Universitatea Politehnica Bucuresti

Facultatea de Electronica, Telecomunicatii si Tehnologia Informatiei

Proiect curs SOA

**Sisteme De Operare Grid - XtreemOS**

Coordonator:

**Conf. dr. ing. S. Stancescu**

 Masterand:

**Cerga Sebastian**

Master IISC

2010

1. **Cuprins**

[**1.** **Cuprins** 2](#_Toc251768800)

[**2.** **Introducere** 3](#_Toc251768801)

[**3.** **Servicii GRID** 4](#_Toc251768802)

[**4.** **Sisteme de operare distribuite** 5](#_Toc251768803)

[**5.** **XtreemOS** 6](#_Toc251768804)

[5.1 Arhitectura 7](#_Toc251768805)

[5.2 Sistemul de gestiune a memoriei 7](#_Toc251768806)

[5.3 Sistemul de gestiune a proceselor 8](#_Toc251768807)

[5.4 Sistemul de gestiune a fisierelor 9](#_Toc251768808)

[5.7 Exemple de Implementari 10](#_Toc251768809)

[**6.** **Bibliografie** 12](#_Toc251768810)

1. **Introducere**

 Ideea de conectare a mai multor computere si folosirea puterii lor de procesare a aparut in ani 1990, cand pe baza unor publicatii s-au creeat mai multe programe de cercetare, printre care le amintim pe cele initate de grupul SETI (Search for Extra-Terrestrial Intelligence), care datorita numarului extensiv de calcule necesare au apelat la public in speranta ca acesta sa ofere timpul nefolsit din procesoarele personale pentru finalizarea acestor calcule.

 Asadar ce a inceput de la niste simple screen-savere s-a extins in noua directie, cloud-computing. Insa inainte de a se trece complet la ,,o gandire” a aparaturii in retea au fost concepute sisteme care sa ajute la creearea unei infrastructurii pentru asemenea initiative.

 Sistemele de calcul distribuit sau GRID sunt, in general, un tip special de sisteme de calcul paralel, care se bazeaza pe computere (ce au incluse procesoare, spatiu de depozitare, alimentarea cu energie electrica, o interfata de retea, etc), conectate la o retea (privata, publica sau prin Internet) printr-o interfata de retea conventionala, cum ar fi Ethernet. Acest lucru este in contrast cu notiunea traditionala de supercomputer, care are mai multe procesoare conectate printr-un bus de mare viteza.

 Avantajul principal al sistemlor de calcul distribuit este ca fiecare nod poate fi achizitionat separat si sa aiba hardware simplu, fapt care poate duce, atunci cand sunt combinate resursele de calcul similare, la un supercomputer multiprocesor, dar la costuri mai mici. Acest lucru se datoreaza economiilor de productie, ce produc hardware de baza, in comparatie cu eficienta mai mica de proiectare si construire a unui numar mic de supercomputere particularizate. Dezavantajul principal de performanta este faptul ca diversele procesoare si spatiile de memorie locale nu au conexiuni de mare viteza. Acest tip de aranjament al nodurie este, prin urmare, potrivit pentru aplicatii in paralel cu mai multe calcule care poate avea loc in mod independent, fara a fi nevoie de a comunica rezultatele intermediare intre procesoare.

 Scalabilitate sistemelor dispersate geografic este, in general favorabila, datorita necesitatii scazute de conectivitate intre noduri in raport cu capacitatea accesului la Internetul local.

1. **Servicii GRID**

 Sistemele de calcul Grid functioneaza pe principiul de resurse partajate in comun. O simpla analogie ar suna astfel. Sa presupunem ca tu si un cuplu de prieteni decideti sa meargeti intr excursie de camping. Ai un cort mare, asa ca te-ai oferit voluntar sa-l impartiti cu totii. Unul dintre prietenii tai se ofera sa aduca produsele alimentare si un altul spune ca va duce transporta intregul grup in SUV-ul lui. Odata proniti in excursie, trei dintre voi, impartasindu-si cunostintele si abilitatile, pot face calatoria mai distractiva si confortabila. Daca ati fi facut excursia pe cont propriu, ati fi avut nevoie de mai mult timp pentru a aduna resursele de care ai nevoie si ar fi trebuit, probabil, sa lucrezi mult mai mult in excursia in sine.

