

Yüksek Performanslı Hesaplama Sistemlerinde

Yeni Yaklaşımlar: GRID Hesaplama

İrfan MACİT, imacit@cu.edu.tr,

Çukurova Üniversitesi Endüstri Mühendisliği Bölümü 0322 3387360

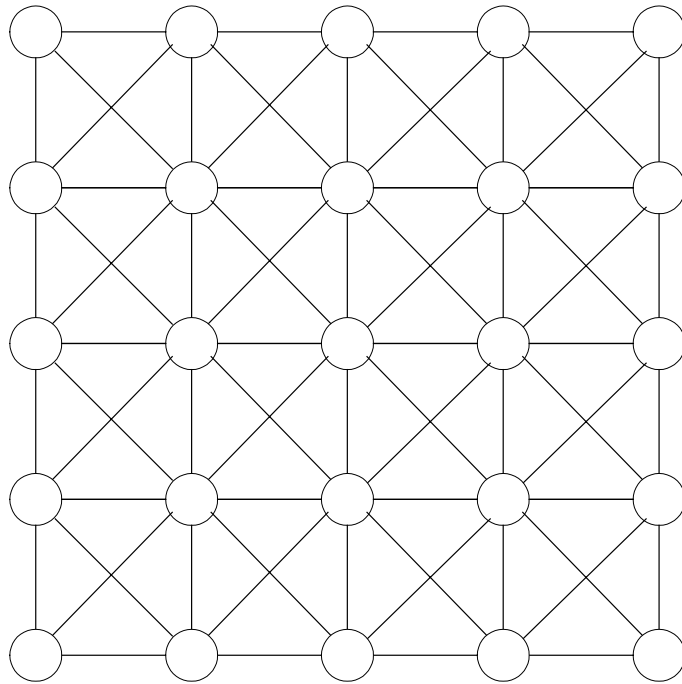
Özet

Hesaplama sistemlerinin gücündeki ve veri paylaşımındaki artan talep teknolojik avantajlar ile bileştiğinde özellikle de bilgisayar ağlarındaki hızın artmasıyla geleneksel hesaplama sistemlerini klasik hesaplama sistemlerinin ötesine taşımaktadır. Kişisel bilgisayar kullanıcıları kendilerine ait bilgisayarların atıl durumlarından başkalarının faydalanmasını isteyebilir veya bunları belirli kurallar dahilinde kiralayabilir. Yararlanma veya kiralama kullanıcılar tarafından seçilen kriterlere ve özelliklere göre belirlenebilir. Kişisel bilgisayar kullanıcıları günlük mesai dışında kullanılması bu tip sistemlerinin geliştirilmesine örnektir. Kişisel bilgisayarlara ait atıl kapasite geliştirilen algoritmalar ve yazılımlar ile uygun şekilde bir araya getirilir veya kümelenir. Bir araya gelen bilgisayarlar veya kümeleri bu uygun yapı ile yüksek hesaplama gücüne ulaşabilmektedir. Grid hesaplama sistemi geliştirilen şeffaf arabirimler sayesinde birçok bilgisayarın bir arada çalışmasına olanak sağlamaktadır. Bu sayede bir çok bilgisayar kolay kullanılabilir ve istenen işlem gücüne rahatlıkla erişilir. Erişilen bu işlem gücü ile birlikte uzun süren hesaplamalar kısa zaman içerisinde bitirilir.

