Office de la Propriété Intellectuelle du Canada
Un organisme d'Industrie Canada

Canadian Intellectual Property Office
An agency of Industry Canada

CA 2741088 C 2017/07/11

(11)(21) **2 741 088**

(12) **BREVET CANADIEN**
CANADIAN PATENT

(13) **C**

(86) Date de dépôt PCT/PCT Filing Date: 2009/10/21
(87) Date publication PCT/PCT Publication Date: 2010/04/29
(45) Date de délivrance/Issue Date: 2017/07/11
(85) Entrée phase nationale/National Entry: 2011/04/18
(86) N° demande PCT/PCT Application No.: US 2009/061521
(87) N° publication PCT/PCT Publication No.: 2010/048316
(30) Priorité/Priority: 2008/10/21 (US61/107,172)

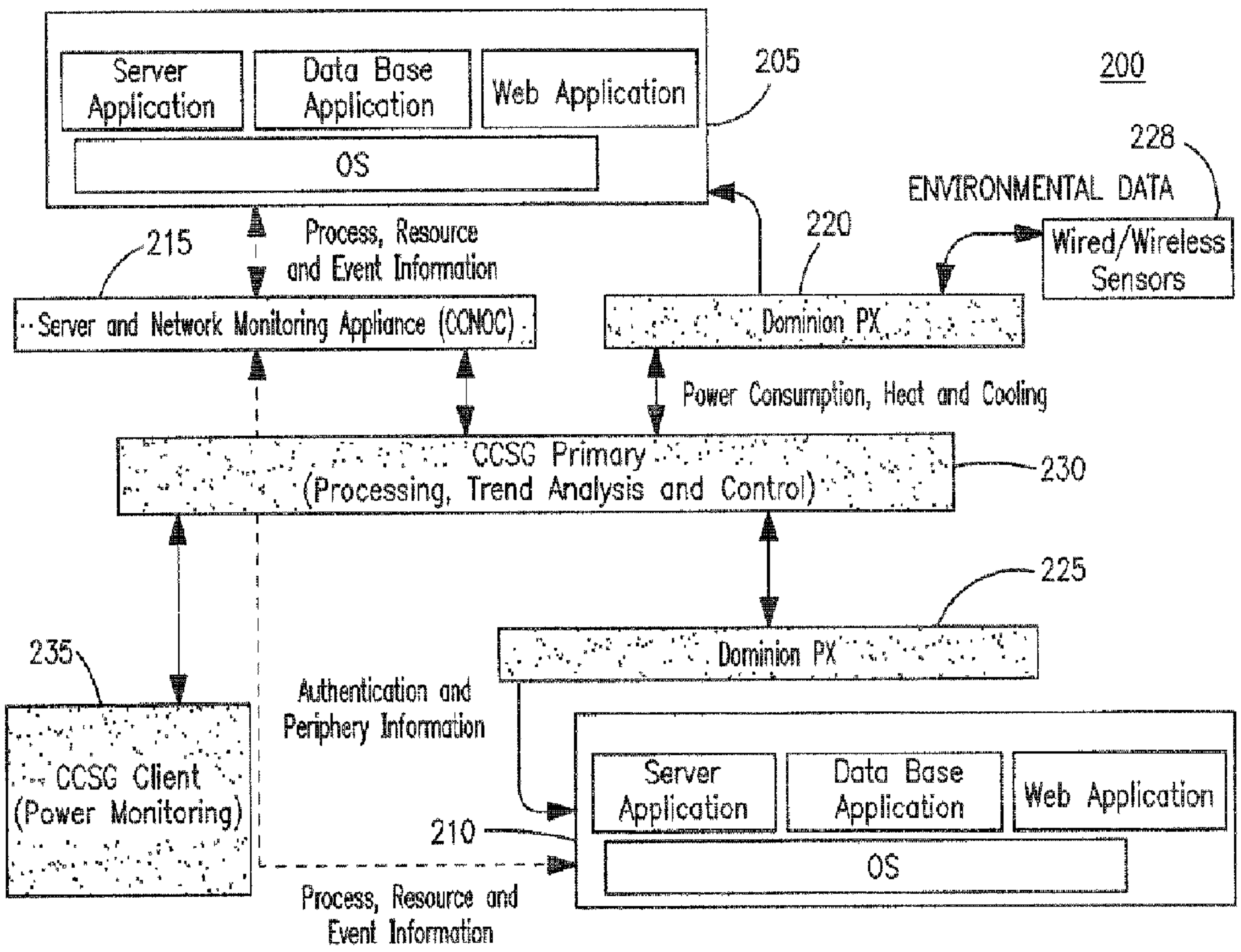
(51) Cl.Int./Int.Cl. *G06F 1/32* (2006.01)

(72) Inventeurs/Inventors:
MALIK, NAIM, US;
PAETZ, CHRISTIAN, DE;
WEINSTOCK, NEIL, US;
YANG, ALLEN, US;
ONYSHKEVYH, VSEVOLOD, US;
SOMASUNDARAM, SIVA, US

(73) Propriétaire/Owner:
RARITAN AMERICAS, INC., US

(74) Agent: KIRBY EADES GALE BAKER

(54) Titre : PROCEDES D'OBTENTION D'UNE GESTION D'ENERGIE CONSCIENTE
(54) Title: METHODS OF ACHIEVING COGNIZANT POWER MANAGEMENT



(57) Abrégé/Abstract:
A system and method of increasing the efficiency of overall power utilization in data centers by integrating a power management approach based on a comprehensive, dynamic model of the data center created with integrated environmental and computational power monitoring to correlate power usage with different configurations of business services utilization, with the techniques of CPU level power management.

