

### Universidad Nacional Autónoma de México Facultad de Ciencias **Diplomado en Minería de Datos**

Módulo 3. Procesamiento analítico de datos

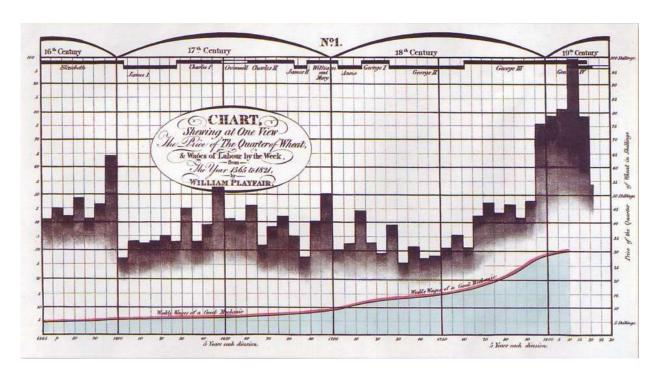
# Visualización de datos en BI

Gerardo Avilés Rosas gar@ciencias.unam.mx



- Las primeras visualizaciones (además de las tablas) fueron las series de tiempo y las gráficas de barras.
- Aunque existen versiones anteriores, este tipo de gráficas se creó en la forma que ahora estamos familiarizados con William Playfair a finales de 1700.



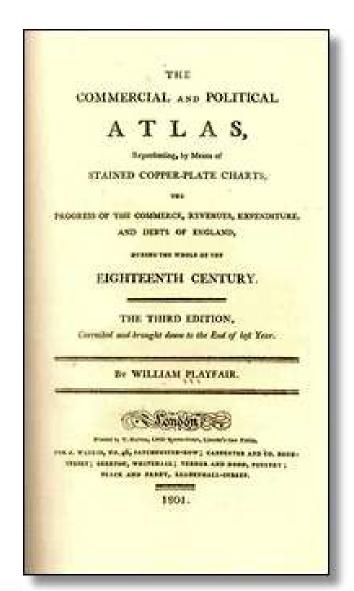


Ingeniero mecánico y economista político escocés. Trabajó más de 36 años en el diseño de gráficos estadísticos. Se le considera pionero en el uso de las gráficas lineales para representar series de tiempo. Creador de las gráficas circulares, de sectores y de barras.



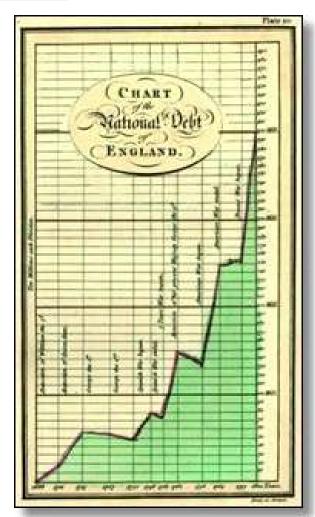
- En 1786 publicó The Commercial and Political Atlas donde incluía 43 series de tiempo y una gráfica de barras sobre diversos aspectos de la economía de Inglaterra, especialmente sobre comercio exterior.
- Según Playfair:

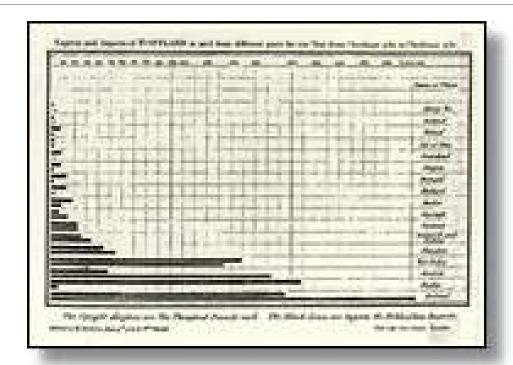
...un buen gráfico proporciona una explicación más adecuada de los hechos que una mera lista de datos o tablas. Sirve para simplificar lo complejo, permite al cerebro una mayor retención y es un instrumento visual de ayuda a hombres ocupados. Por último, los gráficos nos permiten ver relaciones aparentemente inexistentes entre variables, que suelen quedar ocultas entre la multitud de datos y cifras, de difícil comparación de otro modo.

