



DMC ONLINE

#YoMeQuedoEnCasa

Taller:

SQL SERVER FOR ANALYTICS

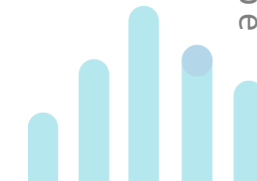




Soy
Victor Gutiérrez

Arquitecto de Datos

#YoMeCapacitoEnCasa





Agenda

- 1 Presentación
- 2 Base de datos Relacional
- 3 Datos Gobernados
- 4 Data Visualization
- 5 Preguntas y Respuestas



PERFIL PROFESIONAL

He asumido roles de Gerente General, Jefe de Proyectos, Arquitecto de datos, Líder Técnico, Analista Senior, Analista Funcional, Analista de Sistemas y Desarrollador; participando en proyectos del Sectores Privados (Telco, Banca, Retail, Metalúrgica, Pesquero); y en el sector público. Especialista en elaborar, formular, revisar y optimizar procesos ETL-ELT Warehouse, complementados con conocimientos de Big Data y Data Science.

He participado como consultor técnico en las POC de Oracle Exadata Database Machine al lado del fabricante Oracle. Miembro de DAMA.org

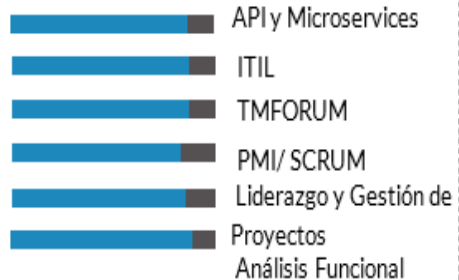


EXPERIENCIA

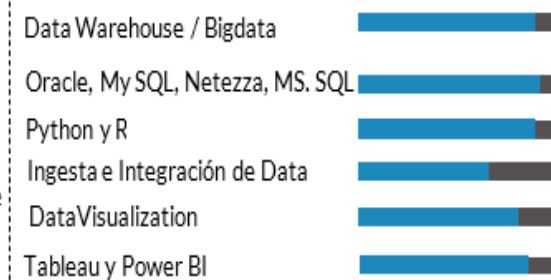


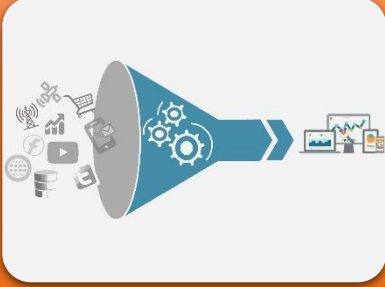
COMPETENCIAS

GESTION



TÉCNICAS





Analítica

- Obtener información
- Conjunto de técnicas para obtener y procesar datos



Porque



A donde vamos

- Integración de datos (Real Time + DWH)
- Analítica Avanzada (Casos de Usó: **CHURN**)

¿Qué son los datos oscuros?

Los datos oscuros son todos los datos desconocidos y sin explotar en toda la organización, generados por los sistemas, los dispositivos y las interacciones.

El cincuenta y cinco por ciento de los datos de una organización son "oscuros":

*sin cuantificar y sin explotar - de acuerdo con el nuevo global investigación de **TRUE Global Intelligence**, patrocinada por **Splunk** Los 1.300 líderes empresariales encuestados reconocen esos datos son clave para el éxito, ahora y en el futuro, aún muy pocos dicen que sus organizaciones pueden aprovechar con éxito el valor de todos sus datos. O encuéntralo*

Fuente: State of Dark Data Report by Splunk (2019)



Lo que impide que las empresas recuperen datos oscuros



Los datos son la clave del éxito corporativo y profesional...



Por donde empezar?

- 1 Descubrir ¿**Qué** datos necesitamos para obtener el mejor retorno de la inversión?



