

# Big Data con Hadoop y Spark

## Módulo 02

- Hue
- Data Governance

# Hue (Hadoop User Experience)

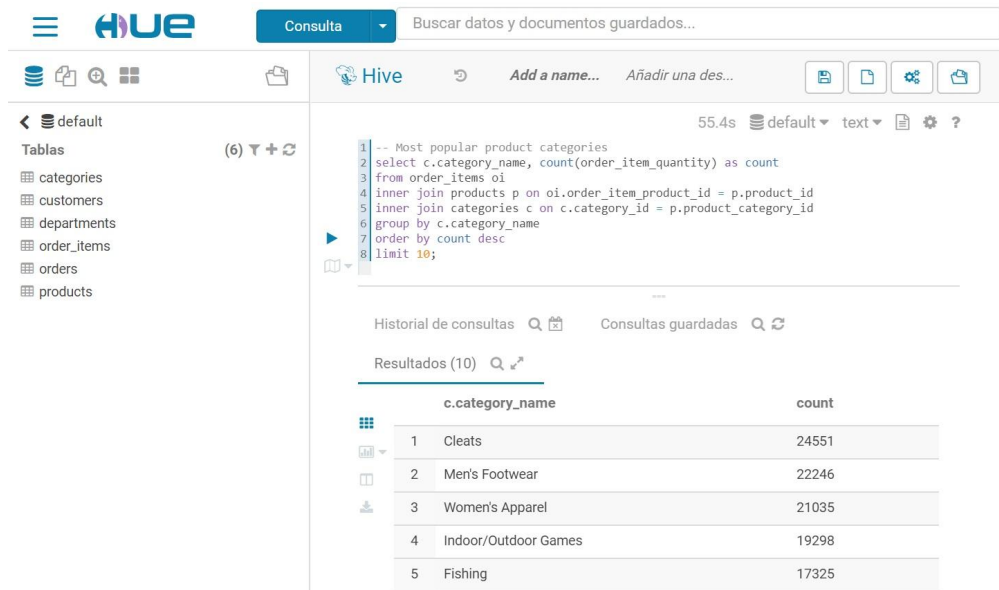
Es una interfaz web que permite ejecutar consultas SQL hacia diferentes motores de bases de datos, principalmente relacionados a Big Data.

Bases de datos soportadas

- [Conectores](#)

Entorno de prueba gratuito

- [Demo](#)



The screenshot displays the Hue web interface. On the left, a sidebar shows a 'default' view with a 'Tablas' (Tables) section containing a list of database tables: categories, customers, departments, order\_items, orders, and products. The main area is titled 'Consulta' (Query) and features a search bar. Below the search bar, a Hive query is entered and executed, taking 55.4s. The query is a SQL statement to find the most popular product categories. The results are displayed in a table with 10 rows, showing the category name and its count.

```
-- Most popular product categories
select c.category_name, count(order_item_quantity) as count
from order_items oi
inner join products p on oi.order_item_product_id = p.product_id
inner join categories c on c.category_id = p.product_category_id
group by c.category_name
order by count desc
limit 10;
```

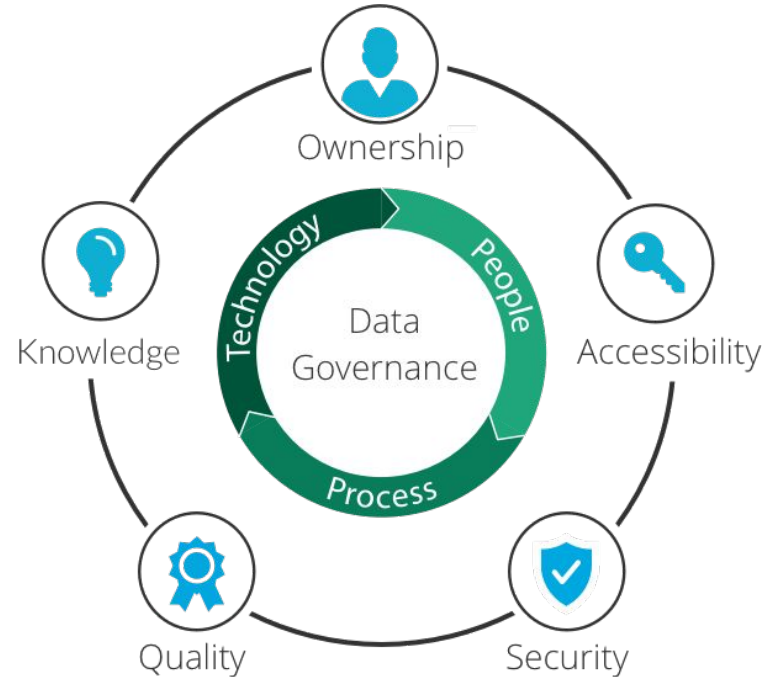
	c.category_name	count
1	Cleats	24551
2	Men's Footwear	22246
3	Women's Apparel	21035
4	Indoor/Outdoor Games	19298
5	Fishing	17325

# Data Governance

Es un concepto que propone considerar a los datos como activos de una empresa y su gestión debe estar alineada con los objetivos estratégicos.

