Hibrit Veri Modeli

A.Erkan ÇELİK Formalis Bilgi Teknolojileri CRM Ürün Geliştirme Yöneticisi





Formalis Hakkında





Formalis Hakkında

Kuruluş Yılı: 2006

Aktif Müşteri Sayısı: 200+ Aktif SaaS Kullanıcı: 40.000+

Aktif On-Premise Kullanıcı: 30.000+

Formalis Personel Sayısı: 40

Deloitte Son 5 Yıllık Büyüme Oranı: %382

2018 Bilişim 500: CRM/BPM Kategorisinde 2. en büyük

yazılım üreticisi

Deloitte'ten 5. ödül

2013, 2015, 2017 yıllarında Türkiye'de *en hızlı büyüyen ilk 50 teknoloji şirketi*.

2015, 2017 yıllarında EMEA bölgesinde (Avrupa, Afrika, Ortadoğu) **en hızlı büyüyen ilk 500 teknoloji şirketi**.

Business Development and Sales

12655 W. Jefferson Blvd. 4th Floor Los Angeles CA 90066, USA P. +1 310 666 9425

R&D and Sales

Sahrayıcedid Mah. Halk Sk. Pakpen Plaza No.40/4 Kadıköy İstanbul, Turkey T. +90 216 361 5438

50 Technology Fast 50
2013
2015
2017

500 Technology Fast 500 2015 2017



Deloitte.

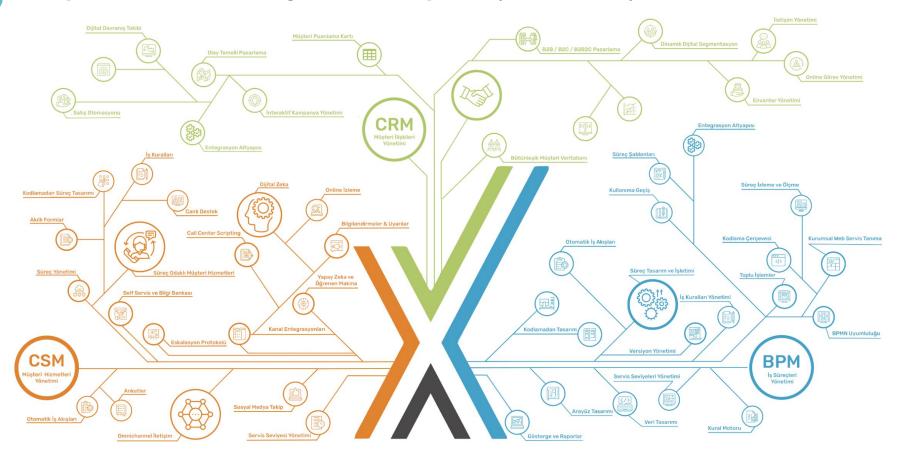








Müşteri Merkezli Dijital Dönüşüm (Next4biz)







































































































































































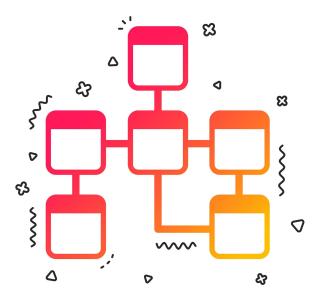






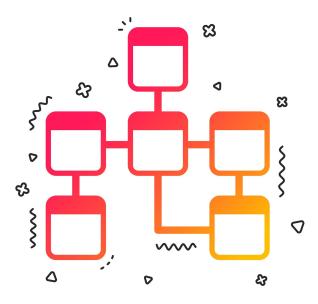






lişkisel veri tabanı, birbirinden farkı tablolara yerleştirilmiş olan verilerin birbirleri ile belirli alanlara göre ilişkilendirilerek düzenlenen veri tabanlarıdır. İlişkisel veri tabanları birden fazla tablodan oluşabilir.





- Veri tekrarından kaçınır
- Veriler yazılırken parçalanır, okunurken birleştirilir
- Daha az fiziksel depolama kullanmayı amaçlar
- Veriyi gerçek zamanlı olarak işlemek ve doğruluğunun kontrolü çok önemlidir.



