Apache Hadoop Kullanarak Doküman Yönetim Sistemi Geliştirme

Atakan Taşsümer[1], Cumali Yaşar[2]

[1] Çanakkale Onsekiz Mart Üniversitesi, Fen Bilimleri Enstitüsü Bilgisayar Mühendisliği

[2] Çanakkale Onsekiz Mart Üniversitesi, Eğitim Fakültesi BÖTE Bölümü

Özet: Çok kullanıcı ortamlarda üretilen belgelerin paylaşımını, güvenliğini, saklanmasını ve taşınmasını sağlamaktır. Dağıtık dosya sistemlerinin kullanımı ve yönetilmesinin yapılması ile ilgili çalışmaları yöntemlerini belirlemek; uygun dağıtık dosya yönetim sistemini seçmek, üretilen belgelerin dağıtık dosya sistemlerine göre uygulanmasını yapmak çalışmamızın özünü teşkil etmektedir. Çalışmamızda Dağıtık dosya sistemlerinin yapılarını ve kullanım alanlarını belirleyerek kullanıcılara belirleyici bir yön vermektir.

Giriş

Teknolojinin ilerlemesi ve Bilgi toplumunun geldiği düzeyin en belirgin özelliği ürettiği verilerin saklanması ve gelecek nesillere aktarılması ile neticelenmektedir. Üretilen bilgi ve belgelere erişim zaman açısından çok önemli olmaktadır. Zamanında ulaşılamayan belge, değerini yitirmektedir. Kurumların ürettikleri veriler, Entelektüel bilgi olarak kurumların piyasa değerini artırmaktadır. Sermaye değeri ürettiği bilgi ve belgenin değeri ile artmaktadır.

Günümüzde çok bilgi ve belge üretilmektedir. Ancak üretilen bilgilerin ekseriyeti belge haline dönüşmemektedir. Bazı bilgiler çok ehemmiyetli olmasına rağmen sadece söz olarak uçup gitmektedir. Bilgi toplumu yapısının en belirgin özellikleri bizce "Belgeye dayalı üretim tekniğidir. Bu teknik yükte hafif, kıymet pahasında ağır olan tekniktir." Çok basit bir harita belgesi, bir tasarım deseni tonlarca buğdayın fiyatına denk olmaktadır. Bu tezimizi iphone tasarımını örnek verebiliriz. Bu tasarım deseninin fiyatı 1000 milyon dolar olarak lisanslamaktadırlar. Bir otomobil tasarımını düşünürsek üretilen bilginin ve pahasını karşılatılabiliriz.

Bilgi toplumu olma sürecinin en önemli özelliklerinden bir diğeri iş süreçlerinin birbirleriyle bağımlılığıdır. Bu bağımlılık proje tabanlı çalışma şeklini ortaya çıkarmıştır. İş yapma kavramı yerine proje kavramını ortaya çıkarmıştır. Bireysel çalışma kişiye bağımlı iken yeni düşünce belgeye dayalı olmaktadır. Çalışanların ürettikleri nesnel varlık artık proje olarak tanımlanmaktadır. Projenin iş süreçlerinin her aşaması belgelere dökülmektedir. Belgelerin servis edilmesi ve korunması, aktarılması bu bildirinin özünü oluşturmaktadır.

Problem Cümlesi

Problem cümlemiz bir kamu veya özel kurumlarda üretilen belgelerin saklanması, tekrar erişilebilir olunması, transfer edilmesi, güvenli bir şekilde yeni nesillere aktarılmasının nasıl olması gerektiğidir.

