

**TÜRK HAVA KURUMU ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ**

**SANALLAŞTIRILMIŞ GRAFİK İŞLEMCİ BİRİMLERİNİN
MALİYET ETKİNLİK ANALİZİ**

YÜKSEK LİSANS TEZİ

Yusuf YAZICI

Bilişim Teknolojileri Anabilim Dalı

Bilişim Teknolojileri Programı

EYLÜL 2016

**TÜRK HAVA KURUMU ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ**

**SANALLAŞTIRILMIŞ GRAFİK İŞLEMCİ BİRİMLERİNİN
MALİYET ETKİNLİK ANALİZİ**

YÜKSEK LİSANS TEZİ

Yusuf YAZICI

1403650028

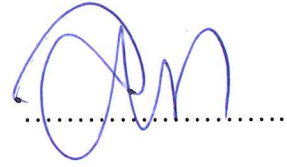
Bilişim Teknolojileri Anabilim Dalı

Bilişim Teknolojileri Programı

Tez Danışmanı: Yrd. Doç. Dr. Tansel DÖKEROĞLU

Türk Hava Kurumu Üniversitesi Fen Bilimleri Enstitüsü'nün 1403650028 numaralı Yüksek Lisans öğrencisi, Yusuf YAZICI ilgili yönetmeliklerin belirlediği gerekli tüm şartları yerine getirdikten sonra hazırladığı “Sanallaştırılmış Grafik İşlemci Birimlerinin Maliyet Etkinlik Analizi” başlıklı tezini, aşağıda imzaları olan jüri önünde başarı ile sunmuştur.

Tez Danışmanı : Yrd. Doç. Dr. Tansel DÖKEROĞLU
Türk Hava Kurumu Üniversitesi



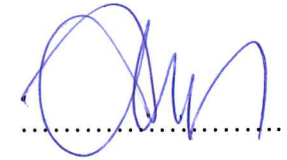
Jüri Üyeleri : Yrd. Doç. Dr. Erhan MENGUSOĞLU
Türk Hava Kurumu Üniversitesi



: Yrd. Doç. Dr. Murat KARAKAYA
Atılım Üniversitesi



: Yrd. Doç. Dr. Tansel DÖKEROĞLU
Türk Hava Kurumu Üniversitesi



Tez Savunma Tarihi: 07 Eylül 2016

**TÜRK HAVA KURUMU ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ MÜDÜRLÜĞÜ'NE**

Yüksek Lisans Tezi olarak sunduğum, “Sanallaştırılmış Grafik İşlemci Birimlerinin Maliyet Etkinlik Analizi” adlı çalışmamın, tarafımdan akademik etik ve kurallara aykırı düşecek bir yardıma başvurmaksızın yazıldığını ve yararlandığım kaynakların kaynakçada gösterilenlerden oluştuğunu, bunlara atıf yapılarak yararlanılmış olduğunu belirtir ve bunu onurumla doğrularım.

01.09.2016

Yusuf YAZICI



ÖNSÖZ

Yüksek Lisans tez çalışma sürecinde beni yönlendiren, karşılaştığım zorlukları bilgi ve tecrübesi ile aşmamda yardımcı olan desteğini ve yardımını hiçbir zaman esirgemeyen tez danışmanım değerli Yrd. Doç. Dr. Tansel Dökeroğlu ile test ortamlarının kurulum ve yapılandırması konularında verdikleri teknik destekten dolayı 4S Bilgi Teknolojileri firmasının kıymetli mühendislerinden Evren Parladiroğlu ve Ahmet Gül'e teşekkürlerimi sunarım.

Yoğun tez çalışmalarım sürecinde her zaman yanımda olan ve anlayış ve desteğini esirgemeyen aileme teşekkürlerimi sunarım.

Ağustos 2016

Yusuf YAZICI

İÇİNDEKİLER

ÖNSÖZ	iv
İÇİNDEKİLER	v
TABLO LİSTESİ.....	vii
ŞEKİL LİSTESİ.....	viii
KISALTMALAR	ix
ÖZET.....	x
ABSTRACT.....	xii
BİRİNCİ BÖLÜM.....	1
1. GİRİŞ	1
1.1 Varsayımlar	1
1.2 Neden GPU Sanallaştırma?.....	2
1.3 Çalışmamızın Amacı	2
1.4 Çalışmada Kullanılan Araçlar	3
1.5 Elde Edilen Dolaylı Kazanımlar	3
İKİNCİ BÖLÜM	8
2. ÖNCEKİ ÇALIŞMALAR.....	8
2.1 Sanallaştırmanın Doğuşu.....	8
2.2 Sanallaştırma Çeşitleri.....	11
2.2.1 Uygulama Sanallaştırma.....	11
2.2.2 Ağ Sanallaştırma	11
2.2.3 Depolama Sanallaştırma.....	13
2.2.4 Sunucu Sanallaştırma	14
2.2.5 Masaüstü Sanallaştırma.....	14
ÜÇÜNCÜ BÖLÜM	16
3. GPU SANALLAŞTIRMA TEKNOLOJİSİ	16
3.1 GPU Nedir?.....	16
3.2 Nvidia Grid Kartları	17
3.3 Citrix XenServer	19
3.4 Masaüstü Sanallaştırma.....	23
DÖRDÜNCÜ BÖLÜM	27
4. DENEY ORTAMI VE ELDE EDİLEN SONUÇLAR	27
4.1 Kullanılan Donanım ve Yazılımlar	27
4.1.1 GPU-Z	28
4.1.2 Fraps	30
4.1.3 Unigine Heaven.....	31
4.2 Testler ve Sonuçları.....	32
4.2.1 Fiziksel İş İstasyonu Testleri.....	32
4.2.2 Sanal İş İstasyonu Testleri.....	34
4.2.2.1 Tek sanal makine testi.....	35
4.2.2.2 Stres Testleri	36
4.3 Maliyet-Performans Analizi.....	38

4.4 Fiziksel ve Sanal İş İstasyonlarının Enerji Tüketimi Karşılaştırması	44
BEŞİNCİ BÖLÜM	45
5. SONUÇLAR	45
KAYNAKLAR	47



TABLO LİSTESİ

Tablo 1.1 : IBM ve Oracle Örnek Sunucu Kapasiteler	4
Tablo 3.1 : Nvidia Grid Serisi Kartlar ve Özellikleri.....	21
Tablo 3.2 : Nvidia Grid Kartları Sanal Profilleri.	18
Tablo 3.3 : İstemci enerji tüketimleri (Monitörler dahil).....	24
Tablo 4.1 : Sanal Platformun Kurulduğu Fiziksel Sunucu Özellikleri	27
Tablo 4.2 : Fiziksel İş İstasyonu Özellikleri	28
Tablo 4.3 : Giriş Seviye Fiziksel İş İstasyonu Fiyat Bilgisi.....	38
Tablo 4.4 : Orta Seviye İş İstasyonu Fiyat Bilgisi	39
Tablo 4.5 : Yüksek Seviye İş İstasyonu Fiyat Bilgisi	39
Tablo 4.6 : Fiziksel ve Sanal İş İstasyonlarının Maliyet Karşılaştırması.....	40
Tablo 4.7 : Fiziksel İş İstasyonları Enerji Tüketim Maliyeti	44
Tablo 4.8 : Sanal İş İstasyonları Enerji Tüketim Maliyeti.....	44

ŞEKİL LİSTESİ

Şekil 1.1	: Örnek bir LPAR profili	5
Şekil 2.1	: Almanya'da bir müzede sergilenen S360.	9
Şekil 2.2	: Network Sanallaştırma Bileşenleri.....	12
Şekil 2.3	: Depolama Sanallaştırma.....	13
Şekil 3.1	: GPU Market Pazar Payı Oranlarının Yıllara Göre Dağılımı	17
Şekil 3.2	: nvidia-smi komutu ile alınan örnek bir çıktı.	21
Şekil 3.3	: Sanal GPU atama işlemi.....	22
Şekil 3.4	: Sanal GPU sürücü bilgileri görüntülenmesi.....	22
Şekil 4.1	: GPU-Z Ekran Kartı Özellikleri.	29
Şekil 4.2	: GPU-Z Performans Görüntüleme.....	30
Şekil 4.4	: Quadro K2000 Ekran Kartı Özellikleri.	33
Şekil 4.5	: Fiziksel ekran kartı test sonucu.	34
Şekil 4.6	: K2000 Fiziksel Kart FPS değerleri	34
Şekil 4.7	: Sanal K220Q GPU profili FPS değerleri.	35
Şekil 4.8	: Unigine Heaven K220Q Test Sonucu.	36
Şekil 4.9	: Sanal İş İstasyonları Toplu Stres Testi.....	36
Şekil 4.10	: Fiziksel GPU testi ile Stres testi değerleri.....	37
Şekil 4.11	: Fiziksel ve Sanal İş İstasyon Maliyetleri.....	44
Şekil 4.12	: Fiziksel ve Sanal İş istasyon maliyetleri (64 İş İstasyonu).	41
Şekil 4.13	: Fiziksel ve Sanal GPU performans karşılaştırması.	42
Şekil 4.14	: GPU sanallaştırma maliyet/performans kazanım alanı.	43

KISALTMALAR

CPU	: Central Processing Unit
GPU	: Graphical Processing Unit
CMS	: Console Monitor System
CUDA	: NVIDIA GPU üzerindeki her bir paralel işlemci birimi
ICA	: Independent Computing Architecture
IO	: Input/Output
IOPS	: Input/Output per Second
LPAR	: Logical Partition
NPIV	: N-Port ID Virtualization
RAM	: Random Access Memory
SAN	: Storage Ara Network
WAN	: Wide Area Network
WWN	: World Wide Name
VLAN	: Virtual Local Area Network
vGPU	: Virtual GPU

ÖZET

SANALLAŞTIRILMIŞ GRAFİK İŞLEMÇİ BİRİMLERİNİN MALİYET ETKİNLİK ANALİZİ

YAZICI, Yusuf

Yüksek Lisans, Bilişim Teknolojileri Anabilim Dalı

Tez Danışmanı: Yrd. Doç. Dr. Tansel DÖKEROĞLU

Ağustos 2016, 49 sayfa

Günümüz hesaplama ortamlarında kullanılan küçük ve orta ölçekli sunucular yüksek kapasiteli kaynaklar barındıran hesaplama ortamları haline gelmiştir. Basit bir sunucu gündelik yaşamda kullandığımız birçok uygulamanın kapasite gereksinimlerinin çok üzerinde kaynaklara sahip olabilmektedir. Bu gelişmelere paralel olarak elimizde bulunan bilgisayar kaynaklarının (CPU, RAM, harddisk, network, I/O genişliği vb.) daha verimli olarak kullanılması günümüz hesaplama ortamlarında yük dengelemesi açısından önemli bir eniyileme problemleri haline gelmiştir. Bu bağlamda, sanallaştırma teknolojileri, sundukları etkili çözümlerle birçok sunucu altyapısında kullanılmaya başlanmış olup, maliyet-etkinlik ve yönetim kolaylıkları gibi özellikleri ile de ön plana çıkmaktadırlar. Yakın zamanda geliştirilen başarılı uzaktan erişim protokolleri ile sanallaştırmayı destekleyen grafik kartların (GPU) üretilmesi, bu alandaki sanallaştırmanın da önünü açmış ve yeni nesil sanallaştırma teknikleri arasına girmiştir. Bu teknoloji sayesinde başta yüksek grafik işlem gücü ihtiyacı duyan kuruluşlar ile dağıtık lokasyonlarda personel çalıştıran kurumsal firmalar merkezden yönetilebilir, güvenli ve esnek bir sistem altyapısına kavuşma imkânı elde etmişlerdir. Bu çalışma ile sanallaştırma desteği bulunan NVIDIA GRID K2 kartlarının kullanıldığı sanal iş istasyonları ile NVIDIA Quadro serisi kartların kullanıldığı fiziksel iş istasyonları üzerinde benchmark problem

kümeleri ile yapılan maliyet ve performans analizleri ve elde edilen deney sonuçları ile GPU sanallaştırmanın olumlu etkileri kapsamlı olarak ortaya konulmuştur.

Anahtar Kelimeler: GPU, Sanallaştırma, Maliyet, Performans



ABSTRACT

COST-EFFECTIVENESS ANALYSIS OF VIRTUALIZED GRAPHICS PROCESSING UNITS

YAZICI, Yusuf

Master, Department of Information Technology

Thesis Supervisor: Asst. Prof. Dr. Tansel Dokeroglu

August-2016, 49 pages

Small and medium-sized servers used in today's computing environment have become high-capacity computing resources. A simple server we use in our everyday life may have several resources on the capacity requirements of many applications. After increase of capacity of computers, researches have been made for efficient use of resources. Virtualization technology has made significant benefits on capacity utilization. Especially server, storage and network virtualization is often used in main data centers. However, GPU (Graphical Processing Unit) virtualization has not been commonly used yet due to various restrictions. Up to 2012 there was no virtualization support for graphic cards. Nvidia announced Grid K2 graphics cards with virtualization support in October 2012. Furthermore, until recently, network bandwidth for WAN infrastructure was not wide enough and remote desktop protocols was not adequately successful to transport graphical data. With development of efficient remote desktop protocols like HDX and PCoIP by virtualization companies and increase of network bandwidth, GPU virtualization technology has become feasible. In this study, cost-effectiveness analysis of GPU virtualization was performed. We focused on comparing graphics operations performance and costs of physical and virtual workstation. We have concluded that

under certain conditions specified in the following sections GPU virtualization exhibits better performance with a lower cost.

Keywords: GPU, Virtualization, Cost, Performance



BİRİNCİ BÖLÜM

GİRİŞ

Çok farklı lokasyonlarda bulunan yüzlerce masaüstü sistemin güncellenmesinin ve yönetiminin zorluğu, yerinde destek verecek teknik personel ihtiyacı, sıklıkla karşılaşılan donanım arızaları, kısa zamanda bu donanımların tamamının yenilenmeleri gereksinimi gibi ciddi maliyet ve işgücü kayıpları masaüstü sanallaştırma teknolojilerine olan talebi artırmaktadır. Sanallaştırma teknolojileri, sunucu tipi sistemlerde sıklıkla kullanılmasına rağmen, yakın zamana kadar grafik kartların sanallaştırma desteğinin olmaması ve uzaktan erişim protokollerinin yetersizliği masaüstü bilgisayarlarda bu çözümü imkânsız hale getirmekteydi. Sanallaştırma desteği sunan ekran kartlarının üretilmesi, Citrix, Vmware, Microsoft gibi firmaların, dar ağ bandı ile (WAN) etkili erişim için verimli görsel arayüz transferi sunan protokolleri geliştirmesi, “GPU Sanallaştırma” teknolojisinin hayatımıza girmesinde etkili olmuştur. Örneğin, CITRIX tarafından geliştirilen Independent Computing Architecture (ICA) protokolü sadece fare ve klavye çıktıları ile ekran güncellemelerini aktardığından, düşük ağ genişliğinin olduğu ortamlarda bile akıcı bir uzak masaüstü hizmeti verebilmektedir. GPU sanallaştırma, sadece masaüstü bilgisayarların merkeze alınması için değil, özellikle bilimsel hesaplamalarda ekran kartı üzerindeki çekirdekleri kullanarak çok iyi sonuçlar elde eden araştırma merkezlerinin sanal platformlarda paralel işlem yaptırabilme ihtiyacına karşılık vermesi açısından da ihtiyaç duyulan bir yöntemdir [1].

1.1 Varsayımlar

Maliyet analizi yapılırken odaklandığımız firmaların hali hazırda kullandıkları bir bilgi işlem odasının bulunduğu varsayılmıştır. Ayrıca mevcut iş istasyonlarını

tamamiyle yenileyecekleri ya da yeni bir yatırım yapacakları kabul edilerek hesaplamalar yapılmıştır. Böylelikle, deney ortamı, ideale yakın ve diğer unsurların etkilerinden uzak sadece GPU hesaplama analizlerinin olduğu bir çevre dikkate alınarak oluşturulmuştur.

1.2 Neden GPU Sanallaştırma?

Sanallaştırma teknolojileri sağladığı en iyi kaynak kullanımı, yönetim ve bakım kolaylıkları, düşük maliyet, enerji tasarrufu gibi sunduğu avantajlar ile başarılı çözümler ortaya koymaktadır. Sağladığı bu kolaylıkları aşağıda alt başlıklar halinde biraz daha detayla incelediğimizde sanallaştırmanın özellikle veri merkezlerinde kullanımının ne kadar önemli olduğu daha iyi anlaşılacaktır.

Grafik tasarımı, işlenmesi gibi uygulamalar üreten firmalar ile simülatör gibi yoğun grafik işlemlerin yapıldığı ürünleri geliştiren firmalar tarafından çok sayıda iş istasyonu kullanılmaktadır. Bugün iş istasyonları masaüstü bilgisayarlara kıyasla oldukça pahalıdır ve bu anlamda grafik işleme ihtiyaçlarına karşı yüksek maliyetli çözümler sunmaktadır. Bugün sanallaştırma desteği sunan güçlü grafik kartlarının üretilmesi ile birlikte birçok iş istasyonunu bir sunucu üzerinde çalıştırma imkânı elde edilmiş durumdadır.

1.3 Çalışmamızın Amacı

Bu tez çalışmamızda GPU sanallaştırma sonucunda normal iş istasyonlarına göre düşük maliyet/yüksek performans elde edilebileceğini ortaya koyduk. NVIDIA GRID K2 kartlarının Citrix platformu ile sanallaştırılması ile elde edilen profillerle oluşturduğumuz sanal grafik kartlarını Fraps, GPU-Z, Unigine Heaven gibi uygulamalarla FPS değerlerini gözlemledik. Sanallaştırma platformunun sağladığı en iyi kaynak tüketimi ile kaynak havuzundaki fiziksel grafik işleme ünitelerinin tamamının kullanılması sayesinde belli şartlarda daha yüksek performans sergilenebileceğini tespit ettik. Bu tespitlerimiz sayısal verilerle ortaya koyarak grafiksel olarak da sunulmuştur. Fiziksel grafik kartları ile sanal grafik kartlarından elde edilen değerleri karşılaştırmalarını da çalışmamıza ekledik. Yüksek performansın yanı sıra maliyet adına da önemli kazanımlar sağlanabileceğini ortaya koyduk. Kısaca kazanılan bu temel kazanımlar şu şekildedir.