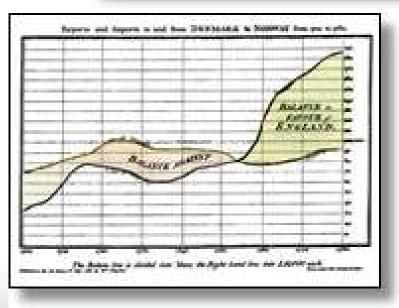




# ...Introducción







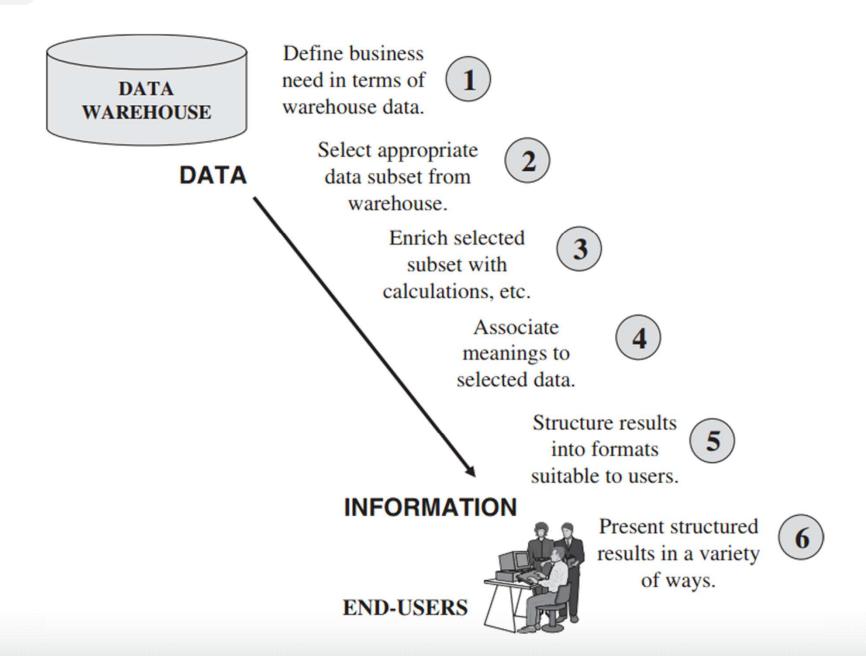




- Cuando hablamos acerca de la visualización de datos, es importante entender que cualquier representación de datos distinta de un simple texto es ya una visualización.
- La primera visualización fue una representación tabular de números, y las tablas siguen siendo una visualización muy potente, de hecho, la más común.
- Las tablas no siempre son la visualización más apropiada para todos los tipos de visualización de datos, podemos encontrarnos: gráficas de barras, columnas y líneas; scorecards e indicadores clave de rendimiento; mapas de red; o los gráficos personalizados dibujados por un ilustrador
- Se trata de técnicas de visualización que, cuando se usan adecuadamente, transmiten el significado de los datos mejor que una simple tabla.









- Una visualización es una forma de ver datos en un informe o a través de una vista que permita a alguien reunir ideas y realizar análisis para tomar mejores decisiones (empresariales).
- Si bien es tentador dejar que los desarrolladores y usuarios finales utilicen cualquiera de las muchas herramientas disponibles para ello, lo más recomendable es tener una herramienta de BI estándar.
- La herramienta de visualización debiera destacar aspectos importantes (una tendencia o un valor atípico).





- Entonces, el objetivo de la visualización de datos es presentar datos para proporcionar una comprensión más intuitiva de los datos o mostrarlos de manera que se puedan ver una gran cantidad de datos en una área más pequeña.
- Existe también la llamada visualización artística, la cual está diseñada para presentar un conjunto de datos de una forma que atraiga a las personas y por lo tanto genere interés en los datos que se presentan.



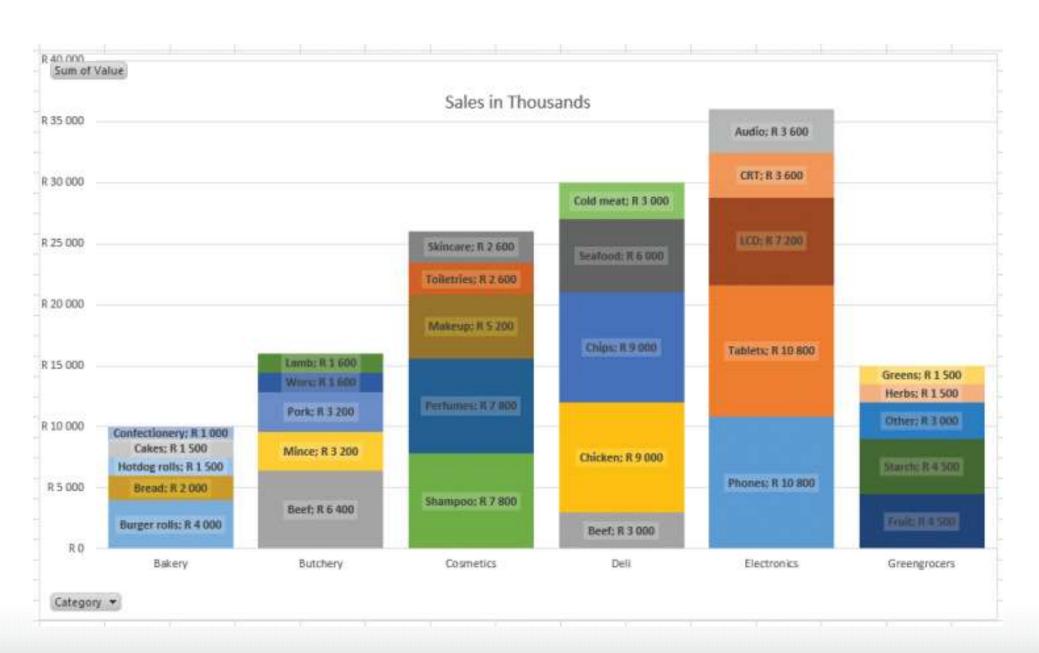


 Se pueden utilizar este estilo visualizaciones, pero enriquecidas a partir de gráficas basada en datos:



 Como se puede observar, se trata de visualizaciones "muy bonitas" pero que consumen mucho espacio. Se suelen utilizar otro tipo de gráficos (más tradicionales), desde la perspectiva de BI, para representar los mismos datos.







