

# BIG DATA SCIENCE & ANALYTICS®



Enzo Roccasalva

Head of Practice: Advanced Analytics & Risk



UNIVERSIDAD  
**AUSTRAL**

INGENIERÍA



**+40**

Años de  
Experiencia



**+75,000**

Instalaciones



**+140**

Países



**1<sup>o</sup>**

Empresa privada de  
software más grande  
del mundo.

Foco en “**Business  
Analytics**”



**+ 3.02**

Billones de USD  
en Facturación



**+13,773**

Empleados



**25%**

Inversión en R&D

## ¿QUÉ DICEN LOS ANALISTAS SOBRE SAS?

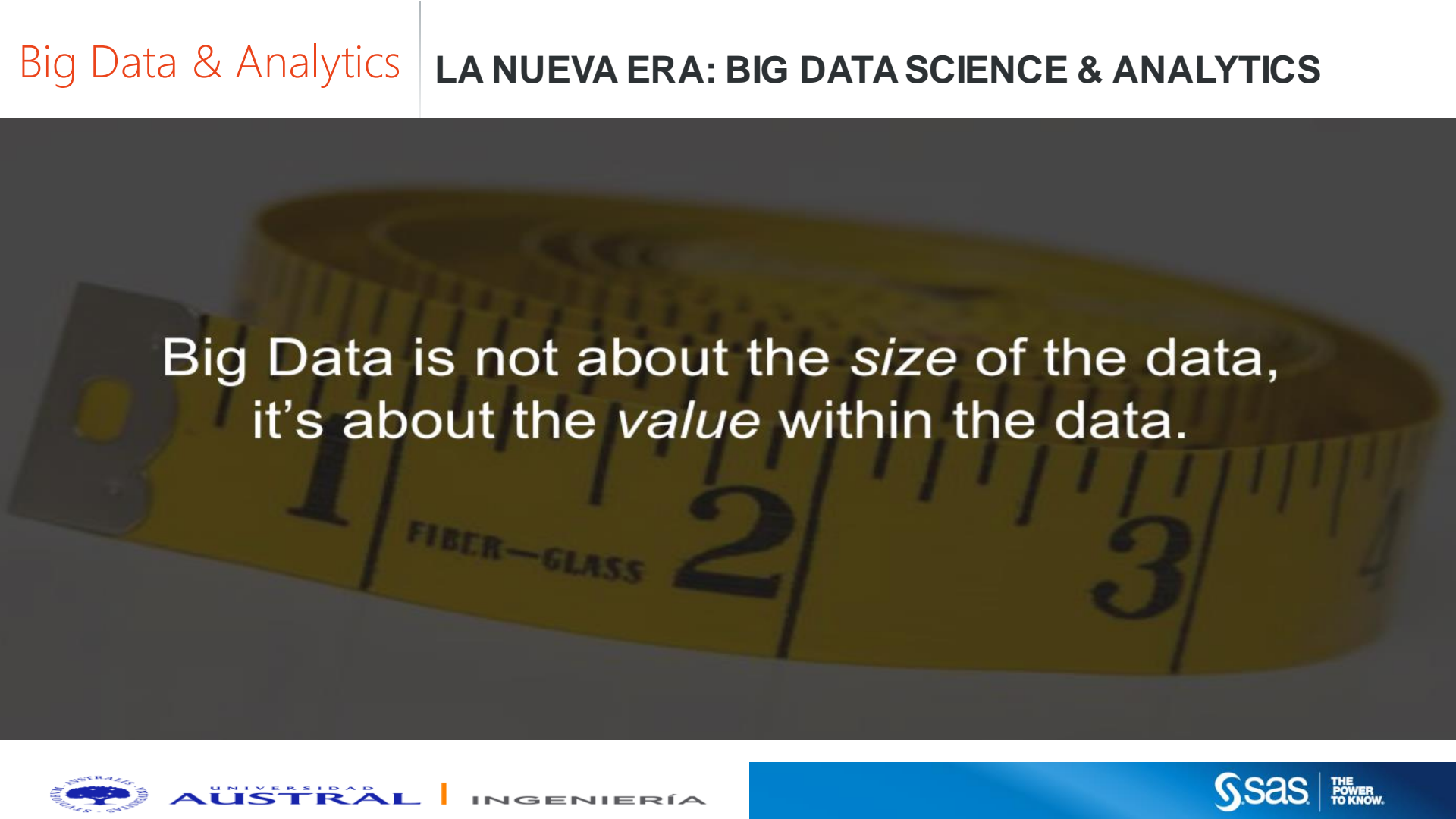
SAS a Leader in 2017 **Forrester**  
Wave for Predictive Analytics and  
Machine Learning