1.Giriş

Sayısal hesaplama yöntemleri karmaşıklaşan günümüz problemlerini daha kısa sürede ve etkin çözmek için yeni yaklaşımlara ihtiyaç duymaya başlamıştır. Kompleks problemler, sayısal yöntem çözümleri, matematiksel modellere kadar bir çok problem karmaşık şekillerde ifade edilmektedir. Bunun durumunda yeni, hızlı ve güvenilir algoritmalara ve yeni teknolojilere ihtiyaç duyulmaktadır. Geliştirilen algoritmalar yardımıyla yazılan bilgisayar programları ise karmaşık hesaplama yöntemleri dolayısıyla daha fazla işlemci gücüne ihtiyaç duymaktadır. Bununla birlikte diğer önemli bir faktörde hesapların yüksek işlemci kapasitesinin yanı sıra en kısa zamanda ihtiyaç duyulan anda bitirilmesidir. Problemler ihtiyacın ortaya çıktığı anda çözümü elde edilmelidir. Aksi takdirde geçerliliği kalmayabilir. Örneğin bir bilgisayar geçici hafıza (RAM) tasarımı yapılırken kurulan matematiksel modelin çözümü 12-18 ay içerisinde bulunuyorsa değişen teknoloji bu çözümü gereksiz kılabilir. Hızla değişen bilgi ve teknolojisi problem karmaşası zaman maliyetini ve çözüme ulaşma sırasındaki gecikmelerin maliyetini de arttırmaktadır. Anlık olarak değişim gösterdiğinden dolayı maliyet faktörleri de

çözümün en kısa zamanda olmasını gerektirmektedir. Zaman maliyetini düşürmek için yapılacak yeni yatırımlar yerine elde var olan kaynakları etkin kullanmak da değişen koşullara uymak yeteneği olarak karşımıza çıkmaktadır. Genel olarak Grid hesaplama sistemleri birbirine bağlı veya birbirinden bağımsız olarak tasarlanabilmektedir, Şekil 1. her bir bağlantı bir düğüm (node) olarak tasarlanır. Her düğüm birbirine bağlı veya birbirinden bağımsız olabilir. Geliştirilen yazılımlar bu konuda çeşitli seçenekler sunmaktadır. Bu sistemlerin tasarlanmasındaki genel amaç kısa sürede ve yüksek maliyetlere ulaşmadan yüksek başarılı hesaplamaları başarabilmektir. Bu sistemlerin daha önce yapılmış olan maliyetleri bilgisayar maliyetleri değerlendirildiğinde atıl durumdaki kapasitelerini veya bekleme durumundaki konumlarından yararlanarak katma değer oluşturmaktır.

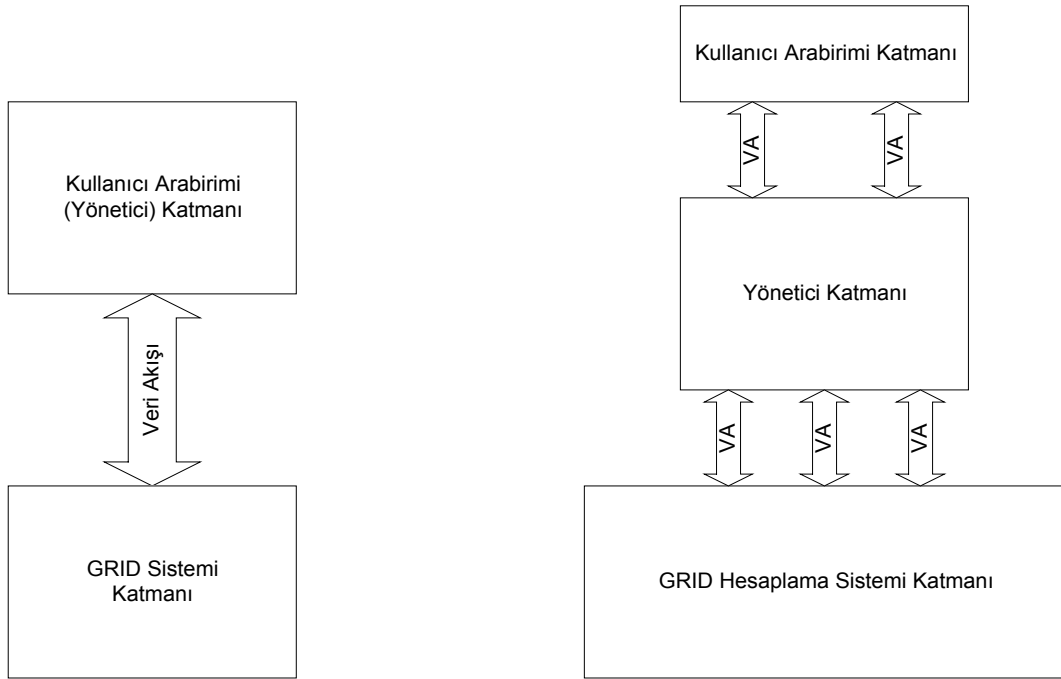


Şekil 1. GRID (Izgara)

