www.stratebi.com 91.788.34.10







BIG DATA – BUSINESS INTELLIGENCE



CONTENIDO

1.	INTRODUCCIÓN	3
R	REFERENCIAS	3
2.	LOGUEARSE EN DREMIO	5
3.	INTERFAZ DE USUARIO DREMIO	10
4.	ESPACIO DE TRABAJO	13
5.	CONECTAR FUENTES DE DATOS	15
6.	CREAR Y UTILIZAR CONJUNTOS DE DATOS	17
F	Ísicos	
۷	/IRTUALES	
7.	EXPLORAR Y SELECCIONAR DATOS	20
8.	CONECTARSE A POWER BI	23
9.	DREMIO Y VERTICA	24
10.	DREMIO VS DATABRICKS	26
11.	CONCLUSIONES	27



1. INTRODUCCIÓN

El acceso a los datos es una faceta crítica de cualquier entorno BI y cada vez más difícil de resolver. Esto es porque hoy en día hay muchas formas de datos estructurados y no estructurados y nuevos repositorios. **Dremio** busca revertir esta tendencia haciendo posible el autoservicio de datos para los profesionales de datos, apoyando así las operaciones de inteligencia empresarial.

Dremio fue creado para tener una capa unificada que permite a cualquier profesional de datos acceder a los datos que necesita en un solo jugar, independientemente de dónde se almacenen los datos subyacentes.

Dremio es una plataforma **Lakehouse** de código abierto, donde fusiona la facilidad de acceso y el soporte de las capacidades de análisis empresarial que se encuentran en las tecnologías "Data Warehouse", con la flexibilidad y el costo relativamente bajo que tienen las plataformas "Data Lake".

Dremio tiene 2 pilares que se detallan a continuación:

- Sonar es el motor de SQL que utiliza Dremio para proveer capacidades y rendimiento a
 nivel de datawarehouse en un datalake, además ofrece una experiencia de autoservicio
 sin necesidad de conocer las tablas físicas previamente optimizadas y agregadas. La
 desventaja de Sonar es que se debe tener una cierta experiencia en SQL, lo que hace
 que los analistas funcionales deban tener esa habilidad. Para mitigar riesgos, Dremio
 tiene conectores con herramientas de análisis de datos como power Bl y Tableu.
- Arctic es el área de trabajo de Dremio, está pensado para que trabaje de forma similar a la herramienta GIT; aquí se podrán optimizar, transformar, realizar modelos de aprendizaje automático, etc, sin afectar a las cargas de trabajo puestas en producción. Una vez realizado se podrán fusionar los cambios.

Dremio se puede desplegar en Cloud o Enterprise. Ofrece una versión gratuita (Cloud) que es adecuada para la mayoría de las pequeñas empresas. En caso de necesitar más funciones de seguridad y control de datos **Dremio** ofrece una versión de pago.

Referencias

Página principal web: https://www.dremio.com/

Documentación oficial de aprendizaje: https://docs.dremio.com/



Cursos https://university.dremio.com/

A continuación, mostraremos una demo en la cual nos conectaremos a una fuente de datos, navegaremos por la interfaz de usuario, crear data y conectarse a Power BI.



2. LOGUEARSE EN DREMIO

Tenemos distintas formas de presentar la demo de **Dremio** y ver sus funcionalidades, la primera es tener una nube de prueba con AWS, ya que **Dremio** se conecta muy bien y por defecto a esta nube.

Para poder hacer esta conexión tendremos que ingresar a <u>https://www.dremio.com/</u> y darle click a get Started.



Y nos redireccionará a



← → C @ app.eu.dremio.doud/signup?_ga	=2.87322460.512726873.1650877484-2037439	9.1650877484		塑 炙 ☆
v				
	Sign up via Email		Or sign up using one of these services	
	First Name	Last Name	G Sign up with Coogle	
	Work Email		Sign up with Microsoft	
	Password	Confirm Password	Sign up with GitHub	
	By clicking sign up, you agree to the Dremic	Terms of Service and Privacy Policy.		
	Already have an account? Log in.			

Procedemos a crear una cuenta o iniciar Sesión.

Una vez iniciada la sesión, nos aparecerá para crear un proyecto en Dremio.