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
29 April 2010 (29.04.2010)

PCT

(10) International Publication Number
WO 2010/048316 A1

(51) International Patent Classification: *G06F 1/32* (2006.01)

(21) International Application Number: PCT/US2009/061521

(22) International Filing Date: 21 October 2009 (21.10.2009)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
61/107,172 21 October 2008 (21.10.2008) US

(71) Applicant *(for all designated States except US)*: **RARITAN AMERICAS, INC.** [US/US]; 400 Cottontail Lane, Somerset, NJ 08873 (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: **MALIK, Naim** [US/US]; 404 Coventry Lane, Somerset, NJ 08873 (US). **PAETZ, Christian** [DE/DE]; Gert-Froebe-Str. 6a, 08064 Zwickau (DE). **WEINSTOCK, Neil** [US/US]; 11 Cedar Ridge Lane, Randolph, NJ 07869 (US). **YANG, Allen** [US/US]; 699 John Christian Drive, Bridgewater, NJ 08807 (US). **ONYSHKEVYH, Vsevolod** [US/US]; 69 Bayberry Road, Princeton, NJ 08540 (US). **SOMASUNDARAM, Siva** [US/US]; 356 Ridge Road, Apt. E-11, Dayton, NJ 08810 (US).

(74) Agents: **EPSTEIN, William** et al.; Gibbons P.C., One Gateway Center, Newark, NJ 07102-5310 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PE, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *with international search report (Art. 21(3))*

(54) Title: METHODS OF ACHIEVING COGNIZANT POWER MANAGEMENT

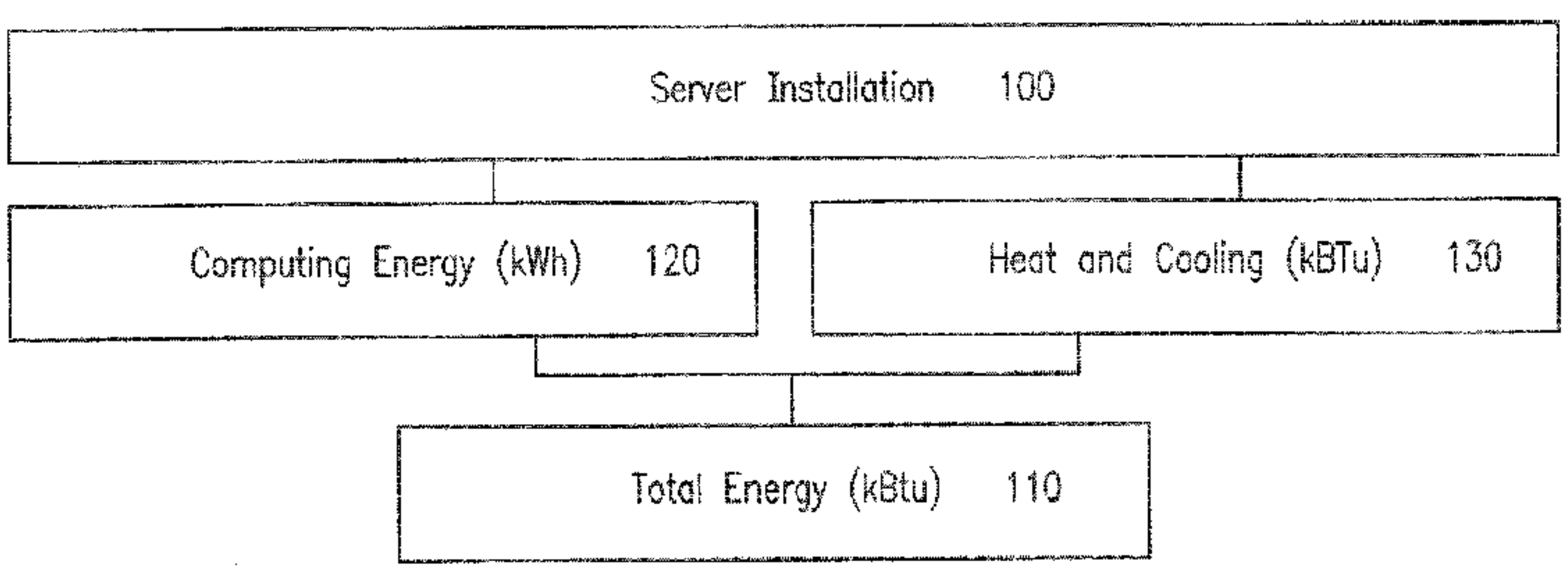


FIG. 1

(57) Abstract: A system and method of increasing the efficiency of overall power utilization in data centers by integrating a power management approach based on a comprehensive, dynamic model of the data center created with integrated environmental and computational power monitoring to correlate power usage with different configurations of business services utilization, with the techniques of CPU level power management.