Materyal ve Yöntem

Biz bu çalışmada bilgi ve belgelerin "Dağıtık Dosya Sistemleri" yöntemini kullanarak nasıl yapılması gerektiğini açıklamaya çalışacağız. Bu yöntemin bilimsel temellerini ve uygulama alanlarını yine çalışmamızda açıklamaya çalışacağız. Öncelikle tanımlamamızda çok sık kullandığımız Dağıtık Dosya sistemi nedir? Bu kavramı açıklayalım. Dağıtık Dosya Sistemi. Üretilen bilgi belgelerin manyetik platformlarda saklanırken her belgenin bulunduğu yeri değiştirmeden saklanması ve yeni alanların ek-

lenmesi durumunda farklı alanlarda olmasına rağmen sanki tek bir fiziksel alanda bulunması olarak tanımlayabiliriz. Bir örnek vermek gerekirse 100 tane manyetik ortamı tek bir manyetik ortam olarak tanımlama becerisidir. Dosya Sunucusu Yönetimi, dosya sunucunuzun yönetimi için merkezi bir araç sağlar. Dosya sunucusu, dosyalarınızı depolamanız ve ağınızdaki kullanıcılarla paylaşmanız için ağ üzerinde merkezi bir konum sağlar.

Literatür Çalışması

Büyük ve kurumsal kimliği oluşmuş Avrupa birliği, NASA, CERN, Google, Microsoft, YÖK, Merkez bankası, Ticaret Odaları, IBM, HP, Hastaneler gibi kurum yada kuruluşların ürettiği terabaytlarca veriyi nasıl saklıyorlar. Sakladıkları verilere nasıl erişiyorlar, Kullanıcılarını bu karmaşık yapıya nasıl uyum sağlatıyorlar şeklinde araştırma yaptığımızda bilgi işlemcilerin bu sorunu çözmek için geliştirdikleri kavramların Dağıtık dosya sistemleri olarak tanımlamalarıdır.

Bu konuda ilk çalışmalar 1980 yılında Carnegie Mellon İBM şirketi ile yaptığı bir düşünce birliğinden ortaya çıkıyor. Ana fikir kullanıcıların bilgisayarlarının güçlerini birleştirerek tek ve süper bir bilgisayar yapmaktır. Bu düşüncesini hayata geçirmek için Andrew File System projesini geliştiriyorlar. Bu projenin temelinde kullancıların çalıştırdıkları bilgisayarların güçlerini birleştirerek disklerini ortak kullanma projedir (IEEE Computer May 1990, Vol. 23, No. 5 page 9.).

Open AFS şu an IBM tarafından ücretsiz olarak açık kaynak olarak kullanıma sunulmuştur. OpenAFS bilinen en eski ve günümüz teknolojileri içinde kurumların en çok tercih ettiği dağıtık dosya yönetim sitemidir. Bütün platformlarda çalışmaktadır. Kurulumu, kullanımı , dokümantasyonu çok açık bir şekilde yapınlanmaktadır. Dileyen kullanıcıların ağa bilgisayar ekleyerek dosya paylaşımını yapabilirler. http://openafs.org/ adresinden indirebilirler.



Hadoop : Apache tarafından "Bilgisayar Tarlası" mantığı ile kullanıcıları bir tarlada toplamak ve kullanıcıları yetkilendirmek en sonunda ortaya tek bir ürün çıkarmak mantığı ile geliştirilmiş ve üzerinde map reduce gibi bir yazılım geliştirme aracı bulunan dağıtık dosya sistemidir. Günümüzde bir standart olarak kabul edilmektedir.