Düşük Maliyet: Standart masaüstü bilgisayarlarının yeterince grafik işleme gücünün olmaması, profesyonel kullanımda uzun süre yüksek yoğunlukla kullanılmaya karşı yeterince dayanıklı olmamaları gibi nedenlerle iş istasyonlarının kullanımı gerekmektedir. Bugün giriş seviyesi iş istasyonlarının 4 bin dolardan başlayan maliyeti söz konusudur ve iş istasyonu sayısı ile doğru orantılı olarak maliyet artmaktadır. GPU sanallaştırmada ise ilk kurulum maliyeti yüksek olsa da onlarca iş istasyonunu içinde barındıracağından toplam maliyette ciddi avantaj sağlamaktadır. Bakım, yedek parça, güncelleme vs. gibi ekstra işçilik gerektiren hususlarda göz önünde bulundurulduğunda maliyet avantajı da artmaktadır.

Performans: GPU sanallaştırma ile kaynaklar optimize edilerek kullanıldığından grafik kartı atıl duruma düşmemektedir. Sanal grafik kartı kendisine ayrılan kaynak yetersiz kaldığında fiziksel GPU katmanında atıl duran grafik işlemci birimlerini kullanabilmektedir. Dolayısıyla belirli bir sanal iş istasyonuna kadar daha verimli bir GPU kullanımı olmakta, dolayısıyla performans konusunda da ciddi kazanım elde edilebilmektedir. Bu kazanımlar ilerleyen bölümlerde grafiklerle ortaya konmuş durumdadır.

1.4 Çalışmada Kullanılan Araçlar

Bu çalışmamızda grafik işleme performansını ölçmek için en önemli kriterlerden biri olan saniyede işlenen kare sayısına (FPS) odaklandık. Aynı senaryoların koşturulduğu performans testlerinde uygulamada bilinen Fraps, Unigine Heaven, GPU-Z araçlarını kullandık. Fiziksel ortam testlerini, üzerinde Quadro K2000 ekran kartı bulunan dört çekirdekli fiziksel iş istasyonu üzerinde, sanal iş istasyon testlerini ise üzerinde sanallaştırma desteği olan NVIDIA GRID K2 grafik kartı bulunan 16 çekirdek sahibi fiziksel sunucu üzerinde oluşturulmuş sanal iş istasyonlarında yaptık.

1.5 Elde Edilen Dolaylı Kazanımlar

GPU sanallaştırma ile sadece maliyet avantajı değil sanallaştırma teknolojilerinin getirdiği standart kazanımlarda elde edilmiş olmaktadır. Bu kısımda genel manada sunucu sanallaştırma ile elde edilen kazanımlar kısaca ortaya konmuştur. Sunucu sanallaştırma ile elde edilen kazanımlar aynen GPU

sanallaştırmanın uygulandığı iş istasyonları için de geçerli olduğundan bu kazanımlarında tercihlerde değerlendirilmesinde büyük avantajlar bulunmaktadır.

En iyi Kaynak Tüketimi: Donanım maliyetleri günden güne düşmekte ve bir kutu içine çok geniş kaynaklar makul fiyatlara sığdırılabilmektedir. Tablo 1.1’de görüldüğü üzere bir adet IBM P795 sistemi 256 işlemci çekirdeği, 16TB bellek kapasitesi, eklenen yeni IO modülleri ile toplamda 640 PCIe slot desteği sunabilmektedir [2]. Sparc T7-4 sunucunun yine 2TB bellek ile 128 çekirdek işlemci desteği bulunmaktadır [3]. Bu kapasiteler göz önünde bulundurulduğunda yüzlerce sistemin kaynak gereksinimlerini karşılayabilecek donanımlara sahip oldukları görülmektedir.

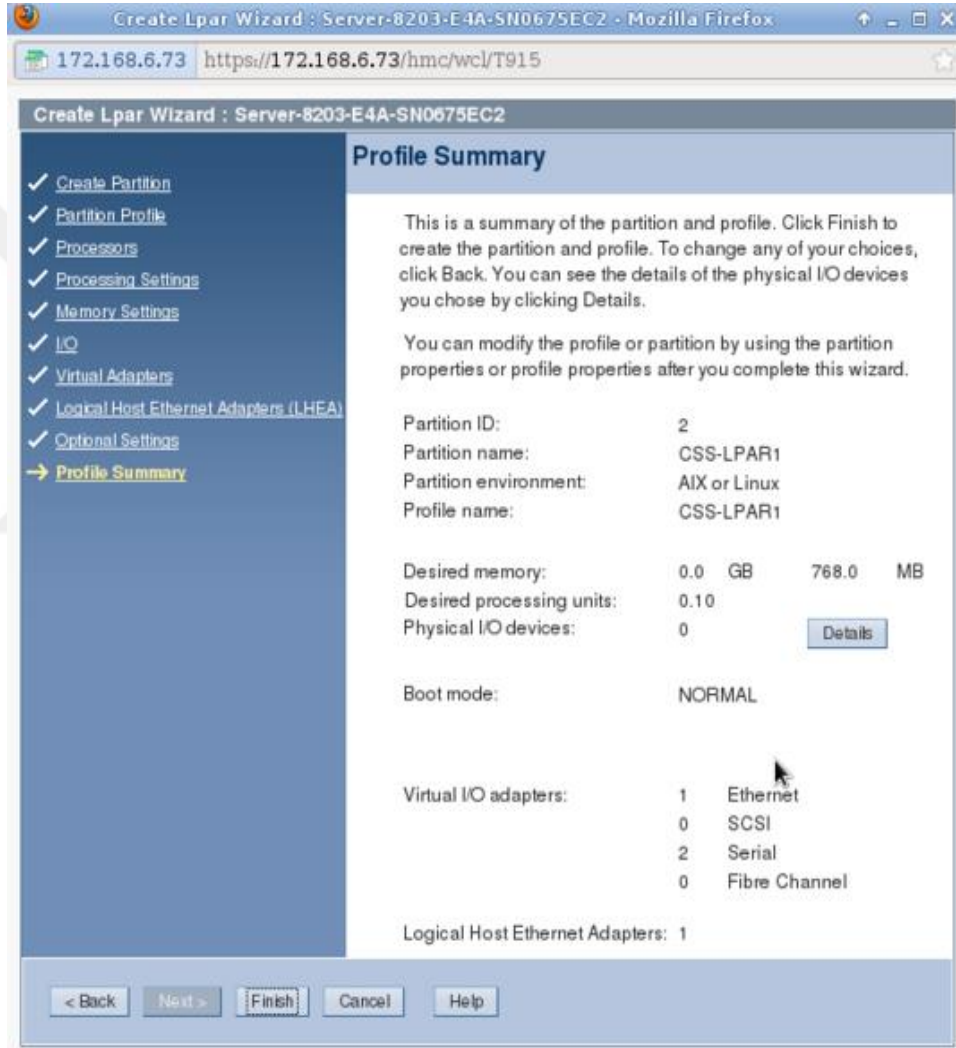
Tablo 1.1: IBM ve Oracle örnek sunucu kapasiteleri.

	IBM Power 795 Sunucu Özellikleri	OracleSparc T7-4 Sunucu Özellikleri
Maximum Çekirdek Sayısı	192 çekirdek 3.72 GHz 256 çekirdek 4.00 GHz 128 çekirdek 4.25 GHz	128 çekirdek 4.13 GHz
İşlemci Modülleri	8	Yok
İşlemci Soket Sayısı	32	4
Bellek Miktarı	16TB	2TB
Adaptör Slot Sayısı	Toplam 640 Slot Desteği	4x10GBe ve 16xPCIe Toplam 20 Adet Slot Desteği

Güncel uygulamalarda fiziksel sunucuların çoğunun üzerlerindeki kaynaklar etkin kullanılamamakta, bazı sistemler ise ciddi kaynak ihtiyacı problemi ile karşılaşmaktadırlar. Fiziksel sunucuların kaynak yükseltilmesi ve taşınması işlemleri, satın almalar ve yönetimsel zorlukları nedeniyle uzun soluklu ve maliyeti yüksek işlemlerdir. Sanallaştırmanın kullanıldığı sistemlerde ise kapasite ve kaynak yönetimi çok esnek olduğundan çok hızlı ve minimum maliyetle bu işlemler tamamlanabilmektedir.

İşlemci Havuzları: İşlemci maliyeti en yüksek olan donanım olup, tüm sanallaştırma platformlarında optimum verimle kullanılması adına önemli algoritmalar geliştirilmiştir. Burada fikir vermesi adına IBM PowerVM sanallaştırma

platformunda işlemcilerin etkin kullanımının nasıl yapıldığını görmekte fayda vardır. Fiziksel kutular üzerinde sistemler oluşturulurken belli profiller ile oluşturulmakta ve bu profillerde belirtilen kaynak atamalarına göre sistemler aktif hale getirilmektedir. Örneğin IBM Power kutularda oluşturulan her bir sanal sisteme LPAR (Logical Partititon) olarak adlandırılmaktadır. Her LPAR kendine ait kaynak atamalarının yapıldığı bir profil ile açılmaktadır. Bu profilde sisteme atanmış işlemci sayısı, bellek miktarı ve slot bilgileri yer almaktadır. Şekil 1.1’de örnek bir profil görülmektedir.



Şekil 2.1: Örnek bir LPAR profili

Kutu üzerindeki en kritik kaynak işlemci olduğundan atanması da dikkatle yapılmalıdır. IBM PowerVM sanallaştırmada işlemci atanırken ilk seçenek, işlemcinin çalışma modu doğru seçilmelidir. Sanallaştırmanın tüm avantajlarını kullanmak için “Shared” seçilmelidir. Bu durumda işlemciler ortak bir havuzda

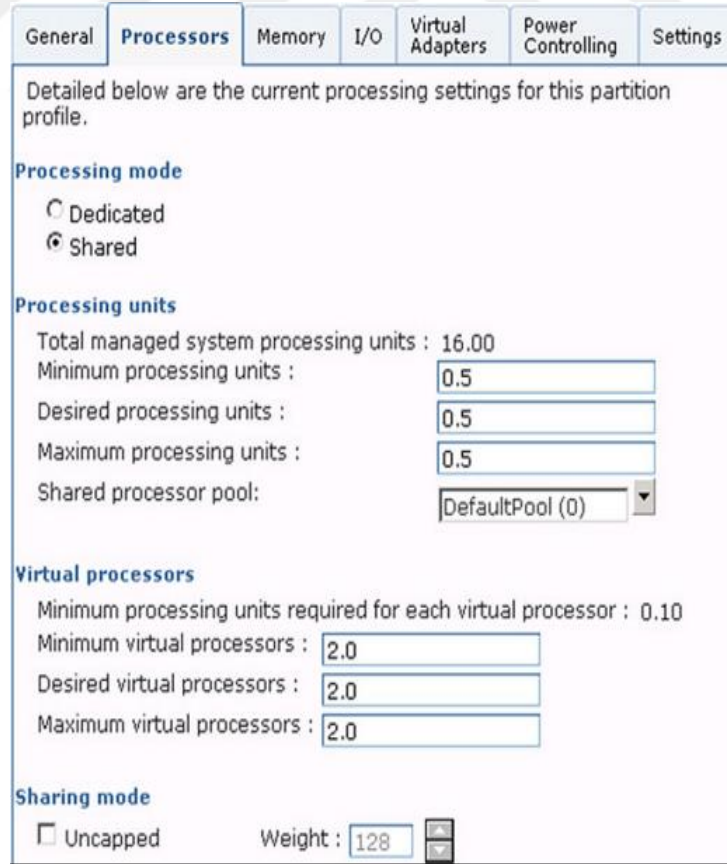
yönetilmektedir. “Dedicated” seçildiğinde belirtilen sayıda işlemci sadece ilgili sisteme atandığından diğer sistemler bu kaynakları kullanamayacaktır. Şekil 1.2’de görülen ikinci seçenek işlemci birimlerinin sayıları. Burada belirtilen “Minimum” ve “Maximum” değerleri sistem çalışır haldeyken işlemci birim sayısını artırıp azaltabileceğimiz esneme sayılarıyken, “Desired” sayılar ise sistemin istenilen işlemci birim sayısıdır. Bu terimler “Virtual processors” içinde geçerlidir. İşlemci birimi ile sanal işlemciler arasında ise farklı bir ilişki vardır. “Desired” kısmına atanan değerler sistemin açılışta kullanılan varsayılan değerlerdir.

Şekil 1.2’de örnek olarak gösterilen profilde sisteme,

Desired processing units : 0,5

Desired virtual processors : 2

olacak şekilde işlemci kaynağı atandığı bir durumda sistemin 0,5 çekirdek üzerinde birinci önceliği varken işlemci havuzunun durumuna göre diğer sistemlere atanmış ve ya atanmamış iki çekirdeğe kadar kaynak kullanabilecek anlamına gelmektedir [4].



The screenshot shows the 'Processors' tab of a system configuration window. It displays the following settings:

- Processing mode:** Shared (selected), Dedicated (unselected).
- Processing units:** Total managed system processing units: 16.00. Minimum processing units: 0.5. Desired processing units: 0.5. Maximum processing units: 0.5. Shared processor pool: DefaultPool (0).
- Virtual processors:** Minimum processing units required for each virtual processor: 0.10. Minimum virtual processors: 2.0. Desired virtual processors: 2.0. Maximum virtual processors: 2.0.
- Sharing mode:** Uncapped (selected). Weight: 128.

Şekil 1.2: LPAR işlemci ataması.

Sanal Girdi/Çıktı Kartları: Sanallaştırma ortamlarında fiziksel kartlar birden fazla sistem arasında paylaştırılabilir. Bu sayede tüm sanal sistemler yedekli olarak çalışabilmekte ve olası tek nokta hatası (single point of failure) sebebiyle kesinti yaşamamaktadır. Olası fiziksel kart arızası durumunda sağlam bir kart ile değiştirilmesi sadece bağ ilişkisi (mapping) değiştirilerek yapılabilmektedir. Bu ise maksimum esneklik anlamına gelmektedir. Ağ tarafında sanal kart kullanımı genel olarak şu şekilde yapılmaktadır. Fiziksel katmanda birden fazla ağ kartı farklı ağ anahtarlarından gelen kablolarla birleştirilerek (Etherchannel, Teaming, Bonding) paylaşımına açılmakta (Shared Ethernet Adapter), tüm sanal makinelere bu yedekli ağ altyapısından sanal kartlar atanmaktadır. Sanal makinenin ekstra bir kart ve ya farklı bir ağa (VLAN) erişim gereksinimi olduğunda hiçbir kesinti yaşanmaksızın, çok kısa bir zaman içinde bu ihtiyaçlar karşılanabilmektedir [5]. Fiziksel yapıda yeni kart satın alınması, donanım olarak makineye takılması, yeni bir anahtar portunun atanarak kablo çekilmesi gibi süreçler ciddi iş gücü ve zaman kayıplarına yol açmaktadır.

Depolama alan ağlarında (SAN) ise sanal makinede bulunan kartlar fiziksel katmanda bulunan fiber kartlar ile eşleştirilmektedir. IBM PowerVM sanallaştırmada kullanılan NPIV(N_Port ID Virtualization) teknolojisi sayesinde her bir sanal kartın kendine ait WWN numarası olduğundan depolama üniteleri tarafından direk ilgili sanal makineye disk atama işlemi yapılabilmektedir. Bu sayede fizikselden sanala geçiş ve ya sanal makinelerin farklı bir donanıma geçişi çok rahat bir şekilde yapılabilmektedir.

Canlı Sistem Geçişleri: Günümüzde bilgisayarlar hayatımızın vazgeçilmezleri durumuna gelmiş durumda. Gerek kamusal, gerek özel birçok işlem 7/24 kesintisiz hizmet vermek durumuna gelmiştir. Dolayısıyla sistemleri kapatmak mesai saatleri dışında da artık mümkün olamamaktadır. Sistemlerin bakımı ve göç işlemleri olabildiğince hızlı ve problemsiz olmak zorundadır. Sanallaştırmanın verdiği esneklikle birlikte bugün hiçbir kesinti yaşamaksızın, sistemler canlı bir şekilde göç edebilir duruma gelmişlerdir [6]. Sanallaştırmanın da yaygınlaşmasıyla artık fiziksel donanımlar üzerinde çok sayıda sanal makine çalışabilmektedir. Makinenin kapatılması gerektiğinde üzerindeki tüm sistemlerin de kapatılması hem ciddi hizmet kesintilerine yol açmakta hem de çalışma planlamalarını çok daha karmaşık hale getirmektedir. İşte bu durumlarda da canlı sistem geçişleri hayat kurtarmaktadır.

İKİNCİ BÖLÜM

ÖNCEKİ ÇALIŞMALAR

Sanallaştırma teknolojileri, bilişim sistemlerine kattığı esneklik, ekonomi ve çeviklikle şüphesiz çağımızın en önemli yeniliklerinden biri olmuştur. Günümüzde sunduğu esneklik ve yönetim kolaylıkları ile bilişimde vazgeçilemez hale gelen sanallaştırma aslında bilgisayarlar kullanılmaya başladığından beri gündemde olmuş ve devamlı gelişerek günümüzdeki haline ulaşmıştır [7].

2.1 Sanallaştırmanın Doğuşu

1960'ların başlarından itibaren IBM ilk nesil bilgisayarları üretmeye başladı. Fakat üretilen her yeni bilgisayar mimari olarak birbirinden farklı olmakta ve önceki mimarilerle uyum problemi olduğundan müşteriler uygulama ve verilerin taşınması işlemini yapamamaktaydı. IBM farklı sistemler arasındaki uyumsuzluğu kaldıran ve müşterilerin ihtiyaçlarına göre kaynaklarını artırabilecekleri bir altyapı sunan "IBM System 360" donanımını 1964 de duyurdu. S 360 interaktif olmayan, tek kullanıcı bir sistemdi fakat bu ürün ile donanım ve yazılım arasındaki organik bağ kırılmış oldu [7]. Şekil 2.1'de örnek bir S360 sistemi görülmektedir.



Şekil 2.1: Almanya'da bir müzede sergilenen S360 [7].