Örnek Problem:

Bir e-ticaret sitesindeki ürünleri, ilişkili markalar ve kategoriler ile birlikte depolayacak bir ilişkisel veritabanı yapısı tasarlayınız.

Her bir kategorinin, alt kategorileri olabilir.

Ürünler en uç kategori ile ilişkili olacaktır.

Ayrıca ürünlerin, özellikleri de veritabanında tutulacaktır.

Ürün özellikleri:

- Metin
- Liste Elemanı
- Çoklu Liste Elemanı
- Sayı

Türlerinde olabilir.



Örnek Problem:

Örneğin, ürünün bir "Dizüstü Bilgisayar" olduğunu hayal edelim

Kategori	Elektronik ve Bilgisayar > Bilgisayar > Dizüstü bilgisayar	
Marka	Lenovo	
Ürün Adı	Ideapad S145-14IWL	
Özellikler	Bluetooth : Var (Liste Elemanı) Ekran Boyutu: 14 (Sayı) Ram : 4 (Sayı) Görüntü Çıkışı: HDMI, VGA (Çoklu Liste Elemanı)	



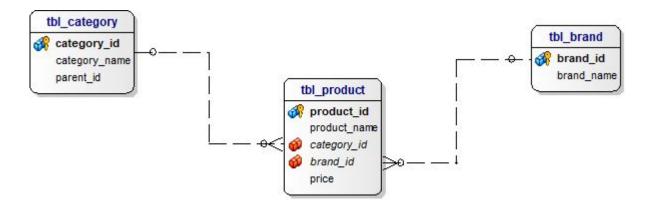




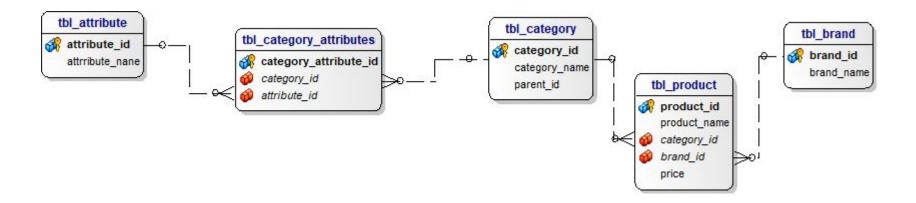




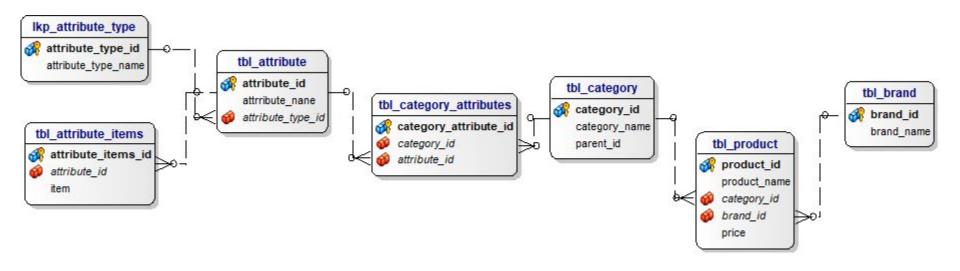




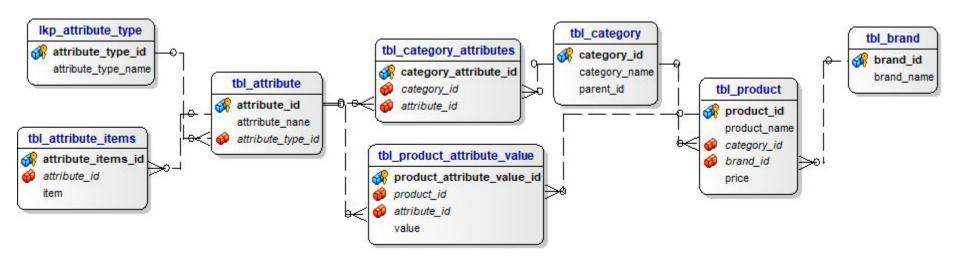














Örnek Problem:

Dizüstü bilgisayar kategorisinde, ağırlığı 2 kg'nin altında, RAM'i 8 GB ve üzerinde olan Diski 256 Gb ve üzeri olan, disk türü SSD olan, işlemcisi intel i7 olan bir bilgisayarları listeleyelim.