Geleneksel hesaplama sistemlerinde kullanılan ana çatı (main frame) donanımın maliyetleri çok yüksektir. Ayrıca geleneksel hesaplama sistemlerinde bakım ve onarım maliyetleri ise özel durumlar oluşturduğundan önerilen sistemlere göre çok pahalıdır. Bir karşılaştırma yapmak gerekirse elektrik kesilmesinde veya havalandırma fanının arızalanması geleneksel hesaplama sistemleri (vektör bilgisayarları, main frame) yüksek hata oranları ile çalışmaya başlayabilir veya hesaplama yarıda kesilebilir, grid hesaplama sistemlerinde ise durum farklıdır. Hesaplama yapan bilgisayarlar birbirinden coğrafik olarak bağımsız olduğundan bir düğümde veya bilgisayarda oluşabilecek hata durumunda programın çalışan

diğer parçacıkları oluşan hata durumundan etkilenmeyerek hesaplama işlemine devam edecektir.

Hesaplama sisteminde tüm sistemi kontrol edebilecek yazılımlara ihtiyaç vardır. Bilgisayarlar coğrafik olarak birbirinden bağımsız olduğu için bir verinin işlenmesi sırasında veya çalışan bir işlemin kontrol edilmesi için bilgisayarların başına gitmek imkansız olabilmektedir. Grid sistemlerindeki kullanıcılar gönderilen verilerin hangi projede veya verinin programın hangi parçasını oluşturduğunu karmaşık olduğundan bilemezler. Bu durumdan dolayı bir yönetim programına ihtiyaç duyarlar. Bu programlar grid yönetim sistemini oluşturan programlardır. Yönetim yazılımları tek katmanlı veya çok katmanlı mimari ile yazılabilmektedir.



a) İki Katmanlı Yapı

b) Üç Katmanlı Yapı

Şekil 2. GRID Yönetim Yapısı ve Katmanlar

2. Kısıtlı Kaynaklardan Yararlanma ve Kaynak Problemleri

Coğrafi olarak birbirinden bağımsız olan bilgisayar kümelerindeki işlem yükleri ve çalışma zamanları, sosyal ve ekonomik koşullar gereği değişmektedir. Kıtalar arasındaki saat farkları boşa bekleyen bilgisayar kaynaklarını ortaya çıkarmaktadır. Uzakdoğu kıtasında mesaisine başlamakta olan firmalarda işbaşı yapılırken Avrupa kıtası geceyi yaşamaktadır.

Diğer bir deyişle mesai olmadığından dolayı insanlar evlerinde veya eğlence yerlerinde zaman geçirmektedirler. Bu durumda iş yerlerinde kullanılan bilgisayarlar boşta beklemekte veya kapalı durmaktadır. Gündüz mesai saatlerinde belirli bir çalışma oranı ile çalışan bilgisayarlar akşam olduğunda mesai haricinde uyku halinde veya kapalı beklemektedir. İşte bu durumlarda yapılan yatırımın burada bilgisayar ve alt yapı yatırımları için zaman maliyetini ortaya çıkarmaktadır. Yapılan yatırımlar sadece belirli zaman dilimleri içerisinde kullanılmakta olduğundan kaynakların kullanım maliyetleri yüksek çıkmaktadır. Bilgisayar için yapılan yatırımlar çabuk eskijen teknoloji olduğundan teknolojik ömürleri sırasında yüksek kullanım oranları ile kullanılmalrı gerekmektedir. Kaynakları kullanılmadıkları veya atıl kaldıkları sürelerde yapılan yatırımları geri ödeme sürelerini ve maksimum fayda sağlama diğer bir deyişle yararlanma süreleri uzatmaktadır.

