

**Ruben Pertusa**

**Miguel Egea**

**@RPertusa @MiguelEgea**

**Tecnicas avanzadas de modelado**

# Sponsors



Microsoft Azure

kabel 

au 

  
SolidQ

東京  
TOKIOTA

encamina  
PIENSA EN COLORES



zaltor  
Software Distribution

alesonITC  
data thinking



Universidad  
Politécnica  
de Madrid

ETSI SISTEMAS  
INFORMÁTICOS



POLITÉCNICA

UNIVERSIDAD  
POLITÉCNICA  
DE MADRID

**SentryOne**®

Thanks!  


# Rubén Pertusa Lopez

@rpertusa

[www.rpertusa.com](http://www.rpertusa.com)

[rpertusalopez@gmail.com](mailto:rpertusalopez@gmail.com)

Microsoft MVP Data Platform

BI BigData Lead Architect & Global

BI Manager @DUFRI

[www.sqlpass.es](http://www.sqlpass.es) , Sql Saturday

Madrid/Barcelona Founder

# Miguel Egea Gómez

@miguelegea

[www.portalsql.es](http://www.portalsql.es)

[megea@solidq.com](mailto:megea@solidq.com)

EX Microsoft MVP Data Platform

Mentor & technical advisor at

SolidQ

Trying to help Ruben at much as  
posible ☺ both in Madrid and  
Barcelona



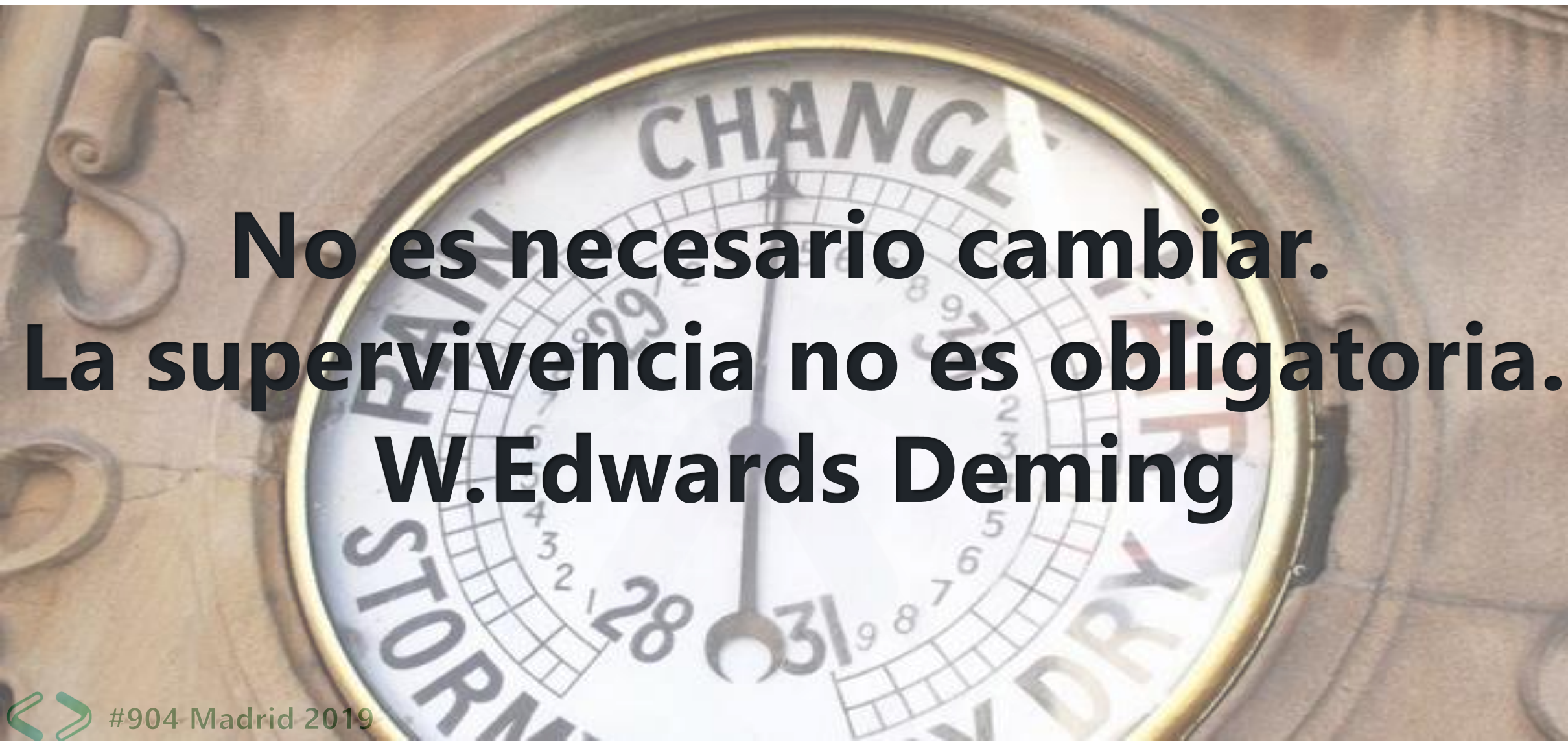
## Modelo Modelo Modelo

- ..... sin modelo no hay vida .....
- Dimensions, hechos, cubos .. Ralph Kimball

## Tipos de hechos, problemas y sus soluciones

- Tablas de hechos sin hechos
  - Ejemplo Posts StackOverflow
  - Ejemplo Stock
- Distinct Count
  - Simple distinct count
  - Distinct Count complejo
  - Otro distinct count complejo
- Bonus track: Snapshot
  - Convirtiendo tablas snapshot a deltas.

# Tratar con datos



**No es necesario cambiar.  
La supervivencia no es obligatoria.  