İlerleyen zamanlarda özellikle üniversitelerden gelen laboratuvarlarda kullanılacak çok kullanıcı sistem taleplerine karşılık GE (General Electric) “time-sharing” bilgisayar mimarisi geliştirirken, IBM müşteriye hiç satılmayan sadece araştırmalarda kullanılmak üzere CP-40 bilgisayarını üretti. CP-40 sanallaştırmayı destekleyen ilk bilgisayar özelliği taşımaktadır (1968). CP-40 üzerindeki işletim sistemi CP(Control Program)/CMS(Console Monitor System) üzerine inşa edildi. CP sanal makinelerin oluşturulduğu tek kullanıcı interaktif bir sistem, CMS ise oluşturulan sanal makineler tarafından çalıştırılan ve uç kullanıcı tarafından kullanılan arayüzdür. Bu sayede bir sistem üzerinde kaynakların paylaştırıldığı birden fazla işletim sistemi hizmet vermekteydi. Bu mimari GE tarafından üretilen “time-sharing” bilgisayarlarına göre daha çok tutuldu. GE tarafından geliştirilen mimaride kaynaklar kullanıcılara paylaştırılmakta ve zaman döngüsü içinde sırayla kaynaklar kullanılmaktaydı. Ortada tek bir sistem olduğundan olası panik durumunda tüm kullanıcılar etkilenmekte ve sistem komple çökmekteydi. IBM mimarisinde ise sanal makineler kaynak tüketimi konusunda birbirlerinden daha izole durumdaydılar [8].

Artık en azından aynı üretici tarafından üretilen donanım değişikliklerinden kaynaklı uyumsuzluklar ortadan kaldırılmıştı. Zamanla TCP/IP, RFC, http gibi gerek donanım gerekse yazılım alanında geliştirilen uluslararası standartlar sayesinde farklı

donanımlar arasında da uyum temin edilmiş oldu. Fakat geliştirilen uygulamalar farklı işletim sistemleri arasında taşınabilir değildi. Uygulamalar makine dilinde (Assembly) yazılmaktaydı ve işletim sistemi ve donanım değişikliklerini desteklememekteydi. Uygulamaların farklı sistemlere kolayca aktarılması ihtiyacı söz konusuydu. Derleyici desteği olan bir dille yazılan ilk işletim sistemi olarak karşımıza çıkan Unix, bu konuda büyük kolaylıklar getirdi. Unix, C dili kullanılarak yazıldı. Farklı platformlarda geliştirilen uygulamalar C derleyicisi ile derlenerek farklı işletim sistemi versiyonları ve donanımlar arasında taşınabilir duruma geldi. Yine de her işletim sistemi için ayrıca derleme ihtiyacı dolayısıyla ihtiyacı karşılayan bir esneklik yoktu.

1991 yılında James Gosling, Mike Sheridan ve Patrick Naughton tarafından platform bağımsız bir dil geliştirme projesini başlattı. Projenin adını birkaç defa değiştirilse de (Greentalk, Oak) 1995 yılında “JAVA” olarak son halini aldı [8]. JRE (Java Runtime Environment) sayesinde kurulu tüm platformlarda kodlar tekrar derlenme ihtiyacı olmaksızın çalıştırılabilir duruma gelmiş oldu. JRE nin en önemli bileşeni JVM (Java Virtual Machine), küçük bir sanal işletim sistemi gibi geliştirilen java kodlarını çalıştırmaktadır. Bu sayede yazılımcı her platform için ayrı kod geliştirmesine gerek kalmamıştır. JVM i uygulama sanallaştırmanın ilk örneği olarak gösterebiliriz. “Process Sanallaştırma” olarak da tanımlanmaktadır.

Donanım sanallaştırma konusunda yukarıda belirtildiği gibi çözüm IBM tarafından üretilmiştir. Fakat “Mainframe” sistemlerinde sunduğu bu çözüm hem pahalı, hem de karmaşık olduğundan yaygınlaşmamıştır. 1987 yılında Insignia Solutions tarafından geliştirilen SoftPC ürünü, Unix workstation üzerinde DOS uygulamalarını çalıştırabilme imkânı sunmuştur. 1994 itibariyle SoftWindows ve SoftOS/2 ürünlerinin satışını yapmaya başlamıştır. Bu sayede uç kullanıcıya hitap edebilecek sanallaştırmanın önü açılmış oldu. 1997 yılında Apple tarafından geliştirilen “Virtual PC” programı zamanının başarılı ürünleri arasında yer almıştır. 1999 da ise bugün sanallaştırma teknolojilerinin önde gelen firmalarından VMware, ilk ürünü VMware Workstation 1.0 ürünü ile sanallaştırma dünyasına giriş yapmıştır [9]. Bu ürün kullanımı kolay, basit bir arayüze sahip oluşu ve lisans maliyetlerinin düşüklüğü ile herkesin kullanabileceği bir platform hazırlayarak sanallaştırmanın hızla yayılmasına katkıda bulunmuştur. 2001 yılında ESX ürünü ile kurumsal bir sanallaştırma çözümü sunmuştur. Sanallaştırma konusunda bugün piyasanın bir diğer

önemli aktörü Citrix, 2003 yılında geliştirilmesine başlanan XenSource açık kaynak yazılımını 2007 de satın alarak güçlü bir rakip olarak ortaya çıkmıştır [9].

2.2 Sanallaştırma Çeşitleri

Sanallaştırmanın sunduğu esneklik, maliyet düşüklüğü, yönetim kolaylıkları gibi avantajlardan dolayı bilişim teknolojilerinin hemen her alanında kullanılır duruma gelmiştir. Bilişim teknolojilerinin en temel bileşenleri olan sistem, ağ, depolama, uygulamalar ile uç kullanıcılara ait masaüstü sistemlerin hepsi için bugün çok başarılı çözümler mevcuttur.

2.2.1 Uygulama Sanallaştırma

Uygulama sanallaştırma, bugün daha çok bulut bilişim aracılığıyla verilen merkezi uygulama kullanma hizmeti olarak kullanılan şekli akla gelmektedir. Burada işletim sisteminden bağımsız uygulamaların çalıştırılabileceği bir ara katman olarak dünyamıza giren sanallaştırma biçimi kastedilmektedir. Yazılımlar çok önce assembly dili ile yazılırdı. Zamanla daha zengin kütüphaneye sahip ve yazımı kolay Pascal, Fortran, C/C++ gibi programlama dilleri ortaya çıktı. Fakat bu dillerde yazılan diller sadece derlendiği sistemler üzerinde çalışabilmekte, çalıştırılmak istenilen her platform için ayrıca derlenmesi ihtiyacı söz konusuydu. Yukarıda da değinildiği üzere ilk JAVA tarafından geliştirilen sanallaştırma katmanı JVM sayesinde uygulama sanallaştırma teknolojisi uygulamaya konmuştur [10].

İlerleyen zamanlarda Microsoft benzer bir arabirim olarak “.NET Framework” yazılımını geliştirmiştir. Uzun yıllar sadece Windows işletim sistemleri için geliştirilen .NET, yakın zamanda Linux platformu içinde geliştirme sürecine girmiştir [11].

2.2.2 Ağ Sanallaştırma

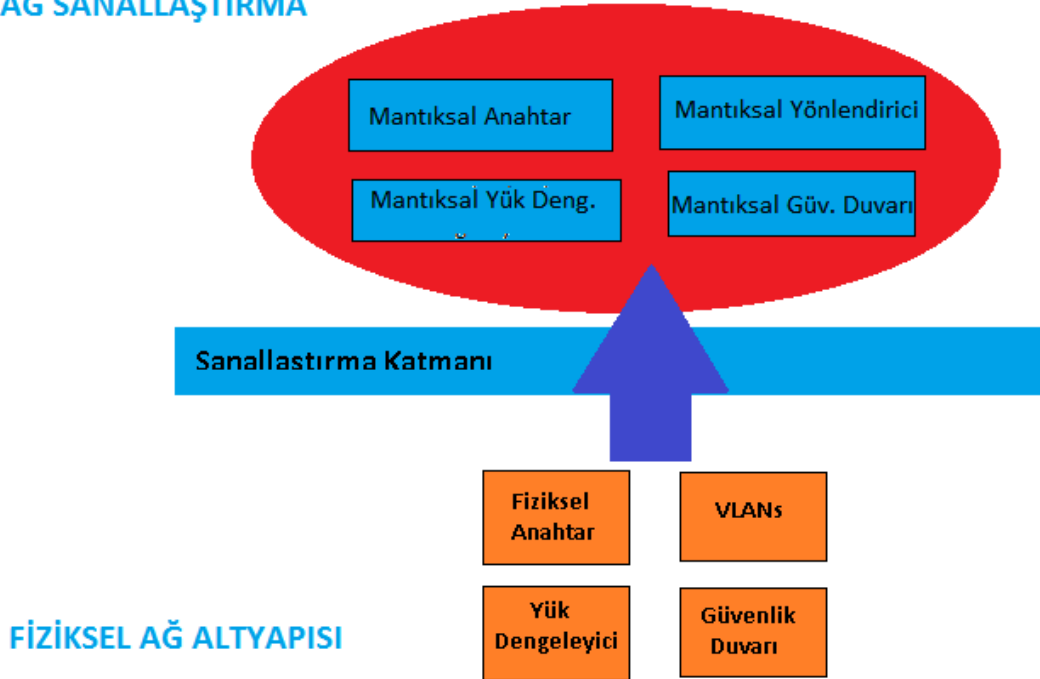
Sistemler ile kullanıcıların birbiriyle iletişimini sağlayacak altyapı, kesintisiz ve güvenli erişimin en önemli unsurlarından biridir. Güvenli ve kesintisiz erişimi sağlayan çok sayıda ağ bileşenleri mevcuttur. Bu ürünlerin başında Şekil 2.1’de görüldüğü gibi ağ anahtarları (Switch), yönlendiriciler (Router), güvenlik duvarları (Firewall), yük dengeleyicileri (Load Balancer) gibi ürünler gelmektedir. Bugün

sistem odalarının çoğunda yumak haline gelmiş ağ kablolarına rastlanmaktadır. Özellikle sanallaştırmanın kullanılmadığı yerlerde her bir sistem için en az bir ağ kablosu olduğu düşünülürse çok karmaşık bir ağ yapısı kaçınılmazdır. Bugün ağ kaynaklarının kapasiteleri de çok gelişmiş durumdadır. Çok yakın bir zamana kadar 100Mbit hatlar kullanılırken bugün 10Gbit ağ arayüzleri yaygınlaşmaktadır. Sistemlerin ortalama ağ kullanımları düşünüldüğünde yüzlerce sistemin ihtiyacı 10Gbit tek bir arayüzden sağlanabilir durumdadır.

Diğer fiziksel katmanlarda olduğu gibi ağ katmanında da bir birinden farklı donanımlar, bu donanımların birbiriyle uyumluluğu donanım değişiklikleri ve ya altyapıya yeni eklenecek cihazlar açısından hep problem olmuştur. Ağ cihazlarının sanallaştırılması ile birlikte ağ teknolojilerinin geleceği çok farklı bir noktaya doğru kaymaya başlamıştır [12].

Sanallaştırmanın sunduğu avantajlar sayesinde altyapı değişiklikleri, güvenlik duvarı ayarları, değişimleri gibi yönetimsel işlemleri çok daha esnek bir şekilde yapılabilmektedir. Optimum kaynak tüketimi ile bazı sistemlerin ağ genişliği nedeniyle yaşadığı problemler etkili ve hızlı bir şekilde çözümlenebilmektedir.

AĞ SANALLAŞTIRMA



Şekil 2.2: Network sanallaştırma bileşenleri [13].

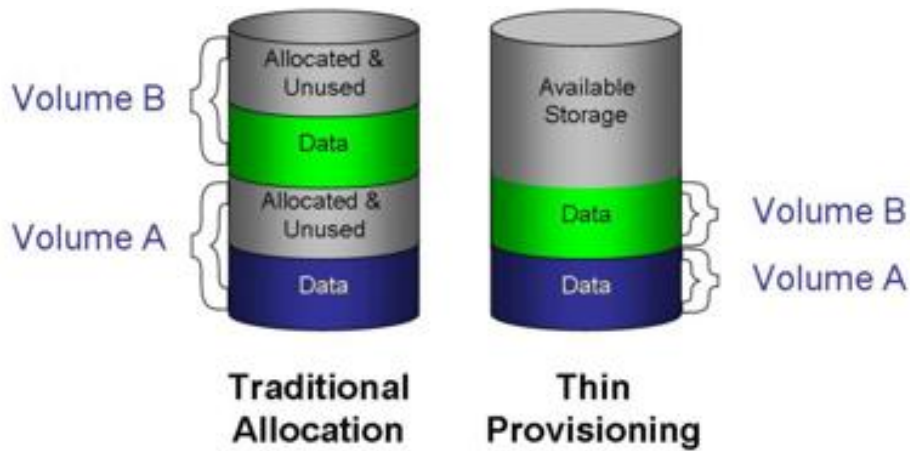
2.2.3 Depolama Sanallaştırma

Disk kapasite ve performansları artarken fiyatlarının düşmesi, ayrıca bazı yasal yaptırımlar depolama miktarının zamanla artmasına neden olmuştur. Görüntü ve seslerin uzun süreli tutulması, uzun süreli geçmiş kayıtların saklanması, alınan yedek sayısının çoklanması gibi sebepler depolama kapasitesini mevcudun çok daha üzerine çıkaracaktır.

Belirlenen ön gereksinimler doğrultusunda sistemlere kapasite atanmaktadır. Çoğu zaman yıllık gereksinimlere göre hareket edilmekte, yıl sonuna gelindiğinde ise mevcut kapasitenin çok altında bir kullanım olduğu görülmektedir. Dolayısıyla toplam kapasite atıl olarak kullanılmakta ve disk kutusu değişimlerinde değişik problemlerle karşılaşılmaktadır.

Depolama sanallaştırma teknolojilerinin geliştirilmesiyle birlikte sistemlere atanan diskler sanal bir havuzdan atanmakta, kullanılmayan alanlar diğer sistemler tarafından kullanılabilir duruma getirilmektedir (Bakınız: Şekil 2.2). Disk kapasite artış ve ya azaltma gibi operasyonlar son derece kolay ve esnek bir hale getirilmiştir.

Depolama sanallaştırmada temelde dört katmanlı hiyerarşik bir yapı söz konusudur. En alt katman disk havuzu, üzerinde hata toleranslı raid disk grupları, depolama havuzları ve en yukarıda gayet esnek sanal disk havuzları (Vdisks) bulunmaktadır. Bu sanal diskler fiziksel katmandan soyutlanmış olup, sanallaştırma yazılımları sayesinde otomatik kapasite artırımına ya da azalmasına gidecek durumdadırlar [14].



Şekil 2.3: Depolama sanallaştırma [15].

2.2.4 Sunucu Sanallaştırma

Sunucu sanallaştırma piyasada en sık kullanılan temel sanallaştırma teknolojisidir. Yüksek kapasitede donanıma sahip olan sunucular üzerindeki işlemci, bellek ve io kartlarının hipervizör (kaynakların sanallaştırıldığı yazılım katmanı) aracılığı ile üzerinde kurulan işletim sistemlerine paylaştırılmaktadır. Bu sayede bir donanım üzerinde birden fazla işletim sistemi çalıştırıldığı gibi birbirinden farklı platformlar da çalıştırılabilmektedir. Örneğin bir fiziksel makine üzerinde Windows Exchange mail sunucusu ile Redhat üzerinde çalışan bir veritabanı sunucusu birlikte hizmet verebilmektedir.

Sunucu sanallaştırma ile birlikte kutu üzerindeki kaynaklar en iyi verim ile dağıtılabilmekte, sunuculardan birinin olası ekstra kaynak ihtiyacında hiçbir kesinti yaşamaksızın kapasite artırımına gidilebilmektedir.

2.2.5 Masaüstü Sanallaştırma

Sanallaştırma teknolojileri arasında uygulamada en az kullanılan, yeni bir sanallaştırma teknolojisidir. Neredeyse tüm operasyonel işlemlerin dijital ortamda yapıldığı günümüzde özellikle kurumsal firmaların çok farklı lokasyonlarda personel hizmetine sunulan masaüstü bilgisayarları bulunmaktadır. Bu bilgisayarların yönetim ve bakımı ile ilgili çok farklı sorunlar ile karşılaşmaktadır. Yerinde bakım destek ihtiyacı, donanım arızaları, veri güvenliği, yüksek enerji tüketimi ve daha birçok sebepten dolayı masaüstü sanallaştırma gereksinimi günden güne artmaktadır. Masaüstü sanallaştırma çözümü için GPU sanallaştırma bir gereksinim olduğundan yaptığımız çalışmamız ile doğrudan ilgilidir.

GPU sanallaştırma yukarıda da belirtildiği üzere nvidia grid kartların piyasaya sürülmesi ve gerekli hipervizör katmanının Citrix, VMware gibi sanallaştırma geliştiren firmalar tarafından üretilmesi ile yaygınlaşmışsa da sanallaştırma hayatımıza girdiğinden beri grafik kartların sanal makinelerde kullanımı üzerine bazı akademik araştırmalar yapılmıştır. Dowty ve Sugerman yaptıkları çalışmada, ekran kartının “Hardware Acceleration” özelliği aktive edildiğinde bunun VMware sanal makine grafik özellikleri üzerinde olumlu tesirlerinden bahsetmişlerdir [16]. Bu çalışmanın neticesinde saniyede 18 ila 44 karenin işlendiği görülmüş olup, fiziksel

ekran kartının %12 ile %86 arasında performans iyileşmeyi sağladığı gözlemlenmiştir.

Bulut bilişim teknolojisi kullanılarak yapılan HPC (High Performance Cluster) sistemlerde de GPU sanallaştırmanın gerekliliği öne çıkmaktadır. HPC üyesi tüm sistemlerde fiziksel GPU işlemcisinin olması hem maliyet hem de tüketilen enerjinin ciddi şekilde artmasına sebep olmaktadır. Geliştirilen bir API sayesinde GPU üzerindeki çekirdekler uzaktan kullanılma imkânı sağlanarak ciddi verim elde edilmektedir [17].

Montella, Agrillo, Coviello yaptıkları çalışmada GPGPU (General-purpose computing on graphics processing units) sanallaştırma üzerinde durmuşlardır [18]. GPGPU yüksek işlemci gücü gerektiren özellikle bilimsel hesaplamalarda öne çıkan farklı bir teknolojidir. GPU üzerinde normal işlemcilere göre daha düşük frekansta çalışan fakat sayıca çok daha fazla olan çekirdeklerin (CUDA) gündelik hayattaki işlemlerde kullanılması amacıyla ortaya çıkmıştır. Bu sayede normal işlemcilere göre oldukça yüksek hesaplama performansı ortaya çıkarılmıştır. Fakat bu gücün bulut bilişimde kullanılabilmesi için sanal platform ihtiyacı söz konusudur. Montella, Agrillo ve Coviello bu çalışmalarında gVirtuS (GPU Virtualization Service) adını verdikleri bir bileşenle GPGPU hizmetini sanal platformda sunmuşlardır. Benzer çalışmalarda GVIM (GPU-accelerated virtual machines) Gdev[19] gibi sistemler tasarlanarak, GPU kaynakları bir miktar performans kayıplarıyla birlikte sanal makineler tarafından kullanılabilir hale getirilmiştir [20]. 2012 Aralık' ta yayınlanan bir çalışmada açıklanan bir araştırmada ise HPC Cluster sistemlerde GPU sanallaştırmanın gerekliliğine değinilmiş, uzaktaki grafik kartlarına ait çekirdekleri kullanma imkânı sunan mevcut rCUDA yazılım katmanına gönderilen grafik işleme kodları üzerinde, CU2rCU adını verdikleri platform ile dinamik olarak değişiklikler yaparak performansta kazanımlar sağladıklarını belirtmişlerdir [21].

