

İŞ ZEKASI PROJESİ

коми: Federal Gayrimenkul Profil Yönetim Sistemi için Veri ambarı oluşturma

HAZIRLAYANLAR:

SINAN YAZICI- 20574050 FERDI KENDIR- 21574005

Proje ve amacı

• Federal Gayrimenkul Profil Yönetim Sistemi (FRPP MS)

FRPP MS (önceden FRPP olarak bilinirdi) ilk olarak 13327 sayılı "Federal Gayrimenkul Varlık Yönetimi" Yürütme Emri uyarınca, Federal Hükümet'in "tüm yürütme organı kurumlarının gözetimi ve kontrolü altındaki tüm gayrimenkullerin veri tabanı olması için oluşturulmuştur. aksi takdirde ulusal güvenlik nedenleriyle gereklidir Acenteler, FRPP veri sözlüğü olarak da bilinen Federal Gayrimenkul Konseyi Gayrimenkul

Envanteri Raporlama Rehberi'nde belirtilen kılavuza göre, gayrimenkul varlıklarının niteliği, kapsamı ve kullanımı hakkında açıklayıcı bilgileri yıllık olarak FRPP MS'ye sunar.

Amaç

FRPP MS'ye sunulan gayrimenkul bilgilerini (adres, şehir, eyalet vb.) daha anlamlı hale getirerek karmaşıklığını azaltmaktır.

Veri ambarı nedir?

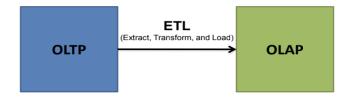
- Veri ambarı, bir işletmenin ya da kuruluşun değişik birimleri tarafından toplanan bilgilerden değerli olanlarının, gelecekte analiz işlemlerinde kullanılması amacıyla işletimsel sistem veritabanından farklı bir ortamda birleştirilmesinden oluşan büyük çaplı bir veri deposudur.
- Bir veri ambarı ilgili veriyi kolay, hızlı ve doğru biçimde analiz etmek için gerekli işlemleri yerine getirir.
- Veri ambarı kullanıldığında, günlük işletimsel görevlerle yeterince meşgul olan veritabanı kullanılmadan, analiz işlemleri farklı bir ortamda yapılır.

Oltp ve olap kavramları

- Veri tabanı işlemlerinin çoğu On-Line Transaction Processing (OTLP) işlemleridir.
- Az sayıda satırı ilgilendiren kısa, basit, ve sıkla uygulanan sorgular ve/veya güncellemeler.
- örnek: internet sitesinde alış verişler, uçak bileti satışı, bankaya para yatırma işlemi.
- On-Line Analytical Processing (OLAP) daha kompleks ve geniş kapsamlı analitik sorguları içerir.
- Daha az sayıda ama daha kompleks ve saatler boyu sürebilirler
- Genellikle bu sorgular anlık güncel veri gerektirmez.

Oltp ve olap kavramları

- OLTP kullanıcılar ile etkileşen uygulama veritabanları
 - o Tüm işlemlerin aktivitelerini kaydeder
 - o Periodik ETL (örneğin, her gece)
- Extract-Transform-Load (ETL)
 - o Veri çekimi: Kaynaklardan veri çekme
 - O Dönüştürme: veri temizliği, bütünlük kontrolü, özetler oluşturma(aggregate), vs.
 - o OLAP veritabanına yükleme
- OLAP veri ambarı amaçlı kullanılan veritabanı
 - o İş zekası: raporlama, ad hoc sorgular, veri madenciliği, vs.
 - o Geri dönüşüm ile OLTP hizmetlerinin gelişimi