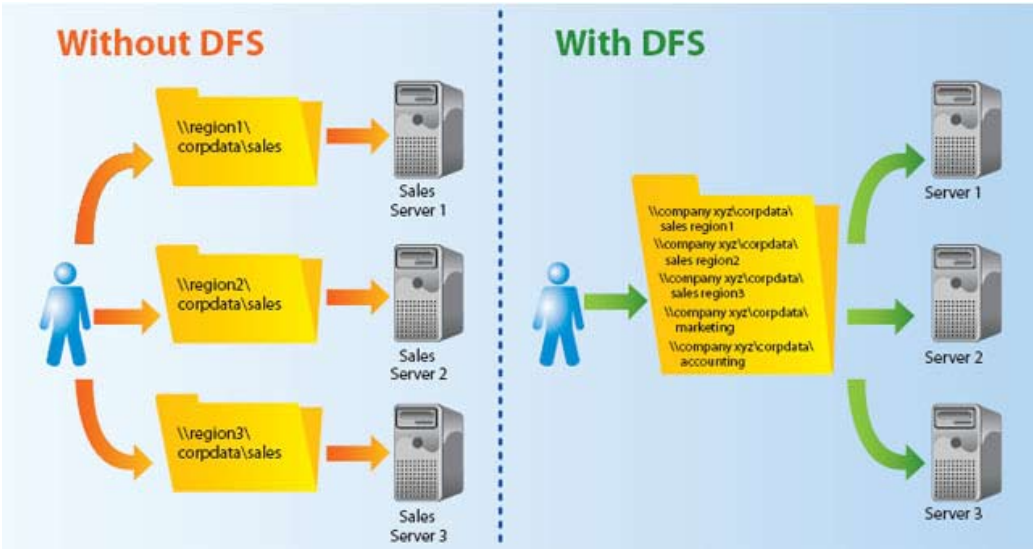
Microsoft DFS : Windows 2000 server ve Aktif Directory ile kullanıcıların hizmetine sunulmuş Dağıtık dosya sistemidir. Aktif Directorydeki her kullanıcının yetkileri kadar erişim iznine sahiptir. Kurulum ve kullanımı klasik Windows mantığı ile çalışmaktadır. Senkronizasyon ve yönetim mantığı 2008 ile bulut sistemleri mantığı ile gelişme göstermiştir. Ortak dosya ve paylaşım, yerel ve uzak bilgisayar arasından senkronizasyon gibi temel özellikleri içerisinde bulundurmaktadır.

Windows 2003 ile gelen DFS Management servisi 2 temel yapı altında incelenebilir.

DFS Namespaces: Dfs'in bu bileşeni bize farklı sunucular üzerindeki paylaştırılmış klasörlerin sanal bir yapı altında toplanmasına izin vermektedir.

DFS Replication:DFS'in bu bileşeni paylaştırılmış klasörlerin içeriklerinin eşitlenmesi için kullanılan Replication olarak düşünülebilir.

DFS'in bu kadar kullanışlı ve File Server mimarisini tek çatı altında olma özelliği barındırmasına rağmen kurulumu ve yönetimi Microsoft ürünlerinin bir özelliği olan kolay kurulum ve yönetilebilir özelliğini bünyesinde barındırmaktadır.Kurulum için Manage your Server Wizard bize yardımcı olacaktır.



Neden Dağıtık Dosya Sistemleri

- Dosya sunucuları eklemeyi veya dosya konumlarında değişiklik yapmayı düşünüyorsunuz,
- Hedeflere erişen kullanıcılar bir bölgede veya çeşitli bölgelerde dağınık durumdaysa,
- Kullanıcıların çoğu birden çok hedefe erişmek istiyorsa,
- Sunucu yükünü dengeleme, hedeflerin yeniden dağıtılmasıyla geliştirilebiliyorsa,
- Kullanıcılar, hedeflere kesintisiz olarak erişmek istiyorsa,
- Kuruluşunuzda iç veya dış kullanıma yönelik Web siteleri varsa,

Biz bu durumda yapacağımız tek şey dağıtık dosya sistemini kurmaktır.

Dosyalara kolay erişim

Dağıtılmış dosya sistemi, kullanıcıların dosyalara daha kolay erişmesini sağlar. Dosyalar fiziksel olarak çok sayıda sunucuya dağılmış olsa bile, kullanıcıların dosyalara erişmek için ağda tek bir konuma gitmeleri yeterlidir.

Ayrıca, bir hedefin fiziksel konumunu değiştirdiğinizde, kullanıcıların klasöre erişimi bundan etkilenmez. Dosyanın konumu aynı görüneceğinden, kullanıcılar klasöre önceden eriştikleri gibi erişeceklerdir.

Kullanıcıların dosyalarına erişmek için birden çok sürücüyü eşlemeleri gerekmez.