Zaman ve kullanım maliyetini düşürmek için yapılan bilgisayar yatırımlarının kullanım oranlarını teknolojik ömürleri sırasında yükseltmek gerekmektedir. Örneğin, geceleri uyku durumunda veya kapalı bekleyen bilgisayarları deęişik yöntemlerle belirlenen işlere tahsis etmek gibi. Diğer önemli bir yatırım olan altyapı yatırımları ise bu yatırımların değerlendirilmesi sırasında çok önemli ve yüksek maliyetli bir yatırım türüdür. Bu yatırımlardan da en fazla faydayı elde etmek gerekmektedir. Zaman ve kullanım maliyetlerini belirleyen çeşitli unsurlar vardır. Maliyetin hangi birimden ve nasıl oluşturulacağını da belirlenmesi gerekmektedir ve bu maliyeti hesaplarken nasıl hesaplanacağı ile birlikte ortak bir birim cinsinden yapmak gerekmektedir. Bu ortak birim yapılan işgücü, bakım onarım, işlemci süresi, kaynakları kullanım oranı gibi faktörleri kapsayacak birim olmalıdır. Bu ortak hesaplama birimine ise grid doları denmektedir. Grid doları kullanılan tüm kaynakların maliyetini ortak birim cinsinden ödenmesine olanak sağlamaktadır. Finanssal sistemlerde ortak ödeme birimi ödeme ve değerlendirme işlemlerini kolaylaştırmaktadır.

Grid hesaplama sisteminde mevcut uygulamaları farklı bilgisayarlarda çalıştırılabilir olmalıdır. Bir uygulamayı normal olarak çalıştıran bilgisayar da oluşacak ani işlem yükünde uygulama grid üzerinde başka bir bilgisayara gönderilir. Bunun için en az iki ön şart gereklidir, ilki uygulama çalıştırılabilir olmalı, ikincisi uzak bir bilgisayar uygulamanın çalıştırılması için uygun yazılım ve donanım veya benzeri kaynak sağlamalıdır.

Uzak bilgisayarlara erişim ve veri haberleşmesi için band genişliği ve sistemin işlemci gücü önemli bir etkidir. Band genişliği grid sistemi için önemli bir kavramdır, Brown, 2003. Band genişliğini etkileyen faktörlerden en önemlileri işlemci ve ağ kullanım oranlarıdır. Bu oranların dağıtımı doğru ve ayarları oluşturulan sisteme göre spesifik olarak oluşturulmalıdır. Aksi takdirde sistemde kayıplar ve istenmeyen güvensiz durumlar ortaya çıkacaktır.

Bir çok firma veya kurum çalıştırdığı kişisel bilgisayarların toplama işlemci kapasitesinin ortalama % 5 kullanmaktadır. Benzer şekilde toplam disk kapasitesinin % 30'unu kullanmaktadır. Bilgisayarlar aylak yani boş kaldığında uygulamaları çalıştırıp grid hesaplama için gerekli olan veri yığınları oluşturulabilir. Bu durum sonucunda aylak halde bulunan kaynaklardan etkin şekilde yararlanılmış olur.

3. Uygulama Alanları ve Kaynakların Yönetimi

İlaç, Tıp, finansal modelleme, hava tahminleri v.b. yoğun bilgisayar tarafından işlenecek verilerin sonucuna ihtiyaç duyular. Bu sektörlerde bilgisayar işlemci gücüne çok ihtiyaç vardır. Geliştirilen uygun algoritmalar ve teknolojik imkanlar ile birlikte alt yapı ve programları ile bahsedilen sektörlerdeki uygulamalara göre parçalara ayrılan program parçacıkları değişik bilgisayarlarda çalıştırılarak sonuçlar üretilebilir. Tek başına çalıştırılan bilgisayarlardan üretilen sonuçlardan çok daha kısa zamanda sonuçlara ulaşılabilir.