Örnek Problem Çözümü:

tbl_category

category_id		category_name	
	100	Dizüstü bilgisayar	



Örnek Problem Çözümü:

tbl_attribute

attribute_id	attribute_name
8	ağırlık
15	bellek
18	disk kapasitesi
22	disk türü
25	işlemci türü



Örnek Problem Çözümü:

tbl_attribute_items

attribute_item_id	attribute_id	item
3	25	intel i7
2	22	SSD



```
SELECT
 tbl brand.brand name,
 tbl product.product name,
 tbl product.price
FROM
  tbl product
  INNER JOIN tbl brand ON (tbl brand.brand id = tbl product.brand id)
  INNER JOIN tbl category ON (tbl category.category id = tbl product.category id)
  INNER JOIN tbl product attribute value AS agirlik ON (agirlik.product id = tbl product.product id)
  INNER JOIN tbl product attribute value AS bellek ON (bellek.product id = tbl product.product id)
  INNER JOIN tbl product attribute value AS disk kapasitesi ON (disk kapasitesi.product id = tbl product.product id)
  INNER JOIN tbl product attribute value AS disk turu ON (disk turu.product id = tbl product.product id)
  INNER JOIN tbl product attribute value AS islemci turu ON (islemci turu.product id = tbl product.product id)
WHFRF
 agirlik.value::real <= 2 AND
 bellek.value::integer >=8 AND
 disk kapasitesi.value::integer >= 256 AND
 disk turu.value::integer = 2 AND
 islemci turu.value::integer = 3
```



Bu örnekten yola çıkarak İlişkisel veri modelinde gözlemlediğimiz problemler:

- ürün tablosunda 1 milyon ürün olsa
- bir ürün için ortalama 10 özellik olsa
- ürün özellikleri tablosında 10 milyon kayıt olmalı
- Sorgulama sırasında 10 kez INNER JOIN yapıldığında 10 milyon kayıt içeren aynı tabloyu 10 kez okumak gerekecektir!



İlişkisel Olmayan Veri Modeli (NoSQL)

NoSQL: Not Only SQL

NoSQL veritabanları, belirli veri modelleri için özel olarak tasarlanmıştır ve modern uygulamalar oluşturmaya yönelik esnek şemalara sahiptir. NoSQL veritabanları uygun ölçekte geliştirme kolaylığı, işlevselliği ve performansıyla geniş çaplı kabul görmüştür. document,graph,key-value, in-memory ve search dahil olmak üzere çeşitli veri modelleri kullanır.

Bu tür veritabanları, özellikle büyük veri hacmi, düşük gecikme süresi ve esnek veri modelleri gerektiren uygulamalar için optimize edilmiştir. Bu gereksinimler, diğer veritabanlarının <u>veri tutarlılığı kısıtlamalarının bir kısmı esnetilerek karşılanır</u>.

(kaynak: aws)



İlişkisel Olmayan Veri Modeli (NoSQL)

Aynı problemi NoSQL ile çözmeye çalışırsak Ürün Dökümanı şöyle olur:

```
"id": "5771953cac3790b64d8b4567",
"product name": "Huawei Mate 20 Lite 64 GB",
"brand": {
  "brand id": "6871151cbd3701b64d8b4510",
  "brand name": "Huawei"
"category": {
   "category name": "Cep Telefonları"
"attributes": {
  "Ağırlık": "172 g",
   "Dahili Hafıza": "64 GB",
   "RAM Kapasitesi": "4 GB RAM",
   "Ön (Selfie) Kamera": "24 MP + 2 MP",
   "İşlemci Kapasitesi": "2,2 GHz Quad Core + 1,7 GHz Quad Core",
   "Kamera Çözünürlüğü": "20 MP + 2 MP",
 "price": "2169.99"
```



İlişkisel Olmayan Veri Modeli (NoSQL)