WO 2010/048316 A1

CLAIMS:

1. A method for cognizant power management in a distributed computing system having a plurality of processing units, comprising the steps of:

a) gathering first values of process level information from the plurality of processing units;

b) gathering first values of environmental metrics for the distributed computing system;

c) generating a predictive behavior power utilization model based on the process level information and the environmental metrics;

d) predicting a power utilization of the computing system according to the predictive behavior power utilization model and a distributed load;

e) setting a state in at least one of the plurality of processing units based on the predictive behavior power utilization model, the state selected from the group consisting of power management performance states and sleep states of the at least one processing unit;

f) distributing an application load to the computing system based on the predictive behavior power utilization model;

g) determining an actual power utilization by gathering second values for the process level information and the environmental metrics; and

h) updating the predictive behavior power utilization model according to the actual power utilization when a difference between the actual power utilization and the predicted power utilization exceeds a predetermined threshold.

2. The method of claim 1, wherein the plurality of processing units includes at least one of a plurality of central processing units, a plurality of functional units within central processing units, and a plurality of hypervisors managing virtual machines.

3. The method of claim 2, wherein setting one of a plurality of performance states includes changing the frequency of a processing unit.

4. The method of claim 2, wherein setting one of a plurality of performance states includes changing the voltage of a processing unit.

5. The method of claim 2, wherein setting one of a plurality of performance states includes changing the frequency and the voltage of a processing unit.

6. The method of claim 2, wherein one of a plurality of performance states includes at least a first performance state and a second performance state, wherein the first performance state can bear a greater computational load than the second state.

7. The method of claim 1, wherein one of a plurality of sleep states includes at least a first sleep state and second sleep state, wherein the first sleep state has a lower wake up time than the second sleep state.

8. A method for cognizant power management in a distributed computing system having a plurality of processing units, comprising the steps of:

a) gathering first values of process level information from the plurality of processing units;

b) gathering first values of environmental metrics for the distributed computing system;

c) generating a predictive behavior power utilization model based on the process level information and the environmental metrics;

d) predicting a power utilization of the computing system according to the predictive behavior power utilization model and a distributed load;

e) setting a state for at least one of the plurality of processing units based on the predictive behavior power utilization model, the state selected from the group consisting of power management performance states and sleep states of the at least one processing unit and the setting step including the step of issuing a software call to an operating system running on the at least one processing unit;

f) distributing an application load to the computing system based on based on the predictive behavior power utilization model;

g) determining an actual power utilization by gathering second values for the process level information and the environmental metrics; and

h) updating the predictive behavior power utilization model according to the actual power utilization when a difference between the actual power utilization and the predicted power utilization exceeds a predetermined threshold.

9. The method of claim 8, wherein the plurality of processing units includes at least one of a plurality of central processing units, a plurality of functional units within central processing units, and a plurality of hypervisors managing virtual machines.

10. The method of claim 9, wherein setting one of a plurality of performance states includes changing the frequency of a processing unit.

11. The method of claim 9, wherein setting one of a plurality of performance states includes changing the voltage of a processing unit.

12. The method of claim 9, wherein setting one of a plurality of performance states includes changing the frequency and the voltage of a processing unit.

13. The method of claim 9, wherein one of a plurality of performance states includes at least a first performance state and a second performance state, wherein the first performance state can bear a greater computational load than the second state.

14. The method of claim 8, wherein one of a plurality of sleep states includes at least a first sleep state and second sleep state, wherein the first sleep state has a lower wake up time than the second sleep state.

CLAIMS:

1. A method for cognizant power management in a distributed computing system having a plurality of processing units, comprising the steps of:

a) gathering first values of process level information from the plurality of processing units;

b) gathering first values of environmental metrics for the distributed computing system;

c) generating a predictive behavior power utilization model based on the process level information and the environmental metrics;

d) predicting a power utilization of the computing system according to the predictive behavior power utilization model and a distributed load;

e) setting a state in at least one of the plurality of processing units based on the predictive behavior power utilization model, the state selected from the group consisting of power management performance states and sleep states of the at least one processing unit;

f) distributing an application load to the computing system based on based on the predictive behavior power utilization model;

g) determining an actual power utilization by gathering second values for the process level information and the environmental metrics; and

h) updating the predictive behavior power utilization model according to actual power utilization when a difference between the actual power utilization and the predicted power utilization exceeds a predetermined threshold.

2. The method of claim 1, wherein the plurality of processing units includes at least one of a plurality of central processing units, a plurality of functional units within central processing units, and a plurality of hypervisors managing virtual machines.

3. The method of claim 2, wherein setting one of a plurality of performance states includes changing the frequency of a processing unit.

4. The method of claim 2, wherein setting one of a plurality of performance states includes changing the voltage of a processing unit.

5. The method of claim 2, wherein setting one of a plurality of performance states includes changing the frequency and the voltage of a processing unit.

6. The method of claim 2, wherein one of a plurality of performance states includes at least a first performance state and a second performance state, wherein the first performance state can bear a greater computational load than the second state.

7. The method of claim 1, wherein one of a plurality of sleep states includes at least a first sleep state and second sleep state, wherein the first sleep state has a lower wake up time than the second sleep state.

8. A method for cognizant power management in a distributed computing system having a plurality of processing units, comprising the steps of:

a) gathering first values of process level information from the plurality of processing units;

b) gathering first values of environmental metrics for the distributed computing system;

c) generating a predictive behavior power utilization model based on the process level information and the environmental metrics;

d) predicting a power utilization of the computing system according to the predictive behavior power utilization model and a distributed load;

e) setting a state for at least one of the plurality of processing units based on the predictive behavior power utilization model, the state selected from the group consisting of power management performance states and sleep states of the at least one processing unit and the setting step including the step of issuing a software call to an operating system running on the at least one processing unit;

f) distributing an application load to the computing system based on based on the predictive behavior power utilization model;

g) determining an actual power utilization by gathering second values for the process level information and the environmental metrics; and

h) updating the predictive behavior power utilization model according to actual power utilization when a difference between the actual power utilization and the predicted power utilization exceeds a predetermined threshold.