Masaüstü sanallaştırmada görüntünün işlenerek iletilmesi söz konusu olduğundan mimari içinde sistemler arası ağ bandı genişliği ve paket gecikme süresi kritik öneme sahiptir. Bunun için bazı mimarilerde infiniband (FDR, QDR) gibi altyapılar kullanılarak iyileştirmeler elde edilmeye çalışılmıştır [22].

ÜÇÜNCÜ BÖLÜM

GPU SANALLAŞTIRMA TEKNOLOJİSİ

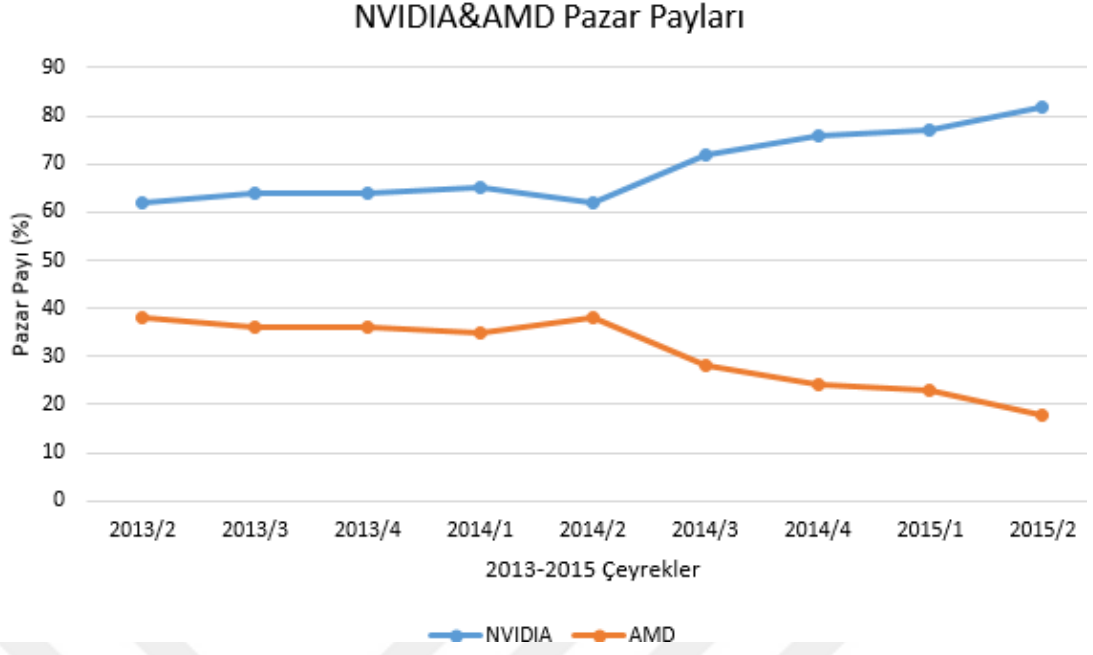
Bu bölümde GPU teknolojisi ile kullanılan ürünleri hakkında bilgilendirme yapılmıştır. Çalışmamızda kullanılan NVIDIA GRID K2 kartının fiziksel özellikleri ile üzerinde oluşturulabilen sanal profiller ve Citrix sanallaştırma platformu gereksinim ve kurulumları ile ilgili bilgilere yer verilmektedir.

3.1 GPU Nedir?

GPU (Graphics Processing Units) grafik verilerinin işlenerek iki boyut formatında ekran üzerindeki piksellere ileten kendine ait mikro işlemci ve belleği olan bir donanımdır [23]. Standart micro işlemciler yüksek frekansta çalışan her türlü hesaplamayı gerçek zamanlı olarak yapmakta, özellikle interaktif işlemlerde başarılı sonuçlar vermektedir. GPU ise grafik tabanlı işlemler yapmak üzere tasarlanmış, üzerinde “CUDA” olarak adlandırılan normal mikroişlemcilere göre düşük frekansta yüzlerce grafik işleme birimlerinden oluşmaktadır. Kendi alanındaki hesaplamalarda normal işlemcilere göre çok daha yüksek bir performans sergilemektedir.

Günümüzde özellikle HPC (High Performans Clusters) sistemlerde bilimsel amaçlı yapılan bazı hesaplamalar GPU tarafından işlenebilir duruma çevrilerek çok daha hızlı sonuçlar elde edilebilmektedir. GPU'nun bu şekilde farklı amaçlarla kullanılması terminolojiye GPGPU (General-Purpose computing on Graphics Processing Units) olarak girmiştir.

GPU üretimi konusunda dünyanın önde gelen iki firması NVIDIA ve AMD firmaları olup, pazar liderliğini Tablo 3.1'de görüldüğü üzere 2015 ikinci çeyrek verilerine göre %82 ile NVIDIA üstlenmektedir [24].



Şekil 3.1: GPU market pazar payı oranlarının yıllara göre dağılımı [24].

NVIDIA sanallaştırma desteği bulunan Grid kartları üreterek bu alanda da teknolojiye önemli katkılar sunmuştur. Çalışmamızda nvidia grid serisi K2 kartlarını kullandık. Bu kartların gerek esnek kullanımı, gerek performans bakımından ciddi avantajları bulunmaktadır.

3.2 Nvidia Grid Kartları

Nvidia tarafından özellikle sanal platformlarda kullanılmak üzere üretilen K1 ve K2 serisi iki çeşit grid kart bulunmaktadır. K1 serisi kartlar daha çok giriş ve orta seviye masaüstü sanallaştırmada kullanılmak üzere tasarlanmıştır. Üzerinde daha az çekirdek, frekansı daha düşük ama kapasitesi yüksek bellek bulunmaktadır. K2 serisi kartlar ise yüksek performans ihtiyacı olan özellikle profesyonel grafik işleme faaliyetinde bulunan kullanıcılar için üretilmiş bir kart olup, üzerinde çok sayıda CUDA işlemci birimi ile yüksek frekans ve geniş veri yoluna sahip bellek bulundurmaktadır (Bakınız Tablo 3.1).

Tablo 3.1: Nvidia grid serisi kartlar ve özellikleri.

	GRID K1	GRID K2
GPU Sayısı	4 Kepler GPUs	2 High-End Kepler GPUs
Toplam CUDA Çekirdek Sayısı	768	3072
Bellek Kapasitesi	16 GB DDR3	8 GB GDDR5
Max Güç	130W	225W
Uzunluk	10.5"	10.5"
Yükseklik	4.4"	4.4"
Ekstra Güç Girişi	6-pin	8-pin
Slot Tipi PCIe	x16	x16
Slot Ölçüsü	Gen3	Gen3
Fan Sistemi	Pasif	Pasif

Bu kartların sunduğu farklı sanal GPU profilleri sayesinde 32 sanal makineye kadar GPU paylaşımı yapılabilmektedir. Örneğin Grid K2 kartı, K220Q profili kullanılarak 192 çekirdek, 512MB bellek sahibi 16 adet sanal GPU olarak paylaştırılabilmektedir. Sanallaştırma desteği olmayan ve piyasada sıklıkla kullanılan Quadro 2000 grafik kartların 192 çekirdek sahibi olduğu düşünüldüğünde, GPU sanallaştırma ile güçlü grafik işleme yeteneğine sahip sanal makinelerin oluşturulabildiğini rahatlıkla söyleyebiliriz. Tablo 3.2'de Grid kartlar üzerinde oluşturulabilen sanal profil tipleri ile imkan sağladıkları kapasite ve maksimum kullanıcı sayıları belirtilmektedir [25].

Tablo 3.2: Nvidia grid kartları sanal profilleri.

NVIDIA GRID Grafik Kartı	Sanal GPU Profili	Bellek Kapasitesi	Kullanıcı Başına Ekran Sayısı	Maksimum Çözünürlük	Mak. VM Sayısı	Tavsiye edilen Kullanıcı Tipi
GRID K2	K280Q	4 GB	4	2560×1600	2	Designer
	K260Q	2 GB	4	2560×1600	4	Designer/Power User
	K240Q	1 GB	2	2560×1600	8	Designer/Power User
	K220Q	512 MB	2	2560×1600	16	Designer/Power User
GRID K1	K180Q	4 GB	4	2560×1600	4	Entry Designer
	K160Q	2 GB	4	2560×1600	8	Power User
	K140Q	1 GB	2	2560×1600	16	Power User
	K120Q	512 MB	2	2560×1600	32	Power User

Sanal GPU Profilleri:

Grid kartlar üzerindeki kaynaklar sanal makinenin grafik işleme gücü gereksinime göre farklı profillerde vGPU oluşturularak atanabilir. K1 ve K2 kartların her birinde 4 farklı kapasitede profiller vardır. K2 kartı üzerinde toplamda 3072 CUDA işlemci birimi ile 8GB bellek bulunmaktadır. Bu kart üzerinde oluşturulabilecek en düşük profil K220Q seçeneğidir. Bu profil ile sanal makineye 192 CUDA işlemci birimi ile 512MB bellek verilmiş olur. K220Q profili K2 kartın üzerindeki kaynakların 1/16 kısmını üzerinde bulundurmaktadır. Sırasıyla K240Q profili 1/8, K260Q 1/4, K280Q ise K2 kartın iki ayrı fiziksel GPU'dan birine ait kaynakları kullanabilir (Bakınız Tablo 3.2).

K1 kartı ise 4 fiziksel Kepler GPU dan oluşmakta olup toplamda 16 GB Bellek ile 768 CUDA işlemci biriminden oluşmaktadır. Bu kart içinde 4 farklı sanal GPU profil seçeneği vardır. K120Q minimum kapasiteye sahip olup 512MB bellek ile 24 CUDA işlemci birimden oluşmaktadır. Bu profil fazla GPU gücü gerektirmeyen sistemlerde kullanılmaktadır ve tek karta ait kaynaklar ile 32 adet sanal sisteme GPU paylaştırılabilmektedir. K140Q profili 1GB bellek ile 48 CUDA, K160Q 2GB bellek ile 96 CUDA işlemci birimine sahip profillerdir. K1 kartın en güçlü profili olan K180Q ile her bir fiziksel GPU kaynaklarının atanmasıyla 4 adet 192 CUDA ve 4GB bellek sahibi sanal GPU oluşturularak 4 adet sanal sunucuya kaynak aktarılabilmektedir.

3.3 Citrix XenServer

Sanallaştırma yapılabilmesi için öncelikle altyapıdaki donanımların hipervizör desteklerinin olması şarttır. Sistemlerin kurulacağı donanımlar ile birlikte üzerlerinde bulunan grafik kartının sanallaştırma desteğinin olması gerekir. Yukarıda da belirttiğimiz gibi Nvidia Grid kartlarının sanallaştırma desteğinin bulunmakta ve farklı sanal vGPU profil seçenekleri mevcuttur. Altyapı sanallaştırma desteğimiz olduğuna göre donanımları sanallaştırarak yönetimini yapacağımız bir uygulama gerekir. Çalışmamızda sanallaştırma teknolojilerine önemli katma değerler sağlamış Citrix firmasının sanallaştırma platformunu kullanıldı. Grafik işleyen programların çoğu mevcut grafik kartı kaynağı üzerindeki işlemci birimlerini çok kısıtlı kullanmaktadır. Uygulamanın birçok aşamasında GPU atıl kullanılmaktadır. Tüm grafik işlemci birimleri sadece bir işletim sistemine hizmet verdiğinden ciddi

verimsizlik ortaya çıkmaktadır. Dolayısıyla her alanda olduğu gibi GPU işlenmesi aşamasında da sanallaştırma büyük önem arz etmektedir. Citrix Xen Server ürünü sanal GPU profillerini yönetirken etkili kaynak paylaşımı ile fiziksel grafik işlemci birimlerinin tamamının etkin kullanılmasını sağlamaktadır. Her sanal GPU profili kendine ait işlemci birimlerinde yüksek önceliğe sahip olmakla birlikte, kullanılmayan diğer profillere ait işlemci profillerine de erişebilmektedir. Dolayısıyla kaynaklar en verimli bir şekilde kullanılmaktadır.

Citrix GPU Sanallaştırma Gereksinimleri: Citrix platformu kullanılarak GPU sanallaştırma yapılabilmesi için belli gereksinimlerin sağlanması gerekmektedir.

1. Nvidia Grid K1 ve ya K2 kartları
2. XenServer ve NVIDIA Grid kart desteği bulunan sunucu
3. Citrix XenServer sistemine Nvidia Grid vGPU yazılım paketi yüklenmelidir.
4. Citrix XenServer 6.2 SP1 ve ya yukarı bir sürüm
5. Nvidia sanal GPU kullanan sistemlere erişebilmek ve çalıştırabilmek için Citrix XenDesktop 7.1 ve ya daha ileri versiyonların kullanılması gereklidir.

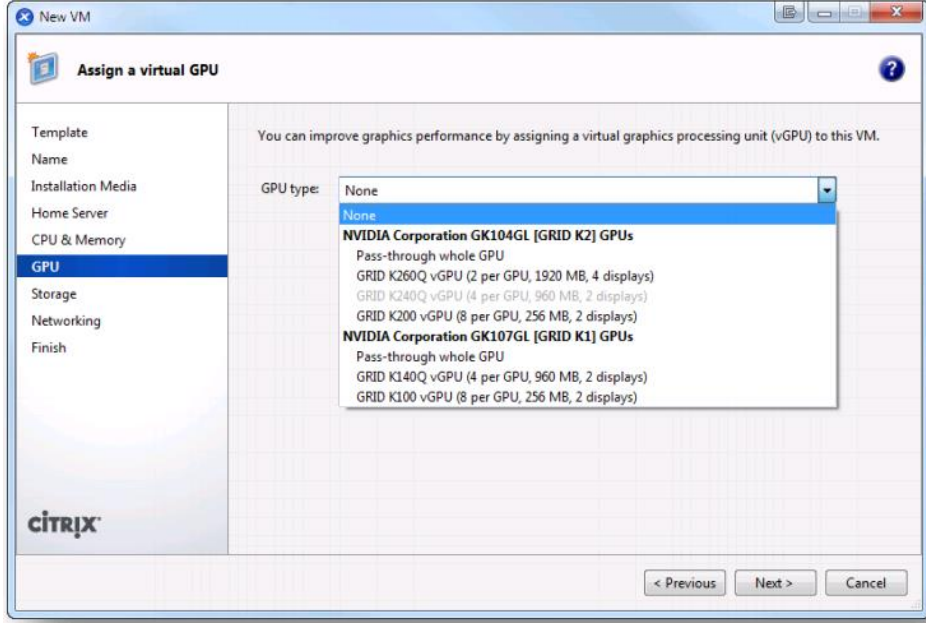
Citrix XenServer vGPU Kurulumu: GPU sanallaştırma için öncelikle vGPU (Virtual GPU) kurulumu gerekmektedir. XenServer üzerinde NVIDIA-vgx-xenserver sürücüsü kurulu olmalıdır.

Sürücüsü kurulumu tamamlandığında “nvidia-smi” komutu ile grid kart üzerindeki fiziksel GPU ve üzerlerindeki sanal profilleri görebilir duruma geliyoruz. Şekil 3.2’de bu komut ile alınan çıktı görülmektedir. Bu çıktıdan GRID K1 kartı üzerinde 8 eşit kaynaklı profil oluşturulduğu görülmektedir.

```
Wed Sep 25 13:05:31 2013
+-----+
| NVIDIA-SMI 4.312.36 Driver Version: 312.36 |
+-----+
| GPU  Name      | Bus-Id      Disp. | Volatile Uncorr. ECC |
| Fan  Temp      | Perf        Pwr:Usage/Cap| Memory-Usage  | GPU-Util    Compute M. |
+-----+-----+-----+-----+-----+-----+
|  N/A  0         | GRID K1     | Off  | N/A |
|  N/A  27C      | P0          | 13W / 31W | 0%   | 9MB / 4095MB | 0%   Default |
+-----+-----+-----+-----+-----+-----+
|  N/A  1         | GRID K1     | Off  | N/A |
|  N/A  25C      | P0          | 13W / 31W | 0%   | 9MB / 4095MB | 0%   Default |
+-----+-----+-----+-----+-----+-----+
|  N/A  2         | GRID K1     | Off  | N/A |
|  N/A  21C      | P0          | 13W / 31W | 0%   | 9MB / 4095MB | 0%   Default |
+-----+-----+-----+-----+-----+-----+
|  N/A  3         | GRID K1     | Off  | N/A |
|  N/A  23C      | P0          | 13W / 31W | 0%   | 9MB / 4095MB | 0%   Default |
+-----+-----+-----+-----+-----+-----+
|  N/A  4         | GRID K1     | Off  | N/A |
|  N/A  24C      | P0          | 13W / 31W | 0%   | 9MB / 4095MB | 0%   Default |
+-----+-----+-----+-----+-----+-----+
|  N/A  5         | GRID K1     | Off  | N/A |
|  N/A  24C      | P0          | 13W / 31W | 0%   | 9MB / 4095MB | 0%   Default |
+-----+-----+-----+-----+-----+-----+
|  N/A  6         | GRID K1     | Off  | N/A |
|  N/A  25C      | P0          | 13W / 31W | 0%   | 9MB / 4095MB | 0%   Default |
+-----+-----+-----+-----+-----+-----+
|  N/A  7         | GRID K1     | Off  | N/A |
|  N/A  25C      | P0          | 12W / 31W | 0%   | 9MB / 4095MB | 0%   Default |
+-----+-----+-----+-----+-----+-----+
+-----+
| Compute processes: GPU Memory |
| GPU  PID      Process name  Usage  |
+-----+-----+-----+-----+
| No running compute processes found |
+-----+
[root@xenserver ~]#
```

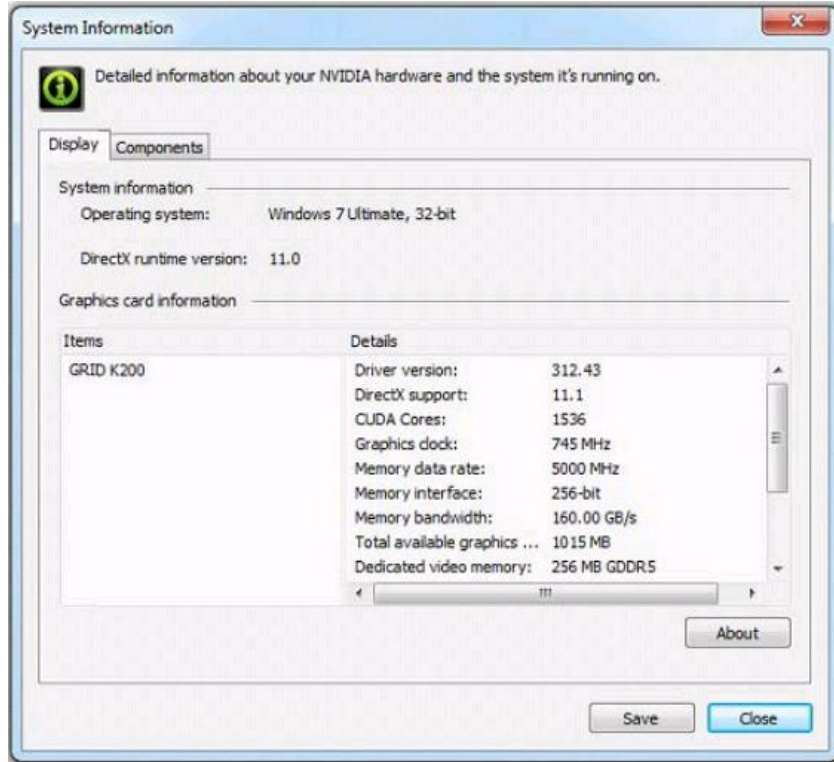
Şekil 3.2: nvidia-smi komutu ile alınan örnek bir çıktı.