Organization Name				
stratebi	Create your	Dremio project		
Project Name My First Project	A project is the analysis. An org and other resou	logical entity for isolating c anization contains one or n rces across projects.	ompute and data needed by a tea nore projects and helps you mana	am for da Ige users,
AWS Region EU West (Paris) eu-west-3	Enter an organiz if needed.	tation and project name to	get started. You can change thes	e after si
> Advanced Configuration	By default Drem the other for run to adjust to you	ilo creates two engines in y nning queries. Dremio autor r workload.	rour cloud account: one for previe matically starts, scales and stops	wing data these enj
	You can specify also change eng	the initial size of these eng gine sizes or add additional	gines under Advanced Configurat engines after you complete sign	ion. You c up.
	also change eng	gine sizes or add additional	engines after you complete sign	up.

Rellenamos los datos y hacemos click en Next.



Set up your cloud connection					
Automatic Automatic Click "launch Stack" to open your AWS console in a n stalilation steps. Once the deployment is successfu tetting up Dremio. Launch Stack Don't have admin access to your company The state intervented it is accessed to access to your company	Manual Manual w tab. You will be guided through the ily completed, you'll be able to continue *'s AWS account?	Connect Dre Your project in I easily connect ID creates cross-ar work with your of Click "Launch St creating your Ch	mio to your Cloud Dremio needs to be linked to remio to your cloud using cocount IAM roles and an S2 data. tack" to open your AWS cor oudFormation stack.	o your cloud account for data ana a CloudFormation template. The t bucket that Dremio uses to creat sole, provide the necessary input	lysis. You can emplate e engines and s and start
https://eu-west-3.console.aws.amazon.com/cloudfo	ormation/home?region=eu-west-3#/stac				

En esta pestaña es donde podremos conectar nuestro Cloud con **Dremio**. Una vez conectado procedemos a darle permiso a los usuarios.

Otra de las formas como podemos probar **Dremio** es a través de una plataforma de software que permite crear y probar implementaciones en esta ocasión hablamos de **Docker** <u>https://docs.docker.com/desktop/</u>, la única limitante es que tendrás que contar con un equipo Mac o Linux.

Se puede descargar una imagen de **Dremio** y el proceso para desplegarla en **Docker** a través del portal https://hub.docker.com/r/dremio/dremio-oss/

Para instalar la imagen de **Dremio** a través de esta línea de comando "*docker pull dremio/dremio-oss*".

Para desplegar la imagen se hace uso de la siguiente línea de comando

"docker run -p 9047:9047 -p 31010:31010 -p 45678:45678 dremio/dremio-oss"

Una vez desplegada la imagen de **Dremio** en **Docker** procedemos a iniciarla.

•	docker		٠	۲	B Sign in			×
۲	Containers / Apps							
•	Images	Containers / Apps 🖝 coords						
-	Volumes	Q Search				So	rt by 🗸	
۲	Dev Environments PREVIEW	naughty_wiles dremio/dremio RUNNING PORT: 31010						



Una vez iniciado, procedemos a ingresar en nuestro explorador predeterminado la dirección localhost:9047 y aparecerá una pestaña para la creación de una cuenta de administración.

Iocalhost:9047/signup				
	gis	We are excited t The first thing ye administrator ac	o have you on board! ou need to do is set up an count.	
	Create Admin Account			
	First Name		Last Name	
	Username		Email	
	Password			
	Confirm Password			
	Next			Privacy

Después de ello se nos presenta la interfaz de usuario

$\leftarrow \rightarrow$	C (i) localhost:9047		॰ 🖄 🛧
۴	$Q_{\!\!c}$ Search Spaces and Datasets		
⊞	Datasets	@leonardo_stratebi Q Search	+ 🗀 🗸
	🖒 leonardo_stratebi	Name ~ Jobs	
>_	Spaces (0) »		
I	You do not have any spaces.		
	Data Lakes (1) »	No Items	
ي ج ا	Samples 0 Add your own data lake: Add Data Lake		



Podemos a través de "University Dremio" la posibilidad de tener una instancia de **Dremio** para probar y hacer cursos. Tener en cuenta que una vez iniciada la instancia, esta dura solo 48 horas. <u>https://university.dremio.com/</u>



3. **INTERFAZ DE USUARIO DREMIO**

Cuando se inicia sesión en **Dremio** aparecerá una página en la cual se tendrá, en la parte izquierda, los espacios de trabajo creados, las conexiones a base de datos y data lake. En el medio se tendrá el resumen de las estructuras creadas.