Big Data is not about the *size* of the data,  
it's about the *value* within the data.



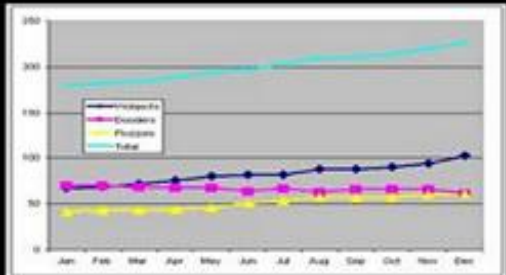
What my friends think I do



What my mom thinks I do



What society thinks I do



What my boss thinks I do

**Gradient vector**

Let  $\theta$  be an  $d$ -dimensional vector and  $f(\theta)$  a scalar-valued function. The gradient vector of  $f(\cdot)$  with respect to  $\theta$  is:

$$\nabla_{\theta} f(\theta) = \begin{bmatrix} \frac{\partial f(\theta)}{\partial \theta_1} \\ \frac{\partial f(\theta)}{\partial \theta_2} \\ \vdots \\ \frac{\partial f(\theta)}{\partial \theta_d} \end{bmatrix} \quad \theta = \begin{bmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_d \end{bmatrix}$$

What I think I do



What I actually do



### Ejecutivos

Realizar nuevos y más rápidos análisis para mejorar los negocios

Tomar decisiones de negocios basadas en los datos cuando y dónde se requieran

Lograr eficiencia operativa, nuevos modelos de negocios y mejores márgenes



### Gerencia de TI

Asegurar la velocidad de los análisis de datos

Administrar el caos

Implementar desarrollos en forma fluida

Asegurar la gobernabilidad de la información



### Data Scientists Usuarios de Negocio

Minimizar el tiempo de preparación de datos

Crear modelos y visualizaciones en forma eficiente

Desarrollar modelos que den beneficios al negocio



## ¿CÓMO SE ADAPTA SAS A LA ERA ACTUAL?

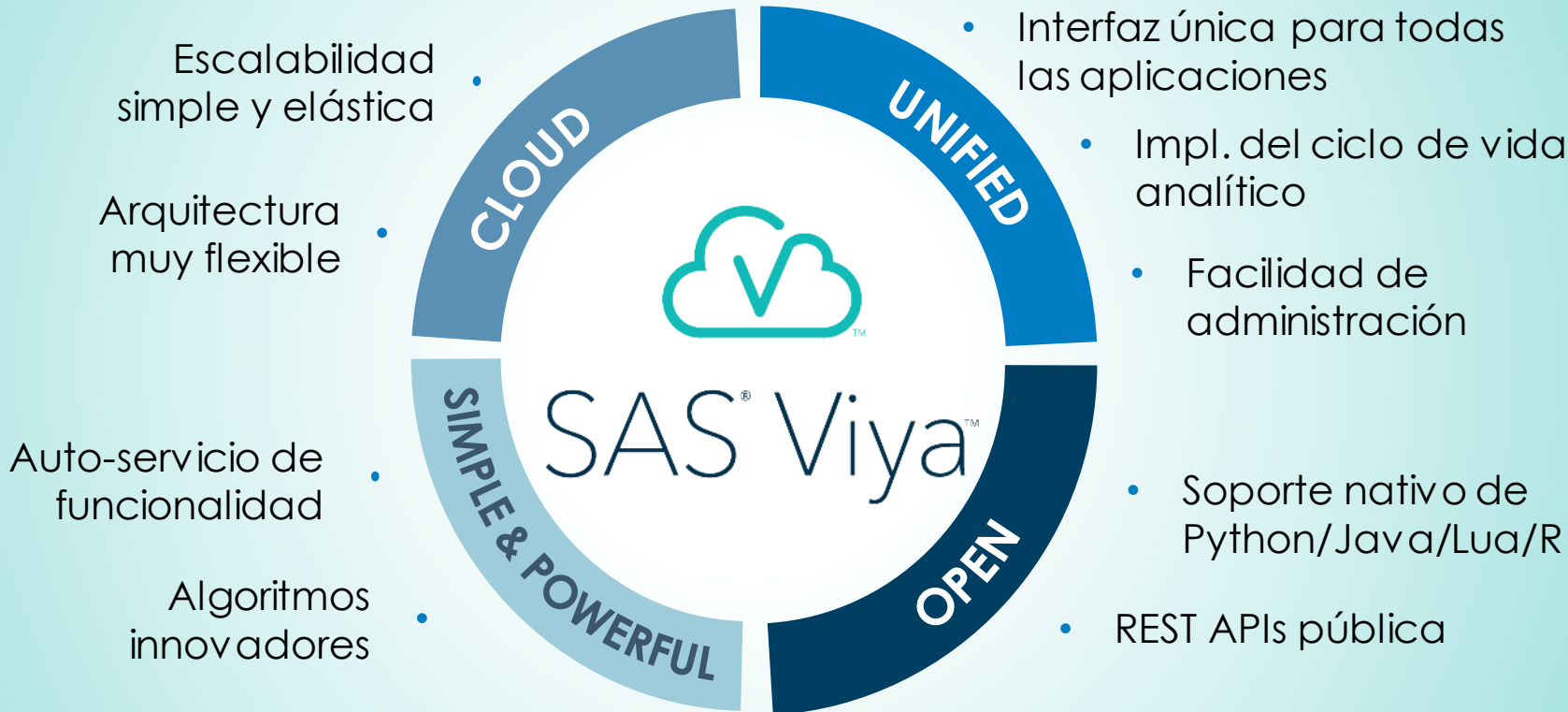


SAS® Viya™

- Que viene de **sas via**
- Que no se pronuncia **sas visha**
- Que es cool pronunciarlo **sas vaiaæ**
- Y lo diremos **sas vaia**