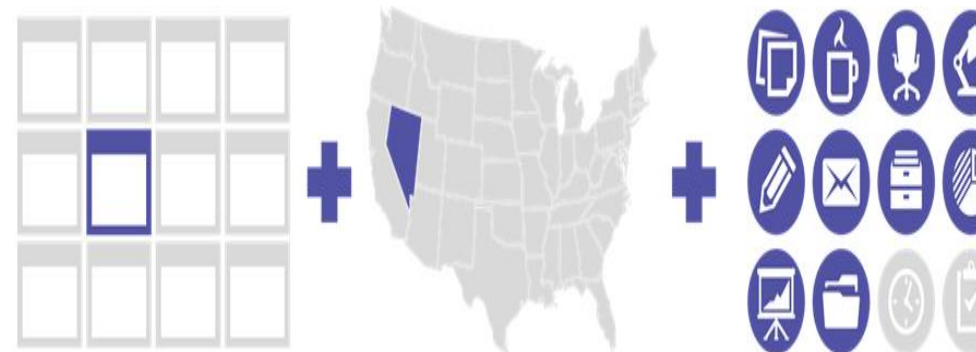
- 2 Encontrar **cuándo** se producen las ventas con mayor recaudación



- 3 Encontrar **dónde** las ventas son más altas



- 4 Encontrar **qué** productos venden más volumen



Soluciones de BI, mas que los dashboard



Base de datos

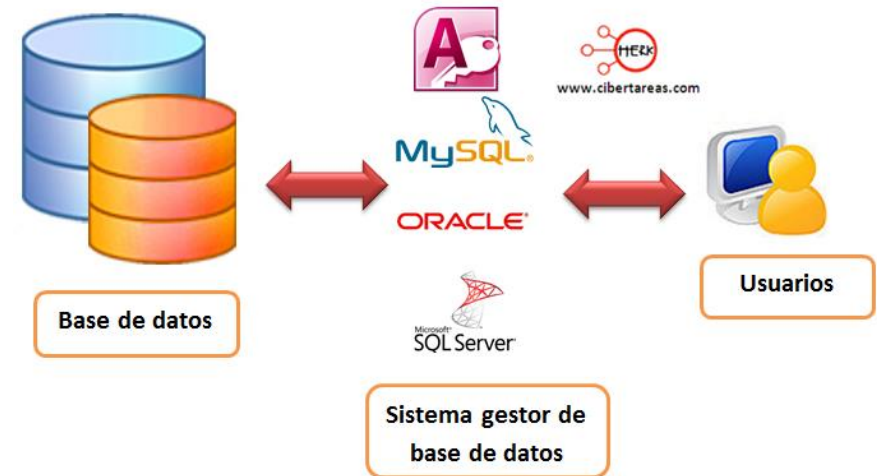
Definición de base de datos

Se define una base de datos como una serie de datos organizados y relacionados entre sí, los cuales son recolectados y explotados por los sistemas de información de una empresa o negocio en particular.

Características

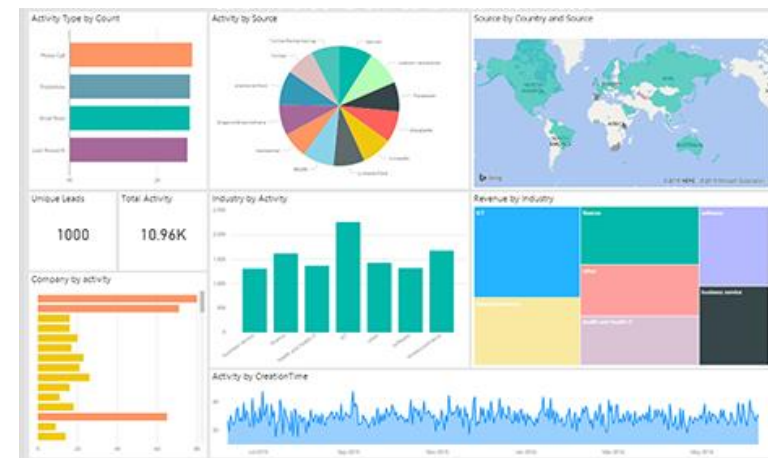
Entre las principales características de los sistemas de base de datos podemos mencionar:

- Acceso concurrente por parte de múltiples usuarios.
- Integridad de los datos.
- Consultas complejas optimizadas.
- Seguridad de acceso y auditoría.
- Acceso a través de lenguajes de programación estándar (Transact-SQL)

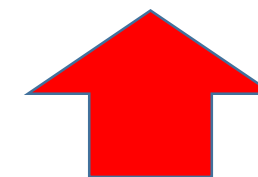
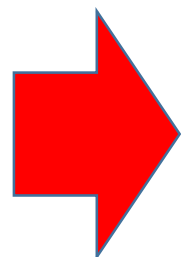