Sürücü kurulumları tamamlandıktan sonra sanal makinelere GPU ekleyebilir duruma gelmektedir. Sanal makine kaynaklarına gelindiğinde GPU menüsünde artık yukarıda değindiğimiz profilleri görebiliyoruz. İstedığımız profili seçerek kaynak ataması tamamlandığında sanal makine üzerinde yeni bir ekran kartı olarak görülecektir. Şekil 3.3’de sanal makineye GPU profili eklerken elde edilen ekran görüntüsü görülmektedir.



Şekil 3.3: Sanal GPU atama işlemi.

Daha sonra sanal makine üzerine ilgili grid kartı sürücülerini yüklenerek vGPU sisteme tanımlanmış olur. Nvidia kontrol panelinde sürücü bilgileri aşağıdaki şekildeki gibi görülecektir. Şekil 3.4'de sanal makine tarafında sanal GPU ve özellikleri görülmektedir.



Şekil 3.4 Sanal GPU sürücü bilgileri görüntülenmesi.

3.4 Masaüstü Sanallaştırma

GPU sanallaştırma teknolojilerinin bilişim sektörüne yansıyan en önemli etkisi klasik masaüstü sistemlerinin sanallaştırılması olmuştur. Bugün müşteriye hizmet veren, aralarında son derece kritik uygulamalarında olduğu birçok sunucu sanallaştırılmış durumdadır. Aslında özellikle yönetim ve bakım ve güvenlik maliyetleri düşünüldüğünde çok daha fazla merkezileştirilmesinde fayda olan masaüstü sistemler teknik gereksinimler oluşmadığından fiziksel olarak kullanılmaya devam etmektedir. Masaüstü sanallaştırma öneminin temel sebepleri şu şekilde sıralanabilir.

Yerinde Destek İhtiyacı:

Bilişim sistemlerinde bugün birçok yönetsel işlemler uzak bağlantı ile farklı lokasyonlardan yapılabilir durumdadır. Fakat makinenin kapatılması, yeniden başlatılması, ağ yönetimi gibi işlemlerde normal yöntemlerle uzaktan erişim mümkün olamayacağından konsol erişimi zorunlu hale gelmektedir. Masaüstü sistemler dağınık lokasyonlarda olduğundan konsol erişim altyapı hazırlanması ekstra maliyet ve yönetim zorluklarını beraberinde getirmektedir. Özellikle donanım değişikliği gibi durumlarda ise yerinde destek kaçınılmaz olmaktadır. Masaüstü sistemler sunucu tabanlı sistemlere göre çok daha fazla arıza verebilmekte ve yerinde destek gereksinimi dolayısıyla daha yavaş çözüm sunulduğundan onarım süresince bazı personel hizmet veremez duruma gelmektedir.

Sistemlerin güncellenmesi, yama geçilmesi gibi durumlarda ise merkez ile şube arasındaki ağ genişliğinin düşüklüğü bununla birlikte güncelleme paket boyutlarının büyüklüğü ile güncellenecek sistemlerin sayısının çokluğu yerinde destek hizmetini zorunlu hale getirmektedir. Dolayısıyla kurumlar zaman ve işgücü maliyeti anlamında ciddi sorunlar ile karşılaşmaktadırlar. Özellikle kritik güvenlik yamalarının yüklenmesi bu sebeple çok gecikebilmektedir.

Sık Sık Karşılaşılan Donanım Arızaları:

Masaüstü bilgisayarları üzerlerindeki mekanik parçalardan (fan, hdisk) dolayı sık sık arıza vermektedir. Arızaların giderilmesi işlemi ve bu süre içinde personelin çalışmaması kurum adına iş gücü ve maliyet kaybı anlamına gelmektedir. Ayrıca zamanla sistemlerden çıkan fan sesleri çalışma ortamının konforuna olumsuz etkiye bulunmaktadır. Masaüstü sanallaştırmada kullanılan ince istemciler üzerinde hiçbir mekanik aksam olmadığından gürültüsüz çalışmakta, kolay kolay

arızalanmamaktadır. Olası arıza durumunda ise farklı bir platform aracılığı ile sisteme çok rahat erişebilmektedir.

Veri Güvenliği Problemi:

Masaüstü sanallaştırmada tüm sistemler ve veriler merkezi, güvenli bir ortamda tutulmaktadır. Dolayısıyla disk çalınması, kaybolması gibi bir durum söz konusu değildir. Ayrıca donanıma, bios ve boot menüsüne direk erişim olmadığından normal masaüstlerindeki gibi ikinci bir işletim sistemi aracılığı makine başlatılamayacaktır. Sistemlerin yönetimi merkezde olduğundan kullanıcıların yaptığı tüm işlemler daha etkin bir şekilde kayıt altına alınabilmektedir. Dolayısıyla veri güvenliği çok daha üst seviyededir.

Enerji Tüketimi:

Sanal masaüstünün önemli avantajlarından biri de düşük enerji tüketimidir. Klasik masaüstü bilgisayarları üzerindeki mekanik parçalar (fan, disk) ile ekran kartı, işlemci gibi diğer parçalar yüksek güç tüketimine neden olmaktadır Tablo 3.3'de ince istemcilerle, masaüstü bilgisayarların enerji tüketimleri gösterilmektedir [26].

Tablo 3.3: İstemci enerji tüketimleri (Monitörler dahil).

İstemci Tipi	Birim Enerji Tüketimi	100 Bilgisayar	1000 Bilgisayar	5000 Bilgisayar
3200	92 watts	9.200 watts	92.000 watts	460.000 watts
3630	24 watts	2.400 watts	24.000 watts	120.000 watts
8230	93 watts	9.300 watts	93.000 watts	465.000 watts
Kişisel Bilgisayar	170 watts	17.000 watts	170.000 watts	850.000 watts

EPDK verilerine göre mesken kwh ücreti 41.17 kr olup bu birim fiyat üzerinden bir hesap yapıldığında karşımıza şu sonuç çıkmaktadır.

5000 desktop ile çalışıldığını düşündüğümüz bir kurumda hafta içi 5 gün ortalama günde 9 saat bilgisayarların açık olduğu varsayılarak yapılacak bir hesaplama yapıldığında şu sonuçla karşılaşırız.

Bir yılda 52 hafta olup, çalışılacak gün sayısı $52 \times 5 = 260$ gün.

Toplam $260 \times 9 = 2340$ saat bilgisayarlar çalışacaktır.

Fiziksel Yapı:

5000 PC bir saatte tükettiği enerji yukarıdaki tabloda belirtildiği üzere 850kw.

Bir yılda tüketilen toplam enerji: $850 \times 2340 = 1989000 \text{ kwh}$

$$1989000 \times 0,4117 = 818.871,3 \text{ TL}$$

Masaüstü Sanallaştırma:

3630 Thin Client ürünü kullanıldığı durumda 5000 adet 3630 cihazı ve monitörler toplamda 120kw enerji tüketmektedirler.

$$\text{Bir yılda tüketilen toplam enerji: } 120 \times 2340 = 280800 \text{ kwh}$$

$$280800 \times 0,4117 = 115605,36 \text{ TL}$$

$$\text{Bir yılda elde edilen kar tutarı: } 818.871,3 - 115605,36 = 703.265,94 \text{ TL}$$

10 yılda toplam 7 milyon TL üzerinde kazanç sağlandığı görülmektedir.

Masaüstü Sanallaştırma için Kullanılan Uç Birimler

Masaüstü sanallaştırmada kullanıcılar sunuculara farklı özelliklerdeki istemcilerle erişeceklerdir. Bunlar, Kalın İstemci, İnce İstemci ve Sıfır İstemcidir. Aslında masaüstüne bağlanmak için bu istemcilere ihtiyaç yoktur. İşletim sistemi taşıyan herhangi bir cihazdan bağlanmak mümkündür. Gerekli istemci yazılımının yüklenmesi yeterlidir.

Kalın istemci (Thick Client): Üzerinde kullanıcı tarafından formatlanabilir işletim sistemi bulunan, veri saklanmak üzere depolama birimi bulunan tablet tipi istemcilerdir. Güvenlik, işletme ve enerji tüketim (150 W) maliyetleri nispeten yüksek, teknolojik hayat süresi kısadır (2-3 yıl) [27].

İnce istemci (Thin Client): Üzerinde düşük performanslı bir işlemci bulunan, kendine ait depolama biriminin bulunmadığı bir istemcidir. Gerek performans düşüklüğü, gerekse disk olmaması nedeniyle farklı amaçlarla kullanılamaz. Alternatif uzaktan erişim protokollerini desteklemektedir. Sadeleştirilmiş bir işletim sistemi çekirdeğine sahip olduğundan uzun ömürlü ve güvenilir bir üründür. Güvenlik, işletme ve enerji tüketim (20 W) maliyetleri nispeten düşüktür. Teknolojik hayat süresi uzundur (5-7 yıl) [27].

Sıfır istemci (Zero-Client): Masaüstü sanallaştırma için son derece optimize edilmiş bir ürün olup, video resim gibi ikili verileri sıkıştırma/açma için özelleştirilmiş işlemciye sahiptir. Yerel veri depolama birimi, işletim sistemi ve uygulaması yoktur. (Donanım üzerinde çalışan sadeleştirilmiş firmware içerir.) Dolayısıyla son derece güvenlidir. Bakım ve güncelleme gereksinimi olmadığından son derece uzun ömürlü ve sorunsuz çalışmaktadır. Farklı masaüstü erişim protokolleri için tasarlanmış özel ürünler olduğu gibi, birden fazla protokolü

destekleyenleri de vardır. Maliyeti görece yüksek ancak işletme ve enerji tüketim (6 W) maliyetleri oldukça düşük, güvenliği son derece yüksektir [27].



DÖRDÜNCÜ BÖLÜM

DENEY ORTAMI VE ELDE EDİLEN SONUÇLAR

Tez çalışmamızda maliyet koşulları da değerlendirilerek fiziksel GPU kullanımı ile sanal GPU kullanımlarının karşılaştırılması yapılmıştır. Maliyet değerlendirilmeden yapılan çalışma objektif olmamakla birlikte elde edilen sonuçların ARGE ve piyasa koşullarında da karşılığı olmayacaktır. Çalışmamızda kullanılan donanımlar ile performans ölçüm araçları hakkında bilgi verilecektir.

4.1 Kullanılan Donanım ve Yazılımlar

Fiziksel grafik kartı testlerini 4 çekirdekli fizisel iş istasyonu, sanal GPU testleri ise 16 çekirdekli sanal platformu üzerinde yapıldı. Tablo 4.1 ve Tablo 4.2’de testlerimizde kullanılan sunucu ve iş istasyonlarına ait donanım bilgileri yer almaktadır.

Tablo 4.1: Sanal platformun kurulduğu fiziksel sunucu özellikleri.

16 çekirdekli Fiziksel Sunucu Özellikleri	
Form Faktörü	2U Raf
İşletim Sistemi	Citrix Xen Server 6.5
İşlemci Ailesi	Intel® Xeon® E5 v3 işlemci
İşlemci	Intel® Xeon® E5-2650
Bellek Kapasitesi	256GB
Bellek yuvaları	24 DIMM
Sürücü denetleyicileri	Tümleşik HP Smart Array B140i Denetleyici
Dahili Depolama	1,6 TB SAS SSD, 4TB 10k SAS Disk
Optik Depolama	İnce SATA DVD-ROM; İnce SATA DVD-RW3
Grafik İşlemci	Sanallaştırılmış GPU: NVIDIA® GRID™ K2 (8 GB)
Güç	2x1400 W %94 verimli FlexSlot
Boyutlar	44,55 x 67,95 x 8,73 cm
Ağırlık	18,54 kg

Tablo 4.2: Fiziksel iş istasyonu özellikleri.

Test Edilen Fiziksel İş istasyonu	
Form Faktörü	Convertible minitower
İşletim Sistemi	Windows 7 Professional 64-bit*
İşlemci Ailesi	Intel® Xeon® E5
İşlemci	Intel® Xeon® E5-1630 V3
Bellek Kapasitesi	32GB
Bellek yuvaları	8 DIMM
Sürücü denetleyicileri	Tümleşik HP Smart Array B140i Denetleyici
Dahili Depolama	1TB SATA
Optik Depolama	İnce SATA DVD-ROM; İnce SATA DVD-RW3
Grafik İşlemci	GPU: NVIDIA® K2000
Güç	2x1400 W %94 verimli FlexSlot
Boyutlar	44.76 x 17.78 x 44.52 cm
Ağırlık	12,5 kg

Grafik işlemcilerin performans ölçümleri, piyasada bilinen GPU-Z, Fraps ve Unigine Heaven uygulamaları kullanılarak yapıldı. Bu uygulamalar ile grafik kartı tarafından saniyede üretilen kare sayısı(FPS) izlendi. GPU-Z uygulaması GPU sıcaklığı, kullanılan bellek miktarı gibi verileri de izleyebilmektedir.

4.1.1 GPU-Z

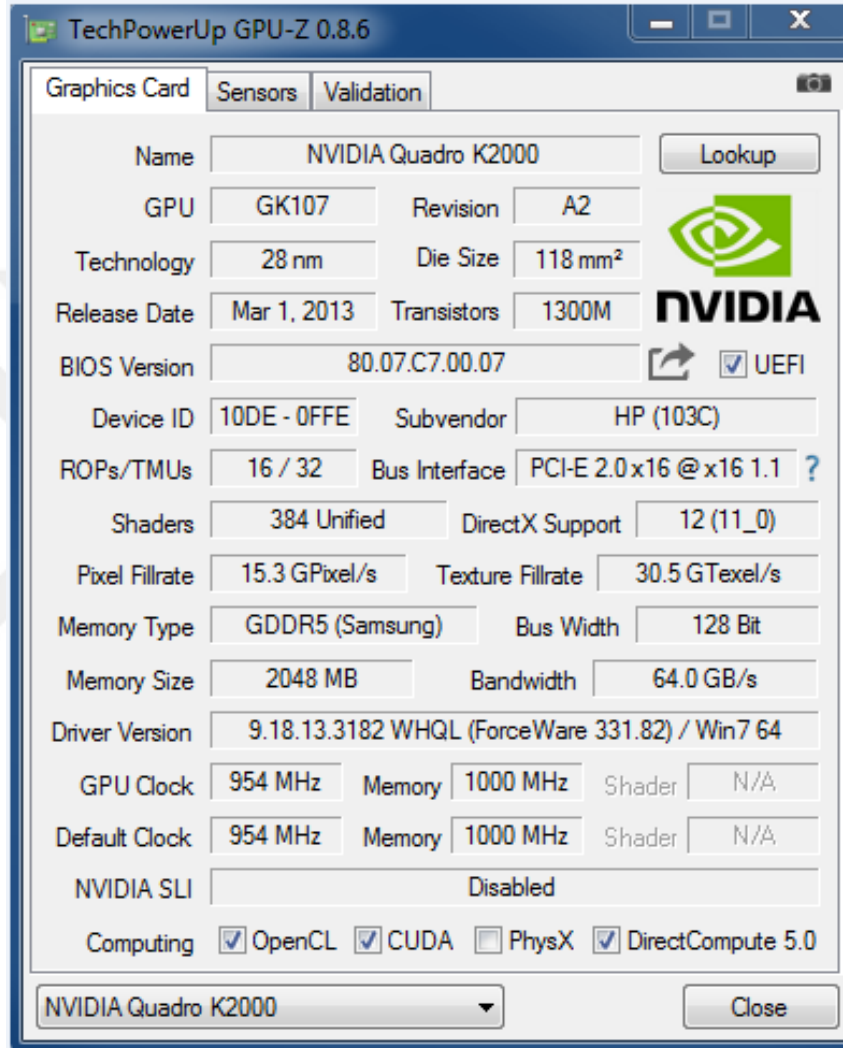
TechPowerUp tarafından geliştirilmiş olan GPU-Z uygulaması grafik kartların özelliklerini göstermenin yanında GPU yük testi, GPU sıcaklığı, fan hızları, bellek kullanımı gibi performans ve bakım amaçlı bilgileri de toplayabilen bir yazılım olup, standart kullanıcılar için ücretsiz olarak elde edilebilmektedir. Kurulum yapılmadan çalıştırılabilir bir dosya olarak da kullanılabilir.