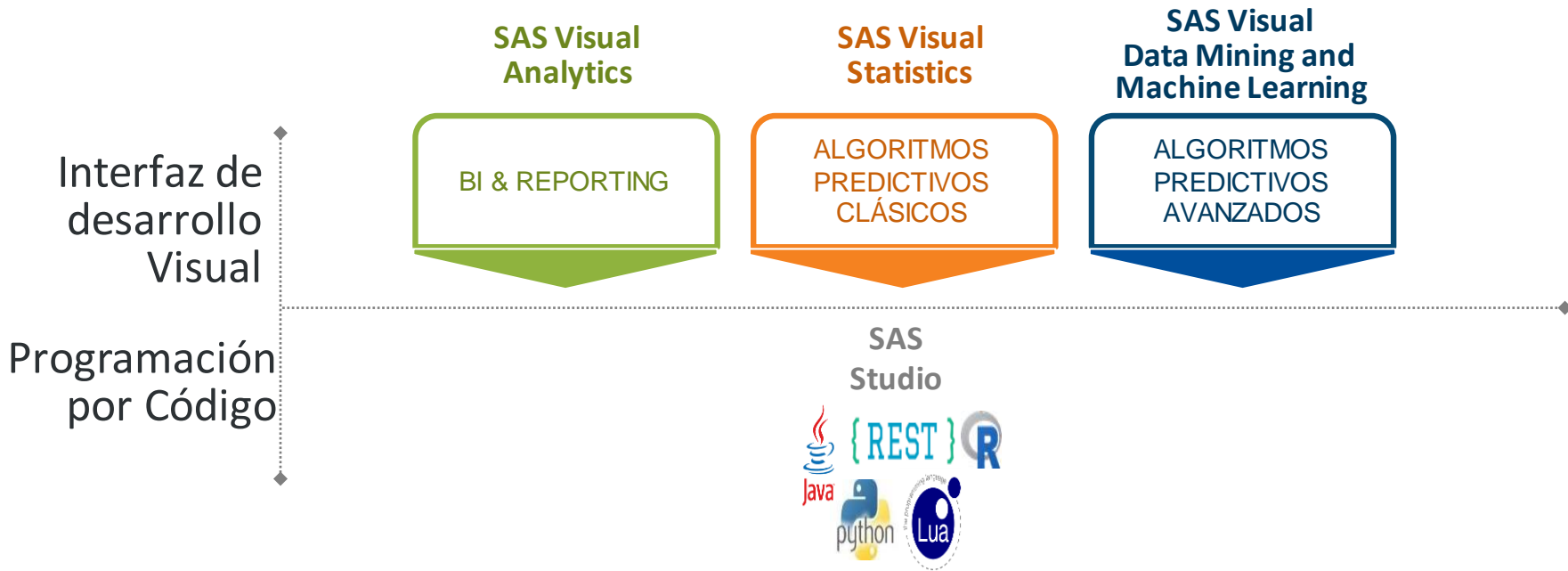
## PLATAFORMA ANALÍTICA SAS





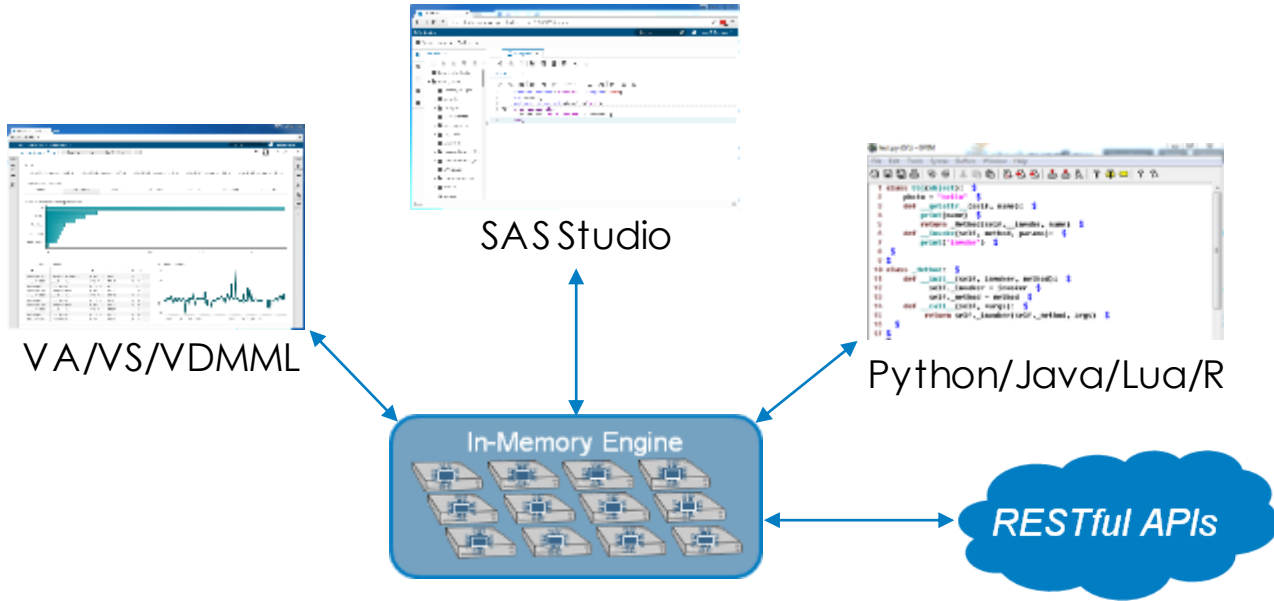
## PLATAFORMA ANALÍTICA SAS

Interfaces



## PLATAFORMA ANALÍTICA SAS

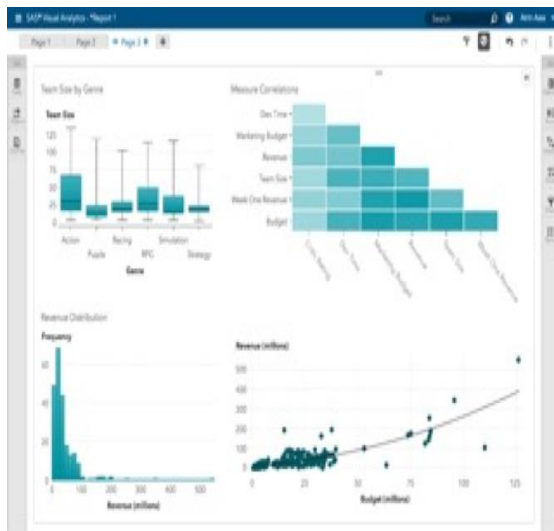
CAS In-Memory Engine Único **Servidor Analítico** para todas las herramientas y soluciones





## Exploración Visual & Análisis Descriptivo

- Descubrir relaciones y tendencias