\leftarrow	C 🛈 localhost:9047				• 6 \$
۴	Q Search Spaces and Datasets				
Ħ	Datasets	@leonardo_stratebi		Q. Search	+ 🗖 🗸
	🖧 leonardo_stratebi	Name v		Jobs	
>	Spaces (0) »				
	You do not have any spaces.				
	Data Lakes (1) »		No Items		
<i>.</i> ???	Samples 0				
ين ?	Add your own data lake: Add Data Lake				

Para manejar Dremio debemos tener en cuenta estos símbolos y sus significados.

Origen: un origen de datos conectado a **Dremio**, como un depósito S3 o ADLS, o un RDBMS.

Conjunto de datos físicos (PDS): un conjunto de datos de origen, como un grupo de archivos en S3 o una tabla RDBMS.

Conjunto de datos virtual (VDS): similar a una vista RDBMS, pero con más funciones, incluido el linaje.

Reflexión: similar a una vista materializada, acelera las consultas de forma transparente para los usuarios.

Espacio: una ubicación que organiza los VDS compartidos con otros usuarios.



😡 Espacio doméstico: un área de trabajo privada para su propio PDS y VDS.

Carpeta: un subdirectorio para organizar elementos en un espacio o fuente.

Otra Interfaz de usuario a destacar es la relativa a la de hacer Querys.

🐔 dremio 🔳 Datasets 🔳 Jobs		+ New Query	⑦ 🌣 fundamenta
New Query		(m) (*	···· · 🖺 · Preview 🧭 Run
SQL Editor			Datasets Functions
1		Browse Search	
		🕨 🎧 @fundamentals	
		🕨 🧐 Samples	
		Contact: Samples	
	lumn filter 0 fields		0

Se puede observar que en la parte izquierda el editor para hacer las consultas a las fuentes de datos cargadas a la derecha.

Otra interfaz importante es la de las capas semánticas. **Dremio** permite organizar los datos en capas semánticas para maximizar el mantenimiento y reutilización de los datos. **Dremio** aconseja a tener 3 capas de datos,

- Capa de Ensayo: es la capa más cercana a los PDS, y es donde se debería realizar las transformaciones y limpieza de datos.
- Capa Empresarial: esta capa debe contener la lógica de Negocio.
- Capa Aplicación: se enfoca en la agregación y presentación de los datos.



ኛ dremio 💷 🛛	atasets 🗮 Jobs 🛛 🔍 Searc	h Catalog		+ New Query		0 .
My First Spaces	Data 📃 Catalog	Graph	C Reflections		Gul) \$	₩
Sources (1)	Parents (1)	Q		Focus "Incidents"		
(İ)	SF_incidents2016 json Samples samples dremio.com Jobs: 3 > Descendants: 1 Fields: 14	n		Jobs: 1 b Descendants: 0 Fields: 13	×	
				Abc IncidntNum Abc Category Abc Descript Abc DayOfWeek Abc Date Abc Time Abc PdDistrict Abc Address Abc Y Abc Y Abc Y Abc PdId		



4. ESPACIO DE TRABAJO

Los espacios de trabajo son ubicaciones compartidas para guardar conjuntos de datos virtuales los cuales permiten agrupar conjuntos de datos por un tema común, como un proyecto o una región geográfica.

Cada espacio se puede configurar para compartir y otros privilegios. Los usuarios no verán espacios para los que no tienen autorización.

Por defecto en **Dremio** cada usuario tiene un área de trabajo privada que puede contener conjuntos de datos tanto físicos como virtuales. Los datos se pueden conservar en el espacio de inicio hasta que estén listos para compartirse moviéndolos a un espacio compartido.

Para crear espacios compartidos:

1. Seleccione el ^O o "Add Space" en la sección "Espacios".



2. Nombra tu nuevo espacio, asigna privilegios y click en "Save".



Add Space							×
Name							
My First Spaces							
Privileges							
Add User/Role							
Username	dd to Privileges						
Users	Select	Alter	Alter Reflection	View Reflection	Manage Grants	Modify	
All Users					5		
						Cancel	Save

3. Crear carpetas en el espacio con el símbolo de

Add Folder	×
Add a Folder to the Space	2
Dremio folders work just like your d	lesktop ones; they add another layer of organization within a Space.
Dremio folders work just like your d Folder Name	lesktop ones; they add another layer of organization within a Space.

Cancel	Save



5. **CONECTAR FUENTES DE DATOS**

En **Dremio** te puedes conectar a data lakes y a fuentes de datos externas para después crear un conjunto de datos Físico (PDS) o virtual (VDS).

Para conectarte a un data lake basta con darle click a "Add Data Lake" en el menú de la parte izquierda de la pantalla.

