

## Dedublication Konsepti

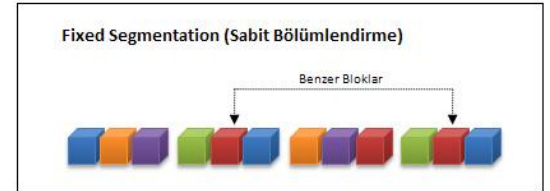
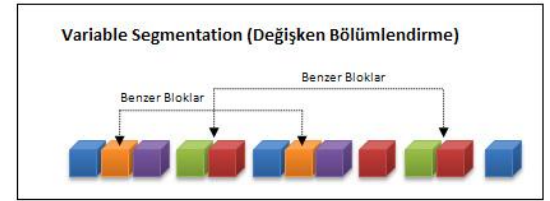
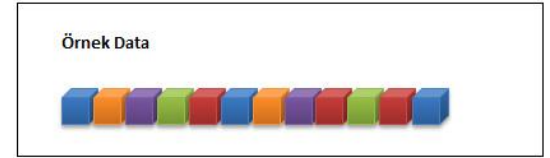
Bu yazıda, 2008 yılında daha fazla duymaya başladığımız ve önümüzdeki yıllarda da yaygınlaşacağını düşündüğümüz "De-duplication (Tekil Yedekleme)" konsepti inceleyerek, endüstrideki çözümler ve yöntemler ile ilgili olarak sizleri bilgilendirmek istemekteyiz. Basit anlamda "De-duplication", verilerin sabit ya da değişken büyüklüklerde parçalara bölünerek yedeklenmesi işlemidir. Aynı parçaların (identical segments) tek kopyası tutulmakta ve gerçek veri için linkler oluşturulmaktadır. De-duplication, Block (byte/bit) seviyesinde yapıldığı takdirde daha iyi sonuçlar alınabilmektedir. Single-Instance Storage, Unique Backup, Delta Encoding gibi tanımlamalar değişik yöntemlerde yapılan de-duplication işlemlerini ifade etmektedir.

Metodoloji olarak "**Fixed Segment**" (Sabit Bölümlendirme) ve "**Variable Segment**" (Değişken Bölümlendirme) olarak iki tip yaklaşım bulunmaktadır. Endüstrideki tüm uygulamalar çoğunlukla bu metodları kullanmaktadır.

**Fixed Segmentation:** Veri, sabit büyüklüklerdeki parçalara (segment) bölünerek tekil parçalar yedeklenmektedir. Tekrarlayan parçalar için de-duplication sistemi üzerinde bağlantılar (link) oluşturulmaktadır. Bu bağlantılar ilişkisel Veri Tabanı (Relational Database) üzerinde tutulduğu gibi dosya sistemi (file system) üzerinde de tutulabilmektedir. Fixed segmentation yönteminde de-duplication oranı "variable segmentation" yöntemine göre daha düşük olabilmektedir.

**Variable Segmentation:** Veri değişken büyüklüklerdeki parçalara (segment) bölünerek tekil parçalar yedeklenmektedir. Tekrarlayan parçalar için de-duplication sistemi üzerinde bağlantılar (link) oluşturulmaktadır. Bu bağlantılar ilişkisel Veri Tabanı (Relational Database) üzerinde tutulduğu gibi dosya sistemi (file system) üzerinde de tutulabilmektedir. Dosya sistemi üzerinde zamanla parçalanmalardan (fragmentation) dolayı performans problemleri yaşanabilmektedir.

Yandaki şemada iki metodolojiye ait de-duplication yaklaşımları görülebilmektedir.



**De-duplication işlemi mantıksal olarak aşağıdaki adımları içermektedir. Çözüm veya uygulama tipine göre bu adımlar değişiklik gösterebilmektedir.**

1- Yedekleme işlemlerinde olduğu gibi değişen dosyalar tespit edilir. Windows sistemlerde iyileştirme zamanı (modification time) ve dosya büyüklüğü (file size), Unix sistemlerde ise değişiklik zamanı (change time) ve dosya büyüklüğü (file size) kullanılabilir.

2- Üzerinde değişiklik yapılmış dosyalar segment'lere bölünerek daha önce yedek alınmış dosya bloklarıyla karşılaştırma yapılır.

3- Bölümlendirme öncesinde data kararlılığı bakımından dosyanın parmak izi çıkartılarak tekil olduğu garanti edilir. (Hash algoritmaları v.s.)

4- Eğer saklanan alanda ilgili blok yok ise (tekrar etmiyorsa) blok, saklanacağı alana transfer edilir ve ilgili blokla dosyanın bütünü arasında mantıksal ilişkilendirme (link) kurulur. (Daha önce bahsettiğim gibi linkler ilişkisel Veri Tabanı 'nda tutulabileceği gibi dosya sistemi üzerinde de tutulabilir.)