- Como se puede observar, no es una gráfica tan llamativa, sin embargo, muestra su utilidad muy rápidamente: las diferentes categorías se pueden comparar entre sí de un vistazo y permite comparaciones de los componentes de las categorías.
- Ahora bien, después de observar estas gráficas, será importante que tengamos en mente las siguientes preguntas cada vez que desarrollemos una visualización:
  - ☐ ¿La visualización contiene más datos que una tabla de tamaño equivalente?
  - ¿Los datos presentados en esta visualización son más fáciles de comprender que una tabla equivalente?
  - □ ¿Los elementos artísticos añaden significado?
  - ¿Se han añadido elementos gratuitos que no agregan significado o distraen del significado, como efectos 3D, transiciones animadas o imágenes?





- Una infografía es un gráfico utilizado para transmitir un mensaje que se conoce antes de la creación de la misma infografía:
  - Difiere de una visualización de datos, en el sentido de que la visualización es una ayuda gráfica utilizada para descubrir un mensaje oculto en los datos.
  - Una visualización de datos puede ser publicada como infografía, sin embargo, las visualizaciones son generalmente interactivas y dinámicas, por lo que si el mensaje cambia, la visualización de datos se actualiza automáticamente y muestra las nuevas cifras.
  - □ Por el contrario, una infografía diseñado específicamente en torno a un objetivo particular seguirá mostrando las mismas cifras, ya que se trata un gráfico plano.
  - La razón de la infografía plana y no interactiva se deriva del hecho de que son a menudo hechas a mano de una sola vez, y el nivel de esfuerzo implicado en la creación de una visualización y eventual actualización, representa una dificultad.
- Cuando se crea una visualización de datos, se habilita el descubrimiento de respuestas a través de los datos presentados, de manera que la presentación de los datos debe ser lo más rica posible.

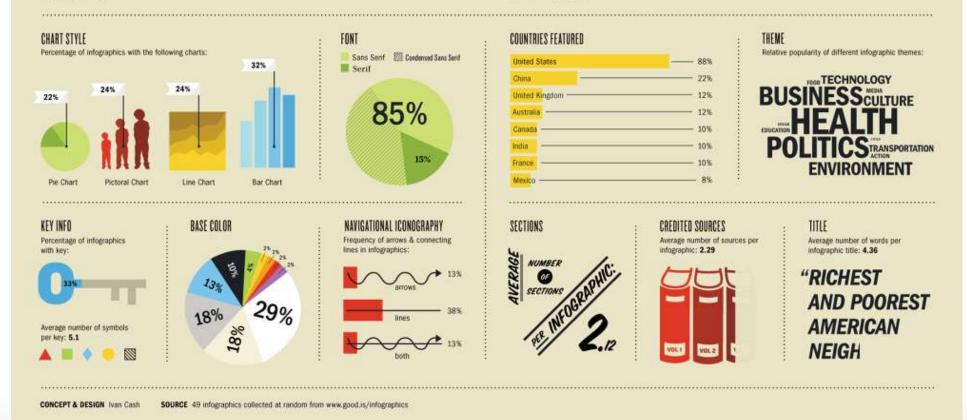


# INFOGRAPHIC INFOGRAPHICS

Data visualization is a popular new way of sharing research. Here is a look at some of the visual devices, informational elements, and general trends found in the modern day infographic.

### DESIGN

### CONTENT





### Identificando a los usuarios

	debe reconocer el número de personas afectadas en la anización cuando se utiliza una <b>herramienta de visualización</b> :
	Usuarios finales que ejecutan informes: Este es el grupo más obvio; requieren características avanzadas y accesibilidad de la herramienta, ya que necesitan sus datos de forma rápida y precisa.
	<b>Desarrolladores de informes:</b> crean los informes para los usuarios finales y necesitan una herramienta que les permita crear esos informes de manero <b>eficiente y precisa</b> .
	<b>Personal de apoyo:</b> Estas personas aseguran que los informes se recuperan en cualquier tipo de escenario de desastre.
	<b>Gestión:</b> Estas personas están preocupadas por el resultado final, incluyendo el número de usuarios, desarrolladores y personal necesario para apoyar la herramienta, y el costo de la herramienta en sí.

- Los aspectos que deben tomarse en cuenta cuando se selecciona una herramienta son: tiempo de desarrollo, tiempo de inactividad, número de recursos y precio.
  - Cuantas más herramientas se utilicen en la organización, lo más probable es que aumente todos estos factores; de manera que se recomienda reducir el número de herramientas de visualización a utilizarse en la organización.