9. The method of claim 8, wherein the plurality of processing units includes at least one of a plurality of central processing units, a plurality of functional units within central processing units, and a plurality of hypervisors managing virtual machines.

10. The method of claim 9, wherein setting one of a plurality of performance states includes changing the frequency of a processing unit.

11. The method of claim 9, wherein setting one of a plurality of performance states includes changing the voltage of a processing unit.

12. The method of claim 9, wherein setting one of a plurality of performance states includes changing the frequency and the voltage of a processing unit.

13. The method of claim 9, wherein one of a plurality of performance states includes at least a first performance state and a second performance state, wherein the first performance state can bear a greater computational load than the second state.

14. The method of claim 8, wherein one of a plurality of sleep states includes at least a first sleep state and second sleep state, wherein the first sleep state has a lower wake up time than the second sleep state.

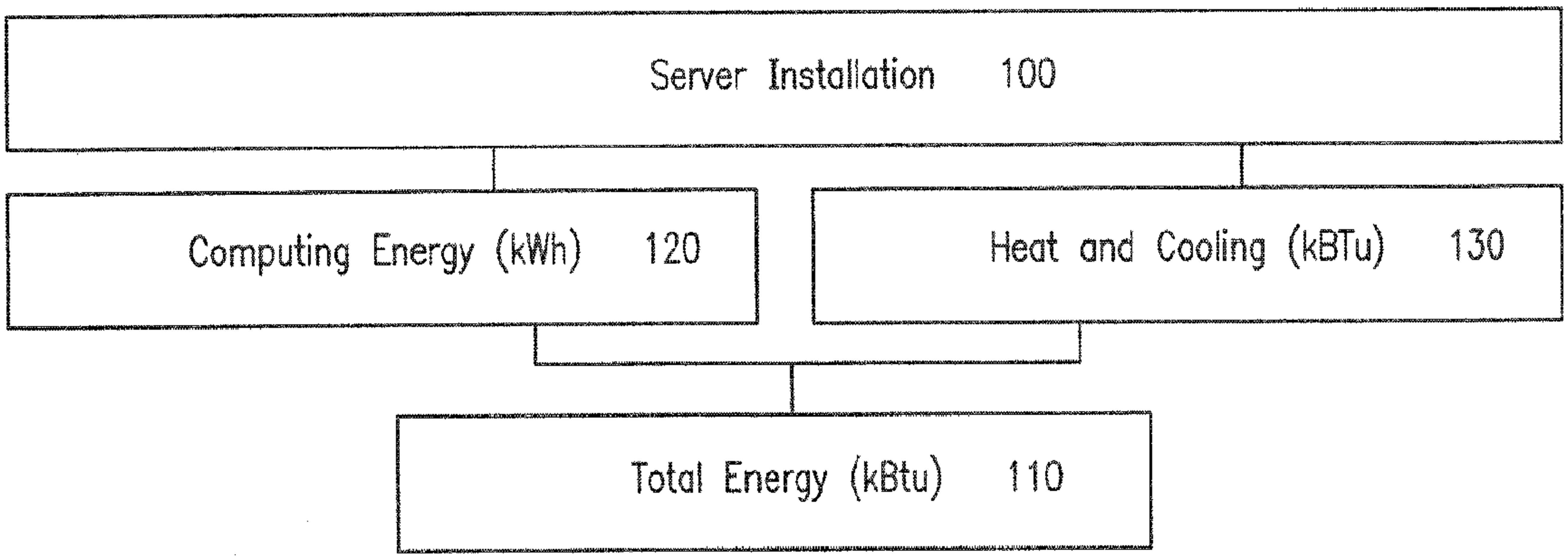

Server Installation 100
Computing Energy (kWh) 120
Heat and Cooling (kBTu) 130
Total Energy (kBtu) 110