- Analytics a través de visualizaciones
- Explorar predicciones



- Estadísticas resumidas
- Clustering
- Matrices de correlación, diagramas de dispersión, diagramas de caja y mapas de calor para visualizar los hallazgos claves





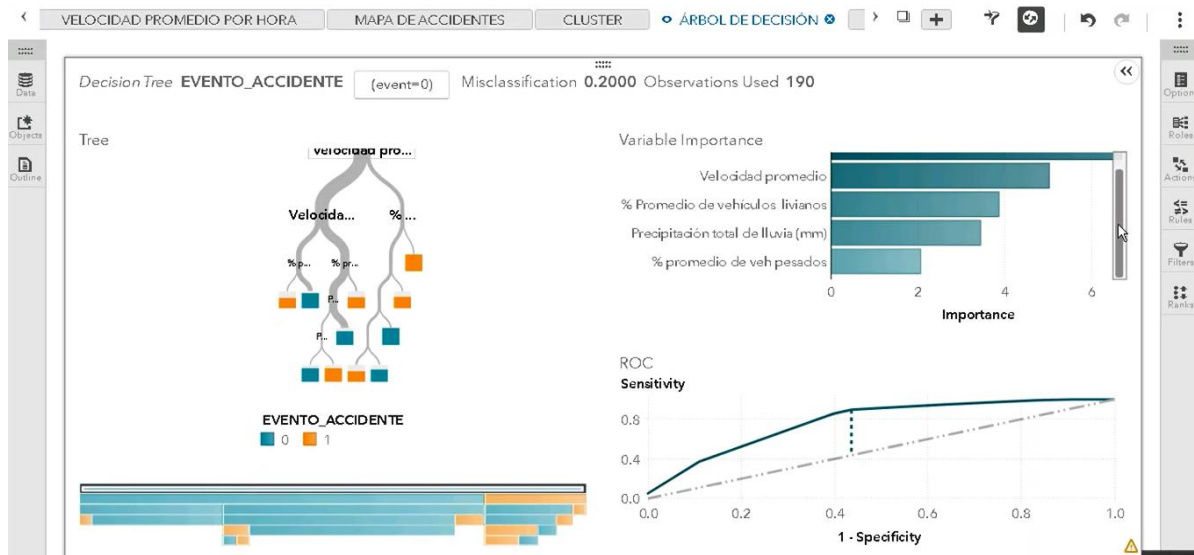


- Regresión Lineal
- Regresión Lógica
- Árbol de decisión
- GLM (Modelos lineales generalizados)
- Forecasting and análisis de escenarios
- Comparación de Modelos