GRID hesaplama uygulamalarını göz önüne alınacak olursak üretilen verilerin taşınması, birleştirilmesi için belirlenecek araçlarında uygulamaları çalıştıracak kabiliyette geliştirilmesi gerekmektedir. Teknik haberleşme alt yapısının oluşturulması bazı durumlarda ek maliyetlere yol açabilir. Önceleri kişisel kullanım için tasarlanan bilgisayarlar daha sonradan grid sisteminde kullanılmak üzere tahsis edildiğinde haberleşme ağının yatırımı ve optimizasyonu gerekebilir.

Sanal kaynaklar ve sanal organizasyonlar grid hesaplama katılımcıları için diğer önemli bir ayrıntı ise kaynak paylaşımının birleştirilmesini mümkün hale getirmektir. Paylaşma ile veri dosyaları diğer bilgisayarlar veya depolama sistemleri üstüne kaydedilecek kopyaları çıkarılır ve bilgilerin otomatik olarak yedeklenmesi sağlanmış olur. Yedekleme sistemi ise geliştirilen çizelgeleme veya bir bağlaç yardımı ile oluşturulabilir. Grid sistemleri daha sonradan geliştirilecek esneklikte tasarlanmaktadır. Sistemler kaynakları , burada kullanılan istemci ve sunucu bilgisayarların çevresel birimlerinden bahsedilmektedir, bilgisayarların donanım güncellemeleri ile de sağlanabilmektedir. Eklenen işlemci ve depolama alanları ile uygulama için yazılan programların çalışacağı ekipmanların kapasitelerini de arttırmaktadır. Örneğin Internet üzerinde aranan bir uygulamayı tüm Internet üzerinde sadece bir bilgisayarın araması ile birkaç bilgisayarın aynı anda araması çok farklıdır.

Grid hesaplama çok büyük miktardaki verileri çeşitli kaynaklara dağıtarak bu kaynaklardan verilere katma değer sağlamaktadır. Büyük veriler işlenirken işlemcinin beklenmeyen bir şekilde yükünün artması durumunda ise uygulamanın değişik kaynaklardan aktarılması

gerekebilir. Bu durumda kaynakların dengelenmesi gerekir. Kaynakları dengelemek için çalıştırılan uygulamalara öncelik verilerek bütün uygulamaların öncelik sırası belirlenir. Karşılaşılan bu problemler grid yönetim katmanında belirlenen kurallar ile aşılmaktadır. Grid yönetim katmanı sistemin genel olarak belirlenecek kuralları uygulamak için kullanılan araçlardan birisidir.

4. Performans ve Performans Hesaplama Yöntemleri

Bilgisayar sistemlerinde yüksek başarımla elde edilebilen birçok parametreye bağlıdır. Kümelenmiş bilgisayar sistemlerinde ise durum ek parametreler ile birlikte daha hassas olmaktadır. Kümelenmiş sistemlerde bilgisayarın kendi hızı yanı sıra diğer bilgisayarlar ile haberleşmesi de bu hassasiyeti arttırmaktadır. Performansı ölçülecek sistemin sonuçlandıracağı işlemin hızı kendi ile birlikte çalışan diğer bilgisayarlar ile haberleşmesine veya ondan gelecek olan işlenen verinin hızına bağlı olarak değişecektir. Gecikme olarak adlandırılan bu işlemler arasındaki geçen süre sistemin performansını direkt olarak etkilemektedir. DeFanti ve Stevens performans gereksinimlerinden dokuz akım tipini yedi ölçülebilir birim ile incelemişler ve aşağıdaki tabloyu elde etmişler. Tablo 1.

Tablo 1 Akım Performans İnceleme Tablosu, DeFanti ve Stevens, S/147.

