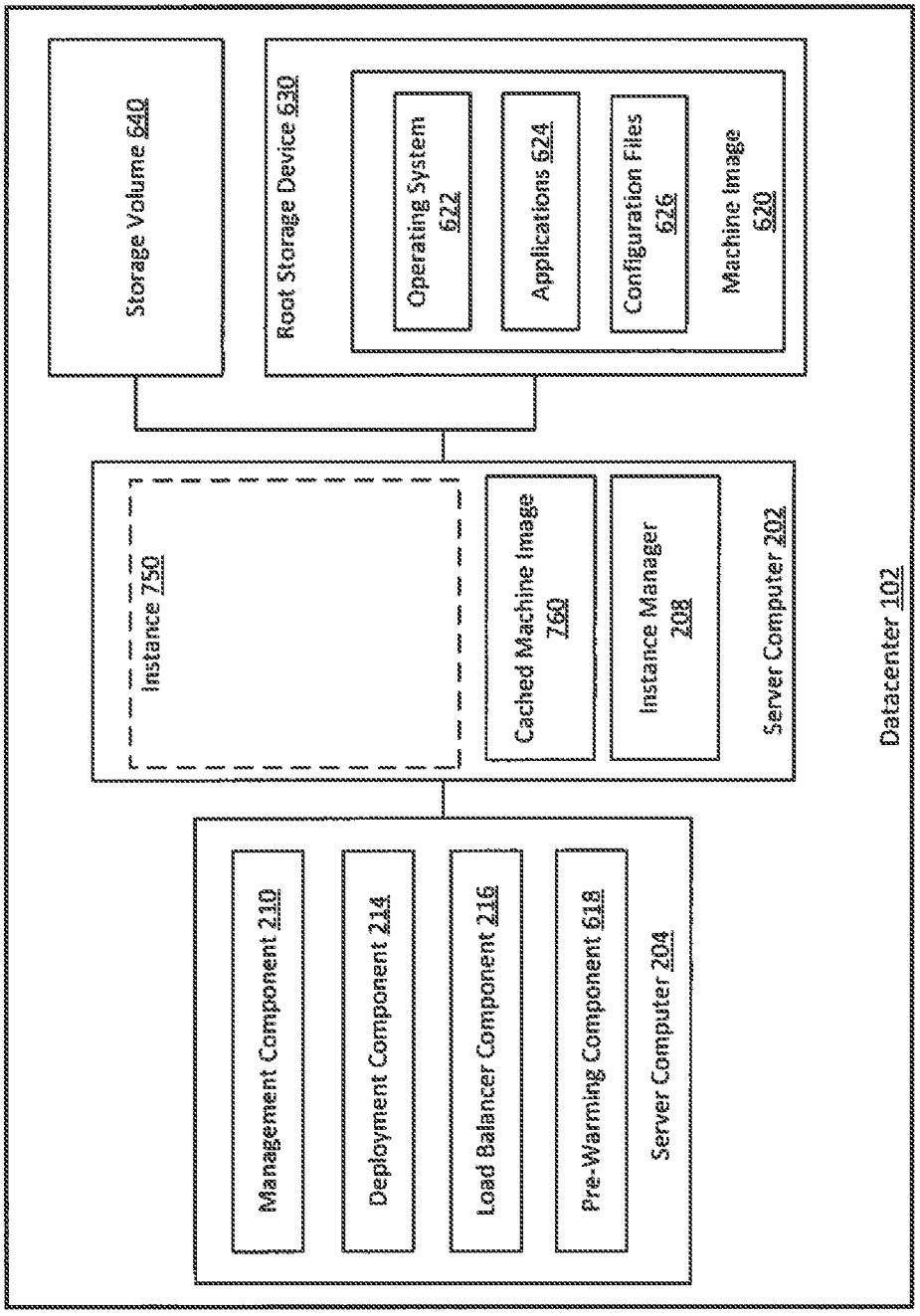
Datacenter 102
Server Computer 204
Management Component 210
Deployment Component 214
Load Balancer Component 216
Pre-Warming Component 618
Server Computer 202
Instance Manager 208
Cached Machine Image 760
Instance 750
Storage Volume 640
Root Storage Device 630
Machine Image 620
Operating System 622
Applications 624
Configuration Files 626