Base de datos

E2E: Desde las aplicaciones al BI



Registran
información de los
Procesos



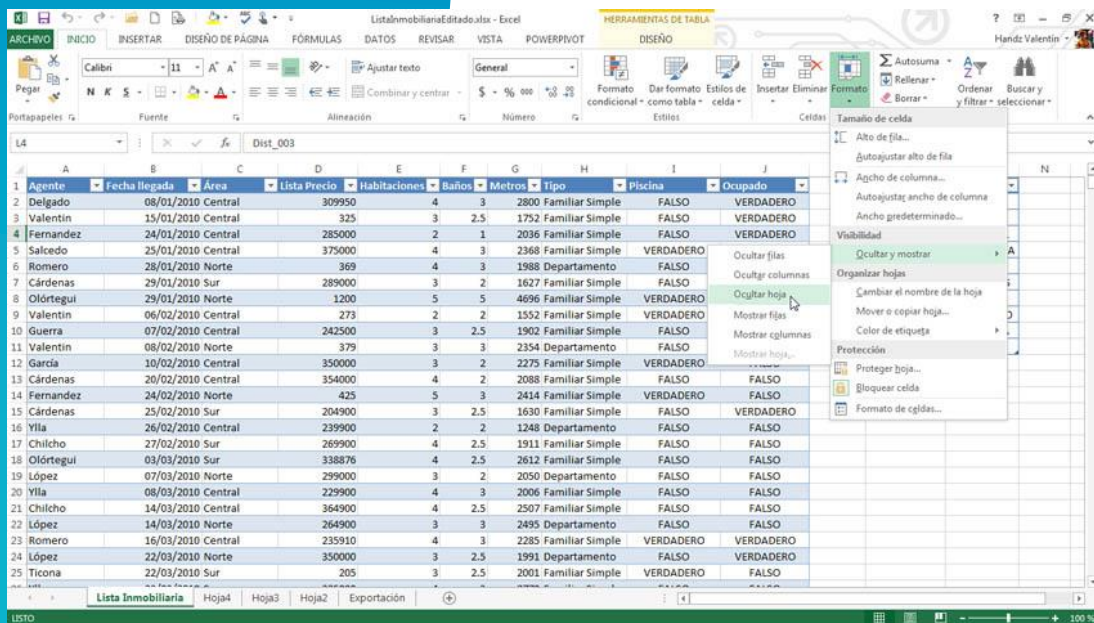
Entregar
información de los
procesos

Base de datos

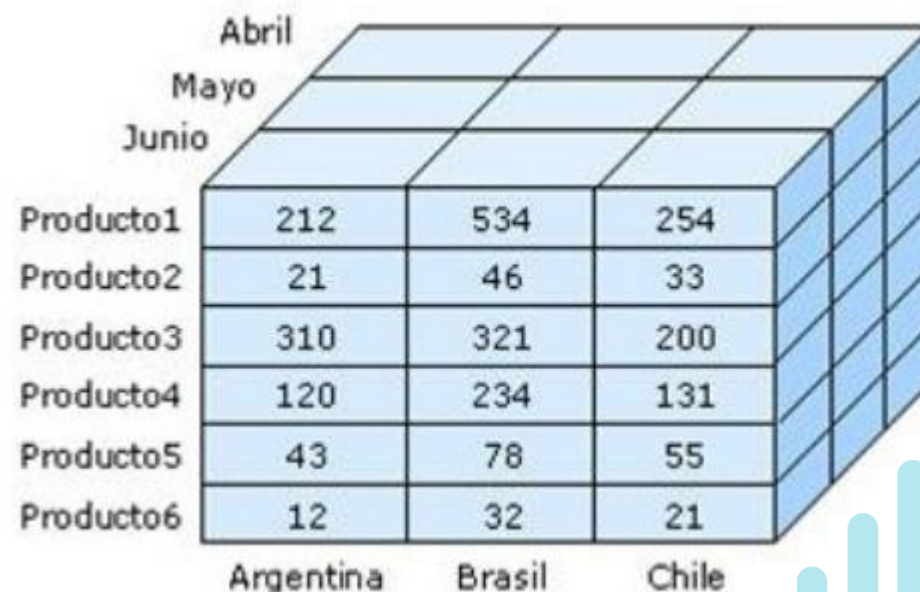
Estructura

Una base de datos, a fin de ordenar la información de manera lógica, posee un orden que debe ser cumplido para acceder a la información de manera coherente. Cada base de datos contiene una o más tablas, que cumplen la función de contener los campos.