ኛ dremio 🖽 Datasets	i 🗐 Jo	bs Q Search Catalog	+ New Query	? ‡
Datasets		Samples."samples.dremio.com"		Q Search
fundamentals		Name 🗸		
Spaces (0) »	θ	Dremio University		
		NYC-taxi-trips		
You do not have any spaces.		SF weather 2018-2019.csv		
Add Space		SF_incidents2016.json		
		tpcds_sf1000		
		zip_lookup.csv		
		🗋 zips.json		
Data Lakes (1) »	Ο			
Samples				
Add your own data lake:				
Add Data Lake				
L				

Luego se abrirá una pestaña emergente con las opciones a conectar,

Table Sto	pres	
÷	Amazon Glue Catalog	
A.	Hive 2.x	
A.	Hive 3.x	
File Store	25	
	Amazon S3	
	Azure Data Lake Storage Gen1	



Puedes conectarte a diferentes orígenes de datos, para hacerlo se tendrá que hacer click al símbolo de [©] al lado de External Source.

ኛ dremio 🖽 Datasets	i≣ Job	s Q Search Catalog	+	New Query
Datasets	S	Samples		Q
合 fundamentals		Name 🗸		
Spaces (0) »	0	samples.dremio.com		
You do not have any spaces.				
Data Lakes (1) »	θ			
Samples				
Add your own data lake: Add Data Lake				
> External Sources (0) »	0			

Se abrirá una ventana emergente con las diferentes tecnologías a la cual te puedes conectar.

Add Exte	ernal Sourc	ce	×
Add External Source Image: Amazon Elasticsearch Service Image: Amazon Redshift Image: Amazon Redshift Image: Elasticsearch Image: Blasticsearch Image: Blasticsea	Í		
		Amazon Elasticsearch Service	
		Amazon Redshift	
		Elasticsearch	
		Microsoft SQL Server	
		MongoDB	
	MySQL	MySQL	
	ORACLE	Oracle	



6. **CREAR Y UTILIZAR CONJUNTOS DE DATOS**

Físicos

Como ya sabemos se pueden cargar data lakes y base de datos externas, pero **Dremio** también permite cargar archivos físicos como CSV, Excel, Json, Parquet y ORC, mediante el botón "Nube" que se encuentra en la parte derecha de la pantalla.

ኛ dremio 🏾 🖽 Datasets 🗧	≣ Jobs Q Search Catalog	+ New Query	
Datasets	@fundamentals	Q Search	+
	Name ~		Jobs Action
Spaces (0) » You do not have any spaces. Add Space	•		
Data Lakes (1) »	•	No Items	
Add your own data lake: Add Data Lake			
> External Sources (0) »	0		

Después de haber realizado el click se mostrará una pantalla emergente para cargar el archivo seleccionado en Dremio.

Add File: Browse for	dd File: Browse for File (Step 1 of 2)					
	Drop a local file here, or browse. 100_Sales_Records.csv 12.4KB					
Name 100_Sales_Records		Cancel	Next			



A continuación, seleccionamos el archivo y le damos a "Next".

Add File: Set Format (St	ep 2 of 2)							×
Format Text (delimited) ~								
Field Delimiter	*	1	Quote Double Quote	~		Comment Number Sign	~	#
Line Delimiter CRLF - Windows	~	\r\n	Escape Double Quote	۷	н	Options Extract Field Names Skip First Line Trim Field Names		
						Back	Sa	ve

Configuramos y le damos a "Save".

Una vez realizado estos pasos aparecerá el conjunto de datos físicos (PDS) en la parte central de la pantalla.

ኛ dremio 🏾 Datasets	≡ Jobs Q Search Catalog	+ New Query	\$
Datasets	@fundamentals	Q Search	
fundamentals	Name \vee		Jol
Spaces (0) » You do not have any spaces.	Interpretation of the second secon		
Add Space			

Virtuales

Los conjuntos de datos virtuales (VDS) se crean sobre uno o más conjuntos de datos físicos u otros conjuntos de datos virtuales. Los conjuntos de datos virtuales tienen las siguientes características importantes:

- Los conjuntos de datos virtuales se definen mediante una instrucción SQL que puede incluir filtros, uniones, uniones, conversiones de tipo, análisis, columnas derivadas, etc. su proceso ETL.
- Los conjuntos de datos virtuales se guardan en su espacio doméstico o en un espacio compartido.