FIG. 1

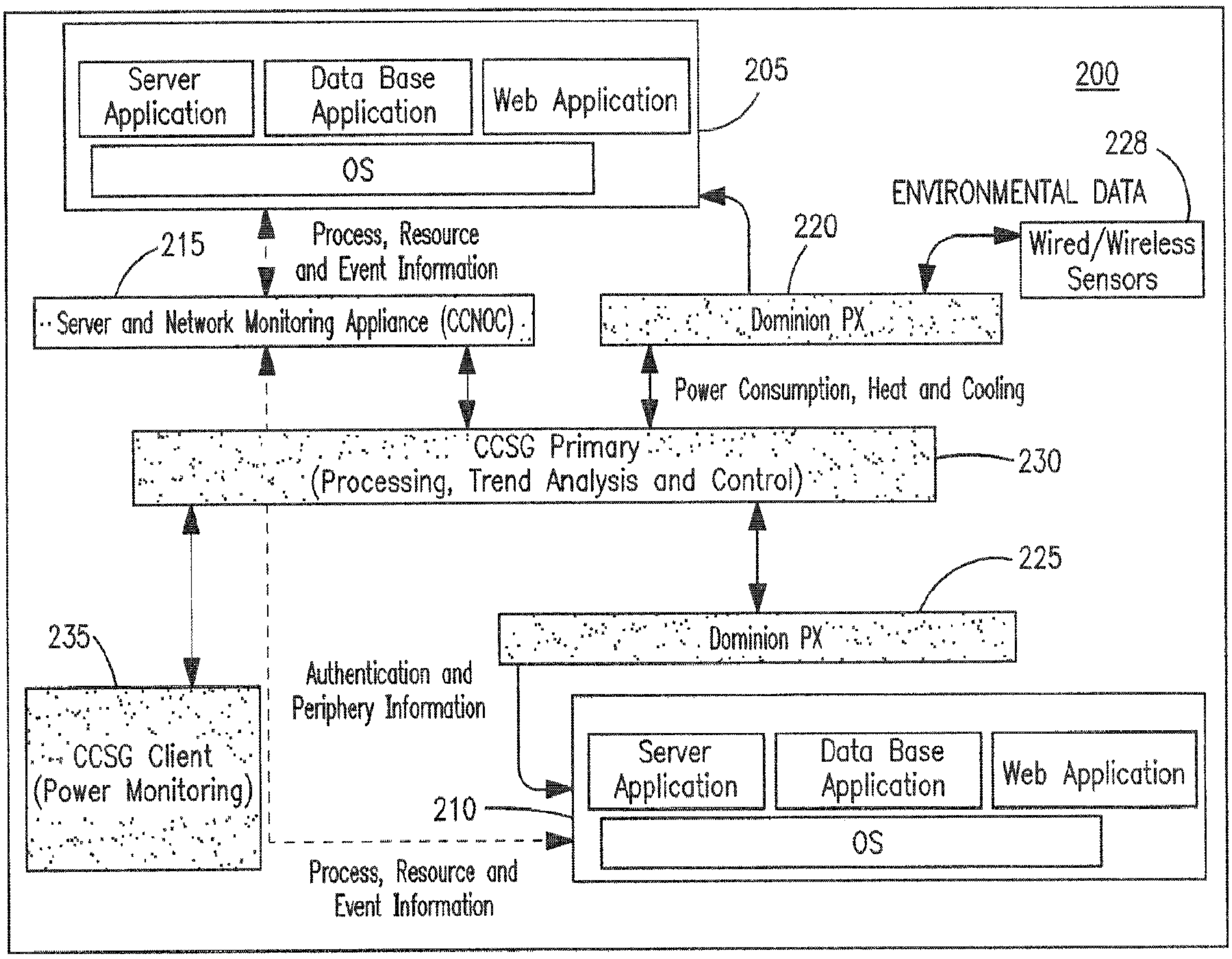

Server Application
Data Base Application
Web Application
OS
205
200
228
ENVIRONMENTAL DATA
Wired/Wireless Sensors
220
215
Process, Resource and Event Information
Server and Network Monitoring Appliance (CCNOC)
Dominion PX
Power Consumption, Heat and Cooling
CCSG Primary (Processing, Trend Analysis and Control)
230
225
Dominion PX
235
Authentication and Periphery Information
CCSG Client (Power Monitoring)
Server Application
Data Base Application
Web Application
210
OS
Process, Resource and Event Information