GPU-Z Özellikleri

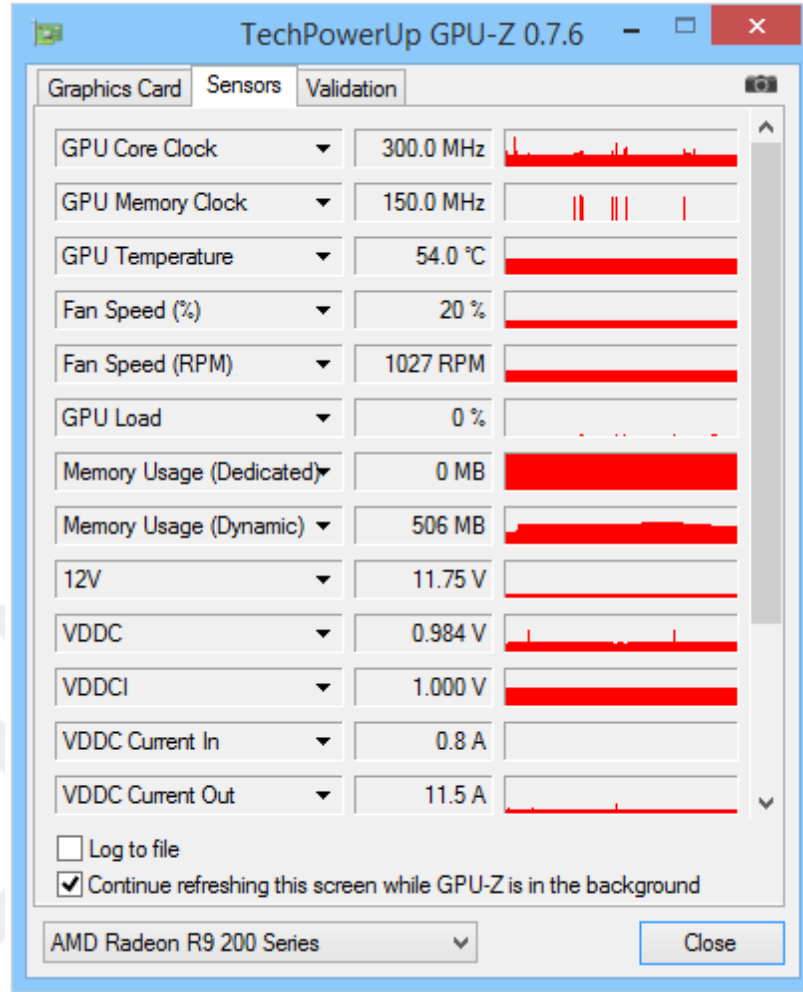
1. NVIDIA, ATI ve Intel grafik kartlarını desteklemektedir.
2. GPU ve Ekran kartının donanım özelliklerini göstermektedir.
3. Overclock, Default Clock, varsa 3D Clock özelliklerini göstermektedir.
4. Yük testi ile GPU performansını ölçme
5. Grafik kartının BIOS bilgilerinin yedeğini alabilmektedir.

6. Kurulum ihtiyacı yoktur, yönetimsel yetki olmadan da kullanılabilir.

Aşağıdaki Şekil 4.1 ve Şekil 4.2’de GPU-Z uygulaması ile ekran kartı özellikleri ve performans bilgilerinin gözlenebildiği gösterilmektedir. Bu bilgileri metin (text) olarak kayıt altına alabilmekte ve bu verilerden grafikler elde edilebilmektedir.



Şekil 4.1: GPU-Z Ekran Kartı Özellikleri.



Şekil 4.2: GPU-Z performans görüntüleme.

GPU-Z elde ettiği verileri bir kayıt (log) dosyasında tutabildiğinden geçmişe dönük performans raporları çıkarabilme konusunda da yardımcı olabilmektedir. Nitekim çalışmamızda bu veriler kayıt edilerek raporlar elde edildi.

4.1.2 Fraps

Fraps bilgisayarın ekranında yaptığınız her şeyi video şeklinde kaydetmeye yarayan bir programdır. Yukarıda yapabileceğiniz 4 ayar mevcuttur. Birincisi genel ayarlar. İkincisi FPS ayarları. Üçüncüsü çekeceğiniz video ayarları, dördüncüsü ise ekran görüntüsü alma ayarı. Burada çalışmamız açısından en önemli ayarlar FPS ayarlarının yapıldığı yer. Burada öncelikle performans testine başlamak için bir kısayol tuşu belirlenmektedir. Varsayılan tuş “F11” tuşudur. Ayrıca hangi metriklerin

toplanacağı ilgili kutucuklar işaretlenerek belirlenir. GPU-Z uygulaması ile FPS değerleri toplanamadığından bu eksiklik Fraps uygulaması ile giderilmiş oldu.

“Folder to save benchmarks in” bölümünde toplanan performans ölçüm değerlerinin saklanacağı dizin belirlenir. Aşağıda Fraps uygulamasına ait FPS ayarlarının yapıldığı formun ekran görüntüsü bulunmaktadır.



Şekil 4.3: Fraps uygulama arayüzü.

4.1.3 Unigine Heaven

Ekran kartlarının performansını test etmek için özel yazılımlar geliştirilmiştir. Bu programlar sayesinde çok farklı senaryolarla kartın grafik üretme yeteneği test edilebilmektedir. Bu test programları arasında başarılı bir yeri olan ve bizim deney ortamımızda kullandığımız Unigine Heaven Benchmark'tır. Bu yazılım ücretsiz olarak indirilebilmekte olup 4.0 versiyonu yayınlanmış durumdadır [28].

Unigine Heaven 4.0 özellikleri:

1. Donanım kararlılık testi
2. %100 GPU tabanlı net sonuçlar
3. DirectX 9, DirectX 11 ve OpenGL 4.0 desteği
4. Değişen ışık şartları ortamını oluşturabilme
5. Değişen gökyüzü ve yansıyan ışık süzmeleri.
6. Advanced SSAO (screen-space ambient occlusion)
7. Fiziksel formüllerle hesaplanan kitlesel bulut yoğunluğu.

8. Test modları üstünde etkileşimli dolaşabilme.

9. Üç adet tesellation test modu.

Unigine Heaven performans hesaplama aracı toplam 26 farklı görsel senaryo işleterek elde ettiği değerler üzerinden bir puan vermektedir. Sonuç ekranına elde edilen minimum ve maksimum ile ortalama FPS değerleri, testin yapıldığı sistem üzerindeki GPU ve CPU bilgileri ile test sonucunda verilen puan belirtilmektedir.

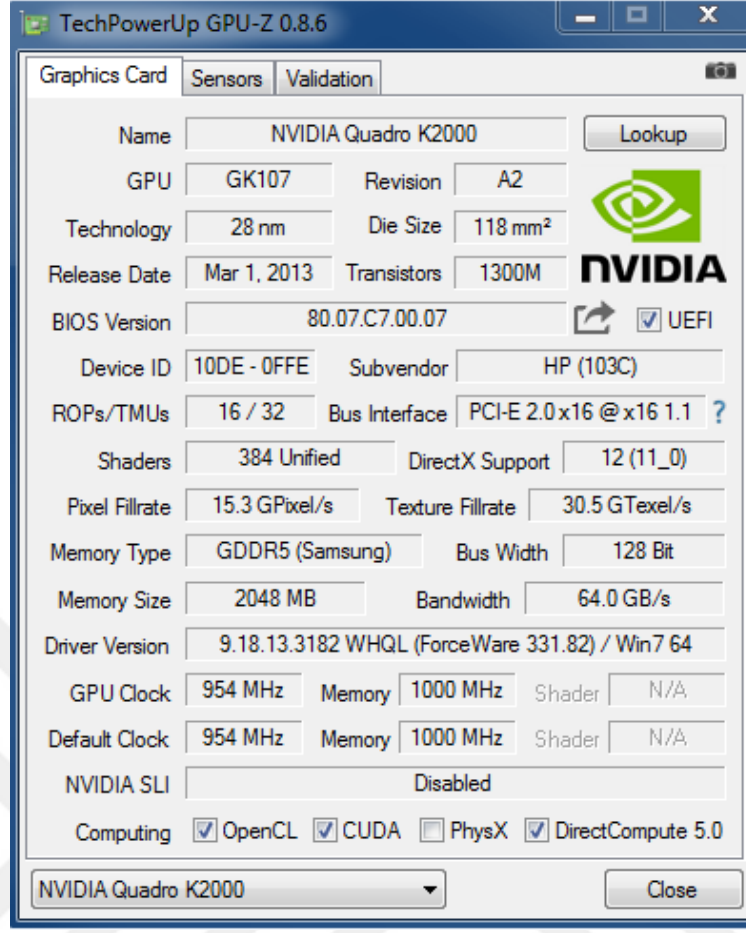
Bu çalışmamızda yapılacak testlerde üzerinde NVIDIA K2000 ekran kartı bulunan iş istasyonları ile her bir iş istasyonu için K240Q profili olan sanal iş istasyonlarının grafik performansları karşılaştırıldı. Karşılaştırmada en önemli veri olarak FPS değerleri referans alındı. FPS değeri bize GPU tarafından saniyede işlenen kare/ekran sayısını vermektedir. Bu değer yüksek olması grafik kartının başarısını ortaya koymaktadır.

Aslında insan gözünün saniyede 30FPS algılayabildiği kabul edilir. Ekranın saniyede yenilenme sayısının çok fazla olması ekran akışkanlığı açısından bir şey ifade etmeyecektir. Fakat normal bir film karesi değil de oyun ve ya grafik tasarım programından bahsedildiğinde GPU tarafından işlenen kare sayısının yüksek olması GPU hesaplama gücünü gösterdiğinden performans adına önemli bir ayrıçtır. FPS sayısı sabit değildir, işlenen grafiklerin karmaşıklığına göre yükselme ve düşmeler görülecektir. Dolayısıyla yüksek hesaplama ihtiyacı olan kareler işlenirken FPS değerinde düşmeler yaşanacaktır.

4.2 Testler ve Sonuçları

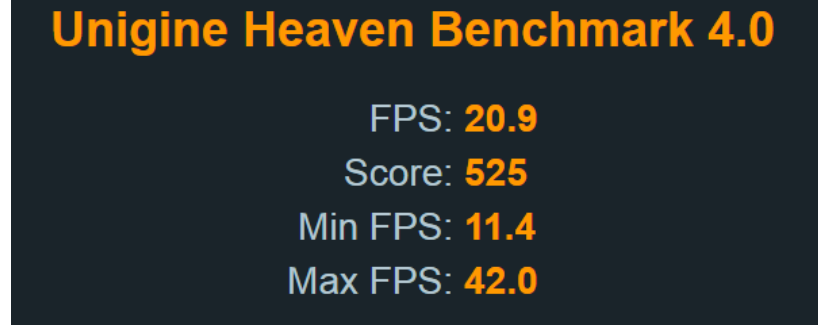
4.2.1 Fiziksel İş İstasyonu Testleri

Testde kullandığımız fiziksel iş istasyonu üzerinde orta seviye diyebileceğimiz profesyonel bir ekran kartı bulunmaktadır. Bu ekran kartına ait GPU-Z uygulamamızdan çektiğimiz özellikleri aşağıdaki Şekil 4.4'de görülmektedir. Sistem üzerinde NVIDIA Quadro K2000 grafik kartı olup, 384 CUDA, 1000MHz 2048MB Bellek,128bit veri yoluna sahiptir.



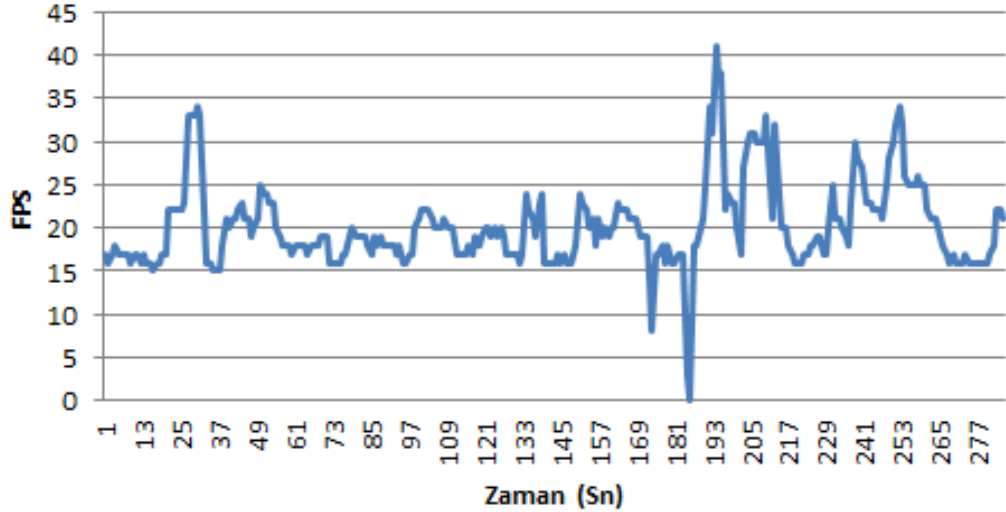
Şekil 4.4: Quadro K2000 ekran kartı özellikleri.

Fiziksel iş istasyonu olarak yukarıda da özelliklerini belirttiğimiz sistem üzerinde yaptığımız testlerde elde ettiğimiz değerlerde ortalama 20.5 FPS değerini gözlemledik. Test boyunca en düşük kare işleme sayısı 6, maksimum işlenen kare sayısı ise 40 değerini elde ettik. Bu değerleri Unigine Heaven 4.0 performans testi uygulaması ile gözlemledik. Unigine Heaven 4.0 Benchmark, 26 farklı senaryo ile grafik kartını test etmektedir. Her senaryoda farklı grafik işlemleri yapılarak homojen bir sonuç ortaya konmaktadır. Bazı senaryolarda aydınlık, parlaklık testleri, bazılarında gölgeler ve ışıklar, bazılarında ise aero-dinamik vs gibi farklı hesaplama tekniklerinin kullanıldığı senaryolar bulunmaktadır. Yüksek çözünürlük durumunda işlenecek piksel sayısı da arttığından FPS sayısına direk etkisi olmaktadır. Çözünürlük arttıkça saniyede işlenen kare sayısı düşecektir. Dolayısıyla testler ve karşılaştırmalar aynı çözünürlük seviyesinde yapılarak değerlendirilmiştir. Unigine Heaven aracı ile aldığımız test sonucu Şekil 4.5’de görülmektedir.



Şekil 4.5: Fiziksel ekran kartı test sonucu.

Ayrıca Fraps uygulaması üzerinden çektiğimiz FPS raporu ile tüm test boyunca elde edilen performans verilerini grafik ortama dökme imkanı elde etmiş olduk. FPS raporundan anlaşılacağı üzere test senaryosuna göre FPS değerleri de değişiklik arz etmektedir. Quadro K2000 fiziksel kartın test süresince ürettiği FPS değerleri Şekil 4.6’da görülmektedir.



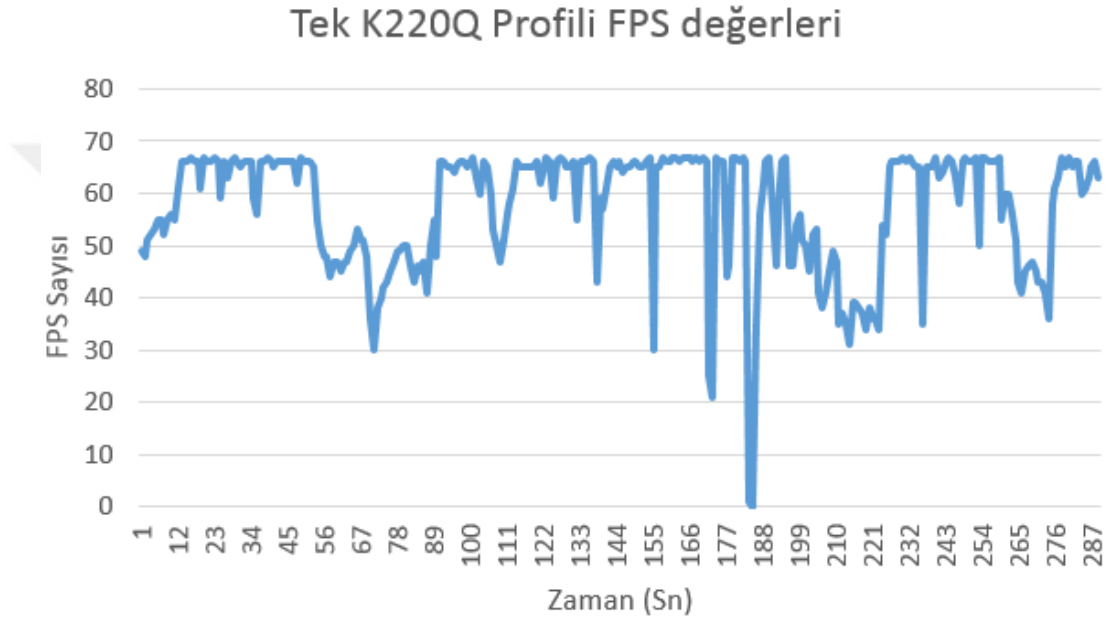
Şekil 4.6: K2000 Fiziksel kart FPS değerleri.

4.2.2 Sanal İş İstasyonu Testleri

16 çekirdekli fiziksel sunucu üzerine kurduğumuz sanal iş istasyonlarının üzerinde grafik işlemek için K220Q sanal GPU profilleri kullanılmaktadır. K220Q profili, makine üzerindeki fiziksel grafik kartında bulunan iki fiziksel GPU dan biri üzerinde çalışmakta olup, bir fiziksel GPU toplamda 8 adet K220Q profili desteklemektedir.

K220Q profili yukarıda da belirtildiği üzere Citrix Xen Server tarafından yönetilmektedir. GPU sanallaştırılırken maksimum verimi elde edebilmek için fiziksel katmandaki tüm işlemci birimleri kullanılmaktadır. Şekil 4.7’de görülen değerler, diğer sanal iş istasyonları grafik işlem yapmadıkları esnada elde edilmiştir. Fiziksel kaynaklar müsait olduğundan tüm kaynaklar bu sanal makineye hizmet vermiş, dolayısıyla yüksek FPS değerleri gözlemlenmiştir.