Pare crear un VDS nos dirigimos a la ventana de "New Query" y procedemos con la siguiente instrucción "*CREATE VDS <virtual-dataset-path> AS <sql-query>*"

ኛ dremio 📱		s Q Search Catalog	+ New Query	\$	fundamentals
New Query	Gremio Datasets Jo New Query Image: Data G SQL Editor Image: REATE VDS "My First Spaces Select * from Samples."samples.dremio.com	Reflections		Preview	🧭 Run
▼ SQL Editor				Datasets	s Functions
1 EREATE VDS " 2 select * fro 3 Samples."sam	My First Spaces" m ples.dremio.com"	"Incidents" as "SF_incidents2016.json"	Browse Search		

Para ver el VDS creado procedemos a ingresar al espacio donde se haya guardado en este caso "My First Spaces".

ኛ dremio	🖽 Datasets		Q Search Catalog	+ New Query		⊘¢	ŧ
Datasets		My	First Spaces		Q Search		
🎧 fundamentals		N	ame 🗸				J
Spaces (1) »		0	Incidents				
My First Spaces		1 2	My first Folder				



7. **EXPLORAR Y SELECCIONAR DATOS**

Dremio proporciona herramienta de limpieza de datos sin necesidad de utilizar herramientas ETL o scripts de SQL, todo se hace desde el asistente gráfico que guía cada paso requerido por cada conversión.

En este caso de ejemplo tenemos nuestro PDS previamente cargado y observamos que la fecha "Order Date" está en un formato de tipo string. Mediante el asistente gráfico podemos convertir este campo a formato fecha haciendo click en "Order Date" seguido de "Convert Data Type".

ኛ New Query - Dre	rmio 🗙 🎸	@fundamentals.100_Sales_Re-	or × +				~ -	σ	×
← → C 🔒	9047-0-b28768d5.univ	versity-labs.dremio.com/ho	me/%40fundamentals/"10	00_Sales_Records"?jobId	=1d959386-c2e3-4fc3-fc	d2c-c7367dc19b00&tipV	ler Q 🖻 🛣	•	•
ኛ dremio 🔳 Data	isets 📃 Jobs 🔍 Search (Citalog.	+ New Query				0 💠	fundam	entals ~
00_Sales_Records Orbitamentals	Data 💽 Cata	ilog 🔓 Graph 🌀 Refle	tions			(a) (\$)	~ 🖺 - Preview	(7) B	un i
* SQL Editor							Data	ets Functi	ons
1 SELECT * FROM "10	0_Sales_Records"							-	
					Sort Ascending Sort Descending Sort Multiple				
					Rename Drop Convert Data Type				
					Group By				
					Convert Case Trim Whitespace Calculated Field				
					Extract Text Replace Text Split				
📑 Add Field 🛛 🗐 Group By	D Join Q. Column filter	14 fields			Keep Only Exclude		Job: Run Records: 100	Time: <1s [2
And Region	- Ann Country	- _{Alt} Item Type	$_{\rm Abc}$ Sales Channel $_{\odot}$ \sim	📖 Order Priority 🗠	Ale: Order Date 🗢	Also Order ID 🗸 🗸	ALL Ship Date	10 AD1 1	211
Australia and Oceania	Tuvalu	Baby Food	Offline	н	5/28/2010	669165933	6/27/2010	9925	
Central America and the Ca	irshi Grenada	Cereal	Coline	c	8/22/2012	963881488	9/15/2012	2804	
Europe	Russia	Office Supplies	Offline	L	5/2/2014	341417157	5/8/2014	1779	
Sub-Saharan Africa	Sao Tome and Principe	Fruits	Online	c	6/20/2014	514321792	7/5/2014	8102	
Sub-Saharan Africa	Rwanda	Office Supplies	Offline	L	2/1/2013	115456712	2/6/2013	5862	
Australia and Oceania	Solomon Islands	Baby Food	Online	c	2/4/2015	547995746	2/21/2015	2974	
Sub-Saharan Africa	Angola	Household	Offline	н	4/23/2011	135425221	4/27/2011	4187	
Sub-Saharan Africa	Burkina Faso	Vegetables	Online	н	7/17/2012	871543967	7/27/2012	8002	

Una vez realizado este paso se desplegará el asistente de **Dremio** para convertir ese campo.