Ecosistema Hadoop:

- Ranger: políticas de seguridad
- Atlas: catálogo de datos



# Ranger

## Policy Details :

Policy Type	<b>Row Level Filter</b>	
Policy Name *	<input type="text" value="rowFilter: cust.customer table"/>	<b>enabled</b>
Hive Database *	<input type="text" value="cust"/>	
Hive Table *	<input type="text" value="customer"/>	
Audit Logging	<b>YES</b>	
Description	<input type="text" value="Restrict employees to access only country-specific customer records"/>	

```
[john@localhost ~]$ beeline -u jdbc:hive2://localhost.localdomain:10000/cust
0: jdbc:hive2://localhost.localdomain:10000> select * from cust.customer;
+-----+-----+-----+-----+-----+-----+
| id | name_first | name_last | addr_country | date_of_birth | phone_num |
+-----+-----+-----+-----+-----+-----+
| 1 | Mackenzy | Smith | US | 1993-12-18 | 123-456-7890 |
| 2 | Sherlyn | Miller | US | 1975-03-22 | 234-567-8901 |
| 3 | Khiana | Wilson | US | 1989-08-14 | 345-678-9012 |
| 4 | Jack | Thompson | US | 1962-10-28 | 456-789-0123 |
+-----+-----+-----+-----+-----+-----+
```

## Row Filter Conditions :

Select Group	Select User	Access Types	Row Level Filter	
<input type="text" value="us-employees"/>	<input type="text" value="Select User"/>	<input type="button" value="select"/> <input type="button" value="edit"/>	<input type="text" value="addr_country = 'US'"/> <input type="button" value="edit"/>	<input type="button" value="x"/>
<input type="text" value="uk-employees"/>	<input type="text" value="Select User"/>	<input type="button" value="select"/> <input type="button" value="edit"/>	<input type="text" value="addr_country = 'UK'"/> <input type="button" value="edit"/>	<input type="button" value="x"/>
<input type="text" value="de-employees"/>	<input type="text" value="Select User"/>	<input type="button" value="select"/> <input type="button" value="edit"/>	<input type="text" value="addr_country = 'DE'"/> <input type="button" value="edit"/>	<input type="button" value="x"/>

# Atlas

Apache Atlas

Q SEARCH CLASSIFICATION GLOSSARY

Basic ☒ Advanced ☐

Search By Type  
hive\_column x ▼

Search By Classification  
PII x ▼

Search By Term  
Search Term ▼

Search By Text  
Search by text

Clear Search

Favorite Searches  
Save Save As  
PII columns ...

Results for: ( Type: **hive\_column** ) AND ( Classification: **PII** )  
If you do not find the entity in search result below then you can [create new entity](#)

Showing 8 records From 1 - 25 ☐ Exclude sub-types ☐ Exclude sub-classifications ☐ Show historical entities Columns ▼

<input type="checkbox"/>	Name	Owner	Type	Type	Classifications	Table	QualifiedName
<input type="checkbox"/>	<a href="#">providename</a>	hive	<a href="#">hive_column</a>	string	<a href="#">VENDOR_PII</a> +	<a href="#">prov_view</a>	claim.prov_view.providename@cl1
<input type="checkbox"/>	<a href="#">emailaddress</a>	hive	<a href="#">hive_column</a>	string	<a href="#">PII</a> x +	<a href="#">ww_customers</a>	hortoniabank.ww_customers.emailaddress@cl1
<input type="checkbox"/>	<a href="#">ccnumber</a>	hive	<a href="#">hive_column</a>	string	<a href="#">PII</a> x +	<a href="#">ww_customers</a>	hortoniabank.ww_customers.ccnumber@cl1
<input type="checkbox"/>	<a href="#">nationalid</a>	hive	<a href="#">hive_column</a>	string	<a href="#">PII</a> x +	<a href="#">ww_customers</a>	hortoniabank.ww_customers.nationalid@cl1
<input type="checkbox"/>	<a href="#">nationalid</a>	hive	<a href="#">hive_column</a>	string	<a href="#">PII</a> x +	<a href="#">us_customers</a>	hortoniabank.us_customers.nationalid@cl1
<input type="checkbox"/>	<a href="#">ssn</a>	hive	<a href="#">hive_column</a>	string	<a href="#">FINANCE...</a> +	<a href="#">tax_2015</a>	finance.tax_2015.ssn@cl1
<input type="checkbox"/>	<a href="#">providename</a>	hive	<a href="#">hive_column</a>	string	<a href="#">VENDOR...</a> x +	<a href="#">provider_summary</a>	claim.provider_summary.providename@cl1
<input type="checkbox"/>	<a href="#">ssn</a>	hive	<a href="#">hive_column</a>	string	<a href="#">FINANCE...</a> x +	<a href="#">tax_2010</a>	finance.tax_2010.ssn@cl1

Page Limit : 25 ▼

# Links de referencia

- **Ranger** <https://ranger.apache.org/>
- **Atlas** <https://atlas.apache.org/>
- **Hue** <https://gethue.com/>

# Revisión

- ¿Qué es Sqoop?
- ¿Dónde están alojados los datos en Hive?
- ¿Qué herramientas de Data Governance ofrece el ecosistema Hadoop?
- ¿Cuáles son los tipos de datos complejos que soporta Hive?
- ¿Qué es una partición en Hive?



# Gracias