### Selección de la herramienta

- Derivado de la necesidad de reducir el número de herramientas de visualización, los pasos básicos a seguir para realizar una elección son los siguientes:
  - 1. Entender los criterios importantes para la organización.
  - 2. Entrevistar a las partes interesadas (usuarios finales, desarrolladores, personal de apoyo y administración) para identificar las **prioridades y los criterios** adicionales.
  - 3. Identificar e investigar las herramientas de BI disponibles.
  - 4. Elaborar una matriz de evaluación que registre toda la información que se recolecte.
  - 5. Completar el proceso verificando los resultados
- El primer paso en el proceso de selección de la herramienta de visualización es entender cada criterio de selección e identificar la importancia de su departamento u organización.
- Los criterios de selección se dividen en dos categorías: capacidades técnicas y capacidades empresariales.
- Cada capacidad debe describirse en términos que tengan más sentido para su ambiente.



# Capacidades empresariales

- Aunque a menudo se pasa por alto, los requisitos de los usuarios empresariales son fundamentales.
- Elegir una herramienta equivocada puede hacer resultar en dificultades para que el usuario complete la tarea, lo que puede resultar en trabajo adicional para los usuarios o el departamento de TI.
- Algunos de los requisitos que la empresa pueden incluir:
  - Creación de informes estándar
  - Creación de informes ad-hoc
  - □ Capacidad de explorar y descargar datos
  - ☐ Facilidad general de uso
  - ☐ Funcionalidad (visualización) y funciones avanzadas (análisis)
  - Capacidades técnicas



# Capacidades técnicas

- Los requisitos a considerar desde el lado técnico incluyen:
  - □ Facilidad de desarrollo
  - Facilidad de mantenimiento
  - Precio
  - Seguridad
  - Personalización
  - Capacidad móvil
  - Integración de datos



### Reunir información necesaria

- Para reunir la información, se debe revisar la documentación existente y hablar con la gente de la organización, a fin de averiguar sus principales preocupaciones. No se debe asumir que se conocen las respuestas intuitivamente:
  - Identificar y leer cualquier documentación de estándares para software, servidores, etc. Si es posible, leer cualquier plan de largo plazo o documentos de estrategia para la empresa y el departamento de TI.
- El siguiente paso consiste en entrevistar a personas de cada uno de los grupos de interesados:
  - Usuarios finales: ¿Qué tan cómodo está al diseñar sus propios informes?, ¿qué tipos de gráficas y gráficos necesita?
  - Desarrolladores: ¿Cuál es su nivel de habilidades técnicas?, ¿cuánto tiempo tiene que dedicar a los informes?
  - Personal de apoyo: ¿Qué tipo de tiempo de inactividad es permisible para un informe?
  - ☐ Gestión: ¿Cuál es el presupuesto disponible para reportar las necesidades?



### Matriz de evaluación

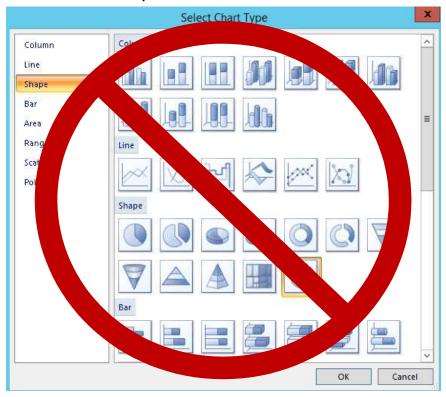
- Después de entender los criterios de selección de visualización y de conocer las diferentes herramientas disponibles, se debe crear y completar la matriz de evaluación.
- Se deben considerar las siguientes características: finalizar los criterios de selección, priorizar los criterios y evaluar las puntuaciones de los criterios.
- Se debe lograr que la matriz refleje las necesidades y metas de la organización.

Organizational Priority	16%	16%	18%	10%	10%	10%	100%		
		Business	S						
Visualization Tool	Ease of Use	Functionality	Advanced Features	Ease of Development	Ease of Maintenance	Price	Ability to Customize	Data Integration	Score
Reporting Services	1	1	1	1	1	1	1	1	1
Power View	3	3	3	3	3	3	3	3	3
Power Map	5	5	5	5	5	5	5	5	5



### Diseñando una visualización

 Diseñar la visualización va más allá de elegir un tipo de una lista que soporte una herramienta en particular:

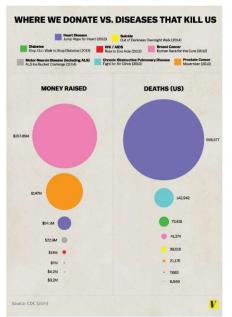


Se trata más bien de elegir la visualización correcta que permita transmitir el mensaje correcto. Una visualización errónea puede confundir el mensaje que está intentando enviar o incluso transmitir el mensaje incorrecto.