Benzer şekilde category, brand ve category attribute dökümanları da oluşturulacaktır Buradaki problem veri bütünlüğü bozulacaktır. Bir ürüne ait marka bilgisi hem product tablosunda hem de brand tablosunda olmalıdır.

```
{
    "id": "6871151cbd3701b64d8b4510",
    "brand_name": "Huawei"
    }
}
```

Bir marka isminde düzenleme yapılacak olursa, bu markaya ait tüm ürünlerde de bu güncelleme yapılması gereklidir. Bu maliyetin büyüklüğü bu tür işlemlerin geri planda yapılmasını zorunlu kılar.

Örneğin bankacılık işlemleri için bu hoş bir durum değildir.



İlişkisel Olmayan Veri Modeli vs İlişkisel Veri Modeli

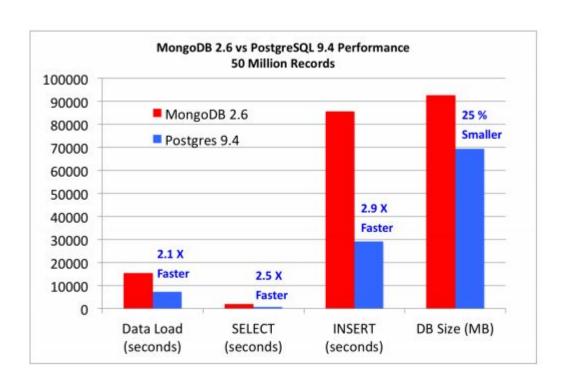
Table 1: MongoDB and Postgres Comparison (Absolute)

	MongoDB 2.6	PostgreSQL 9.4
Data load (s)	15391	7319
Inserts (s)	85639	29125
Selects (s)	1929	753
DB Size (GB)	92.63	69.36

Figure 10: Relative Performance Comparison of MongoDB 2.6 with PostgreSQL 9.4



İlişkisel Olmayan Veri Modeli vs İlişkisel Veri Modeli





Bir e-ticaret sitesinde Bir milyon ürün olabilir. Bu ürünlerin her birinin detay sayfası vardır. Bu detay sayfalarına ziyaretçiler, tıklar. Her bir tıklama kayıt altına alınmalıdır. Bu tıklamaların bazıları, kimliği belli tıklamalardır. Yani sitenin üyesi olan, oturum açmış müşterilerin tıklamalarıdır.





Bir e-ticaret sitesinde Bir milyon ürün olabilir. Bu ürünlerin her birinin detay sayfası vardır. Bu detay sayfalarına ziyaretçiler, tıklar. Her bir tıklama kayıt altına alınmalıdır. Bu tıklamaların bazıları, kimliği belli tıklamalardır. Yani sitenin üyesi olan, oturum açmış müşterilerin tıklamalarıdır.





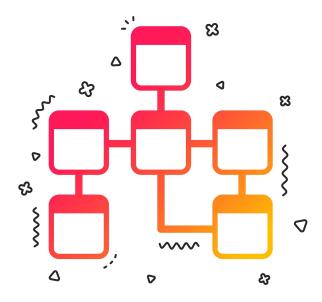
Böyle bir tabloda milyonlarca kayıt bulunur ama hiç birisi tek başına bir anlam ifade etmez. Ancak bir araya geldiklerinde bazı soruların yanıtı olurlar:

- X ürününe kaç kişi bakmış?
- Bir üye X ürününü farklı zamanlarda kaç kez incelemiş?
- En çok hangi ürünler ziyaret edilmiş?
- X ürününü ziyaret edenler başka hangi ürünleri en çok ziyaret etmiş? Buradan ürün önerileri de çıkar. Bu ürünü inceleyenler şu ürüne de baktılar gibi.



Görüldüğü üzere bu tablodaki verilerin kullanılabilmesi için bir işleme tabi tutulmaları gerekir. Devasa büyüklükte olacağı tahmin edilebilen bu tablo ilişkisel yapıda olduğunda veri tabanını da şişirecektir! Bu tablonun ilişkisel bir veri tabanında olmasına gerek yoktur. Ancak işlem sonrasında ortaya çıkan rakamlar ilişkisel veri tabanına yazılabilir.