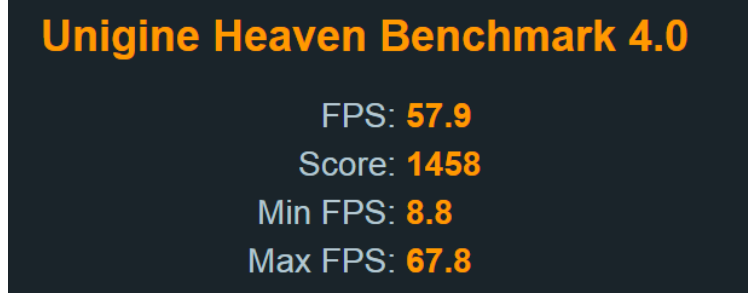
4.2.2.1 Tek sanal makine testi



Şekil 4.7: Sanal K220Q GPU profili FPS değerleri.

Unigine Heaven Benchmark 4.0 ile yapılan performans testi ise Şekil ortalama 57.9 FPS değeri ve 1458 puan ile sonuçlanmıştır (Bakınız Şekil 4.8).

Tek sanal iş istasyonunda yapılan test sonucunda sanal GPU tarafında fiziksel iş istasyonuna göre çok daha yüksek bir performans elde edildiği tespit edilmiştir. Bu performans farkı sanallaştırmanın sağladığı optimum kaynak kullanımı ile ihtiyaç durumunda GRID K2 kartının tek fiziksel GPU ünitesinde bulunan 1536 grafik işlemci biriminin de devreye alınmasıyla elde edilmiştir.

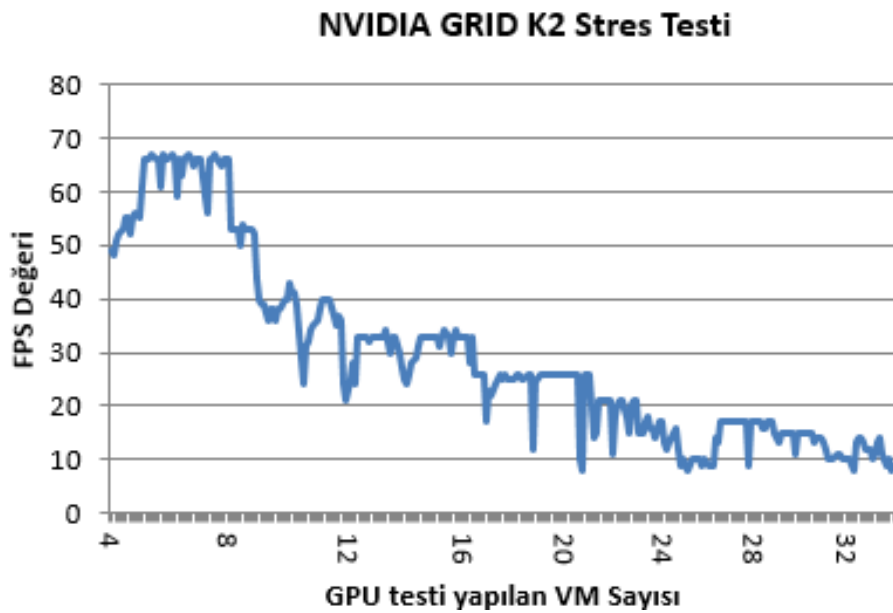


Şekil 4.8: Unigine Heaven K220Q test sonucu.

4.2.2.2 Stres Testleri

Tek sanal iş istasyonu yukarıda belirttiğimiz üzere fiziksel grafik işlemci ünitelerinin hepsine erişebilir durumda olduğundan sürdürülebilir bir performans olarak değerlendirilmemektedir. Dolayısıyla yoğun GPU kaynağı tüketen sanal makine sayısı artırılarak maksimum iş yükü altında elde edilecek performans değerlerinin gözlemlenebilmesi için stres testi yapılmıştır.

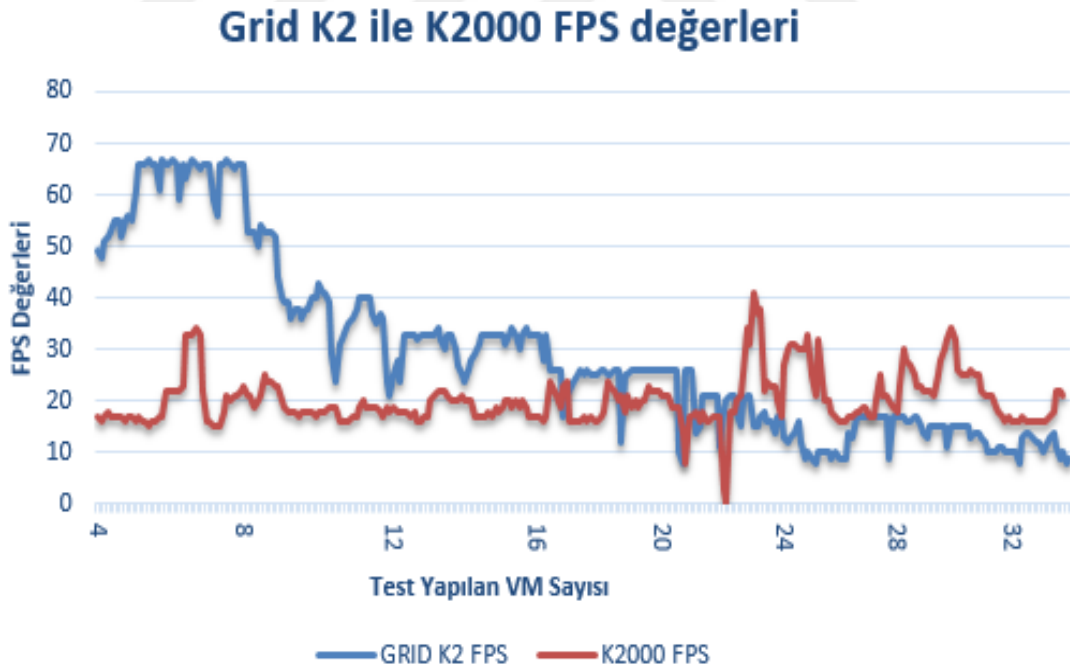
Sanal makinelerden biri üzerinde performans testini başlattıktan sonra aynı fiziksel kart üzerindeki kaynakları kullanan diğer tüm sanal makineler üzerinde de yoğun GPU kullanan uygulamalar çalıştırılarak aşağıdaki sonuç gözlemlenmiştir. 26 senaryodan oluşan Unigine Heaven performans testi başlatıldıktan sonra belli aşamalarda diğer sanal makinelerde de yoğun GPU kullanan işlemler başlatılmıştır. Şekil 4.9’da yaptığımız stres testi FPS değerleri grafik olarak gösterilmektedir.



Şekil 4.9: Sanal iş istasyonları toplu stres testi.

Bu grafiđi dođru yorumlayabilme adına GPU sanallařtırma altyapısını kısaca gözden geçirmekte fayda vardır. Sanal ortamda kullanılan NVIDIA GRID K2 kartı üzerinde iki adet fiziksel GPU bulunmaktadır. Her bir fiziksel GPU üzerinde ise 8 adet sanal K220Q profili oluşturulabilmektedir. Sadece bir K220Q profili üzerinde, son derece iyi deđerler elde edilirken, diđer 7 sanal iř istasyonu üzerinde de belli aralıklarla GPU iřlemleri bařlatıldıkça FPS deđerlerinde düşüşler gözlemlenmiřtir. Sistem üzerinde iki adet Nvidia Grid K2 kartı olduđu varsayıldıđında toplam 4 fiziksel GPU, dolayısıyla toplamda 32 adet K220Q profili oluşturulabilmektedir. Tüm sanal iř istasyonlarındaki profiller yoğun olarak kullanıldığında fiziksel K2000 kartına göre düşük performans verdiđi tespit edilmiřtir.

řekil 4.10'da fiziksel ve sanal iř istasyonları üzerlerinde yapılan GPU performans testlerinden elde edilen deđerler aynı grafikte birlikte gösterilmiřtir. Bu veriler Unigine Heaven 4.0 tarafından geliřtirilmiř aynı test senaryoları sonucunda elde edilmiřtir. Bu grafikte sanal iř istasyonlarının belli bir adede kadar daha iyi performans sergiledikleri görülmektedir.



řekil 4.10: Fiziksel GPU testi ile Stres testi deđerleri.

4.3 Maliyet-Performans Analizi

Fiziksel iş istasyonlarına ait maliyetler, iş istasyonu sayısı ile doğru orantılı olarak artacaktır. Dolayısıyla işletme tarafından ihtiyaç duyulan iş istasyonu sayısı arttıkça maliyetlerde ciddi artışlar gözlemlenecektir. Bu durumda sanal iş istasyonu çözümlerinin sunduğu maliyet avantajı da yükselmektedir. Çalışmamızda kullandığımız iş istasyonuna ait maliyet bilgisi Tablo 4.3’de görülmektedir.

Tablo 4.3: Giriş seviye fiziksel iş istasyonu fiyat bilgisi.

Model	İş İstasyonu (1cpu, 4cores, 8threads)
Power Supply	700W 90% Efficient Chassis
İşlemci	Intel Xeon E5-1630v3 3.7GHz 10MB Cache 2133MHz 4C
Chipset	Intel C612 workstation Chipset
Bellek	32GB (2x16GB) DDR4 2133MHz ECC (max.128GB, 8 DIMM)
Ekran Kartı / GPU	NVIDIA Quadro K2000 4GB
Harddisk	256GB SATA 1st Solid State Drive
Harddisk-2	256GB SATA 2nd Solid State Drive
Harddisk-3	1TB 7200 RPM SATA 3rd Hard Drive
Harddisk-4	1TB 7200 RPM SATA 4th Hard Drive
Controller	LSI 9217-4i4e 8-port SAS 6Gb/s RAID Card
Ethernet	361T PCIe Dual Port Gigabit NIC
Fiyat	\$4.560/adet

Görüldüğü üzere 1 adet fiziksel iş istasyonunun fiyatı \$4560 olup, 16 adet iş istasyonu ihtiyacı durumunda toplam maliyet \$72960 olacaktır. Ayrıca test ortamında kullandığımız iş istasyonu, giriş seviyesi bir iş istasyonu olup diğer iş istasyonlarının maliyetleri çok daha yüksektir.

Tablo 4.4’de alternatif orta seviye iş istasyonu fiyat bilgisi görülmektedir. Testlerimizde kullandığımız iş istasyonuna göre %30 daha fazla maliyeti olduğu görülmektedir.

Tablo 4.4: Orta seviye iş istasyonu fiyat bilgisi.

Model	İş İstasyonu (1cpu, 4cores, 8threads)
Power Supply	925W 90% Efficient Chassis
İşlemci	Intel Xeon E5-2637 v4 3.50GHz 9.6GT/s 15MB Cache 2400MHz 4C
Chipset	Intel C612 workstation Chipset
Bellek	32GB (2x16GB) DDR4 2400MHz ECC RAM (max.256GB, 8 DIMM)
Ekran Kartı / GPU	NVIDIA Quadro K2000 4GB
Harddisk	256GB SATA 1st Solid State Drive
Harddisk-2	256GB SATA 2nd Solid State Drive
Harddisk-3	1TB 7200 RPM SATA 3rd Hard Drive
Harddisk-4	1TB 7200 RPM SATA 4th Hard Drive
Controller	LSI 9217-4i4e 8-port SAS 6Gb/s RAID Card
Ethernet	361T PCIe Dual Port Gigabit NIC
Fiyat	\$5.965/adet

Üst seviye iş istasyonları arasında geçen iş istasyonuna ait örnek bir fiyatlandırma bilgisi Tablo 4.5’de sunulmuştur. Maliyetin giriş seviyesi bir iş istasyonuna göre neredeyse iki katına çıktığı görülmektedir.

Tablo 4.5: Yüksek seviye iş istasyonu fiyat bilgisi.

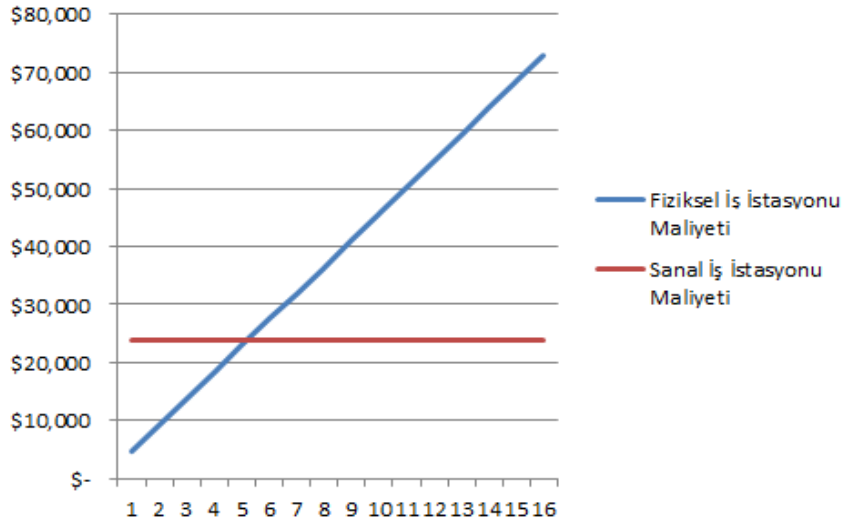
Model	İş İstasyonu (1cpu, 6cores, 12threads)
Power Supply	1125W (1450W / 200V) 90% Efficient rackable Chassis
İşlemci	Intel Xeon E5-2643 v4 3.4GHz 9.6GT/s 20MB Cache 2400MHz 6
Chipset	Intel C612 workstation Chipset
Bellek	32GB (2x16GB) DDR4 2400MHz ECC RAM (max.256GB, 8 DIMM)
Ekran Kartı / GPU	NVIDIA Quadro K2000 4GB
Harddisk	Z Turbo Drive G2 512GB PCIe 1st SSD
Harddisk-2	Opsiyonel
Harddisk-3	1TB 7200 RPM SATA 3rd Hard Drive
Harddisk-4	1TB 7200 RPM SATA 4th Hard Drive
Controller	LSI 9270-8i SAS 6Gb/s ROC RAID Card
Ethernet	361T PCIe Dual Port Gigabit NIC
Fiyat	\$8.363/adet

16 iş istasyonu için maliyet verileri tablo ve grafik olarak aşağıda belirtilmiştir. Görüldüğü üzere 16 adet fiziksel iş istasyonunun toplam maliyeti 70 bin doların üzerine çıkmaktadır. Sanal platformda ise yaklaşık 23 bin dolarlık bir maliyet ile 16 adet iş istasyonu gereksinimi karşılanabilmektedir. Tablodaki verilerden 5 adet ve altında iş istasyonu kullanılan proje ve ya işletmelerde fiziksel çözümlerin maliyet bakımından tercih edilebilir olduğu görülse de sanallaştırmanın sağladığı güvenlik, kolay yönetilebilirlik gibi avantajlarla birlikte değerlendirildiğinde sanal GPU kullanmakta özel fayda görülebilir. Çalışmamız boyunca yaptığımız testlerde 16 iş istasyonuna kadar sanal platformda daha iyi performans değerlerinin alındığı görülmektedir. Dolayısıyla düşük sayıdaki iş istasyonlarında sanallaştırma performans avantajı da sunmaktadır.

Tablo 4.6: Fiziksel ve sanal iş istasyonlarının maliyet karşılaştırması.

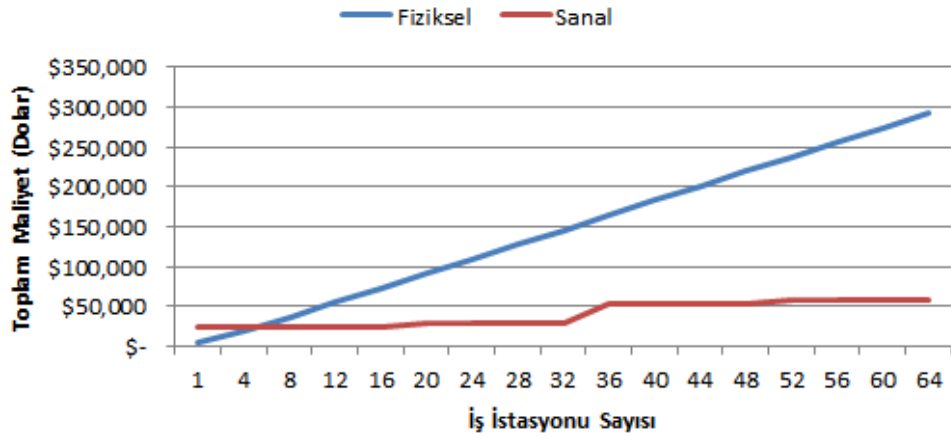
İş İstasyonu Sayısı	Fiziksel İş İstasyonu Maliyeti (4 çekirdek)	Sanal Platform Maliyeti(16 çekirdek)
1	\$4,560	\$23,810
2	\$9,120	\$23,810
4	\$18,240	\$23,810
6	\$27,360	\$23,810
8	\$36,480	\$23,810
10	\$45,600	\$23,810
12	\$54,720	\$23,810
14	\$63,840	\$23,810
16	\$72,960	\$23,810

Tablodaki verilerin grafik olarak gösterimi Şekil 4.11'deki gibidir. Grafikte 5 iş istasyonundan sonra sanal platformun maliyet avantajı sağladığı net bir şekilde görülmektedir.



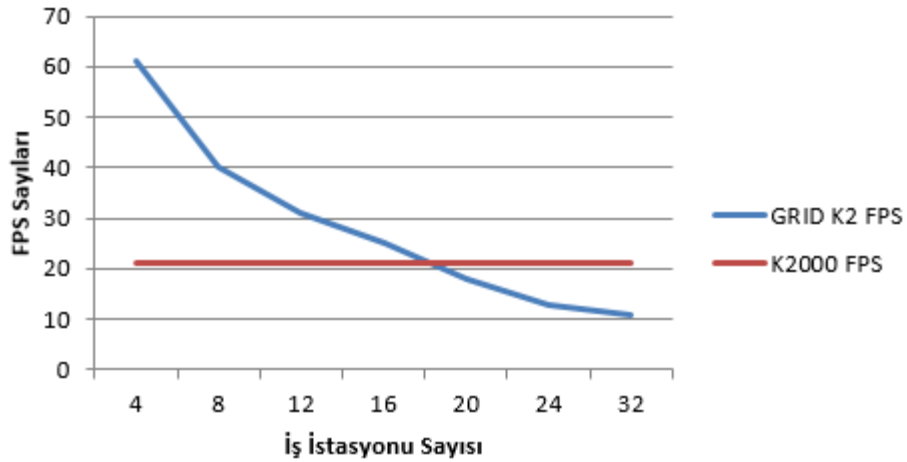
Şekil 4.11: Fiziksel ve sanal iş istasyon maliyetleri.