W. Edwards Deming**

# Modelado dimensional



Es la implementación de un modelo lógico en una base de datos relacional



Estructura los datos en una forma comprensible para los usuarios de negocio



Se usa para la implementación de DWH o DM



Da información sobre

- Cómo las entidades se relacionan entre sí
- Cardinalidad, restricciones, rendimiento

# Trabajando con tablas de hechos



# Retos trabajando con tablas de hechos



## Medidas semi-aditivas

Normalmente no aditivas con la dim tiempo



## Distinct-count

No aditivas en absoluto



## Fact-less complicated scenarios

Como calculo de negativos ...

### Solutions

Usar

Lastnonempty

Lastnonblank

LastDate

Convertir

Medidas  
semiaditivas

en

Aditivas  
cuando sea  
posible

Usar

Distinct count  
functions

Convertir

Distinct count

en

Count + distinct  
count Cuando sea  
posible

Usar

Baja  
granularity

Expandir  
tablas

Expandir  
tablas  
netgativas

Convertir

Hechos en  
delta

Positivos en  
fechas de  
inicio

Negativos al  
final



## Modelo Modelo Modelo

- ..... sin modelo no hay vida .....
- Dimensions, hechos, cubos .. Ralph Kimball

## Tipos de hechos, problemas y sus soluciones

- Tablas de hechos sin hechos
  - Stock basado en Deltas
  - Metricas History To Date
  - Optimizaciones HTD
- Distinct Count
  - Simple distinct count
  - Distinct Count complejo
  - Otro distinct count complejo
- Snapshot
  - Convirtiendo tablas snapshot a deltas.

