**Case**

Bir haber ajansı, büyük boyuttaki haber metinlerini HDFS üzerinde saklamakta ve bu verileri MapReduce ile analiz etmektedir. Yeni işe başlayan bir veri analisti olarak sana, HDFS üzerinde dosya yönetimi görevleri veriliyor.

**1. Aşama: HDFS Dizini Oluşturma**

**Senaryo:** Yeni bir proje başlatıyorsun. Projeye ait veriler için bir klasör oluşturman gerekiyor.

Görev:

* /projects/news dizinini oluştur
* /projects/news/raw ve /projects/news/processed alt dizinlerini oluştur

hdfs dfs -mkdir -p /projects/news/raw

hdfs dfs -mkdir /projects/news/processed

**2. Aşama: Yerel Dosya Sistemi → HDFS'e Veri Yükleme**

**Senaryo:** Elinde birkaç haber metnini içeren .txt dosyaları var. Bunları raw dizinine yüklemelisin.

Görev:

* news1.txt, news2.txt, news3.txt adlı 3 dosyayı yükle

hdfs dfs -put ~/news1.txt /projects/news/raw

hdfs dfs -put ~/news2.txt /projects/news/raw

hdfs dfs -put ~/news3.txt /projects/news/raw

**3. Aşama: Listeleme, Okuma ve Bilgi Alma**

Görev:

* /projects/news/raw dizinini listele
* news2.txt dosyasının içeriğini görüntüle
* Dosyanın replikasyon değerini kontrol et
* Dosya izinlerini ve sahibi bilgisini kontrol et

hdfs dfs -ls /projects/news/raw

hdfs dfs -cat /projects/news/raw/news2.txt

hdfs dfs -stat %r /projects/news/raw/news2.txt

hdfs dfs -stat "%a %u %g" /projects/news/raw/news2.txt

**4. Aşama: Dosya Taşıma ve Yeniden Adlandırma**

Görev:

* news3.txt dosyasını processed klasörüne taşı
* Bu dosyanın adını final\_news.txt olarak değiştir

hdfs dfs -mv /projects/news/raw/news3.txt /projects/news/processed/final\_news.txt

**5. Aşama: İzin ve Sahiplik İşlemleri**

Görev:

* processed klasöründeki tüm dosyaların erişim izinlerini 755 yap
* Dosya sahibini sampleuser1 olarak değiştir

sudo useradd sampleuser1

sudo passwd sampleuser1

sudo usermod -aG hadoop sampleuser1

hdfs dfs -chmod 755 /projects/news/processed/final\_news.txt

hdfs dfs -chown sampleuser1:hadoop /projects/news/processed/final\_news.txt

**6. Aşama: Replikasyon Ayarlama**

Görev:

* raw/news2.txt dosyasının replikasyonunu 2 yap
* Kontrol et

hdfs dfs -setrep -w 2 /projects/news/raw/news2.txt

hdfs dfs -stat %r /projects/news/raw/news2.txt

**7. Aşama: Disk Kullanım ve Sayım**

Görev:

* raw dizininin toplam boyutunu ve dosya sayısını bul
* Genel HDFS kapasitesini görüntüle

hdfs dfs -du -h /projects/news/raw

hdfs dfs -count /projects/news/raw

hdfs dfs -df -h

**8. Aşama: Silme, Çöp Kutusu ve Geri Yükleme**

Görev:

* news1.txt dosyasını sil
* Trash dizinine taşındığını kontrol et
* Dosyayı geri getir

hdfs dfs -rm /projects/news/raw/news1.txt

hdfs dfs -ls /user/sampleuser1/.Trash/Current/projects/news/raw

hdfs dfs -mv /user/sampleuser1/.Trash/Current/projects/news/raw/news1.txt /projects/news/raw/

**9. Aşama: Snapshot Alma ve Geri Dönme**

Görev:

* projects/news klasörü için snapshot alın
* Snapshot içeriğini görüntüle
* Eski bir dosyayı geri kopyala

hdfs dfsadmin -allowSnapshot /projects/news

hdfs dfs -createSnapshot /projects/news snapshot-2025-07-16

hdfs dfs -ls /projects/news/.snapshot/snapshot-2025-07-16/raw

hdfs dfs -cp /projects/news/.snapshot/snapshot-2025-07-16/raw/news2.txt /projects/news/raw/

**10. Aşama: Balance ve Fsck Kontrolleri**

Görev:

* HDFS balancer'ı çalıştır
* Bir dosya için fsck durumu kontrol et

hdfs balancer -threshold 10

hdfs fsck /projects/news/raw/news2.txt -files -blocks -locations