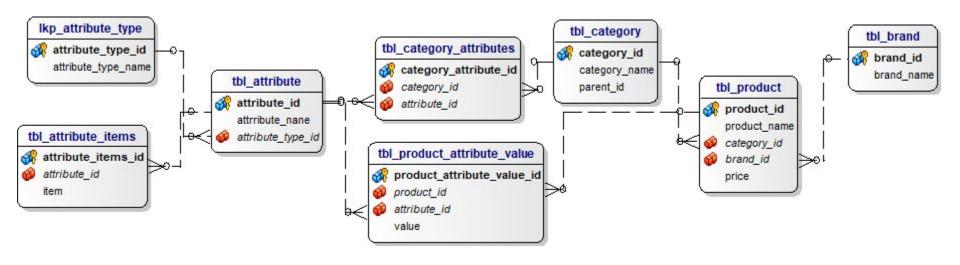




Hibrit data Model bu gibi durumlar için kullanılabilecek bir çözümdür. İlişkisel veritabanında ilişkisel olmayan verileri tutmaya yarar.

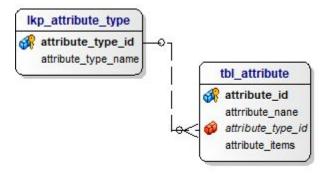


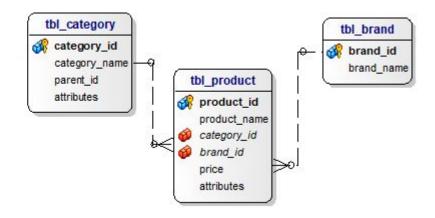
İlişkisel Modeldeki Şu Çözüm :





Hibrit Model İle şu şekilde olabilir :







Avantajları

- İlişkisel veri tabanı olduğu için daha az disk kullanır
- veri bütünlüğü korunur
- okuma yazma hızından taviz verilmez. Hatta jsonb tipindeki bir alanda sorgulama doğal alanlarda sorgulamadan daha hızlıdır.
- SQL kullanarak sorgu yazarsınız



Dezavantajları

- DBMS'nin yerleşik araçları ile import/export yapmak mümkün olmaz.
 Tüm DB'yi yada tüm tabloyu import export yapabilirsiniz ancak bir csv dosyasından bir tablo oluşturamazsınız. Bunun için DB'nin programlama dilinde kod yazmanız gerekir.
- JSONB türündeki alanlar GIN index ile indexlenebilir ancak index'in diskteki boyutu oldukça büyüktür.
- tarih gibi tiplerin JSON karşılığı olmadığı için kendi çözümünüzü üretmek zorundasınız.



PostgreSQL

- Postgresql 9.0 versiyonundan itibaren json veri tiplerini desteklemektedir.
- Oracle, MySQL ve MSSQL de json veri tipini destekliyor ancak Postgresql kadar ileri seviyede veri manevraları henüz ya yapamıyorlar yada uzun fonksiyon isimleri yüzünden karmaşıklaştırıyolar.
- Postgresql GIN index ile bu tipteki alanların sorgulamasını hızlandırmaktadır.







PostgreSQL json ve jsonb operatörleri

Operator	Açıklama	Örnek	Sonuç
->	Bir json içinden key yada index ile belirtilen değeri okur sonuç yine json yada jsonb tipindedir.	'[{"a":"foo"},{"b":"bar"},{"c":"baz"}]'::json-	{"c":"baz"}
->>	Bir json içinden property yada index ile belirtilen değeri okur sonuç text tipindedir.	'[1,2,3]'::json->>2	3
#>	Bir json objesinden, belirlenen bir path deki json'ı alır. Sonuç yine json yada jsonb tipindedir.	'{"a": {"b":{"c": "foo"}}}'::json#>'{a,b}'	{"c": "foo"}
#>>	Bir json objesinden, belirlenen bir path deki json'ı alır. Sonuç text tipindedir.	'{"a":[1,2,3],"b":[4,5,6]}'::json#>>'{a,2}'	3