# PLATAFORMA ANALÍTICA SAS



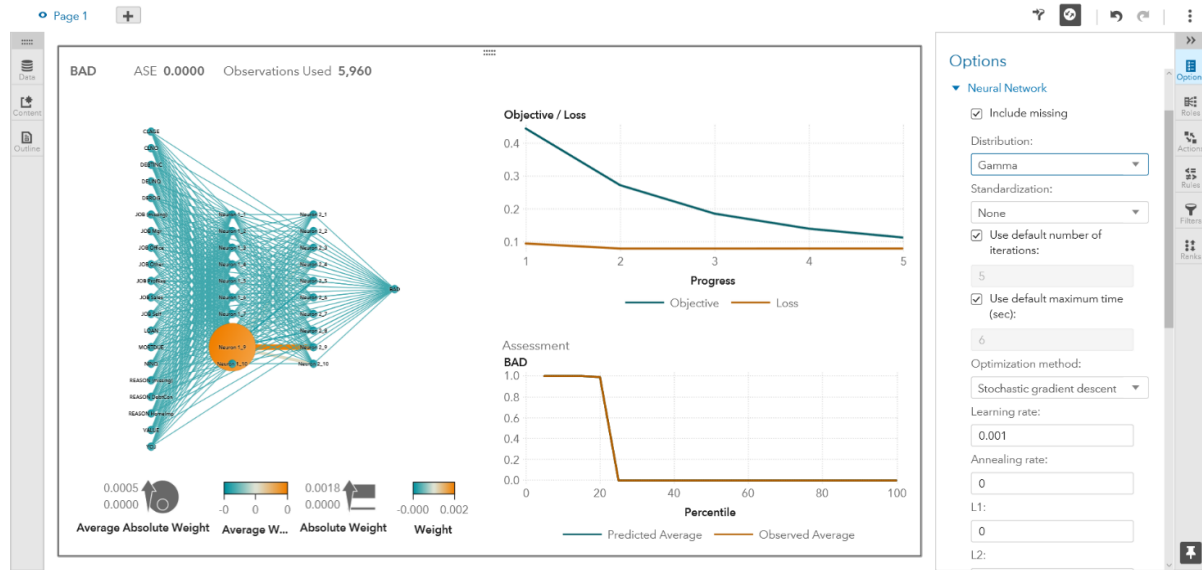
## Análisis Predictivo Clásico





## Análisis Predictivo Avanzado

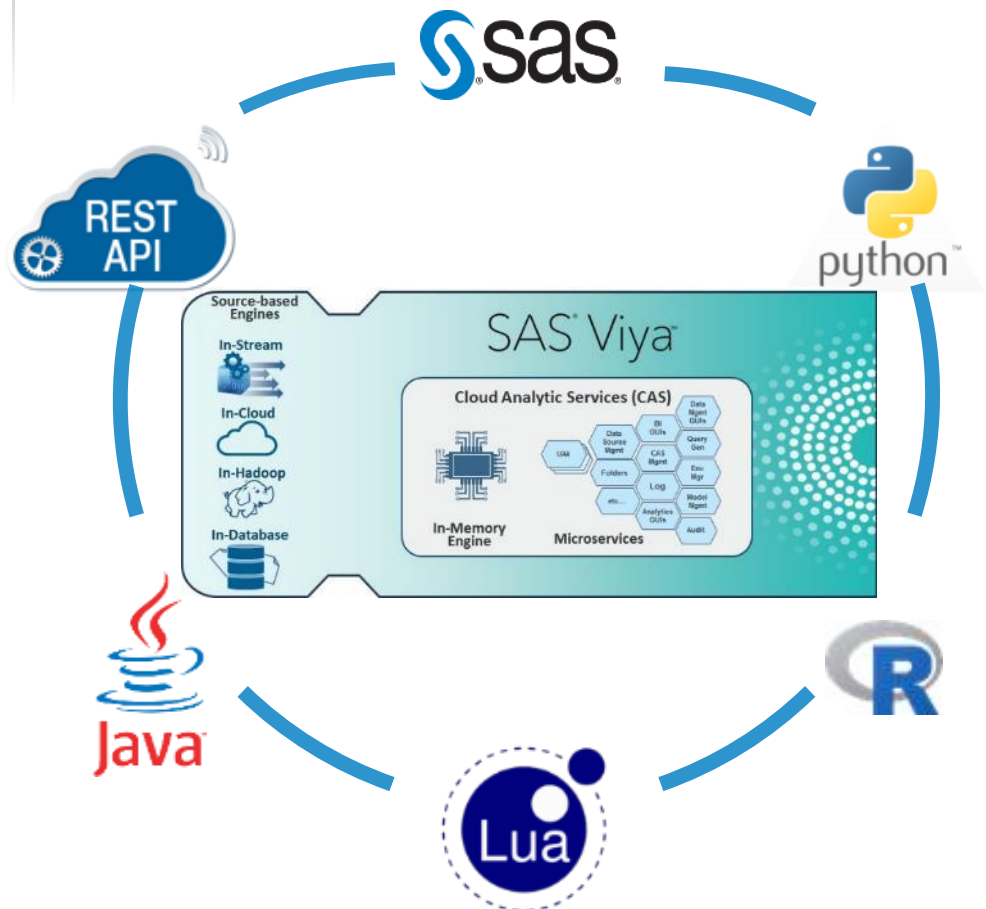
- Factorization Machine Models
- Random Forest Models
- Gradient Boosting Models
- Redes Neuronales
- Support Vector Machine Models
- Text Mining