*Como podemos responder diferentes preguntas de lo que esta pasando:
¿ Cuanto vendimos?, ¿ Que vendimos ?, ¿ A quienes le vendimos?*



Agente	Fecha Llegada	Área	Lista Precio	Habitaciones	Baños	Metros	Tipo	Piscina	Ocupado
Delgado	08/01/2010	Central	309950	4	3	2800	Familiar Simple	FALSO	VERDADERO
Valentin	15/01/2010	Central	325	3	2.5	1752	Familiar Simple	FALSO	VERDADERO
Fernandez	24/01/2010	Central	285000	2	1	2036	Familiar Simple	FALSO	VERDADERO
Salcedo	25/01/2010	Central	375000	4	3	2368	Familiar Simple	VERDADERO	
Romero	28/01/2010	Norte	369	4	3	1988	Departamento	FALSO	
Cárdenas	29/01/2010	Sur	289000	3	2	1627	Familiar Simple	FALSO	
Olórtégui	29/01/2010	Norte	1200	5	5	4696	Familiar Simple	VERDADERO	
Valentin	06/02/2010	Central	273	2	2	1552	Familiar Simple	VERDADERO	
Guerra	07/02/2010	Central	242500	3	2.5	1902	Familiar Simple	FALSO	
Valentin	08/02/2010	Norte	379	3	3	2354	Departamento	FALSO	
García	10/02/2010	Central	350000	3	2	2275	Familiar Simple	VERDADERO	
Cárdenas	20/02/2010	Central	354000	4	2	2088	Familiar Simple	FALSO	FALSO
Fernandez	24/02/2010	Norte	425	5	3	2414	Familiar Simple	VERDADERO	FALSO
Cárdenas	25/02/2010	Sur	204900	3	2.5	1630	Familiar Simple	FALSO	VERDADERO
Villa	26/02/2010	Central	239900	2	2	1248	Departamento	FALSO	FALSO
Chilcho	27/02/2010	Sur	269900	4	2.5	1911	Familiar Simple	FALSO	FALSO
Olórtégui	03/03/2010	Sur	338876	4	2.5	2612	Familiar Simple	FALSO	FALSO
López	07/03/2010	Norte	299000	3	2	2050	Departamento	FALSO	FALSO
Villa	08/03/2010	Central	229900	4	3	2006	Familiar Simple	FALSO	FALSO
Chilcho	14/03/2010	Central	364900	4	2.5	2507	Familiar Simple	FALSO	FALSO
López	14/03/2010	Norte	264900	3	3	2495	Departamento	FALSO	FALSO
Romero	16/03/2010	Central	235910	4	3	2285	Familiar Simple	VERDADERO	VERDADERO
López	22/03/2010	Norte	350000	3	2.5	1991	Departamento	FALSO	VERDADERO
Ticona	22/03/2010	Sur	205	3	2.5	2001	Familiar Simple	VERDADERO	FALSO



	Abril	Mayo	Junio
Producto1	212	534	254
Producto2	21	46	33
Producto3	310	321	200
Producto4	120	234	131
Producto5	43	78	55
Producto6	12	32	21

Datos Gobernados

¿Que Buscamos?



En el lugar
correcto



En el tiempo
correcto



Para todos
los usuarios



En el formato
correcto

¿Que Haremos?



Métricas y Conceptos
Consistentes



Monitoreo
Continuo



Información

- ✓ Consistente
- ✓ Confiable
- ✓ Repetible
- ✓ Única
- ✓ Guía

Importancia

- ✓ Mayor Cantidad
- ✓ Mayor Diversidad

Importancia Negocio

- ✓ Mayores Necesidades
- ✓ Mayores Regulaciones y
Controles

Visualización de Datos

- ❑ La visualización de datos busca representar los 1 y 0 del proceso **ETL**
- ❑ Dependiendo de las herramientas podemos utilizar dashboard, gráficos y mapas
- ❑ Proporcionar la forma accesible de ver y comprender tendencias, valores atípicos y patrones en los datos.

Ingesta



Kinesis Firehose



Kinesis Data Streams



SQS



AWS IoT Core



DMS



Snowball

Almacenamiento



S3



DynamoDB



MCS

Procesamiento



Lambda



Glue



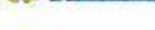
Data Pipeline



EMR

APACHE
Spark

HUE



Hadoop



HIVE

Análisis



Kinesis Analytics



RedShift



Athena



ElasticSearch

Visualización



QuickSight

Datalake



Lake Formation

Seguridad



CloudHSM



KMS



CloudTrail



STS

Data Storytelling: Qué necesitas saber?

Narrativa + Datos = podremos explicar qué ha pasado y por qué un insight puede ser importante. Necesitaremos contexto para entender las conclusiones por completo.