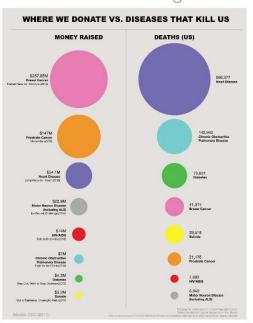


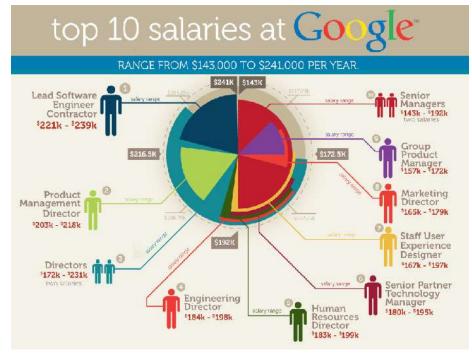
### ...Diseñando una visualización

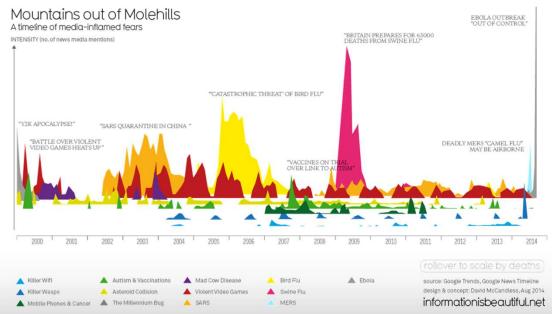
#### Original Design



#### Corrected Design



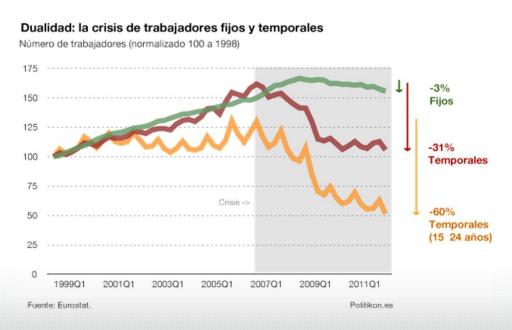






# Objetivo de la visualización

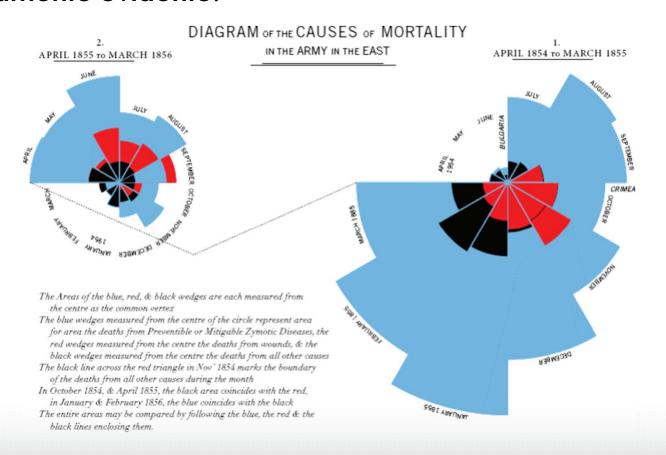
- El objetivo de una visualización es brindar la posibilidad de dar respuesta a preguntas, incluso a aquellas preguntas que no se sabía que deberíamos estar preguntando, sino hasta vimos el patrón de los datos en la visualización:
  - ¿Cuál es mi línea de productos más rentable?, ¿Quién es el mejor vendedor de la empresa?, ¿Cómo afecta la estacionalidad a mis niveles de existencias?, etc.
- El análisis temporal es el lugar donde las ventajas de la visualización se hacen fácilmente evidentes: una persona que lee dos líneas en un gráfico puede predecir dónde se cruzarán o divergirán las líneas con tan solo mirarlo.





# ...Objetivo de la visualización

Las visualizaciones responden a preguntas resaltando patrones y valores atípicos: cambiar el color de un elemento que es significativamente diferente de sus vecinos, el seguimiento de la relación de dos líneas a lo largo del tiempo, comparar dos columnas que están lado a lado, etc., son maneras de ilustrar gráficamente un patrón que puede no ser inmediatamente evidente.





# ...Objetivo de la visualización

- En resumen, podemos decir que las metas de la visualización son:
  - Presentar la mayor cantidad de datos posibles.
  - □ Ilustrar patrones que no son inmediatamente evidentes.
  - Responder a las preguntas planteadas por un espectador.
  - Comparar valores
  - Mostrar cambios con respecto al tiempo.
  - Extraer fácilmente los puntos de datos subyacentes utilizados.
  - Crear un mecanismo rápido para ver un valor.





# Tipos de visualizaciones

- Las visualizaciones estratégicas, tácticas y operativas reflejan una necesidad real de que los diferentes niveles de una organización tengan un tipo diferente de visión de los datos que fluyen dentro de la misma.
- La razón para tener una visión de los datos distinta es la de poder tomar una decisión y/o tomar una acción.
- La diferencia clave entre los distintos tipos de visualizaciones radica en el nivel de detalle requerido y la magnitud o importancia de la decisión que se toma.