30)	_								_
Change Data Type									×
Convert to									
🛅 Date 🗸 🗸									
Format			Formatting Options						
YYYY-MM-DD			YYYY: Four digits of y	/ear					
MM.DD.YY			MM: Month (1-12)	yoor					
DD/MM/YY			MONTH: Full month r	ame (March, October)					
MON DD YYYY			Learn more	31/					
Custom:									
mm/dd/aa									
Action for Non-matching Values									
Replace values with null Delate seconds									
O Delete records									
New Field Name	Options								
Order Date	Drop Source	Field (Order Date)							
Apply Preview	Cancel	A Result based o	n sample dataset						
Rt: Order Priority	Aix Order	Date	Abc Order ID	Air Ship Date	Abc Units Sold	Abc Unit Price	Ab: Unit Cost	Atc Total Revenue	
н	5/28/2010		669165933	6/27/2018	9925	255.28	159.42	2533654.00	1
c	8/22/2012		963881480	9/15/2012	2804	205.70	117.11	576782.88	
L	5/2/2014		341417157	5/8/2014	1779	651.21	524.95	1158502.59	
¢	6/20/2014		514321792	7/5/2014	8102	9.33	6.92	75591.60	
L	2/1/2013		115456712	2/6/2013	5052	651.21	524.95	3295425.02	
¢	2/4/2015		547995746	2/21/2015	2974	255.28	159.42	759282.72	

Hacemos Click en "Apply" y a continuación observamos una consulta SQL donde nos muestra cómo sería la consulta con el dato transformado.

New Query	Data	😵 Reflections						Preview	🔨 Run
▼ SQL Editor								Datasets	Functions
1 SEECT Region, Revenue", "Tot 2 FROM "#DATE 3 WHERE is_DATE	. Country, :al Cost", ntals"."16 :("Order Da	"Item Type", "Sales " "Total Profit" O Sales Records" te", "MUDD/TYYY")	Channel", "Order Priority", TO	_DATE("Order Date", "H4/OC	γγγγγ, θ) AS "Order Date"	", "Order ID", "Ship Data",	"Units Sold", "Unit Price	", "Unit Cost", "Total	
								Context: @fundamer	reals ()
📑 Add Field 😑 Group I	By ① Join	Q, Column filter	14 fields				Job	Preview Records: 100 Time	: <1s 👧
Abc Sales Channel	~)	be Order Priority	✓ □ Order Date ✓	Abc Order ID 🗸 🗸	Abc Ship Date 🔍	Abo Units Sold ~	$_{\rm Abc}$ Unit Price $_{\rm V}$	Abc Unit Cost	Abo To
Offline	н		2010-05-28	669165933	6/27/2010	9925	255.28	159.42	253365
Online	c		2012-08-22	963881480	9/15/2012	2804	205.70	117.11	576782
Offline	L		2014-05-02	341417157	5/8/2014	1779	651.21	524.96	115850
Online	c		2014-05-20	514321792	7/5/2014	8102	9.33	6.92	75591.
Offline	L		2013-02-01	115456712	2/6/2013	5062	651.21	524.95	329642
Online	c		2015-02-04	547995746	2/21/2015	2974	255.28	159.42	759202
Offline	м		2011-04-23	135425221	4/27/2011	4187	668.27	502.54	279884

Dremio permite realizar consultas SQL directamente en sus fuentes de datos (VDS y PDS), incluso cuando esas fuentes de datos normalmente no permitirían este tipo de operaciones, es decir, Elasticsearch o MongoDB.

Para acceder a la fuente de datos desde cualquier lugar de la interfaz de usuario, simplemente seleccione el icono "New Query".

ኛ dremio 🖽 Datasets 📰	lobs Q Search Catalog	+ New Query
Datasets	My First Spaces	
☆ fundamentals	Name ~	
Spaces (1) »	Incidents	
D My First Spaces 3 🖄	My first Folder	



El editor de consultas proporciona:

- 1. Escribir consultas SQL contra sus fuentes de datos o conjuntos de datos.
- 2. Explora datos.
- 3. Descarga conjunto de datos en diferentes formatos.
- 4. Conéctese con 1 click a sus herramientas de BI.
- 5. Guarde el conjunto de datos (si se realizaron cambios).
- 6. Obtener una vista previa o ejecutar una consulta.