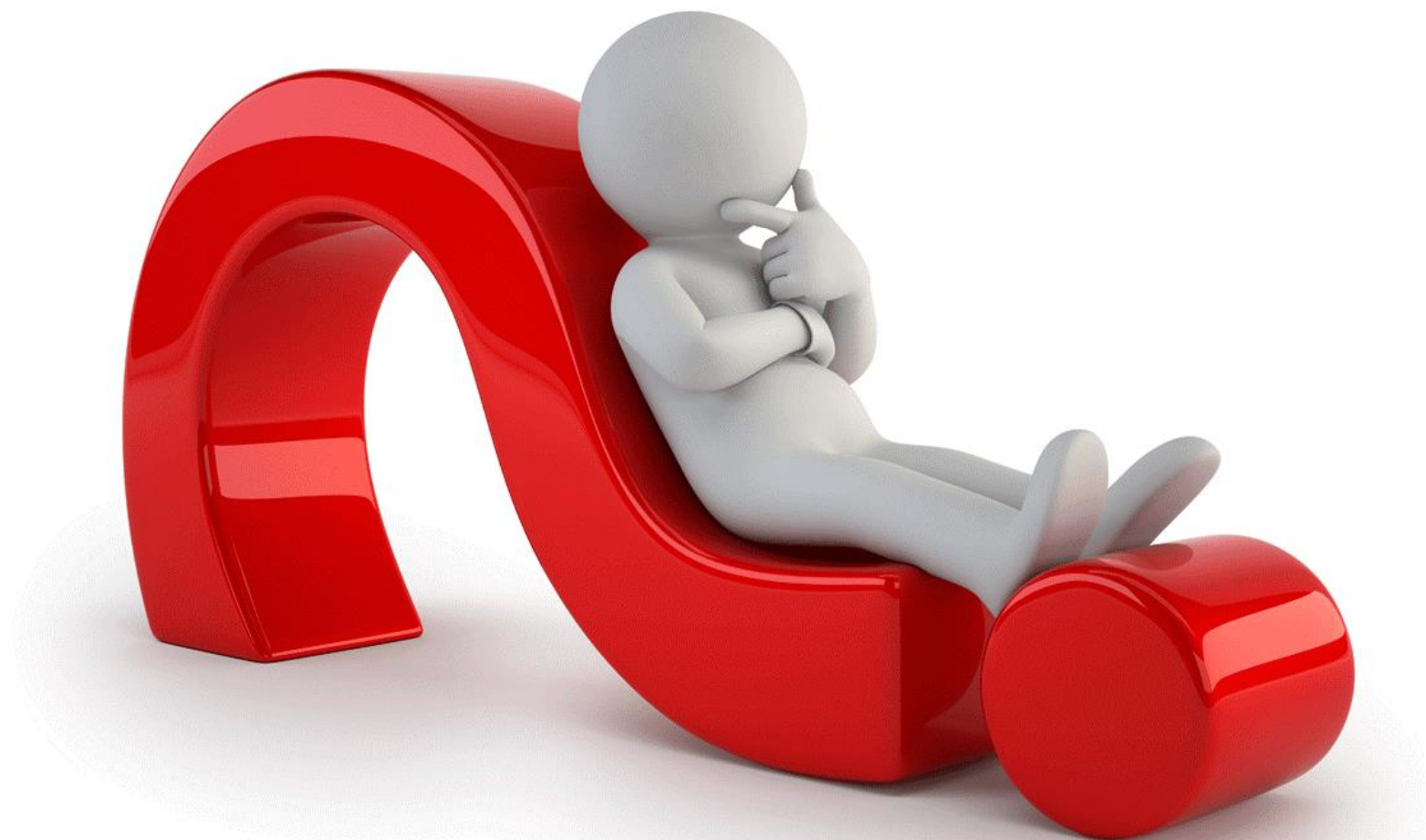
Visualización + Datos = Enlighten. Cuando añadimos una visualización a nuestros datos, podemos iluminar a nuestra audiencia con insights que no habrían visto de otra manera.

Narrativa + Visualización = Engagement. La combinación perfecta para lograr ese interés e incluso para entretener a nuestra audiencia.

Pero, cuando unimos **Visualización + Narración + Datos = Change**, logramos contar una historia con nuestros datos, logramos influenciar y llevar a ese cambio que estábamos buscando.



Fuente: Forbes



Consultas y comentarios.