# Estratégica



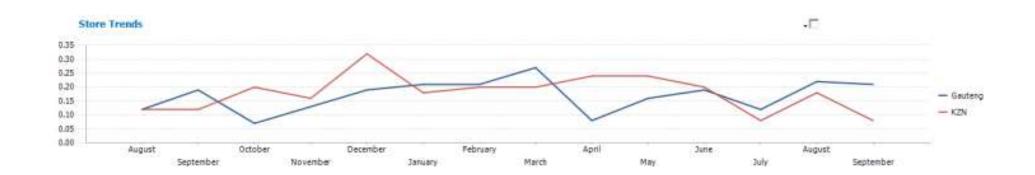
- Imaginemos que deseamos abrir una nueva tienda de conveniencia (decisión estratégica), esta decisión suele ser colaborativa (muchas personas contribuyendo a tomar la decisión):
  - ☐ Una empresa de investigación de mercado puede contratarse para descubrir y comparar la demografía de las posibles áreas en las que la tienda podría abrir.
  - ☐ El departamento de Finanzas evalúa el costo de hacer negocios en esas áreas y hace proyecciones de rentabilidad.
  - □ Los **gerentes de producto** proporcionan su conocimiento de **qué productos funcionaría bien en cada tienda**.
  - ☐ El **CEO** y otros ejecutivos necesitan saber, de un vistazo, que el negocio sigue funcionando de manera óptima, o al menos adecuadamente.
  - □ Los **ejecutivos** necesitan saber cuál de las regiones existentes, y qué tiendas dentro de esas regiones, están haciendo mal su trabajo para que una intervención pueda ocurrir de inmediato.
- La visión estratégica está en un nivel agregado, y ofrece la capacidad de perforar más detalles.



# ...Estratégica

Rating by Store					
	Rating	Target Rating			
Weighted Score	2.086	<u> </u>			
☐ Parent Organisation	2.086	۵			
<b>⊞</b> Gauteng	2.023	۵			
Store 1	2.112	4			
Store 2	1.934				
⊜ KZN	2.211				
Store 3	2.211	<b>A</b>			

	January		February		Harch		April		Hay		June	July		August.		Septemb	er
	Rating		Rating		Rating		Rating		Rating		Rating	Rating		Rating		Rating	
3AI	2.300	0	2.160	4	2.500	0	2.060	4	2,060	۵	1.620	1.847	٠	2.157	4	2.227	6
<b>□</b> CUSTOMER	0.200		0.207		0.247		0.133		0.187	٠	0.193	0.107		0.207		0.167	<
Pending applications	0.107		0.093	٠	0.120		0.067		0.093		0.093	0.053		0.107		0.080	4
Upgrade/New applications	0.047		0.067	٠	0.073		0.033		0.047		0.047	0.027		0.047		0.053	4
Value Added Services	0.047		0.047	٠	0.053		0.033		0.047		0.053	0.027		0.053		0.033	4
⊞ FINANCIAL	0.917		0.643		0.870		0.530		0.900		0.363	0.610		0.733		0.670	4
<b>⊞ INTERNAL PROCESSES</b>	0.990		1.070		1.030		1.257		0.760		1.030	0.820		0.937		0.970	4
<b>⊞</b> PEOPLE	0.283		0.240		0.353		0.140		0.213		0.233	0.310		0.280		0.220	4





# Operativa

- Una vista operativa, por el contrario, está en un nivel detallado:
  - □ Por ejemplo, en la misma organización minorista, un ejecutivo de crédito de una de las tiendas puede considerar una solicitud de un cliente para obtener una extensión de crédito para una compra específica y requerirá un informe que muestre el historial de pagos y la calificación crediticia del cliente.
  - ☐ El estilo de presentación de este tipo de **datos detallados** suele ser muy diferente de la vista de alto nivel utilizada en el nivel estratégico.
  - ☐ En contraste directo con las decisiones estratégicas, las **decisiones** operacionales se toman con frecuencia.
  - □ Individualmente, tienen muy poco efecto en la organización, pero en conjunto significan la diferencia entre el éxito y la bancarrota.



# ...Operativa

stodate	TII.						7/11
Project		(4)					
0.7							
		Values		-			
Duckag	7 Puckage Desc	Originodget	App Charges	Cor Dudget	Committed	Artimi Conf	HY
001	In Pit Crushing System Turnkey	93,511,631	91,340,593	84,852,225	32,763,440	4,986,290	84,832,225
002	Stock Yard Stacker & Reclaimer Systems	86,185,120	28,166,581	58,018,539	58,018,539	10,635,803	58,018,53
003	Samplers (Include design & layout of systems)	25,558,969	-3,118,471	22,440,497	20,087,785		21,577,116
004	Filter Press	92,159,506	17,318,179	09,477,685	82,465,618	- 35	09,447,74
005	Dust Extraction System	63,280,054	21,806,354	41,473,700	31,628,904	1,617,870	41,473,700
006	Flocculant Mixing and Dosing Plant	1,032,177	1,128,802	2,160,979	2,160,979		2,160,975
007	Ash Analysers	9,443,808	2,805,358	12,249,166	10,961,445	(2)	12,249,16
800	Dust Supression System (INCL DESIGN)	3,901,359	5,630,247	9,531,606	2,024,594	22	9,227,344
009	Control Room Architecture/Layout Design	3	9,630,870	9,630,870		4.5	9,630,870
010	33kV overhead Powerlines with Transformers	61,756,665	41,422,272	03,178,937	88,618,408	79,232,855	88,966,06
011	Steel Bassalt Lined Piping Supply & Fabrication - All plant areas- Area 1 and 2	14,631,253	-14,631,253	0		3-1	-
012	Large Cyclones	1,979,280	-450,968	1,528,312	1,180,193		1,528,312
013	Small Cyclones	5,361,924	1,710,354	7,072,278	6,282,698	90	7,072,278
014	Coarse Coal Sizing Screens	19,340,041	-5,625,610	13,714,431	10,558,721		11,490,07
015	Belt Feeders	13,643,549	5,619,718	20,263,267	19,858,002	-	20,263,263
016	Double Roll Crushers	20,231,535	1,109,906	21,341,441	12,815,262		14,027,77
017	Actuators	2,447,018	-1,051,387	1,395,631	493,107	4.0	891,38
018	Hydraulic Power Packs	15.026.594	-8,005,413	7,021,181	6,400,099	140	7,021,181
019	Compressors, including pressure vessels, driers (complete package)	10,901,221		5,252,481	4,172,180	23	5,252,48
020	Overhead Travelling Cranes	47,568,483		29,096,774	9,915,544		14,895,30
100	RESERVED AND THE PROPERTY OF T	7 120 110					