FIG. 2

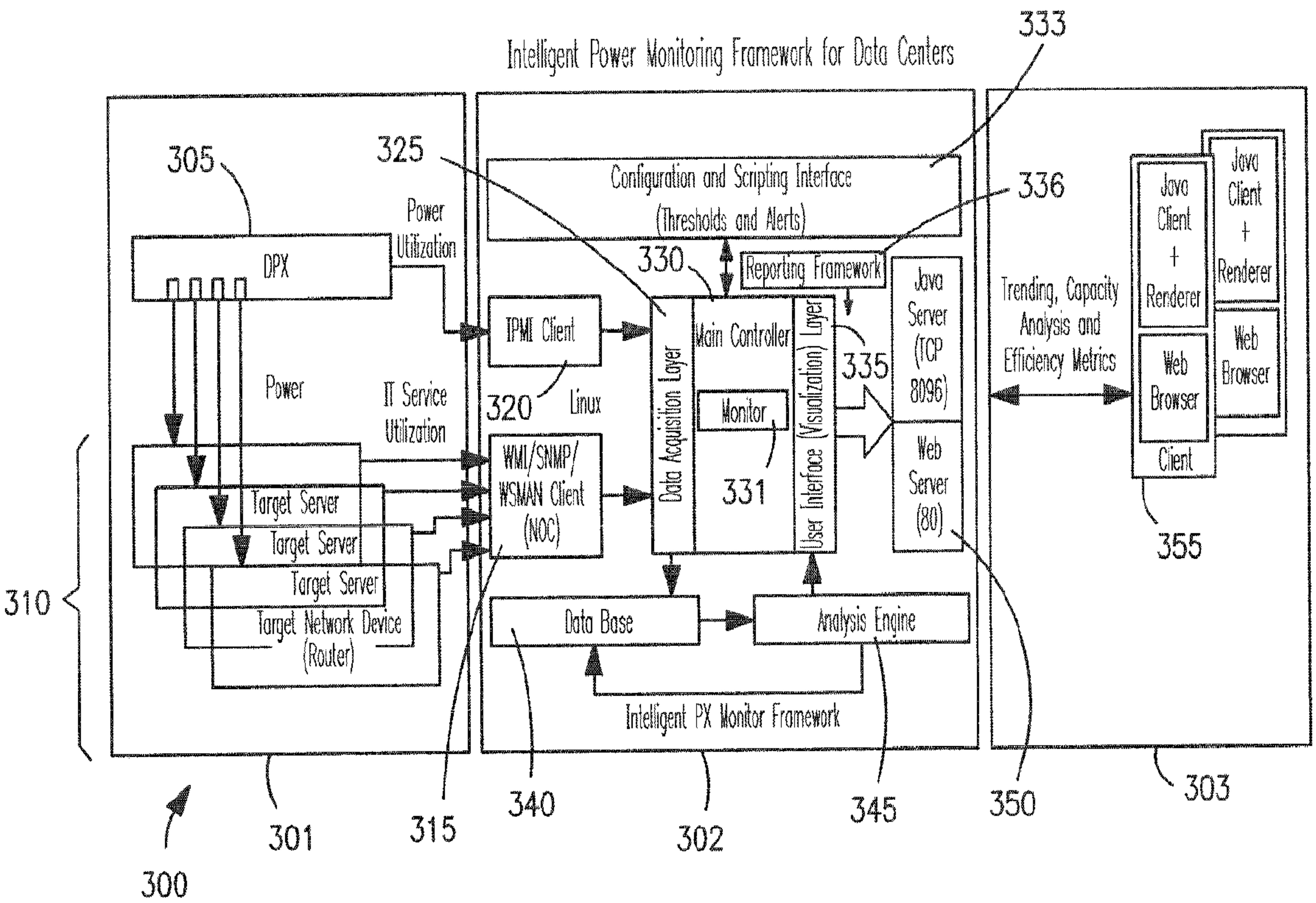

Intelligent Power Monitoring Framework for Data Centers
333
305
325
Power Utilization
Configuration and Scripting Interface
(Thresholds and Alerts)
336
DPX
330
Reporting Framework
Java Server (TCP 8096)
Trending, Capacity Analysis and Efficiency Metrics
Java Client + Renderer
Web Browser
Client
IPMI Client
Main Controller
Data Acquisition Layer
User Interface (Visualization) Layer
335
Power
IT Service Utilization
320
Linux
Monitor
WMI/SNMP/ WSMAN Client (NOC)
331
Web Server (80)
Target Server
Target Network Device (Router)
310
Data Base
Analysis Engine
Intelligent PX Monitor Framework
355
301
315
340
302
345
350
303
300