# Problemas desde-hasta

En la base de datos stackoverflow queremos saber cuantas preguntas hay sin responder



# Retos para responder

## No hay preguntas y respuestas

- La estructura de hechos es un árbol
- Afortunadamente podemos aplanar ese árbol

## Relación con el tiempo

- No nos basta con saber cuantas preguntas hay sin responder ahora, necesitamos saber cuantas había en cualquier momento del tiempo

# Diseño de la tabla de hechos

```
SELECT p.[Id],  
       p.[AcceptedAnswerId],  
       p.OwnerUserId,  
       p.[AnswerCount],  
       p.[CreationDate] FromDate,  
       COALESCE(PostLeft.CreationDate, '20220517') ToDate  
FROM [dbo].[Posts] p  
LEFT JOIN dbo.Posts PostLeft ON  
       p.AcceptedAnswerId = PostLeft.id;
```



# Solución

## Snapshots

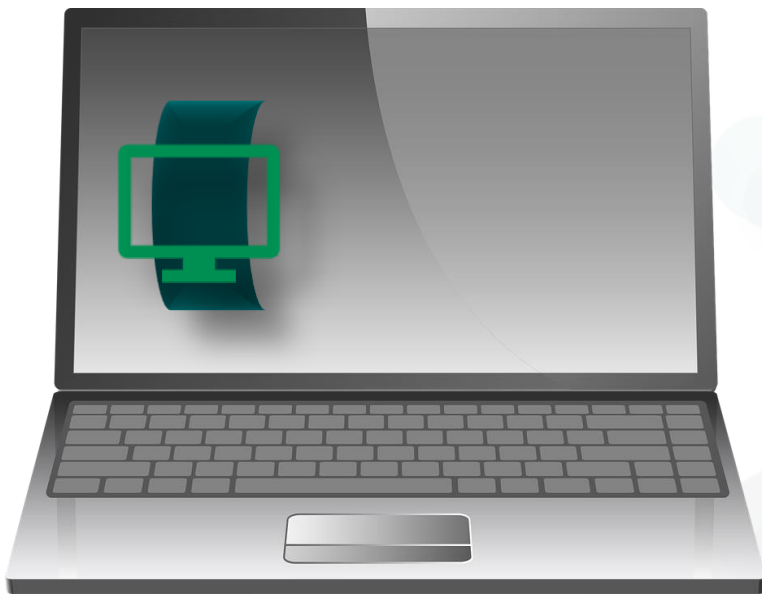


- ☐ Capturar en cada momento del tiempo todas las preguntas sin responder
- ☐ Granularidad... usuario-mes
- ☐ Contar tablas usando medidas semi-aditivas

## Transaccional



- ☐ Crear campos desde- hasta
- ☐ Situar hasta lejano en el tiempo cuando sea nulo
- ☐ Sumar desde el principio de los tiempos hasta el momento actual



# Demo. Desde-Hasta Demo. Snapshot

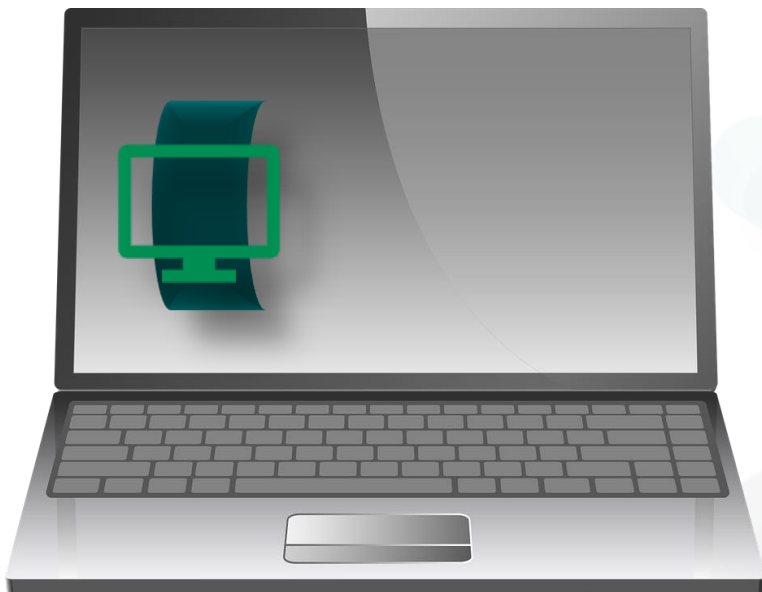
## Planificación de soluciones

Las instantáneas son sencillas a nivel de DWH. ¿y a nivel de Queries?