Bıg data'dan sql'e geçiş

- o MSSQL connection nesnesi ile veribanı bağlantısı oluşturulur.
- o Big data dosyamız okunur ve ilgili class'a ilgili sütunlar gönderilir.

```
You.now|1 author(You)

class ReportingAgencyModel:

def __init__(self, agency_code, agency_name):
    self.agency_code = agency_code
    self.agency_name = agency_name

You.now|1 author(You)

class ReportingAgency:

def createTable(self):
    cursor = self.connection.cursor()
    cursor.execute(self.create_table_sql)
    self.connection.commit()

create_table_sql = '''
    CREATE TABLE ReportingAgencies ( agency_id int primary key IDENTITY NOT NULL, agency_code nvarchar(50),
    agency_name nvarchar(50) ) You, now * Uncommitted changes
    '''

insert_agency_query = "insert into ReportingAgencies (agency_name, agency_code) values (?, ?)"
    get_agency_by_id = "select agency_id from ReportingAgencies where agency_code=? and agency_name=?"

def __init__(self, agency_code_column, agency_name_column, connection):
    self.agency_code_column = agency_name_column
    self.agency_name_column = agency_name_column
    self.agency_name_column = agency_name_column
    self.connection = connection
    # self.createTable()
```

Veritabanı tablo modelimizi karşılayacak class tanımlanır.

- Ardından veri akışını sağlayacak class tanımlanır. Bu class;
 - Tablo oluşturmak
 - Kayıt eklemek
 - Kayıt sorgulamak
 - o için metodlar içerir.

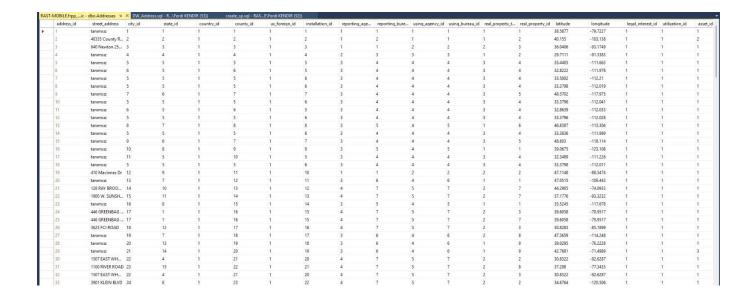
```
def getReportingAgencyId(self, agency_code, agency_name):
   cursor = self.connection.cursor()
   cursor.execute(self.get_agency_by_id, (str(agency_code), agency_name))
   agency_id = cursor.fetchone()
    if agency_id is None:
       return 0
   return agency_id[0]
def saveDatabase(self, reporting_agency_item):
   agency_val = (reporting_agency_item.agency_name,
                 int(reporting_agency_item.agency_code))
   cursor = self.connection.cursor()
   cursor.execute(self.insert_agency_query, agency_val)
   self.connection.commit()
def separateData(self):
   reporting_agency_list = []
   reporting_agency_names = self.agency_name_column.unique()
   reporting_agency_codes = self.agency_code_column.unique()
    for i in range(len(reporting_agency_names)):
       reporting_agency_list.append(ReportingAgencyModel(
            reporting_agency_codes[i], reporting_agency_names[i]))
    for reporting_agency_item in reporting_agency_list:
        self.saveDatabase(reporting_agency_item)
```

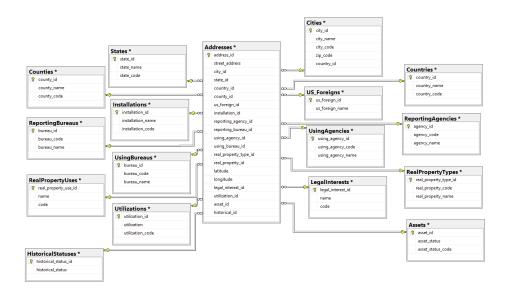
- Veri sorgulamak için yazılmış getReportingAgencyId() metodu tanımlanmıştır.
- O Big data'dan alınan satırın daha önce eklenip eklenmediğini kontrol etmek için separateData() metodu tanımlanmıştır.
- o separateData() metodunda ayrıştırılan veri tek tek tabloya eklenmelidir. Bunun için saveDatabase() metodu yazılmıştır.

Yıldız şemasının oluşumu

- İlgili big data ayrıştırıldığında OLTP veritabınında 17 adet tablo oluşmaktadır. Bu 17 adet tablolardan bir tanesi Fact (gerçek) tablo diğer 16 tanesi Dimension (boyut) tablolarıdır.
- Yıldız şemasının dışında OLAP veri veritabanına geçiş için bir başka tablo bulunmaktadır. Bu tablo 16 adet boyut tablosunun join'ler ile birleştirilmesiyle oluşturulmuştur.
- Fact tablosundaki foreign keyler anlamlandırılmıştır ve tek satırda bütün bilgileri içermektedir.