 Un sistem de calcul de retea foloseste acelasi concept: cota de incarcare pe mai multe computere pentru a finaliza sarcinile mai eficient si mai rapid. Inainte de a merge mai departe, sa aruncam o privire la resursele unui computer:

    \* Unitatea centrala de prelucrare (CPU): Un procesor este un microprocesor care efectueaza operatiuni matematice si directioneaza date la diferite locatii de memorie. Computere poate avea mai mult de un procesor.
    \* Memorie: Memoria mentine date relevante la indemana pentru microprocesor. Fara memorie, un microprocesor ar trebui sa caute si sa preia date de la un dispozitiv de stocare permanenta, cum ar fi o unitate hard disk.
    \* Depozitare: In termeni Grid computing, de stocare permanenta se ocupa la dispozitive de stocare a datelor clasice cum ar fi unitati de hard disk organizate sau nu in baze de date.



**Figura 1. Retea Grid classica**

 In mod normal, un computer poate functiona doar in limitele propriilor resurse. Exista o limita superioara pentru a cat de repede se poate finaliza o operatie sau cat de multe informatii poate stoca. Cele mai multe calculatoare upgradabile, ceea ce inseamna ca este posibila adaugarea de mai multa putere sau capacitate pe un singur computer, dar asta ofera doar o crestere incrementala in performanta.

 Serviciile Grid se folosesc de aceasta structura, fie prin implementarea unui soft specializate pe o platforma deja stabilita (Globus Toolkit peste un server Linux) sau, mai nou, prin folosirea unui sistem de operarea distribuit (ExtreemOS, Legion, PhantomOS, ChromeOS etc. )



**Figura 2. Serviciu Grid accesibil prin o interfata Web**

Pas 1. Cerere de raspuns pentru un set de date / Pas 2. Efectuarea Calcule / Pas 3. Trimitere rezultat

1. **Sisteme de operare distribuite**

 Sistemele de calcul GRID au facut progrese masive in decursul ultimului deceniu. Cu toate acestea, exista unele bariere in calea adoptarii pe scara mare, ce au fost, in mare, create de platformele middleware existente pana in acest punct. In prezent, retelele GRID sunt departe de a fi usor de implementat si dificil de intretinut si sunt necesare imbunatatiri ale tehnologiei GRID.

 O implementare a unei aplicatii in retelele existente Grid presupune instalarea pe echipamente a mai multor straturi de software. Cele mai multe dintre acestea nu ofera o baza pentru toate tipurile de calcule intalnite. Middleware-ul GRID existent pote creea topologii de retea inflexibile. Deseori, aceste topologii nu sunt tolerante la erori asa cum se prevede in retele GRID dinamice orientate spre serviciu. Globus Toolkit, una dintre platformele GRID cele mai apreciate, nu se adreseaza mediilor caracterizate de cereri de aderare, ca membru, foarte mari si extrem de dinamice.

 Componente arhitectural ale unui sistem de operare GRID trebuie sa fie modulare. Componentele trebuie concepute pentru a fi auto-configurabile si plug-and-play, in scopul de a
facilita dezvoltarea rapida (exemplu: adaugarea unui nod la un site implica simpla procedura de autentificare, adaugarea unui site la o regiune, implica o procedura simpla de inregistrare, cu un regiune adiacenta).

1. **XtreemOS**

 Sistemul de operare XtreemOS ofera retelelor GRID ceea ce prevede un sistem de operare traditional pentru un singur computer: abstractizarea partii hardware si partajarea sigura de resurse intre utilizatori diferiti.

 Astfel, se simplifica activitatea utilizatorilor care apartin unor organizatiilor virtuale, oferindu-le iluzia de a folosi un calculator traditional, in timp ce sarcina de gestionare a resurselor mediului (tipic Grid) cade in incidenta sistemului. Atunci cand un utilizator executa o cerere pentru XtreemOS, sistemul de operare gaseste automat toate resursele necesare pentru executare, configureaza accesul utilizatorului cu privire la resursele selectate si porneste aplicatia.

 Sistemul de operare XtreemOS ofera servicii de trei mari distribuite pentru utilizatori: Managementul executiei aplicatiilor (furnizarea de descoperire scalabila a resurselor si de planificare executie pentru aplicatii interactive distribuite), gestionare de date (accesare si stocarea datelor in XtreemFS, un sistem de fisiere similar cu POSIX care se intinde peste GRID) si gestionarea organizatiei virtuale (construirea si exploatarea organizatiilor virtuale dinamice).

  Punerea in aplicare a acestui nou sistem de operare Grila introduce noi provocari:

* **Scalabilitate**: sprijinirea sute de mii de noduri si milioane de utilizatori ce adera si parasesc reteaua dinamic.
* **Transparenta**: ascunderea complexitatii retelei Grid de catre serviciile sistemului de operare distribuit care sa permita rularea de aplicatii noi sau mai vechi perfect.
* **Interoperabilitate**: respectarea standardele de majore, cum ar fi POSIX si Saga,
* **Fiabilitate**: furnizarea de servicii fiabile cu disponibilitate ridicata prin checkpointing si de replicare,
* **Securitate**: acordare de privilegii / restrictii conformitate cu politici de securitate personalizabile.