### Deltas

skVenta	skProd	skFecha	Importe	uds	Dto	Deteccion	fa
1	1	2/01/19	100 €	3		2/01/19	1
	1	2/01/19	100 €	3		4/01/19	-1
	1	2/01/19	200 €	3		4/01/19	1
2	1	3/01/19	50 €	2	5 €	3/01/19	1
		SUM*FA	250 €	5			



DEMO STOCK



## Modelo Modelo Modelo

- ..... sin modelo no hay vida .....
- Dimensions, hechos, cubos .. Ralph Kimball

## Tipos de hechos, problemas y sus soluciones

- Tablas de hechos sin hechos
  - Stock basado en Deltas
  - Metricas History To Date
  - Optimizaciones HTD
- Distinct Count
  - Simple distinct count
  - Distinct Count complejo
  - Otro distinct count complejo
- Snapshot
  - Convirtiendo tablas snapshot a deltas.

# Opciones para Distinct count



Distinct Count puro  
N

No aditiva  
No tan buen rendimiento  
Mucha memoria ocupada  
Mucha lectura



Distinct-count tuneado

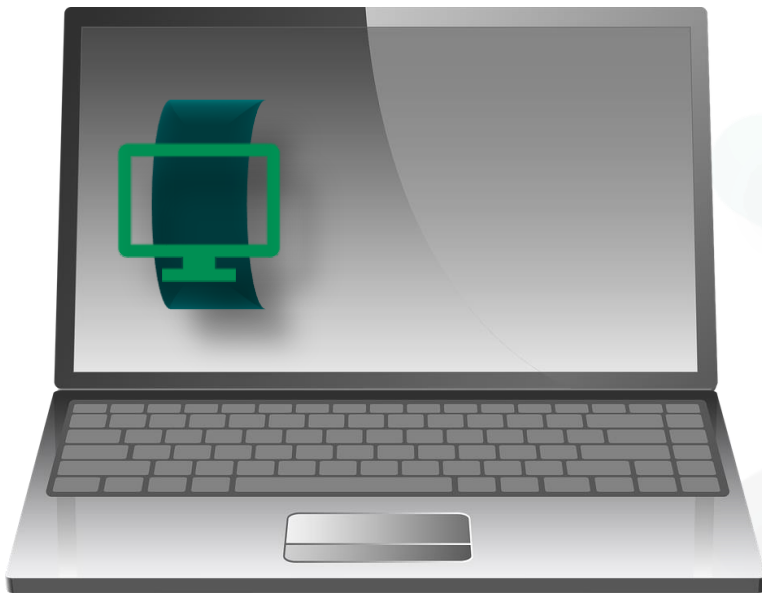
- Crea un atributo que indique si el pedido tiene más de una línea
- Crea una table con solo los ids de cabecera y ese atributo
- Si no se usa ningun atributo específico del detalle... usa count
- Usa distinctcount solo filtrando por las cabeceras de mas de 1 línea



Distinct- count vs MxN

Compresión de relaciones MxN

	MxN Relations					Distinct Count				
	total	SE Cpu	FE	SE		total	SE Cpu	FE	SE	
Ejecución 1	181	531	41	140		166	1031	10	156	
Ejecución 2	162	532	40	122		169	969	8	161	
Ejecución 3	166	500	41	125		184	1141	10	174	
Ejecución 4	173	375	43	130		166	875	9	157	
	<b>170.5</b>	<b>484.5</b>	<b>41.25</b>	<b>129.25</b>		<b>171.25</b>	<b>1004</b>	<b>9.25</b>	<b>162</b>	



# Demo distinct count