Fact ve Dimension Tabloları





```
⊡create TABLE DW_Addresses (
             street_address nvarchar(max),
             agency name nvarchar(50),
             agency_code nvarchar(50),
             bureau name nvarchar(50),
             bureau_code nvarchar(50),
             using_agency_name nvarchar(50),
             using_agency_code nvarchar(50),
             using_bureau_name nvarchar(50),
             using_bureau_code nvarchar(50),
             city_name nvarchar(50),
             city_code nvarchar(50),
             zip code nvarchar(50),
             country_name nvarchar(50),
             country code nvarchar(50),
             county_name nvarchar(50),
             county_code nvarchar(50),
             state_name nvarchar(50),
             state_code nvarchar(50),
             installation_name nvarchar(250),
             installation_code nvarchar(250),
             us_foreign_name nvarchar(50),
             real property name nvarchar(250),
             real_property_code nvarchar(250),
             real_property_use_name nvarchar(250),
             real_property_use_code nvarchar(250),
             legal_interest_name nvarchar(250),
             legal_interest_code nvarchar(250),
             utilization nvarchar(250),
             utilization_code nvarchar(250),
             asset_status nvarchar(250),
             asset_status_code nvarchar(250),
             historical status nvarchar(250),
             latitude nvarchar(50),
             longitude nvarchar(50)
```

```
USE [frpp_public]

ON

PART ASSI_NALLS ON

OSE QUOTED_IDENTIFIER ON

BAILER PROCEDURE [dbo].[getAddresses] Script Date: 28.05.2022 12:52:48 *****/

SET ASSI_NALLS ON

OSE QUOTED_IDENTIFIER ON

BAILER PROCEDURE [dbo].[getAddresses]

AS

DESIRCT TOP[1000000]. Addresses.s.street_address. ReportingAgencies.agency_name, ReportingAgencies.agency_code, ReportingBureaus.bureau_code as using_bureau_code, UsingAgencies.using_agency_name, UsingAgencies.using_agency_name, UsingAgencies.using_agency_name, UsingAgencies.using_agency_name, UsingAgencies.using_agency_name, Cuntries.country_code, Country_code, Cuttes.cty_name, Cuttes.cty_Toped._Countries.country_name, Countries.country_code, Cuttes.cty_Toped._Countries.country_code, Cutter.cty_Toped._Countries.country_code, Cutter.cty_Toped._Countries.country_code, Cutter.cty_Toped._Countries.country_code, Cutter.cty_Toped._Countries.country_code, Cutter.cty_Toped._Countries.country_code, Cutter.cty_Toped._Countries.country_toped._Countries.country_toped._Countries.country_toped._Countries.country_toped._Countries.country_toped._Countries.country_toped._Countries.country_toped._Countries.country_toped._Countries.country_toped._Countries.country_toped._Countries.country_toped._Countries.country_toped._Countries.country_toped._Countries.country_toped._Countries.country_toped._Countries.country_toped._Countries.country_toped._Countries.country_toped._Countries.country_toped._Countries.country_toped._Countries.country_toped._Countries.country_toped._Countries.country_toped._Countries.country_toped._Countries.country_toped._Countries.country_toped._Countries.country_tope
```