##  5.1 Arhitectura

 

**Figura 3. Arhitectura XtreemOS**

 Arhitectura XtreemOS este impartit in mod logic, in doua straturi.

 Stratul de XtreemOS Foundation, XtreemOS-F, ofera o modificare a Kernel-ului Linux prin adoptarea unui suport nativ pentru Organizatii Virtuale. Acest lucru va fi furnizat in trei variante majore: o versiune care vizeaza PC-uri si statii de lucru; o versiune ce are scopul de a oferi un singura imagine de sistem pentru retele de calculatoare dispersate, bazat pe sistemul existent Kerrighed, precum si o versiune care poate fi implementata pe mici dispozitive mobile, cum ar ca PDA.

 Stratul suport XtreemOS Grid, XtreemOS-G, ce ofera servicii de sistem de operare distribuit Grid pentru a gestiona in siguranta resursele de date si de calcul. Principalele servicii furnizate in G-Layer sunt Managementul Executiei Aplicatiilor (AEM), un sistem de fisiere Grid (XtreemFS), precum si aplicarea unui nivel de securitate. In scopul de a dezvolta XtreemOS peste un anumit numar de noduri participante, o infrastructura de servicii de disponibilitate mare vor gestiona aceste noduri. Aceasta ofera, de exemplu, asistenta pentru publica / subscrie servicii, de virtualizare a nodurilor, si la liste de servicii disponibile pentru nod.

##  5.2 Sistemul de gestiune a memoriei

 XtreemOS fiind o distributie Linux are ca baza standardul POSIX (Portable Operating System Interface [for Unix]), managementul Linux standard cu aditia unui sistem de check-pointuri (figura 4.).

 Asadar acest mecanis pe nalga clasicele spatii pentru utilizator si pentru kernel adauga o noua dimensiune prin inglobarea lor intr-un asa numit job ce se va executa cu o anumita secventialitate fata de alte asemenea joburi existente in GRID, datele returnate de ele ducand la rezultatul final al calculelor.



**Figura 4. Arhitectura Checkpointer-ului XtreemOS**

##  5.3 Sistemul de gestiune a proceselor

 Cele trei module sunt implicate in procesul de gestionare a proceslor sunt: Sched, EPM
si Proc.
 Modulul Proc pune in aplicare procesul de gestionare la nivel de cluster. Se ocupa cu
identificatori de proces global, semnalizare, a mentinerii unor legaturi intre proces mama si copil,
respectiv terminarea procesului.

 Modulul EPM pune in aplicare de gestionare a procesului de functii avansate, cum ar fi
procesul de migratie, duplicarea, punctul de control si de garda.

 Modulul Sched are sarcina de a echilibrare a incarcatrii intre nodurile din cluster.

 Modulul EPM (Enhanced Process Management) accepta in prezent procesul de migratie, de fork la distanta.
 Kerrighed pune la dispozitie urmatoarea comanda shell pentru a migra unui anumit proces
(cu pid cunoscut) la un nod dat (cu nodeid cunsocut). Aceasta comanda poate fi executata de la orice nod cluster.
 *migrate <pid> <nodeid>*

 Procesul de migrare se bazeaza pe urmatoarele module:
 **Ghost** - pentru a trece starea procesului in ,,fantoma”
 **Container** - pentru a obtine informatii cu privire la procesele de mama si copiii, sa migreze
procesul de spatiu de memorie la cerere, pentru a face lucra cu fisierele deschise,
 **Dynamic Stream** – pentru a migra extremitatile fluxuri de date.

 Atunci cand un proces cu fisiere deschise este migrat, fisierele deschise sunt pur si simplu re-deschise pe nodul tinta in cazul in care acestea exista acolo.
 Metoda fork de la distanta este pusa in aplicare bazandu-se pe aceleasi mecanisme
ca si procesul de migrare

 Operatiuni EPM sunt executate prin trimiterea unui semnal intern pentru procesul tinta. Operatiuni care trebuie executate de acest proces sunt descris intr-un camp dedicat din stuctura task-ului asociat.