- Las decisiones tácticas son el punto medio entre las decisiones estratégicas y las operacionales.
- Las decisiones tácticas tienen que ver con la exploración de datos más que simplemente monitorear o evaluar una vista de detalle:
  - □ Por ejemplo, un gerente de producto puede necesitar decidir qué productos necesitan mantenerse como una acción adicional durante una temporada festiva.
  - □ El **gerente de producto** examina los datos de ventas del producto y su estacionalidad, y también podría examinar los datos de lo productos de moda y destacar cuáles tendrán fuertes campañas de mercadotécnia.
  - ☐ La clave aquí es **más interactividad y flexibilidad** que en las otras vistas.









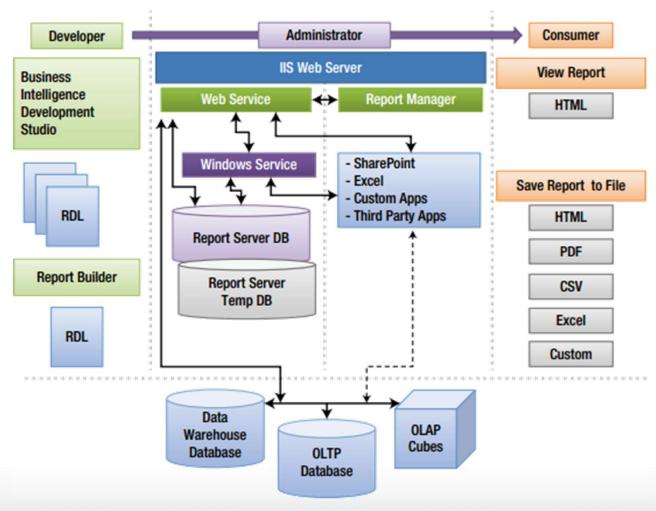
# MS Reporting Services

- Reporting Services es una herramienta que permite generar informes
  Web de manera nativa o en combinación con Microsoft SharePoint, a fin de trabajar en un entorno colaborativo.
- Permite a los usuarios generar rápida y fácilmente informes de bases de datos relacionales o multidimensionales.
  - ☐ Combina las capacidades de **Microsoft Visual Studio**, lo cual permite conectarse a bases de datos SQL y utilizar herramientas de **SSRS** para dar formato a los informes de forma versátil.
  - □ SSRS proporciona una herramienta llamada Report Builder que permite dar formato a los informes de SQL de menor complejidad.
  - ☐ Se administra a través de una interfaz Web, que puede ser utilizada para preparar y entregar una variedad de informes interactivos e impresos.
- SSRS compite con Crystal Reports y Pentaho, además de otras herramientas de BI.





SSRS es el más complejo de los servicios de inteligencia de negocios de SQL Server, ya que tiene muchos componentes diferentes, diseñados para trabajar en conjunto para crear una solución de informes completa.





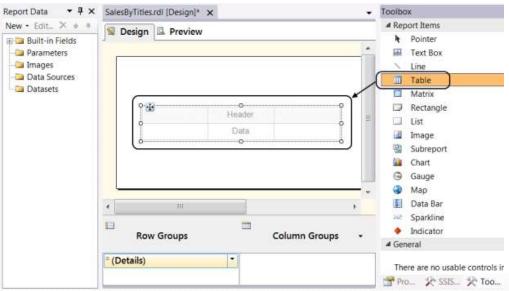
### Herramientas de desarrollo

- Los reportes se pueden crear utilizando herramientas como Visual Studio o Report Builder.
- Estas herramientas crean archivos XML utilizando el formato Report Definition Language (RDL). Este tipo de archivos se han propuesto como un estándar para la creación de informes, de manera que es posible utilizar aplicaciones de terceros para crear archivos RDL.
- Una vez creado el archivo RDL, se puede obtener una vista previa convirtiendo el código XML en una salida HTML.
- Los archivos RDL se cargan en el servicio web de Reporting Services, para que cuando un usuario final solicite un reporte, el código del archivo RDL se lee de la base de datos y se convierte en un formato legible (la salida por omisión es HTML, pero también puede estar en un PDF, CSV, Excel, etc.).