<i>Akım Tipi</i>	<i>Gecikme</i>	<i>Band Genişliği</i>	<i>Güvenirlilik</i>	<i>Çoklu Yayın</i>	<i>Güvenlik</i>	<i>Akım</i>	<i>Servis Kalitesi</i>
<i>Kontrol</i>	< 30	64 Kb/s	Var	Yok	Yüksek	Yok	Düşük
<i>Yazı</i>	< 100	64 Kb/s	Var	Yok	Orta	Yok	Düşük
<i>Video</i>	< 30	N x 128 Kb/s	Yok	Var	Orta	Var	Orta
<i>Ses</i>	< 100	N x 5 Mb/s	Yok	Var	Düşük	Var	Orta
<i>İzleme</i>	< 10	N x 128 Kb/s	Yok	Var	Düşük	Var	Orta
<i>Veritabanı</i>	< 100	> 1 GB/s	Var	Belki	Orta	Belki	Yüksek
<i>Benzetim</i>	< 30	> 1 GB/s	Karışık	Belki	Orta	Belki	Yüksek
<i>Dokunuş</i>	< 10	> 1 MB/s	Karışık	Belki	Yüksek	Var	Yüksek
<i>Kaplama</i>	< 30	> 1 GB/s	Yok	Belki	Düşük	Belki	Orta

Yapılan analizi incelediğimizde maksimum kabul edilebilir gecikme için band genişliği gereksinimini, güvenilirlik durumu, güvenlik seviyesi, haberleşme katarlarının kullanım ve servis kalitesindeki değişim görülmektedir. Bu analiz sonuçları bize grid sistemindeki haberleşme ve performansı etkileyen faktörler arasındaki ilişki hakkında genel

bilgi vermektedir. Tüm grid sistemlerinin performansının ölçülmesi için bir yaklaşım değildir. Tablodan okunan bilgiler performans değerlendirmeleri sırasında kullanılacak olan parametrelerin değişimi veya hangi ölçekte olabileceği yaklaşık olarak verilmiştir.

Band genişliği gereksimi hesaplamak için kullanılan eşitlik (4.1) aşağıdaki gibidir. Bu eşitlikte N_f f tipindeki katar akım tipi, B_f ise f akım tipi, yazı, ses, görüntü, izleme, veritabanı, benzetim, dokunuş ve kaplama ile ilgili ortalama band genişliğidir.

$$B_{f=c,e,a,v,t,d,s,h,r} = \sum N_f \cdot B_f \quad (4.1)$$

Hesaplanan teorik band genişliği gerçekte uygulamada değişebilmektedir. Bu değişim çeşitli faktörlere bağlıdır. Bu faktörler ve etkileri ayrı bir inceleme konusu olduğu için burada bahsedilmeyecektir.

5. Sonuç ve Öneriler

Grid sisteminin yapılarını kontrol etmek için kullanılan yönetim katmanları iki veya üç katmanlı olarak önerilmektedir. Katmanların hangi mimaride olması gerektiğini kullanılacak donanımın tipi ve işletim sistemi belirlemektedir. Sun Microsystems için önerilen mimari Sun GRID Engine kullanan daha çok üç katmanlı mimaridir. IBM bilgisayar sistemlerinde iki ve üç katmanlı kontrol mimarileri de uygulanabilmektedir. Yazılım mimarileri ve yazılım için kullanılan araçlar değiştikçe yeni yaklaşımlar da ortaya çıkmaktadır. Yazılım araçları geliştikçe daha yeni algoritmalarında geliştirilmesi söz konusudur. Sonuç olarak her yeni yaklaşımda sistemleri daha iyiye taşıyacak yöntemler geliştirilecek kullanıma sunulacaktır.

Kaynaklar

- Brown, M., Bandwidth Management in Grids, IBM Developerworks, Ekim 2003.
- Galli, D., Distrubuted Operating Systems Concept & Practices, Prentice & Hall, 2000.
- Coomer, J., Introduction To GRID Computing I, Sun Blueprints, 2002.
- Coomer, J., Introduction To GRID Computing II, Sun Blueprints, 2002.
- Ferriera, L. Ve Diğ., Introduction to GRID Computing With Globus, IBM Redbooks, 2002.
- Baker, M., Buyya, R., Cluster Computing At A Glance, s3-47, NY,1999
- DeFanti, T., Stevens, R., Teleimmersion, s131-157, Morgan & Kaufmann Publishers, 1999.