##  5.4 Sistemul de gestiune a fisierelor

 Sistemul de gestiue al fisierelor in XtreemOS este denumit XtreemFS si pune in aplicare o arhitectura bazata pe obiect a sistemului de fisiere.
Continut este divizat intr-o serie de obiecte de dimensiuni fixa si depozitate in intreaga arie de stocare a serverelor, in timp ce metadatelesunt stocate pe un server separat de metadate. Meta -
server de date organizeaza aceste metadate ca un set de volume, dintre care fiecare
pune in aplicare un spatiu de nume separat al sistemului, in forma de un arbore de director.
 Spre deosebire de sistemele bazate pe blocuri, managementul spatiului de depozitare disponibil si utilizat este descarcat din serverul de metadate la serverele de stocare.
 Mai degraba decat liste inode cu adrese de bloc, sistemul de fisier al metadatelor contine liste de servere de stocare responsabil pentru obiecte, impreuna cu politici de distribuire care sa determien offsetul in baiti si ID-urile de obiect. Acest lucru implica faptul ca
marimile de obiect poate varia de la o implementare la alta, oferind flexibitate.

 XtreemFS contine trei tipuri de de servere care poate rula pe unul sau mai multe masini:
 **DIR - Directory Service** Directorul de servicii este un registru central pentru toate serviciile din XtreemFS. MRC il foloseste pentru a descoperi servere de stocare.
 **MRC - Catalog metadata si replica** MRC contine arborele de directoare metadate, cum ar fi numele, dimensiune sau ora de modificare a fisierelor. Mai mult, MRC autentifica utilizatorilor si autorizeaza accesul la sistemul de fisiere.



**Figura 5. Arhitectura XtreemFS**

 **OSD - Object Storage Device** Un OSD este o colecite arbitrara de obiecte; clientilor citeasc si scriu date pe OSD-uri.

##  5.7 Exemple de Implementari

 **IMA**

 IMA este o aplicatie de mesagerie instata orientata pentru a facilita task-uri specifice XtreemOS, cum ar fi inregistrarea intr-un VO, stocarea de fisiere de configuratie si a logurilor conversatiilor in retea (prin XtreemFS). Este planuita adaugarea de contacte in functie de membrii VO si posibilitatea unui schimb de fisiere intre clienti. Va fi disponibul in XtreemOS 2.1 versiunea pentru dispozitive mobile.



 **Figura 6. Interfata IMA**

 **jCAE**
 jCAE este o platforma CAE extensibila (Computer Aided Engineering). In versiunea sa open source, jCAE este o aplicatie Netbeans dezvoltata dintr-un modelator CAD simplu, cu suport pentru suprafete mesh CAD si cu un view-er interactiv 3D. Modelatorul CAD isi are radacinile in cadrul OpenCASCADE si este capabil sa combine obiecte simple (cutii, cilindri, toruri, ...) prin intermediul operatiunilor de tip boolean, in scopul de a construi obiecte complexe. Modulul Mesher este capabil sa genereze obiecte formate din triunghiuri pe baza de la definitiei CAD. Ceea ce face mesher jCAE unic este capacitatea sa de a lucra pe mai multe nuclee; prin urmare, este capabil sa genereze meshe mari (de pana la cateva sute de milioane de celule). Va fi disponibil cu XtreemOS 2.0.



 **Figura 7. Interfata jCAE**

 **JobMA**
 Aplicatia JobMA ofera un acces intuitiv la management de job-uri ale XtreemOS si al instalatiilor de monitorizare de pe dispozitive mobile. Aplicatia permite utilizatorilor sa lanseze, opreasca, reia si anula joburi din retea intr-un mod simplu si intuitiv. Cu JobMA versiunea 2.0, acesta va fi capabil sa vizualizeze starea jobului, in timp real, sa genereze statistici personalizate si sa efecteze masuratorile asupra acestor joburi din retea. Interfata grafica este special conceputa pentru afisearea pe telefonul mobil. JobMA 2.0 va fi disponibila in versiunea XtreemOS 2.1.



 **Figura 7. Interfata JobMA**

1. **Bibliografie**

http://en.wikipedia.org/wiki/Grid\_computing

http://legion.virginia.edu/presentations/UMBC-Apr2001/index.htm

http://www.springerlink.com/content/539lrllmekbfqbpv/

http://www.xtreemos.eu/

http://www.xtreemos.eu/science-and-research/plonearticlemultipage.2007-05-03.6408426978/copy\_of\_xtreemos-architecture/

http://www.xtreemos.eu/overview/plonearticlemultipage.2006-06-08.9297943452/project-summary

http://www.xtreemos.eu/overview/plonearticlemultipage.2006-06-08.9297943452/project-objectives

http://www.xtreemos.eu/publications/plonearticlemultipage.2008-06-26.0232965573/public-deliverables

http://www.xtreemos.eu/software/application-samples

http://www.xtreemos.eu/demonstrations/plonearticlemultipage.2009-10-19.1771310160/featured-xtreemos-use-cases/

http://users.utcluj.ro/~somodi/so2/index.html

http://www.elcom.pub.ro/discipline/so/