### ...Herramientas de desarrollo





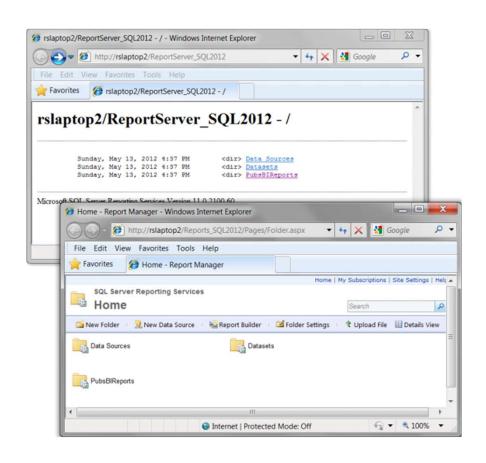


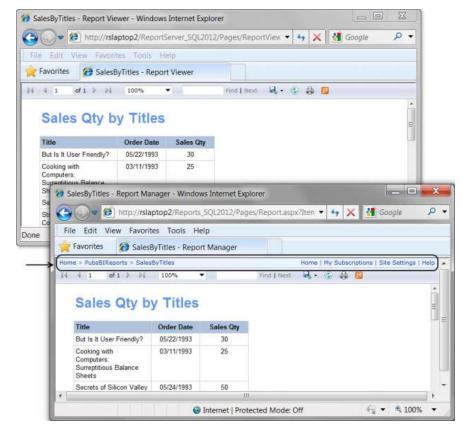
### Servicios de administración

- Después de que los archivos RDL son desarrollados y probados, se cargan en el servicio web de Reporting Services, que a su vez almacena los archivos RDL en una base de datos Microsoft SQL Server.
- El acceso a los informes es a través de un sitio web de ASP.NE, conocido como Administrador de informes, sin embrago, también se puede acceder a través de aplicaciones propias personalizadas.
- Los servicios administrativos de SSRS se dividen en funciones de administración de reportes y soporte al usuario final.
- Las funciones de administración consisten en el servicio web que administra la carga y descarga de archivos de informe, representación, y funciones de administración.
- Cada instalación de Reporting Services incluye dos aplicaciones web, los cuales se construyen utilizando ASP.NET. El primero es el servicio web que se mencionó anteriormente, y el segundo es una aplicación de usuario final para su visualización y administración de informes.



### ...Servicios de administración

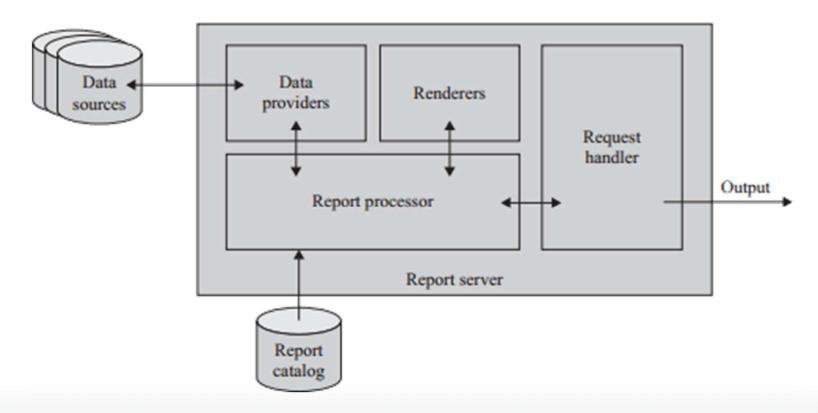






# Servidor de reportes

- Se trata del componente principal de Reporting Services y está implementado como un servicio web y como un servicio de Windows.
- El servicio web se compone de un conjunto de interfaces que las aplicaciones cliente utilizan para acceder a los informes en un servidor web. El servicio de Windows proporciona servicios de programación y entrega.





# ...Servidor de reportes

- El procesador de informes gestiona la ejecución de un informe.
  Recupera la definición del informe, que se realiza en el RDL, y determina lo que se necesita para el informe.
- El procesador de informes también recupera los datos que se utiliza para cargar el conjunto de datos. Después de eso, se selecciona un proveedor de datos que sabe cómo obtener información de la fuente de datos.
- La tarea del proveedor de datos es para conectarse a la fuente de datos, obtener la información para el informe, y devolverlo al procesador.
- Para generar el diseño, el procesador tiene que saber el formato del informe (HTML o PDF, por ejemplo). Los procesadores se utilizan para construir el formato correspondiente.
- El controlador de solicitudes recibe solicitudes de informes y las envía al procesador de informes. También ofrece el informe completo.



# Catálogo de reportes

- Contiene dos bases de datos que se utilizan para almacenar las definiciones de todos los informes existentes que pertenecen a un servidor de informes en particular, la cual incluye: nombres de informes, descripciones, información de conexión de origen de datos, información de credenciales, parámetros y propiedades de ejecución.
- También almacena la configuración de seguridad y la información relativa a la programación y los datos de entrega.
- Reporting Services utiliza dos bases de datos, la base de datos del servidor de informes y la base de datos temporal del servidor de informes, para separar el almacenamiento de datos persistentes de los requisitos de almacenamiento temporal. Por omisión, los nombres de bases de datos son reporterserver y reportServerTempDB.
- La primera se utiliza para almacenar Catálogo de informes, mientras que el segundo se utiliza como almacenamiento temporal. (La base de datos temporal del servidor de informes se utiliza para almacenar datos de sesión, informes en caché, y mesas de trabajo que se generan por el informe.)