İş istasyonu sayısı arttıkça sanal platformun maliyet avantajı da artmaktadır. Fiziksel iş istasyonlarında maliyet doğrusal bir artış eğilimi gösterirken, sanal platformda maliyetin sayı arttıkça düştüğü grafikte görülmektedir. 16 adet sanal iş istasyonunun yetersiz kaldığı durumda ikinci bir GRID K2 kartı takılarak 32 sanal iş istasyonuna kadar kapasite artırımı yapılabilmektedir. 32 sanal iş istasyonu yetersiz kaldığında ikinci bir sanallaştırma platformunun satın alınması çoklu grid kartı desteği sağlayan daha yüksek kapasiteli sunucu alımı yapılarak ciddi bir maliyet avantajı sağlanabilecektir. Şekil 4.12’de 16 adet iş istasyonundan sonra ikinci grid grafik kartı takılması, 32 ve katlarında ise aynı tip sunucudan bir adet daha alınması durumunda elde edilen maliyet grafiği görülmektedir.



Şekil 4.12: Fiziksel ve sanal iş istasyon maliyetleri (64 İş İstasyonu).

Aşağıdaki grafikte yoğun GPU işlemi yapan sanal iş istasyonu sayısındaki artışın performansa etkisi görülmektedir. Fiziksel iş istasyonu yaptığımız testlerde ortalama 21 FPS değeri ile grafik işleme performansı sergilemiştir. Fiziksel iş istasyonlarının sayısının artması maliyeti doğrusal olarak artırmakta fakat performans değerleri doğal olarak aynı çizgide devam etmektedir. Bu değerler sanal iş istasyonlarındaki GPU kaynakları maksimum seviyede kullanılarak elde edilmiş değerlerdir. Normal kullanımlarda iş istasyonu sayısının artmasının performansa etkisi bu seviyede olmayacaktır. Sanal iş istasyonlarının bulunduğu fiziksel sunucu üzerinde iki adet NVIDIA Grid K2 kartı takılı olup, her bir kart üzerinde ikişer adet fiziksel GPU işlemci bulunmaktadır. Toplamda dört ayrı fiziksel GPU kaynağı mevcuttur. Dolayısıyla sanal iş istasyonları üzerindeki GPU profilleri farklı fiziksel GPU kaynaklarına homojen olarak dağıtılsa ilk dört sanal iş istasyonu maksimum performans ile çalışacaktır. Bu bağlamda performans değerlendirmesini 4 ve katları olacak şekilde değerlendirmek, bu şekilde ölçümlene yapmak doğru bir sonuç ortaya koyacaktır. Yukarıda da belirtilen stres testinde elde edilen değerlerin ortalamaları alınarak iş istasyonu sayısı ile GPU performansı arasındaki ilişki Şekil 4.13’de sade ve anlaşılır olarak ortaya koymuştur.



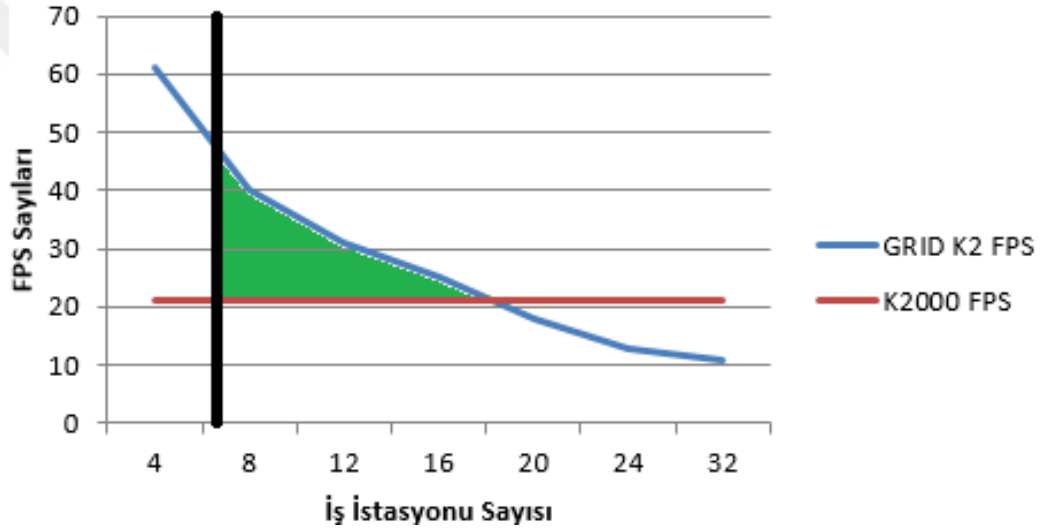
Şekil 4.13: Fiziksel ve Sanal GPU performans karşılaştırması.

Bu değerlerle maliyetleri ortak değerlendirdiğimizde ise Tablo 4.6’da görüldüğü üzere 6 adet iş istasyonu maliyeti GPU sanallaştırmanın kullanıldığı sunucu maliyetinden yüksektir. Performans değerlendirildiğinde ise ilk 4 iş istasyonundan en fazla verim alınmakta, fiziksel istasyonlarının (Z440) çok üzerinde

grafik işleme gücü elde edilmektedir. İş istasyonu sayısı 16 olana kadar fiziksel iş istasyonlarına göre daha iyi performans değerleri vermiştir.

Grafik işleme ihtiyacı bulunan ve orta seviye grafik kartlarının gereksinimi karşıladığı işletmelerde GPU sanallaştırma teknolojileri rahatlıkla tercih edilebileceği gözlemlenmiştir. 6 ile 16 arasında iş istasyonunun kullanıldığı projelerde sanal GPU kullanarak, hem performans, hem de maliyet kazanımı sağlanabilmektedir. 16 ve üzerinde iş istasyonu ihtiyacı varsa ikinci bir DL380 sanal platformu alımı ile performans ve maliyet kazanımları korunabilecektir.

Şekil 4.14'teki grafikte maliyet ve performans kazanımı sağlayan değerler yeşil alanda belirtilmiştir.



Şekil 4.14: GPU sanallaştırma maliyet/performans kazanım alanı.

Test ortamımızda kullandığımız fiziksel iş istasyonu 4 çekirdekli iken sanal platformda kullanılan sunucu 16 çekirdeğe sahiptir. Sanallaştırmanın verdiği esneklikle çekirdekler sistemdeki iş yoğunluğuna göre dağıtılabilmekte, ihtiyaç durumunda canlı olarak diğer sistemlere aktarılabilmektedir. Ayrıca grafik işlemlerde işlemci gücünden ziyade GPU performansı ön plana çıktığından daha az çekirdekli sunuculardan daha iyi grafik performansı elde edilebilmektedir. Fiziksel iş istasyonları kullanıldığında gereğinden fazla işlemci maliyeti çıkmaktadır.

4.4 Fiziksel ve Sanal İş İstasyonlarının Enerji Tüketimi Karşılaştırması

Fiziksel iş istasyonlarının enerji tüketimi sayısı ile doğru orantılı olarak artmaktadır. Her bir fiziksel iş istasyonunun kendine ait bir güç sağlayıcısı olması enerji maliyetini önemli derece de etkilemektedir. Sanal iş istasyonlarının tamamı bir adet fiziksel sunucu üzerinde çalıştığından enerji tüketim maliyeti anlamında da önemli kazanımlar sağlamaktadır. Aşağıdaki tablolarda fiziksel ve sanal iş istasyonlarının maksimum enerji tüketimi ile çalıştıklarında harcadıkları günlük ve yıllık enerji tüketim miktarları ile yıllık maliyetleri belirtilmiştir.

Tablo 4.7: Fiziksel iş istasyonları enerji tüketim maliyeti.

Fiziksel İş İstasyonu Yıllık Enerji Tüketimi (16 Adet)					
	Birim Tüketim(watt)	16 Adet	Günlük Tüketim(kwh)	Yıllık Tüketim (kwh)	Maliyet (TL)
Monitör	30	480	11.52	4204.8	
Kasa	700	11200	268.8	98112	
Toplam	730	11680	280.32	102316.8	42072.6

Tablo 4.8: Sanal iş istasyonları enerji tüketim maliyeti.

Sanal İş İstasyonu Yıllık Enerji Tüketimi (16 Adet)					
	Birim Tüketim(watt)	16 Adet	Günlük Tüketim(kwh)	Yıllık Tüketim (kwh)	Maliyet (TL)
Monitör	30	480	11.52	4204.8	
Sıfır İstemci	20	320	7.68	2803.2	
Sunucu	2800	2800	67.2	24528	
Toplam	2850	3600	86.4	31536	12967.6

Elektrik kwh ücretini 41.17 kr olarak hesapladığımızda fiziksel iş istasyonu kullanmak yıllık yaklaşık 29 bin TL civarı ekstra maliyet çıkarmaktadır.

BEŞİNCİ BÖLÜM

SONUÇLAR

Günümüzde bilgi işlem teknolojileri hızla gelişmekte ve standart bilişim hizmetlerinin maliyetleri de katlanarak artmaktadır. Teknolojinin sunduğu kolaylıklardan faydalanmak ve bunları en düşük maliyet ile elde edebilmek için sanallaştırma ve bulut bilişim araçlarının etkili kullanımı önemli bir fırsat haline gelmiştir. Görsellik ve grafik tasarımlar bu anlamda etkili kullanılır duruma gelmiştir. Bununla birlikte verilerin merkezde tutulması, güvenliği gibi kavramlar kurumlar tarafından yüksek öncelik taşımaktadır. Dolayısıyla sistem ve verilerin merkezden yönetildiği, güçlü grafik desteği veren çözümler firmalar tarafından talep edilmektedir. Bu alanda, GPU sanallaştırma yavaş yavaş kurumsal hizmetlerde yaygınlaşmaktadır.

Bu çalışma ile orta seviye grafik kartı kullanılan işletmelerde, sanal iş istasyonu kullanmanın maliyet ve performans açısından daha avantajlı olduğunu gösteren deneysel bir analiz yapılmıştır. Özellikle grafik işleyen firmalar (simülasyon üreticileri, grafik tasarım firmaları vs.) çalışmamızın odaklandığı alanlardır. Bugün ortalama bir iş istasyonunun maliyeti yaklaşık 4 bin dolar civarındadır. Yaklaşık 15 iş istasyonunun hizmet verdiği bir işletmenin sadece iş istasyonu maliyeti 60 bin doların üzerine çıkmaktadır. Fiziksel iş istasyonu kullanmanın donanım maliyetlerinin yanında daha fazla enerji tüketimi, arıza ve bakım giderlerinin daha yüksek olması, yerinde destek ihtiyacı, verilerin güvenliği ile platform ve lokasyon bağımlılığı gibi etkenlerden dolayı sanallaştırmanın sunduğu avantajları verememekte ve dolayısıyla GPU sanallaştırmanın kurumlar tarafından kullanılmasının birçok yönden maliyet-etkin bir yaklaşım olduğu deneysel sonuçlarla destekli olarak ortaya konulmuştur.

Sanallaştırma teknolojilerinin temel kazanımlarından olan en iyi kaynak tüketimi ile atıl kaynakların aktif hale getirilmesi belli koşullarda önemli performans kazanımlarını da beraberinde getirmiştir. Çalışmamızda kullandığımız NVIDIA GRID K2 kartı üzerinde iki adet fiziksel GPU ünitesi bulunmaktadır. Her bir GPU ünitesinde 1.536 CUDA grafik işlemci birimi bulunmaktadır. Fiziksel iş istasyonlarında kullanılan grafik kartları ise 384 grafik işlemci biriminden oluşmakta olup gereksinimi karşılamaktadır. Sanal iş istasyonlarına atanan GPU profillerinde de benzer kaynak atanmış olsa da, sanallaştırmanın sağladığı en iyi kaynak yönetimi ile ihtiyaç halinde fiziksel GPU da atıl durumda bulunan grafik işlemci birimleri de kullanılmakta, dolayısıyla sistemler müsait olduğu durumda bir sanal iş istasyonu 1.536 grafik işlemci gücüne sahip olabilmektedir. Dolayısıyla düşük iş istasyonu sayısı ya da az yoğunluğun olduğu süreçlerde fiziksel iş istasyonlarına göre daha iyi performans gösterebilmektedir. GPU sanallaştırmanın sunduğu maliyet, performans kazanımları ile genel olarak sanallaştırma teknolojilerinin sunduğu esneklik birlikte ele alınmalıdır. Yukarıdaki bölümlerde detaylı olarak belirtilen fiziksel altyapı kullanımında karşılaşılan zorluklar, donanım, sürücü, bakım zorlukları da değerlendirildiğinde, GPU sanallaştırma teknolojilerinin kullanımında maliyet ve performans açısından önemli kazanımlar tespit edilmiştir.

Günümüz teknolojilerinde paralel/dağıtık hesaplama teknikleri kullanımının kaçınılmaz olduğu ve donanımdaki ilerlemenin bu alanda olacağı açıkça görülmektedir. Sanallaştırmanın da bu anlamda önemli bir katkı sağlayacağı açıkça görülmektedir. Gelecekte yapacağımız çalışmalarda, sanallaştırılmış GPU destekli MPI kütüphaneleri ile dağıtık hesaplama ortamlarında NP-Zor kombinatoriyal eniyileme problemlerinin sezgisel algoritmalar ile sanallaştırma kullanılarak çözülmesi hedeflenmektedir.

KAYNAKLAR

- [1] Chao-Tung Yang, Hsien-Yi Wang, Yu-Tso Liu. Using PCI Pass-Through for GPU Virtualization with CUDA. In book "Network and Parallel Computing" pp 445-452
- [2] <http://www-03.ibm.com/systems/power/hardware/795/specs.html> 11.02.2016
- [3] <http://www.oracle.com/us/products/servers-storage/sparc-t7-4-server-ds-2687051.pdf> 11.02.2016
- [4] <http://www.ibmssystemsmag.com/aix/administrator/systemsmanagement/Configuring-Processor-Resources-for-System-p5-Shar/?page=3> 22.06.2016
- [5] Mel Cordero, Lúcio Correia, Hai Lin, Vamshikrishna Thatikonda, Rodrigo Xavier, *IBM PowerVM Virtualization Introduction and Configuration* in ibm.com/redbooks
- [6] Christopher Clark, Keir Fraser, Steven Hand, Jacob Gorm Hansen, Eric Jul, Christian Limpach, Ian Pratt, Andrew Warfield, *Live migration of virtual machines*, In Proceeding NSDI'05 Proceedings of the 2nd conference on Symposium on Networked Systems Design & Implementation - Volume 2, Pages 273-286
- [7] https://tr.wikipedia.org/wiki/IBM_360 05.11.2015
- [8] <http://www.javatpoint.com/history-of-java#> 28.12.2015
- [9] <http://www.everythingvm.com/content/history-virtualization> 11.01.2016
- [10] Susanta Nanda, Tzi-cker Chiueh {susanta,chiueh}@cs.sunysb.edu, *A Survey on Virtualization Technologies*, Department of Computer Science SUNY at Stony Brook Stony Brook, NY 11794-4400
- [11] <https://msdn.microsoft.com/en-us/vstudio/aa496123>
- [12] N.M. Mosharaf Kabir Chowdhurya, Raouf Boutaba, *A survey of network virtualization*, in Computer Networks Volume 54, Issue 5, 8 April 2010, Pages 862–876

- [13] <https://sreeninet.wordpress.com/2014/03/29/network-virtualization-overview-and-commercial-solutions/> 27.07.2016
- [14] Bill Bao, *Storage virtualization*, Publication Number US20060101204 A1
- [15] http://www.wikibon.org/wiki/v/Thin_provisioning 27.07.2016
- [16] Dowty, M., & Sugerman, J. (2009). GPU virtualization on VMware's hosted I/O architecture. *ACM SIGOPS Operating Systems Review*, 43(3), 73-82.
- [17] Duato, J., Igual, F. D., Mayo, R., Peña, A. J., Quintana-Ortí, E. S., & Silla, F. (2010, January). An efficient implementation of GPU virtualization in high performance clusters. In *Euro-Par 2009–Parallel Processing Workshops* (pp. 385-394). Springer Berlin Heidelberg.
- [18] Giunta, G., Montella, R., Agrillo, G., & Coviello, G. (2010). A GPGPU transparent virtualization component for high performance computing clouds. In *Euro-Par 2010-Parallel Processing* (pp. 379-391). Springer Berlin Heidelberg.
- [19] Shinpei Kato, Michael McThrow, Carlos Maltzahn, and Scott Brandt Department of Computer Science, UC Santa Cruz. Gdev: First-Class GPU Resource Management in the Operating System
- [20] Gupta, V., Gavrilovska, A., Schwan, K., Kharche, H., Tolia, N., Talwar, V., & Ranganathan, P. (2009, March). GVIM: GPU-accelerated virtual machines. In *Proceedings of the 3rd ACM Workshop on System-level Virtualization for High Performance Computing* (pp. 17-24). ACM.
- [21] Reano, C., Pea, A. J., Silla, F., Duato, J., Mayo, R., & Quintana-Orti, E. S. (2012, December). Cu2rcu: Towards the complete rcuda remote GPU virtualization and sharing solution. In *High Performance Computing (HiPC), 2012 19th International Conference on* (pp. 1-10). IEEE.
- [22] Reano, C.; Univ. Jaume I, Castellón, Spain ; Mayo, R.; Quintana-Orti, E.S.; Silla, F. (2013, September) Influence of InfiniBand FDR on the performance of remote GPU virtualization. In *Cluster Computing (CLUSTER), 2013 IEEE International Conference on Indianapolis, IN*
- [23] Minh Tri Do Dinh, *GPUs - Graphics Processing Units*, Vertiefungsseminar Architektur von Prozessoren, SS 2008 Institute of Computer Science, University of Innsbruck July 7, 2008

- [24] <http://www.forbes.com/sites/jasonevangelho/2015/08/19/nvidia-increases-desktop-gpu-market-share-again-despite-multiple-amd-radeon-releases/#190e3eb6250e1c0779bb250e> 19.02.2016
- [25] http://www.nvidia.com/content/grid/resources/AutoCAD_GRID_vGPU_Scalability_Solutions_Guide.pdf 01.07.2016
- [26] http://www.athena.dk/files/userdir/documents/energy_study.pdf 11.12.2015
- [27] Nevzat Melih Tünek, *Sanallaştırma ve Provisioning Teknikleri*, Türk Hava Kurumu Üniversitesi, Ankara
- [28] <https://thgtr.com/directx-11-test-programi-unigine-heaven-benchmark-2-0-yayinlandi> 22.07.2016