Veri çekmek için yazılan storage prodecure

Column Name	Data Type	Allow Nul
street_address	nvarchar(MAX)	
agency_name	nvarchar(50)	
agency_code	nvarchar(50)	
bureau_name	nvarchar(50)	
bureau_code	nvarchar(50)	
using_agency_name	nvarchar(50)	\checkmark
using_agency_code	nvarchar(50)	$\overline{\checkmark}$
using_bureau_name	nvarchar(50)	
using_bureau_code	nvarchar(50)	\checkmark
city_name	nvarchar(50)	\checkmark
city_code	nvarchar(50)	\checkmark
zip_code	nvarchar(50)	\checkmark
country_name	nvarchar(50)	\checkmark
country_code	nvarchar(50)	$\overline{\checkmark}$
county_name	nvarchar(50)	\checkmark
county_code	nvarchar(50)	\checkmark
state_name	nvarchar(50)	\checkmark
state_code	nvarchar(50)	\checkmark
installation_name	nvarchar(250)	\checkmark
installation_code	nvarchar(250)	\checkmark
us_foreign_name	nvarchar(50)	$\overline{\checkmark}$
real_property_name	nvarchar(250)	\checkmark
real_property_code	nvarchar(250)	
real_property_use_name	nvarchar(250)	
real_property_use_code	nvarchar(250)	\checkmark
legal_interest_name	nvarchar(250)	
legal_interest_code	nvarchar(250)	$\overline{\mathbf{v}}$
utilization	nvarchar(250)	
utilization_code	nvarchar(250)	\checkmark
asset_status	nvarchar(250)	
asset_status_code	nvarchar(250)	
historical_status	nvarchar(250)	
latitude	nvarchar(50)	\smile
longitude	nvarchar(50)	

- OLAP veritabanı için kullanılan tablonun özellikleri yandaki gibidir.
- Bu tablonun oluşması için 17 adet tablo join'lenir. Bu join storage procedure ile gerçekleştirilir.

```
ALTER PROCEDURE [dbo].[getAddresses]

AS

Select TOP(100000) Addresses.street_address, ReportingAgencies.agency_name, ReportingAgencies.agency_code, ReportingBureaus.bureau_name, ReportingBureaus.bureau_code, UsingBureaus.agency_name, UsingBureaus.bureau_name, UsingBureaus.bureau_code as using_bureau_name, UsingBureaus.bureau_code as using_bureau_code, Cities.city_name, Cities.city_name, Cities.city_name, Cities.city_name, Cities.city_name, Cities.city_name, Cities.city_name, Countres_country_name, Countres_countres_countres_countres_countres_countres_countres_countres_countres_countres_countres_countres_countres_countres_countres_countres_countres_countres_countres_countres_countres_countres_countres_countres_countres_countres_countres_countres_countres_countres_countres_countres_countres_countres_countres_countres_countres_countres_countres_countres_countres_countres_countres_countres_countres_countres_countres_countres_countres_countres_countres_countres_countres_countres_countres_countres_countres_countres_countres_countres_countres_countres_countres_countres_countres_countres_countres_countres_countres_countres_countres_countres_countres_countres_countres_countres_countres_countres_countres_countres_countres_countres_countres_countres_countres_countres_countres_countres_countres_countres_countres_countres_countres_countres_countres_countres_countres_countres_countres_countres_countres_countres_countres_countres_countres_countres_countres_countres_countres_countres_countres_countres_countres_countres_countres_countres_countres_countres_countres_countres_countres_countres_countres_countres_countres_countres_countres_countres_countres_countres_countres_countres_countres_countres_countres_countres_countres_countres_countres_countres_countres_countres_countres_countres_countres_countres_countres_countres_cou
```

Dashboard

street_address 🔻	agency_name	agency_code v	bureau_name
tanımsız	AGRICULTURE	12	FOREST SERVICE
40335 County Road GG	AGRICULTURE	12	AGRICULTURAL RESEARCH SERVICE
640 Newton 2500 Rd	AGRICULTURE	12	FOREST SERVICE
tanımsız	TRANSPORTATION	69	FEDERAL AVIATION ADMINISTRATION
tanımsız	INTERIOR	14	BUREAU OF RECLAMATION
tanımsız	INTERIOR	14	BUREAU OF RECLAMATION
tanımsız	INTERIOR	14	BUREAU OF RECLAMATION
tanımsız	INTERIOR	14	BUREAU OF RECLAMATION
tanımsız	INTERIOR	14	BUREAU OF RECLAMATION
tanımsız	INTERIOR	14	BUREAU OF RECLAMATION
tanımsız	INTERIOR	14	BUREAU OF RECLAMATION
tanımsız	INTERIOR	14	BUREAU OF RECLAMATION
tanımsız	INTERIOR	14	BUREAU OF LAND MANAGEMENT
tanımsız	INTERIOR	14	BUREAU OF RECLAMATION
tanımsız	INTERIOR	14	BUREAU OF RECLAMATION
tanımsız	INTERIOR	14	BUREAU OF LAND MANAGEMENT
tanımsız	INTERIOR	14	BUREAU OF RECLAMATION
tanımsız	INTERIOR	14	BUREAU OF RECLAMATION
410 MacInnes Dr	AGRICULTURE	12	FOREST SERVICE
tanımsız	INTERIOR	14	FISH AND WILDLIFE SERVICE
128 RAY BROOK ROAD	JUSTICE	15	BUREAU OF PRISONS
1900 W. SUNSHINE ST	JUSTICE	15	BUREAU OF PRISONS
tanımsız	INTERIOR	14	BUREAU OF LAND MANAGEMENT
446 GREENBAG ROAD ROUTE 857	JUSTICE	15	BUREAU OF PRISONS