De-duplication sistemleri entegre donanımsal çözümler (hardware appliance) olabileceği gibi yazılımsal (software appliance) çözümler de olabilmektedir. Bazı de-duplication çözümlerinde yedekleme yazılımına ihtiyaç duyabileceği gibi replikasyon, şifreleme, sıkıştırma ve teyp yedekleme özelliklerini de birlikte sunulabilmektedir.

#### **Donanımsal çözümlerin avantajları;**

- Kolayca mevcut ortama entegre olabilmektedir,
- Kurulum gerektirmemektedir,

#### **Donanımsal çözümlerin dezavantajları;**

- Donanım seçme özgürlüğünü kısıtlamaktadır (sunucu, disk kaynağı, I/O birimleri v.s.)
- Firmaya bağlı çözümler kullanılması zorunludur,
- Genişleme seçenekleri sınırlıdır.

#### **Yazılımsal çözümlerin avantajları;**

- Ortam ihtiyaçlarına göre iyileştirme (modification) daha detaylı yapılabilmektedir,
- Donanım seçme özgürlüğü bulunmaktadır,
- Genişleyebilirliği yüksektir,

#### **Yazılımsal çözümlerin dezavantajları;**

- Kurulum bilgisi gerektirir,

#### **Genelde 3 noktada de-duplication yapılmaktadır.**

1. Kaynakta (Source) yapılan de-duplication: Sunucu de-duplication işlemini yapar. İşlemci bellek yükü getirir. Uzak lokasyonlar için en efektif çözümdür.
2. Sunucu üzerinde yapılan de-duplication : Atanmış bir sunucu üzerinde de-duplication yapılır. Client ın hızlı bir şekilde yedek alması sağlanır. Kaynak üzerinde CPU ve bellek yükü daha azdır.
3. Hedefte (Target) yapılan de-duplication : Hızlı yedek alınır, . Hedef ortamının kaynakları dar boğaz oluşturur.

#### **De-duplication uygulaması ya da çözümüne karar verirken aşağıdaki maddelerin sorgulanması faydalı olacaktır,**

- 1- De-duplication hangi yöntemle yapılmaktadır?
- 2- Segmentasyon (segmentation) işleminde data kararlılığını sağlamak için ne tür çözüm ya da yöntemler kullanılıyor?
- 3- De-duplication sistemi donanımsal ve mantıksal hatalara karşı nasıl korunuyor?
- 4- Dosya sistemi parçalanması (fragmentation) için ne tür bir önlem alınıyor?
- 5- Donanımsal ve yazılımsal çözümler aynı özellikleri sunuyor mu?

- 6- Sanal sunucu ortamları için çözümler var mıdır?
- 7- De-duplication işlemi sırasında encryption ve compression yapılabiliyor mu?
- 8- De-duplication ortamı uzak birimleri replikasyonu mümkün müdür?
- 9- Lisanslanma nasıl yapılmaktadır?
- 10- De-duplicate edilmiş data tape e çıkartılabiliyor mu?
- 11- Yedekleme sistemleri ile entegrasyonu var mı?
- 12- De-duplication ortamlarında recovery nasıl yapılıyor?
- 13- Kullanıcı yetkilendirmeleri ne seviyede nasıl yapılıyor?
- 14- De-duplication ortamının limitasyonları nelerdir? (dosya sistemi, dosya adedi v.s.)
- 15- Client tarafındaki CPU ve memory kullanımı değeri nedir? (Şifreleme ve sıkıştırma değerleri ayrıca tesbit edilmelidir.)
- 16- Raporlama nasıl yapılmaktadır?
- 17- İşlemci, bellek, network kullanımı kısıtlaması yapılabiliyor mu?

#### **Gantek Teknoloji Hakkında:**

Bilgi ve İletişim Teknolojileri'nde (ICT) 22 yıllık deneyime sahip olan ve kurumsal yönetim prensipleri ile faaliyet gösteren Gantek Teknoloji, Türkiye'nin ve bölgenin önde gelen Sistem Entegratörü olarak hizmet vermektedir. Gantek, 9 yıldır yaptığı kesintisiz Ar-Ge yatırımıyla bölgenin en yenilikçi ICT firmalarından birisi olarak konumlanmaktadır. Firma, başta Numara Taşınabilirliği, Gelir Maksimizasyonu ve Katma Değerli Hizmetler (VAS) ile telekom sektörüne, risk yönetimi çözümleri ile finans sektörüne ve çeşitli grafik uygulamaları ile savunma sektörüne odaklı çözümler sunmaktadır. Gantek, Türkiye'deki tüm mobil ve sabit telekom operatörleri, önde gelen finans kurumları ve CIS bölgesindeki telekom operatörleri için uzun süreli iş birliğine dayalı etkili çözümler üretmektedir. Şirket, 56 milyon doları aşkın iş hacmi, alanında uzman çalışanları ve stratejik işbirlikleri ile Deloitte Fast 50 araştırmasında, araştırmanın başladığı yıldan bugüne kadar, Türkiye'nin en hızlı büyüyen teknoloji şirketleri arasında yer almaktadır. Daha fazla bilgi için lütfen [www.gantek.com](http://www.gantek.com) adresini ziyaret edin.