PostgreSQL json ve jsonb operatörleri

Operator	Açıklama	Örnek
@>	Soldaki json/jsonb içinde sağdaki json/jsonb yada değer var mı? (arama en üst seviyede yapılır)	'{"a":1, "b":2}'::jsonb @> '{"b":2}'::jsonb
<@	Sağdaki json/jsonb içinde soldaki json/jsonb yada değer var mı? (arama en üst seviyede yapılır)	'{"b":2}'::jsonb <@ '{"a":1, "b":2}'::jsonb
?	Bu json/jsonb içinde böyle bir key var mı?	'{"a":1, "b":2}'::jsonb ? 'b'
?	Bu json/jsonb içinde böyle bir keylerden birisi var mı?	'{"a":1, "b":2, "c":3}'::jsonb ? array['b', 'c']



PostgreSQL json ve jsonb operatörleri

Operator	Açıklama	Örnek
?&	Bu json/jsonb içinde bu keylerin hepsi var mı? (arama en üst seviyede yapılır)	["a", "b"]'::jsonb ?& array['a', 'b']
II	iki json /jsonb yi birleştirir. birleştirme en üst seviyede yapılır.	'["a", "b"]'::jsonb '["c", "d"]'::jsonb
-	Bir json/jsonb den key-value çifti silmek için kullanılır	'{"a": "b"}'::jsonb - 'a'
#-	Belirlenen bir path 'deki json objexini silmek için kullanılır	'["a", {"b":1}]'::jsonb #- '{1,b}'



PostgreSQL json ve jsonb fonksiyonları

- jsonb_array_length
- to_jsonb
- jsonb_each
- jsonb_each_text
- jsonb extract path
- jsonb_extract_path_text
- jsonb object keys
- jsonb_populate_record
- jsonb_populate_recordset
- array_to_json
- row_to_json
- json**b**_array_elements

- jsonb_array_elements_text
- jsonb_typeof
- jsonb_to_record
- jsonb_to_recordset
- jsonb_strip_nulls
- jsonb_set
- jsonb_pretty
- json**b**_build_array
- json**b**_build_object



PostgreSQL'de yeni bir tablo oluşturalım

```
create table public.tbl_product
(
    product_id bigserial not null
    constraint tbl_product_pk
    primary key,
    product_name varchar(512),
    properties jsonb,
    brand_id bigint,
    category_id bigint,
    price real,
    real integer,
    product_slug varchar(255)
);
```



Bir de index ekleyelim

CREATE INDEX idx_tbl_product_properties
ON tbl_product USING GIN(properties);



Örnek Problem:

Dizüstü bilgisayar kategorisinde, ağırlığı 2 kg'nin altında, RAM'i 8 GB ve üzerinde olan Diski 256 Gb ve üzeri olan, disk türü SSD olan, işlemcisi intel i7 olan bir bilgisayarları listeleyelim.



İlişkisel Veri Modeli

Örnek Problem Çözümü:

```
SELECT
 tbl brand.brand name,
 tbl product.product name,
 tbl product.price
FROM
  tbl product
  INNER JOIN tbl brand ON (tbl brand.brand id = tbl product.brand id)
  INNER JOIN tbl category ON (tbl category.category id = tbl product.category id)
  INNER JOIN tbl product attribute value AS agirlik ON (agirlik.product id = tbl product.product id)
  INNER JOIN tbl product attribute value AS bellek ON (bellek.product id = tbl product.product id)
  INNER JOIN tbl product attribute value AS disk kapasitesi ON (disk kapasitesi.product id = tbl product.product id)
  INNER JOIN tbl product attribute value AS disk turu ON (disk turu.product id = tbl product.product id)
  INNER JOIN tbl product attribute value AS islemci turu ON (islemci turu.product id = tbl product.product id)
WHFRF
 agirlik.value::real <= 2 AND
 bellek.value::integer >=8 AND
 disk kapasitesi.value::integer >= 256 AND
 disk turu.value::integer = 2 AND
 islemci turu.value::integer = 3
```