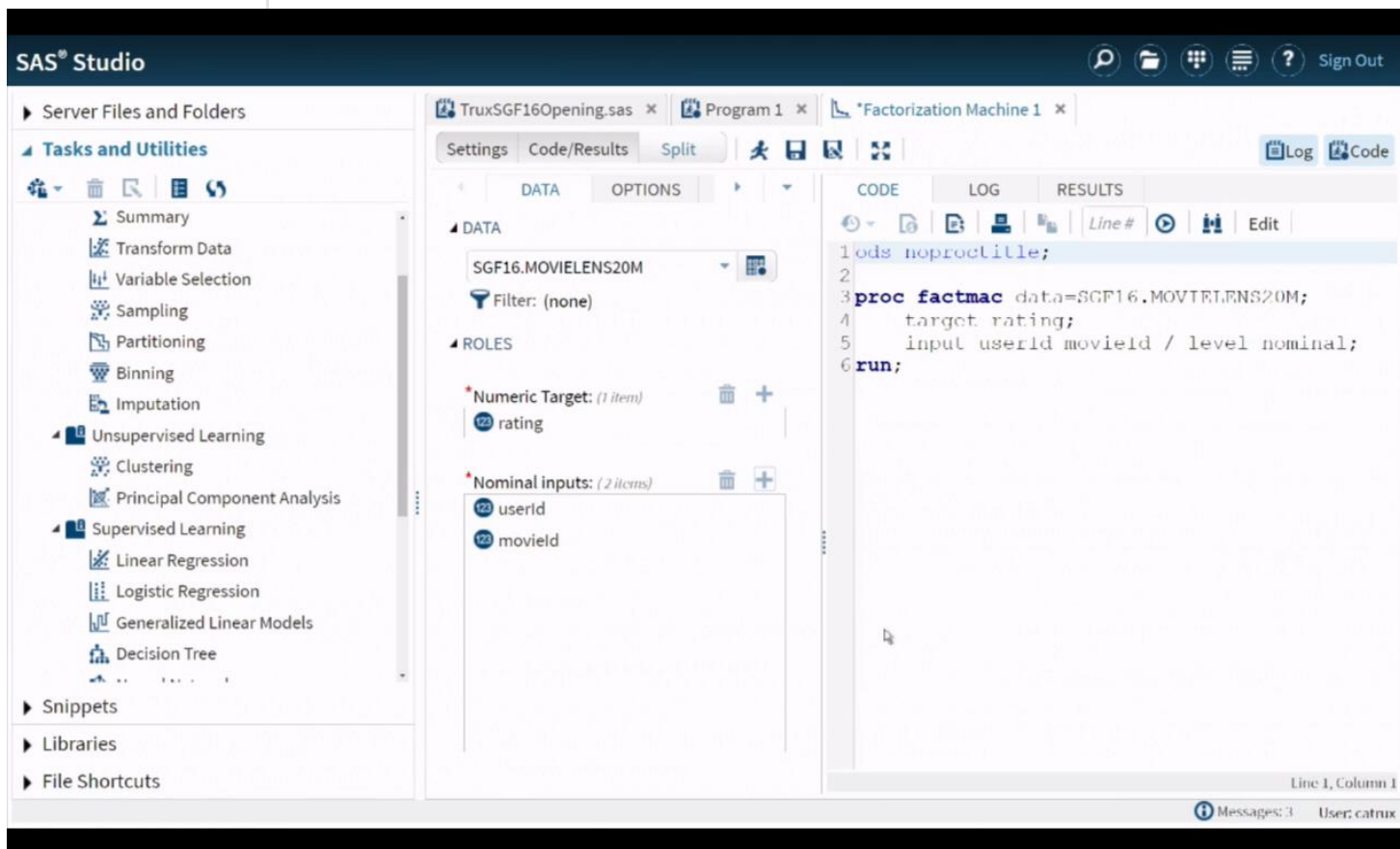


# Big Data & Analytics



Programación en SAS,  
Java, Python, Lua y R





The screenshot displays the SAS Studio interface. On the left, the 'Tasks and Utilities' pane lists various data processing tasks such as Transform Data, Variable Selection, and Machine Learning algorithms. The main workspace is divided into three panes: DATA, OPTIONS, and CODE. The DATA pane shows the dataset 'SGF16.MOVIELENS20M' with a filter of '(none)'. The ROLES pane shows 'rating' as the Numeric Target and 'userid' and 'movieid' as Nominal inputs. The CODE pane contains the following SAS code:

```
1 ods noproctitle;
2
3 proc factmac data=SGF16.MOVIELENS20M;
4   target rating;
5   input userid movieid / level nominal;
6 run;
```

The status bar at the bottom indicates 'Line 1, Column 1' and 'Messages: 3 User: catrux'.

## PROCESO ANALÍTICO

Fuente Datos



Exploración & Análisis



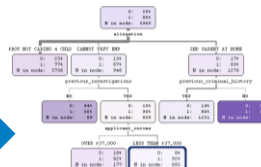
Muestra Desarrollo



Muestra Validación



Elección Algoritmo Predictivo



Verificación de Resultados





### OBJETIVO:

- **Identificar** aquellos clientes con **mayor propensión** a descargar un white paper de una compañía high-tech.
- La variable objetivo (Target) se encuentra definida como:
  - **-SI: Descargo un white paper y -No: No descargo un white paper.**

### DATOS:

- Se utilizarán “**datos de visita**”, para realizar la predicción.
- Datos ( variables dummy):
  - 1) **Visitó las siguientes categorías:** Cloud Computing, Database, Internet of things, Mobile & Wireless, Security y Storage.
  - 2) **Duración de la visita:** de 10 a 30 segundos, de 30 segundos a 5 minutos, etc.

## Demostración de Redes Neuronales en Interfaz Visual



## Referencias

SAS Viya [https://www.sas.com/en\\_us/software/viya.html](https://www.sas.com/en_us/software/viya.html)

SAS Visual Data Mining & Machine Learning [https://www.sas.com/en\\_us/software/visual-data-mining-machine-learning.html](https://www.sas.com/en_us/software/visual-data-mining-machine-learning.html)

SAS Visual Statistics [https://www.sas.com/en\\_us/software/visual-statistics.html](https://www.sas.com/en_us/software/visual-statistics.html)

SAS Visual Analytics [https://www.sas.com/en\\_us/software/visual-analytics.html](https://www.sas.com/en_us/software/visual-analytics.html)

SAS Studio [https://www.sas.com/en\\_us/software/studio.html](https://www.sas.com/en_us/software/studio.html)

## Videos

SAS Viya <http://video.sas.com/#category/videos/sas-viya>