FIG. 7

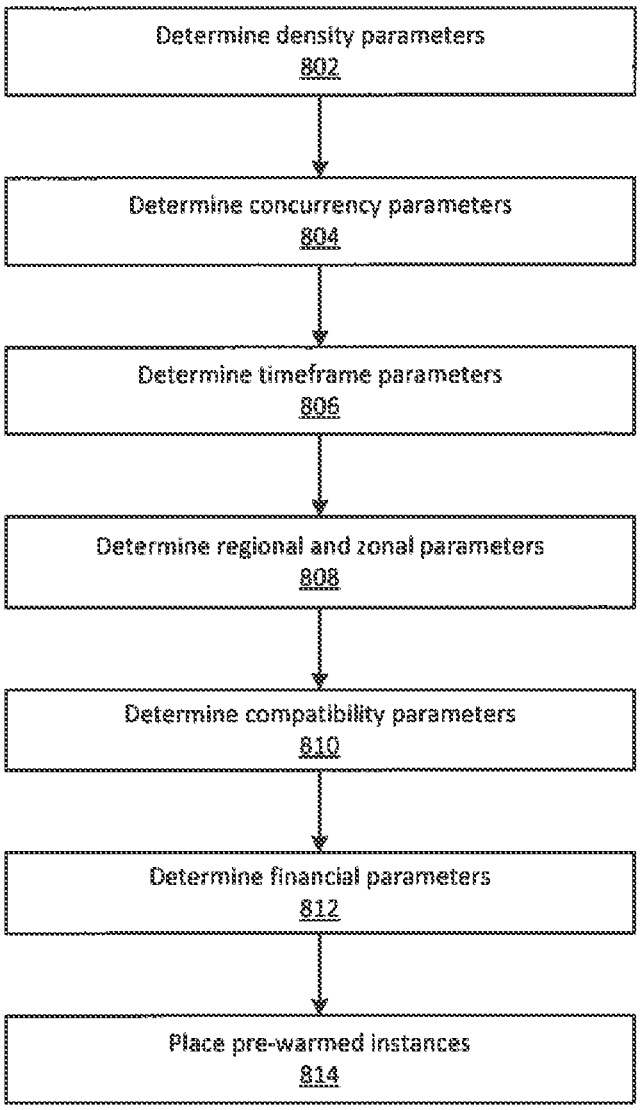
Determine density parameters
802
Determine concurrency parameters
804
Determine timeframe parameters
806
Determine regional and zonal parameters
808
Determine compatibility parameters
810
Determine financial parameters
812
Place pre-warmed instances
814

FIG. 8

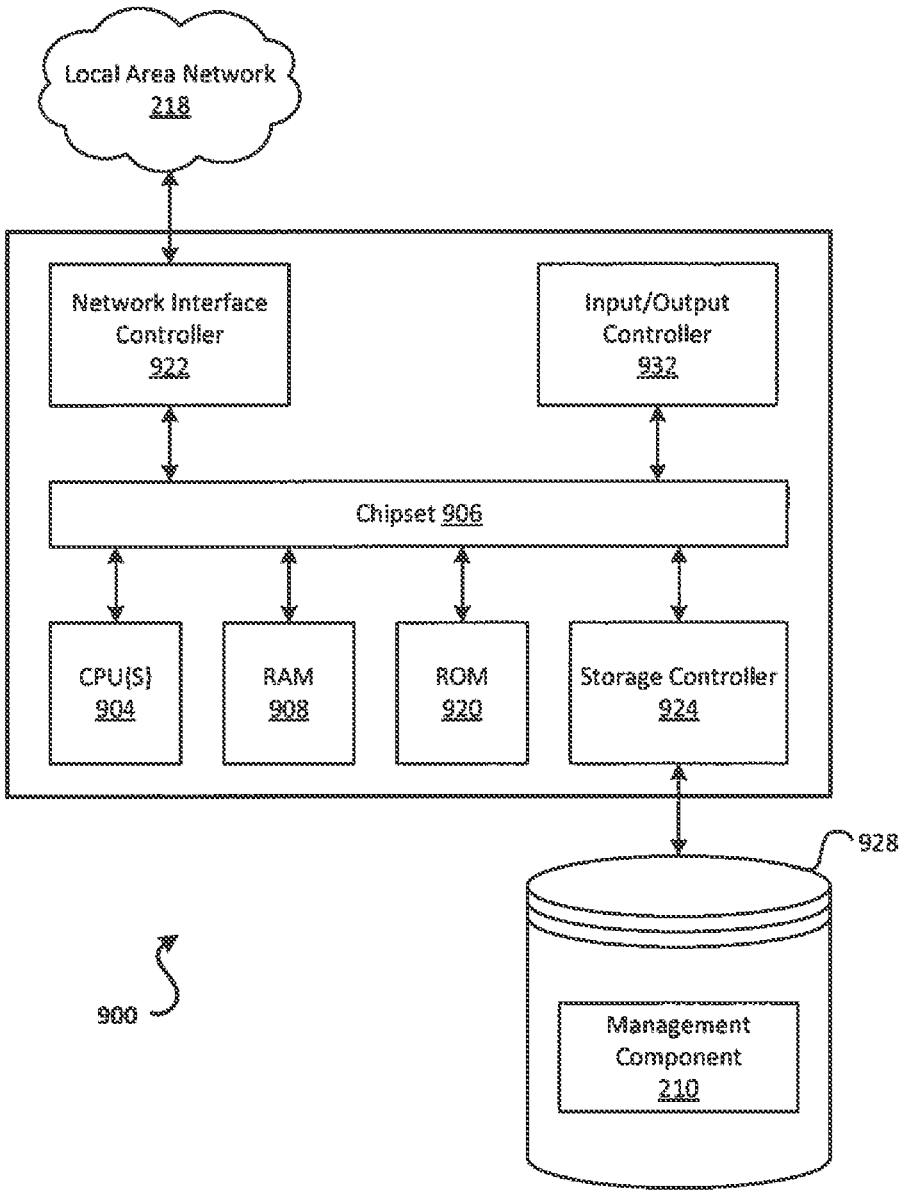
Local Area Network
218
Network Interface Controller
922
Input/Output Controller
932
Chipset 906
CPU(S)
904
RAM
908
ROM
920
Storage Controller
924
928
900
Management Component
210

FIG. 9