						5	
<i>F</i>dremio Datasets J	lobs Q Search Catalog	> New Query				Help 🗸 Admin	dremio 🗸
100_Sales_Records @dremio	Data 🗄 Catalog 🕫	Graph 🥠 Reflections			<i>i</i>) 0	Preview	Run
▼ SQL Editor					Dataset Settings	Datasets Fu	tions
1 SELECT * FROM **100_Sa 1 Add Field 1 Group By	Join Q Column filter 14	fields	= 2		Download JSON CSV	Context: @dn w Status: Completed Time: <1	s 🚉
Abc Region ~	Abc Country ~	Abc Item Type 🗸 🗸	Abc Sales Channel	✓ Abc Orde	Parquet	Abc Order Date 🗸 🗸	Abc Orc
Australia and Oceania	Tuvalu	Baby Food	Offline	н	Analyze With: 🕑 Driver	5/ 8/2010 4	66916
Central America and the Carib	l Grenada	Cereal	Online	с	Tableau	3/22/2012	96388
Europe	Russia	Office Supplies	Offline	L	Power BI	5/2/2014	34141
Sub-Saharan Africa	Sao Tome and Principe	Fruits	Online	с	Qlik Sense	5/20/2014	51432
Sub-Saharan Africa	Rwanda	Office Supplies	Offline	L		2/1/2013	11545
Australia and Oceania	Solomon Islands	Baby Food	Online	с		2/4/2015	54799

Imagen realizada por Dremio Academy



8. **CONECTARSE A POWER BI**

Dremio por defecto tiene una conexión nativa e integradora a Power BI, solo tendremos que darle al icono de Power BI en la sección de querys de **Dremio**.

ኛ dremio 🖩 Dataset	s 📃 Jobs 🛛 🔍 Search Catal	og	+ New Query				? 🌣 🕴 f	undamenta
Incidents My First Spaces	Data 💽 Catalog	Graph 🎸 Reflections				a	Preview	🐔 Run
* SQL Editor Edit Original SQL							Datasets	Functions
1 SELECT * FROM Incide	nts							
							Context: My First Sp	aces D
📋 Add Field 😑 Group By 🕧	Join Q. Column filter 13	3 fields				oL	b: Preview Records: 10.000 Tin	ne: 1s 📋
Abo IncidntNum ~	Abs Category V	Abo Descript \sim	Abs DayOfNeek v	Abc Date 🗸 🗸	$_{\rm Abc}$ Time \sim	$_{\rm Abc}$ PdDistrict \sim	$_{\rm Abc}$ Resolution \sim	Abc Addr
120058272	WEAPON LAWS	POSS OF PROHIBITED WEAPON	Friday	2016-01-29	11:00	SOUTHERN	ARREST, BOOKED	S00 Bloc
120058272	WEAPON LAWS	FIREARM, LOADED, IN VEHICLE,	Friday	2016-01-29	11:00	SOUTHERN	ARREST, BOOKED	800 Bloc
141059263	WARRANTS	WARRANT ARREST	Monday	2016-04-25	14:59	BAYVIEW	ARREST, BOCKED	KEITH ST
160013562	NON-CRIMINAL	LOST PROPERTY	Tuesday	2016-01-05	23:50	TENDERLOIN	NONE	JONES ST
160002740	NON-CRIMINAL	LOST PROPERTY	Friday	2016-01-01	00:30	MISSION	NONE	16TH ST
160002869	ASSAULT	BATTERY	Friday	2016-01-01	21:35	NORTHERN	NONE	1700 Blo

Se abrirá Power BI y a continuación se ingresarán las credenciales de conexión de Dremio.

	2 6	1				Sin título	o - Power BI De	esktop				
Archiv	vo	Inicio	Insertar	Model	ado Ver	Ayuda	Herramier	itas externas				
Pegar	¥ È≞ ∛	Obtener L datos v	ibro de Conjur Excel	itos de datos Power Bl	de SQL Especific Server datos	© ar Dataverse	e Orígenes recientes v	Transformar Actualizar datos v	Nuevo objete visual	O Cuadro de texto	Más objetos visuales 🗸	Nueva Medid medida rápida
Portapape	les				Datos			Consultas		Insertar		Cálculos
000												
▦									Dremio			\times
唱					Username/Pass	word	~ 90)47-0-b28768d5.	university	-labs.dr	emio.com	
					Microsoft Acco	unt	Nombr Contra	e de usuario seña				
											Conectar	Cancelar

Previamente habrá que descargarse el controlador ODBC para crear la conexión entre Power BI y Dremio. El enlace del controlador <u>https://www.dremio.com/drivers/odbc/</u>



9. **DREMIO Y VERTICA**

A la fecha de realizado este documento, **Dremio** no trae por defecto un instalador con Vertica, se tendrá que descargar un controlador en la página oficial de **Dremio** y seguir los pasos para su instalación <u>https://www.dremio.com/hub/vertica/</u>.