Örnek Problem Çözümü:

```
SELECT
 tbl brand.brand name,
 tbl product.product name,
 tbl product.price
FROM
 tbl product
 INNER JOIN tbl brand ON (tbl brand.brand id = tbl product.brand id)
 INNER JOIN tbl category ON (tbl category.category id = tbl product.category id)
WHFRF
 (tbl product.properties->>'agirlik')::real <= 2 AND
 (tbl product.properties->>'bellek')::integer >=8 AND
 (tbl product.properties->>'disk kapasitesi')::integer >= 256 AND
 tbl product.properties->>'disk turu' = 'ssd' AND
 tbl product.properties->>'islemci turu' = 'intel i7'
```



Güncelleme Sorgusu:

```
UPDATE tbl_product
   SET properties = properties || '{"kameralar":["ön","arka"]}'
WHERE product_id=1;

UPDATE tbl_product
   SET properties = properties || jsonb_build_object('kameralar',ARRAY['ön','arka'])
WHERE product_id=1;
```



Sanal Tablo Problemi:

- Bir bulut (SAAS) CRM uygulamasında farklı sektörler farklı veri kümelerine ihtiyaç duyar
 - Sigorta sektörü Poliçe satar
 - Havayolu firması bilet satar
 - Mobilya firması ürün satar
- Tüm bu ihtiyaçları bulut uygulamada yerleşik olarak oluşturmanın maliyeti (zaman, para, iş-gücü) çok fazladır.
- Bu sebeplerle hibrit veri modeli kullanarak, müşterilerimizin sanal tablolar oluşturabilmesine olanak sağladık.



Sanal Tablo Problemi:

tbl_vtable

vtable_id	name	title	channel_id
100	urun	Ürün	89



Sanal Tablo Problemi:

tbl_vtable_row

vtable_row _id	vtable_id	related_id	custom_data	channel_id	relation_kind_id
1	100	NULL	{"urun_adi": "Performance 22VH3021 22",}	89	3



Sanal Tablo Problemi:

tbl_vtable

vtable_id name		title	channel_id	
100	urun	Ürün	89	
101	siparis	Sipariş	89	
102	siparis_urunleri	Sipariş Ürünleri	89	



Sanal Tablo Problemi:

tbl_vtable_row

vtable_row _id	vtable_id	related_id	custom_data	channel_id	relation_kind_id
3	101	427406	{"siparis_no": "55422",}	89	1
2	102	3	{"miktar": 2,"urun_id":3,}	89	3
1	100	NULL	{"urun_adi": "Performance 22VH3021 22",}	89	3



Sanal Tablo Problemi:



Gerekenler:

- Postgresql 11
- DB yönetim aracı (DataGrip)
- Örek DB
 https://github.com/aerkanc/hybrid-db



Cep telefonu kategorisi altında hangi ürün özellikleri var?

```
SELECT
  item ->> 0 as attribute
FROM (
    SELECT
       jsonb_array_elements(attributes) as item
FROM
    tbl_category
    WHERE
       category_id = 2
) as attribute_items;
```



Bir Cep Telefonu Kategorisinde Pil Gücü özelliğinin seçenekleri neler?

```
SELECT
DISTINCT properties->>'Pil Gücü'
FROM
tbl_product
WHERE
category_id = 2;
```



2000 mAh dan büyük pil gücüne sahip telefonlar neler?

```
SELECT
brand_name,
product_name,
price
FROM
tbl_product
INNER JOIN tbl_brand ON (tbl_brand.brand_id = tbl_product.brand_id)
WHERE
category_id = 2 AND
REPLACE(properties ->> 'Pil Gücü', ' mAh', ")::INTEGER > 2000
```



Wifi'si 802.11 ac yi destekleyen telefonlar neler?

```
SELECT

brand_name,

product_name,

tbl_product.properties->>'Wi-Fi' as "Wi-Fi",

price

FROM

tbl_product

INNER JOIN tbl_brand ON (tbl_brand.brand_id = tbl_product.brand_id)

WHERE

category_id = 2 AND

tbl_product.properties->>'Wi-Fi' ~ '^802\.11.*ac.*$'
```



Teşekkürler