Son olarak, planlı dosya sunucusu bakımı, yazılım yükseltmeleri ve normal olarak sunucunun çevrimdışı bırakılmasını gerektiren diğer görevler, kullanıcı erişimini kesintiye uğratmadan yerine getirilebilir. Bu özellikle Web sunucuları açısından yararlıdır. Web sitesinin kökü olarak bir dağıtık dosya sistemleri kökü seçerek, HTML bağlantılarını koparmadan kaynakları dağıtılmış dosya sisteminde taşıyabilirsiniz.

Kullanılabilirlik

Etki alanı dağıtık dosya sistemleri, kullanıcıların dosyalarına erişiminin korunmasını iki yoldan sağlar:

Önce, dağıtık dosya sistemleri ad alanlarının etki alanındaki tüm sunucularda bulunan kullanıcılar tarafından her zaman görülebilmesi sağlanır.

İkincisi, yönetici olarak dağıtık dosya sistemleri köklerinizi ve hedeflerinizi çoğaltabilirsiniz. Çoğaltma, etki alanındaki birden çok sunucuda dağıtık dosya sistemleri köklerinizin ve hedeflerin örneğinin çıkarılmasıdır. Böylece kullanıcılar, dosyaların bulunduğu fiziksel sunuculardan biri devre dışı kaldığında bile dosyalarınıza erişebilir.

Sunucu yükünü dengeleme

dağıtık dosya sistemleri kökü, fiziksel olarak bütün bir ağa dağılmış çok sayıda hedefi destekleyebilir. Örneğin, kullanıcılarınızın çok sık erişeceğini bildiğiniz bir dosya varsa bu yararlı olacaktır. Kullanıcıların, fiziksel olarak, bu dosyaya tek bir sunucudan erişmesi ve bu nedenle sunucunun yükünü artırmasının tersine, dağıtık dosya sistemleri, kullanıcıların dosyaya erişiminin birden çok sunucuya dağıtılmasını sağlar. Kullanıcılar açısından ise, dosya ağdaki tek bir konumda bulunur.

Dosya ve klasör güvenliği

dağıtık dosya sistemleri'nin yönettiği paylaşılan kaynaklar, standart dosya paylaşım izinleri kullandığından, yalnızca yetkili kullanıcıların hassas verilere erişimi olmasını sağlamak için zaten varolan güvenlik gruplarını ve kullanıcı hesaplarını kullanabilirsiniz.

Sonuç

İnsanların hafızası vardır. Bireysel gelişmelerini hafızaları sayesinde ilerletirler. Gelişmeler tecrübelerine, bilgi birikimine yansır. Bu tecrübelerini gelecek nesillere aktarırlar. İnsanlık gelişim evresini devam ettirir. Bu sayede Tarım toplumu, Sanayi Devrimi, Bilgi Devrimi şeklinde insanlık evrelerini sürdürmektedirler.

Aynı süreç kurum ve şirketler içinde aynı süreç devam etmektedir. Kurumların ürettikleri belgeler; kurumların hafızalarını oluşturmaktadır. Kurumsal hafızalarını iyi yönetenler başarıyı yakalama eğilimleri yüksek olmaktadır. Kurumsal hafızalar günümüzde entelektüel birikim olarak sermayeye kayılmaktadır.

Biz bu çalışmada kurumsal hafızaya katkı vermek için, elektronik ortamda geliştirdiğimiz yazılımı tanıtmak ve bunun için gerekli olan ihtiyaçları belirlemeye çalıştık. Önerimiz; dağıtık dosya sistemlerini kurumlar için ideal bir çözüm olduğudur.

Kaynakça

[1] http://www.moreum.com/download/DYS-TerraDocs.pdf (Ocak, 2011)

[2] http://www.tarimreformu.gov.tr/files/EBYS.pdf (Ocak, 2011)

[3] http://www.devletarsivleri.gov.tr/Handlers/hhFile.ashx?Id=9c0cb331-1392-4d19-8034-1941d4d9c7d4

[4] http://www.erisimbil.com/EDYS.pdf (Ocak, 2011)