FIG. 3

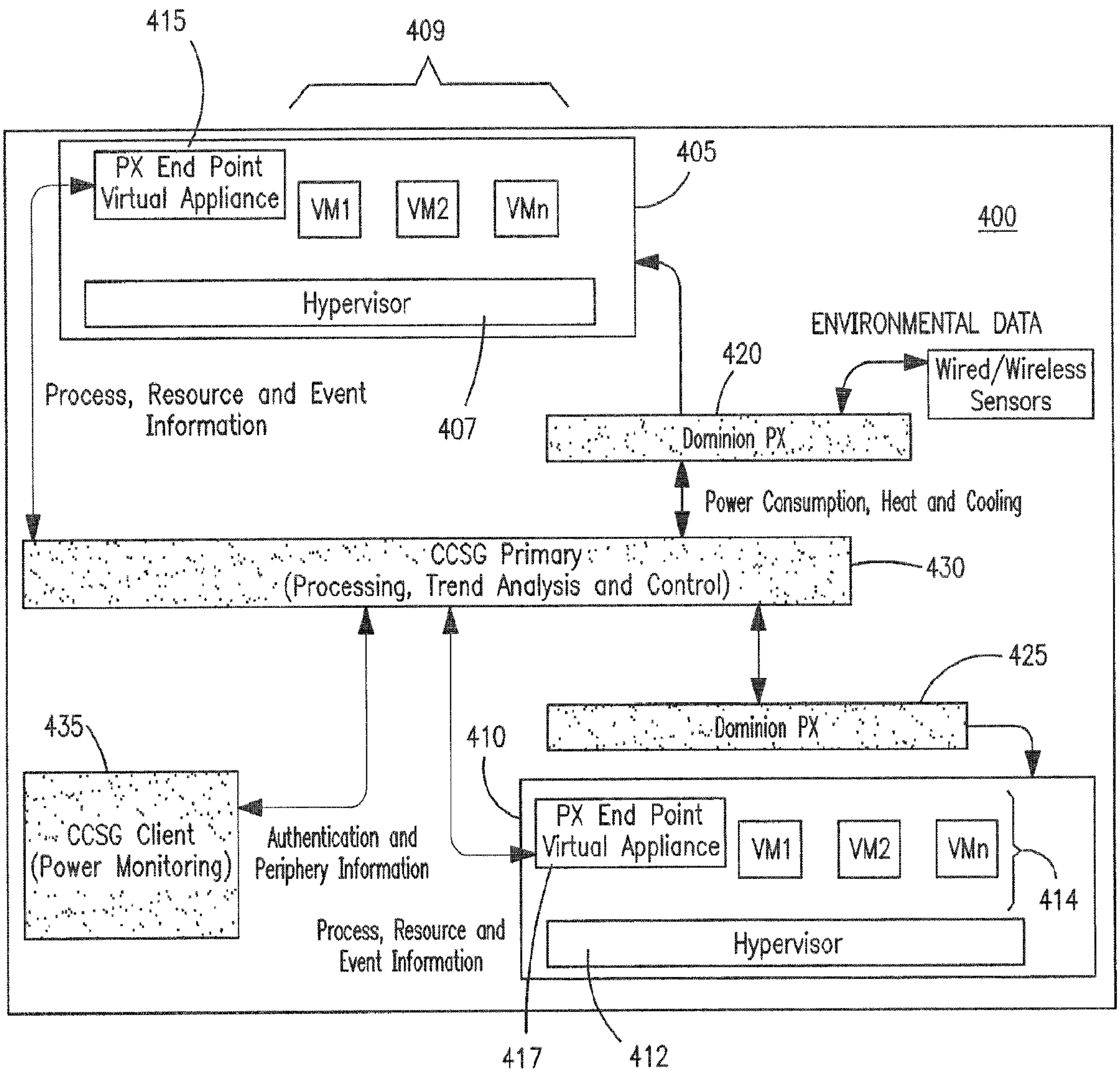

415
409
405
400
PX End Point
Virtual Appliance
VM1
VM2
VMn
Hypervisor
ENVIRONMENTAL DATA
420
Wired/Wireless
Sensors
Process, Resource and Event
Information
407
Dominion PX
Power Consumption, Heat and Cooling
CCSG Primary
(Processing, Trend Analysis and Control)
430
425
435
410
Dominion PX
PX End Point
Virtual Appliance
VM1
VM2
VMn
CCSG Client
(Power Monitoring)
Authentication and
Periphery Information
414
Process, Resource and
Event Information
Hypervisor
417
412