• Join sonucu excel formatındaki OLAP veritabanı

street_address	▼ agency_name	▼ agency_cod	de 🔻 bureau_name	bureau_c	ode 💌 using_agency_name
tanımsız	AGRICULTURE	12	FOREST SERVICE	23	0
40335 County Road GG	AGRICULTURE	12	AGRICULTURAL RESEARCH SERVICE	5	0
640 Newton 2500 Rd	AGRICULTURE	12	FOREST SERVICE	23	AGRICULTURE
tanımsız	TRANSPORTATION	69	FEDERAL AVIATION ADMINISTRATION	5	TRANSPORTATION
tanımsız	INTERIOR	14	BUREAU OF RECLAMATION	23	INTERIOR
tanımsız	INTERIOR	14	BUREAU OF RECLAMATION	23	INTERIOR
tanımsız	INTERIOR	14	BUREAU OF RECLAMATION	23	INTERIOR
tanımsız	INTERIOR	14	BUREAU OF RECLAMATION	23	INTERIOR
tanımsız	INTERIOR	14	BUREAU OF RECLAMATION	23	INTERIOR
tanımsız	INTERIOR	14	BUREAU OF RECLAMATION	23	INTERIOR
tanımsız	INTERIOR	14	BUREAU OF RECLAMATION	23	INTERIOR
tanımsız	INTERIOR	14	BUREAU OF RECLAMATION	23	INTERIOR
tanımsız	INTERIOR	14	BUREAU OF LAND MANAGEMENT	11	INTERIOR
tanımsız	INTERIOR	14	BUREAU OF RECLAMATION	23	INTERIOR
tanımsız	INTERIOR	14	BUREAU OF RECLAMATION	23	INTERIOR
tanımsız	INTERIOR	14	BUREAU OF LAND MANAGEMENT	11	INTERIOR
tanımsız	INTERIOR	14	BUREAU OF RECLAMATION	23	INTERIOR
tanımsız	INTERIOR	14	BUREAU OF RECLAMATION	23	INTERIOR
410 MacInnes Dr	AGRICULTURE	12	FOREST SERVICE	23	AGRICULTURE
tanımsız	INTERIOR	14	FISH AND WILDLIFE SERVICE	36	INTERIOR
128 RAY BROOK ROAD	JUSTICE	15	BUREAU OF PRISONS	19	JUSTICE
1900 W. SUNSHINE ST	JUSTICE	15	BUREAU OF PRISONS	19	JUSTICE
tanımsız	INTERIOR	14	BUREAU OF LAND MANAGEMENT	11	INTERIOR
446 GREENBAG ROAD ROUTE 857	JUSTICE	15	BUREAU OF PRISONS	19	JUSTICE
446 GREENBAG ROAD ROUTE 857	JUSTICE	15	BUREAU OF PRISONS	19	JUSTICE
3625 FCI ROAD	JUSTICE	15	BUREAU OF PRISONS	19	JUSTICE
tanımsız	INTERIOR	14	FISH AND WILDLIFF SERVICE	36	INTERIOR

OLAP veritabanı için oluşturulan DW_Addresses tablosunun dasboard'daki görüntüsü

Kaynakça

- https://www.gsa.gov/policy-regulations/policy/real-property-policy-division-overview
- https://docs.microsoft.com/en-us/sql/connect/homepage-sql-connection-programming?view=sql-server-ver15