Una vez instalado el controlador la conexión con Vertica e la siguiente:

- Click en en "External Sources" ubicado en la parte inferior izquierda de la pantalla.
- Click en Vertica una vez desplegado la ventana emergente.

Add Externa	al Sourc	se la	×
		Microsoft SQL Server	•
		MongoDB	
(MySQL	MySQL	
	RACLE	Oracle	
	(F)	PostgreSQL	
	VERTICA	vertica	

• Ingresar los datos de la conexión y seguidamente guardar.



New vertica Source					×
General Advanced Options	vertica So	urce			
Reflection Refresh Metadata	VERTICA	Name			
	Connection	1			- 1
	Hostname e.g.123.123.12	23.123	Port 5433		
	Database				
	Encrypt co	onnection			
	Username			Cancel Save	



10. **DREMIO VS DATABRICKS**

Dremio Sonar vs. Databricks SQL Feature Comparison

General	Dremio Sonar	Databricks SQL
Environments	Cloud and on-premises	Cloud only
Use case focus	Supports all BI workloads, from ad hoc to mission-critical against the data lake and relational databases.	Focused on providing ad-hoc SQL against the data lake, mainly serving only ad hoc queries in practice.
Bl experience	Dremio Sonar	Databricks SQL
BI acceleration	Transparent reflections to accelerate queries. These are maintained and applied by Dremio Sonar across queries transparent to BI clients.	Doesn't natively offer any similar functionality to accelerate BI workloads.

Data Management	Dremio Sonar	Databricks SQL	
ACID transactions	Uses Apache Iceberg for ACID. • No lock-in, Iceberg is open and community- driven.	Uses Delta Lake for ACID. • Spark is the only engine that can write safely to Delta Lake.	
Multi-statement transactions	Multi-statement transactions supported with Project Nessie.	Delta Lake supports only single-statement transactions. Multi-statement transactions are not supported.	
Data lineage	Yes	No	
Data lakehouse management	A separate data layer with a Git-like experience for tables and views via integration with Project Nessie to branch, tag, and time travel datasets all while automatically optimizing the files to ensure high- performance analytics.	Databricks does not natively offer anything similar.	
Performance	Dremio Sonar	Databricks SQL	
Client throughput	 Standard throughput for ODBC client. High-performance JDBC transfers via Arrow Flight JDBC driver. High-speed data transfers with Arrow Flight to deliver up to 20x throughput increase between clients and Dremio Sonar. 	 Standard throughput for ODBC and JDBC clients. Supports mechanisms for fetching data in parallel via cloud storage such as AWS S3 and Azure Data Lake Storage to bring the data faster to BI tools. 	



11. **CONCLUSIONES**

Dremio es una plataforma **Lakehouse** de código abierto, donde fusiona la facilidad de acceso y el soporte de las capacidades de análisis empresarial que se encuentran en las tecnologías "Data Warehouse", con la flexibilidad y el costo relativamente bajo que tienen las plataformas "Data Lake".

Después de analizar con detenimiento y probar **Dremio** podemos decir que este tiene 2 grandes nucleos:

- Sonar es el motor de SQL que utiliza Dremio para proveer capacidades y rendimiento a nivel de datawarehouse en un datalake, además ofrece una experiencia de autoservicio sin necesidad de conocer las tablas físicas previamente optimizadas y agregadas.
- Arctic es el área de trabajo de Dremio, está pensado para que trabaje de forma similar a la herramienta GIT; aquí se podrán optimizar, transformar, realizar modelos de aprendizaje automático, etc, sin afectar a las cargas de trabajo puestas en producción. Una vez realizado se podrán fusionar los cambios.

Fue creado para tener una capa unificada que permite a cualquier profesional de datos acceder a los datos que necesita en un solo lugar, independientemente de dónde se almacenen los datos subyacentes.

Se ha podido comprobar que **Dremio** permite realizar consultas SQL directamente en sus fuentes de datos (VDS y PDS), incluso cuando esas fuentes de datos normalmente no permitirían este tipo de operaciones.

Dremio por defecto tiene una conexión nativa e integradora a Power BI, solo tendremos que darle al icono de Power BI en la sección de querys de **Dremio**.

Dremio se puede desplegar en Cloud o Enterprise. Ofrece una versión gratuita (Cloud) que es adecuada para la mayoría de las pequeñas empresas. En caso de necesitar más funciones de seguridad y control de datos **Dremio** ofrece una versión de pago.