FIG. 4

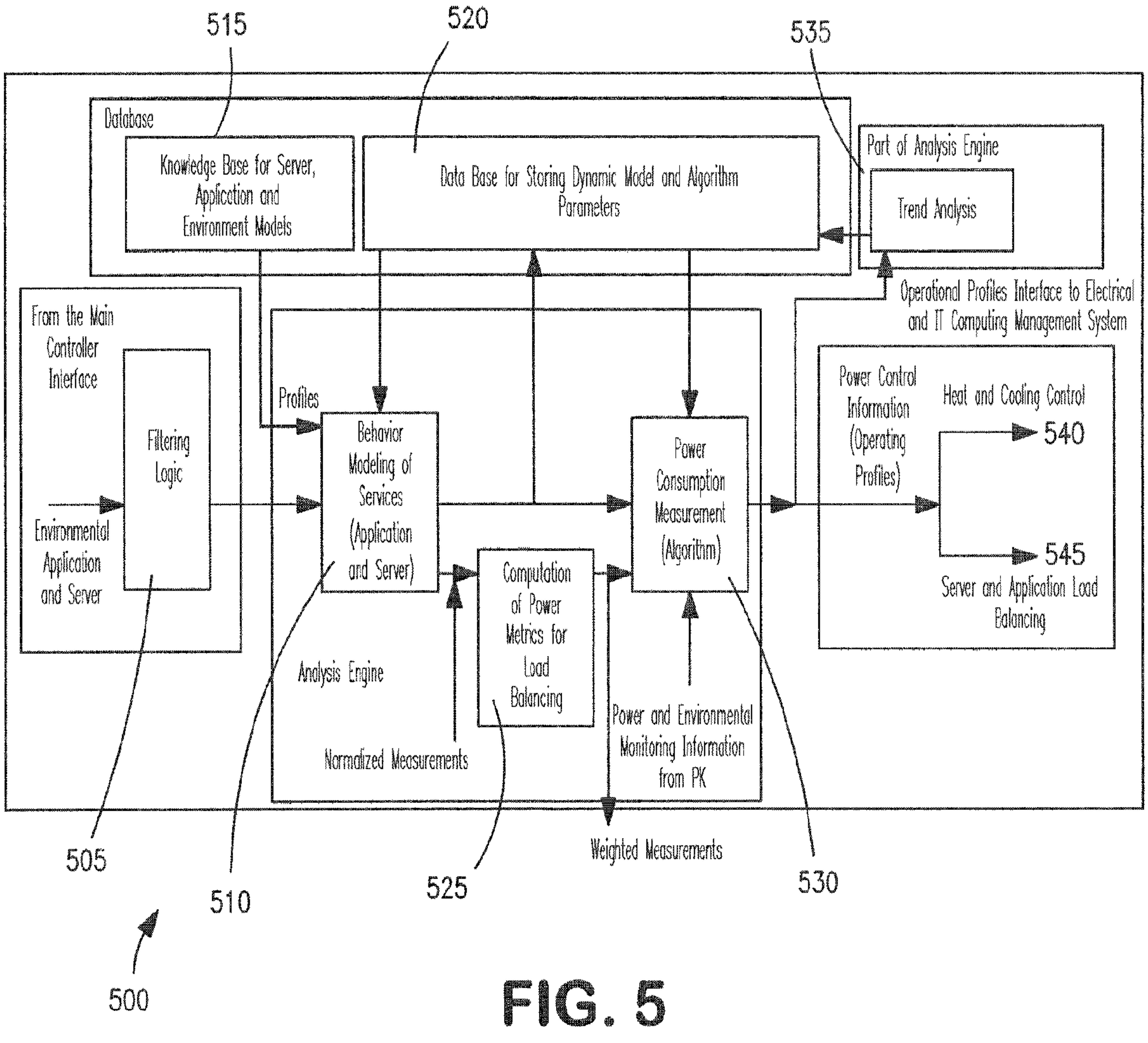
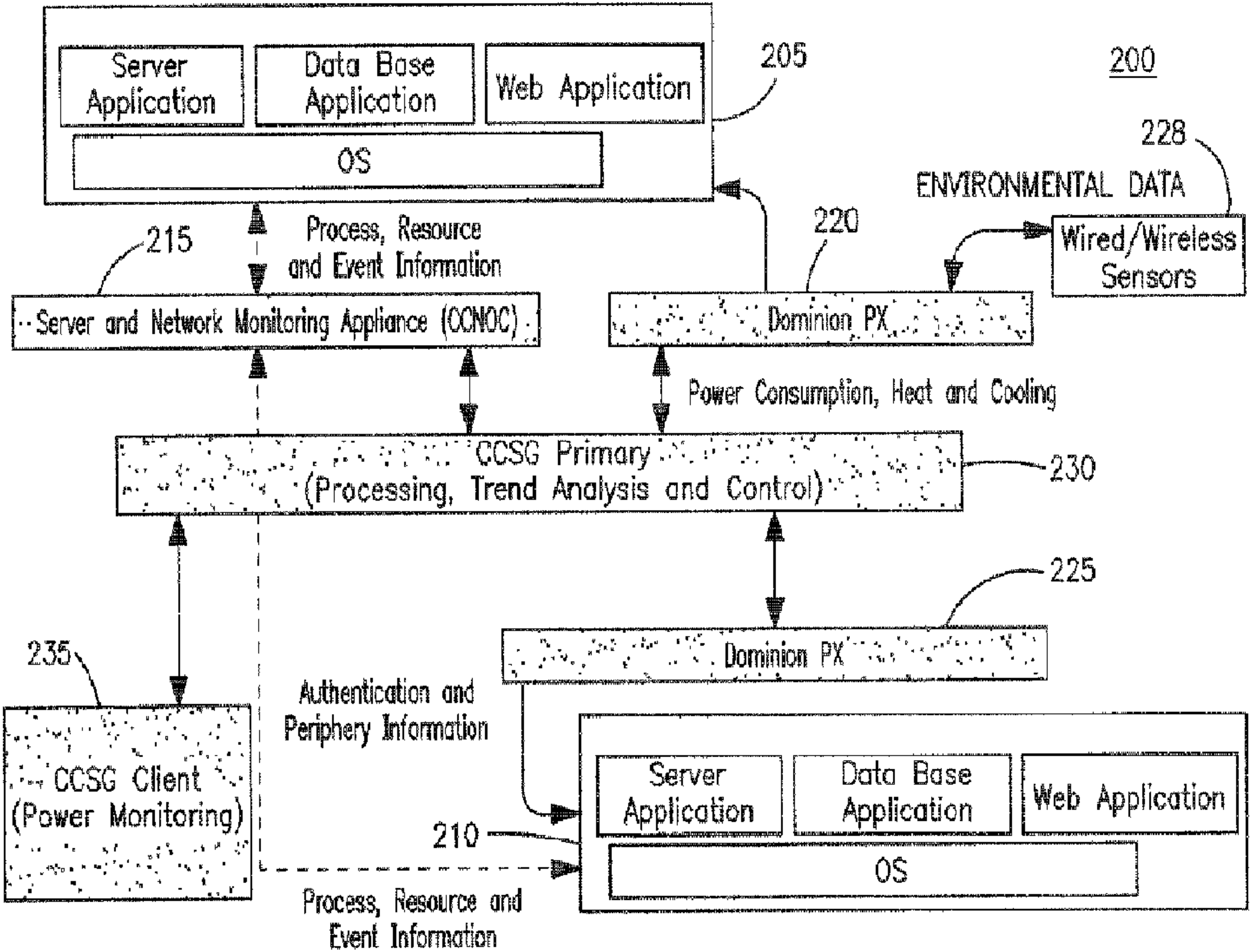

515
520
535
Database
Knowledge Base for Server, Application and Environment Models
Data Base for Storing Dynamic Model and Algorithm Parameters
Part of Analysis Engine
Trend Analysis
Operational Profiles Interface to Electrical and IT Computing Management System
From the Main Controller Interface
Filtering Logic
Environmental Application and Server
Profiles
Behavior Modeling of Services (Application and Server)
Power Consumption Measurement (Algorithm)
Power Control Information (Operating Profiles)
Heat and Cooling Control
540
545
Server and Application Load Balancing
Computation of Power Metrics for Load Balancing
Analysis Engine
Normalized Measurements
Power and Environmental Monitoring Information from PK
Weighted Measurements
505
510
525
530
500

FIG. 5


200
205
Server Application
Data Base Application
Web Application
OS
228
ENVIRONMENTAL DATA
Process, Resource and Event Information
215
220
Wired/Wireless Sensors
Server and Network Monitoring Appliance (CCNOC)
Dominion PX
Power Consumption, Heat and Cooling
CCSG Primary (Processing, Trend Analysis and Control)
230
225
235
Dominion PX
Authentication and Periphery Information
CCSG Client (Power Monitoring)
Server Application
Data Base Application
Web Application
210
OS
Process, Resource and Event Information