

Fecha de liberación oficial: miércoles 17 de abril 07:40

Tarea para el Hogar CINCO

Esta es la última Tarea para el Hogar, hará la primera corrida de Experimentos Colaborativos con un Workflow muy liviano que hemos llamado *Guantes Blancos*. Este corrida demandará dos horas y media de procesamiento, y le permitirá concentrarse en SU experimento colaborativo.

En Experimentos Colaborativos alumnos organizados en grupos optimizarán distintas etapas del workflow y colectivamente se encontrarán las mejores prácticas, que se aplicarán directamente a la entrega final de la Competencia Kaggle.

No se angustie si por primera vez no es sorprendido con ganancias superadoras, la sábana es corta, el workflow Guantes Blancos fue pensado para que corra rápido con un mínimo esfuerzo de operación, así usted se puede concentrar en tareas cognitivas de superior nivel como lo es entender el código del experimento que eligió para estar así en condiciones de elegir una mejor configuración del workflow y a los diez días modificar el script correspondiente a su experimento colaborativo.

Le llevará las clases 06 y 07 terminar de entender conceptualmente la funcionalidad de cada etapa del workflow, sus parámetros y la forma en que cada etapa afecta a las que siguen.

La determinación de los parámetros óptimos de cada etapa será empírica ya que depende de las características de estos datos y de la correlación entre las variables independientes y la clase.

Es la idea que los alumnos que superaron la etapa de la angustia prueben configuraciones distintas de los parámetros de Guantes Blancos, así la planilla colaborativa se completa con enriquecedoras estrategias, y de esos resultados surgen nuevas ideas para mejorarlos.

Sección Pasado (ya lo debería haber hecho)

1. Elección Experimento Colaborativo

Solo podrá usted continuar con el resto de esta Tarea para el Hogar CINCO si ya ha elegido un Experimento Colaborativo escribiendo su nombre en la portada del experimento y ya estan conformados los antagonicos grupos A y B.

Sección Deseable

2. Corrida del workflow *Guantes Blancos*

2.1 De lo que logró entender en los videos sobre el workflow, identifique cual etapa del workflow se corresponde a su Experimento Colaborativo

2.2 Actualice su repositorio GitHub

2.3 Junto a su compañera/o de equipo lea en detalle el script
`./src/workflow01/z501_workflow_guantesblancos.r`


2.4 Haga una copia del mismo, quitándole la letra “z” inicial


2.5 Si en el equipo se animan a ensuciarse vuestros guantes blancos, hagan algun pequeño cambio a los parámetros de la etapa del Workflow que corresponde al experimento colaborativo que eligieron.


2.6 Cada integrante del grupo debe actualizar su repositorio, subiendo su nuevo script modificado a GitHub


2.7 Vayan a la consola de Google Cloud <https://console.cloud.google.com/compute/instancesAdd?&creationFlow=fromTemplate> y marque la opcion temp-08vcpu-064ram para crear una virtual machine de 64 GB de memoria RAM y 8 vCPU


2.8 Presione el boton Continue

 Create an instance


 **New VM instance**
Create a single VM instance from scratch

 **New VM instance from template**
Create a single VM instance from an existing template


 **New VM instance from machine image**
Create a single VM instance from an existing machine image

 **Marketplace**
Deploy a ready-to-go solution onto a VM instance

1 Select template — **2 Customize VM instance**

 **Filter** Filter instance templates

	Name ↑	Machine type	Disk type	Location	Creation time
<input type="radio"/>	temp-08vcpu-016ram	custom-8-16384-ext	Standard persistent disk	global	Mar 28, 2024, 8:39:48 AM UTC-03:00
<input type="radio"/>	temp-08vcpu-032ram	custom-8-32768-ext	Standard persistent disk	global	Mar 28, 2024, 8:39:53 AM UTC-03:00
<input checked="" type="radio"/>	temp-08vcpu-064ram	custom-8-65536-ext	Standard persistent disk	global	Mar 28, 2024, 8:39:58 AM UTC-03:00
<input type="radio"/>	temp-08vcpu-128ram	custom-8-131072-ext	Standard persistent disk	global	Mar 28, 2024, 8:40:03 AM UTC-03:00
<input type="radio"/>	temp-08vcpu-256ram	custom-8-262144-ext	Standard persistent disk	global	Mar 28, 2024, 8:40:07 AM UTC-03:00

CONTINUE 

2.9 Dentro de la pantalla de configuración, llame a esa virtual machina `vm-fantasmita` , y elija como región Las Vegas y como zona `us-west4-a`

✓ Select template — 2 Customize VM instance

Source template

temp-08vcpu-064ram

CHANGE TEMPLATE

Name *

`vm-fantasmita`

MANAGE TAGS AND LABELS

Region *

us-west4 (Las Vegas)

Region is permanent

Zone *

us-west4-a

Zone is permanent

2.10 vaya al fondo de la pantalla y presione el botón azul de CREATE

2.11 Espere a que su virtual machine `vm-fantasmita` se encienda y copie la External IP al portapapeles

Filter Enter property name or value

<input type="checkbox"/>	Status	Name ↑	Zone	Creation time	Machine type	Provisioning model	Internal IP	External IP	Co
<input type="checkbox"/>		desktop	southamerica-west1-a	Mar 28, 2024, 8:40:14 AM	e2-standard-4	Spot	10.194.0.3 (nic0)		S
<input type="checkbox"/>		instance-instalacion	us-west4-a	Mar 28, 2024, 7:23:15 AM	t2d-standard-4	Spot	10.182.0.7 (nic0)		S
<input type="checkbox"/>		vm-fantasmita	us-west4-a	Apr 17, 2024, 5:46:31 AM	custom-8-65536-ext	Spot	10.182.0.9 (nic0)	34.125.123.110 (nic0)	S

Related actions

Copy to clipboard

2.12 Desde un browser en su PC ingresa a `http://(la ip externa que copio)` por favor tenga en cuenta que es http SIN ninguna “s” al final En el ejemplo de la imagen, se debio escribir `http://34.125.123.110/`

2.13 Le aparecerá la pantalla de credenciales para ingresar a RStudio, cargue su usuario y su clave.

2.14 Ya dentro de RStudio , busque SU script al que le quito la “z” inicial y si se animó, modificó [~/labo2023v1/src/workflow01/501_workflow_guantesblancos.r](#)

2.15 Ponga a correr su script, directamente con el botón `Soruce()` de Rstudio

2.16 La corrida demorará apenas dos horas y media ya que los hiperparámetros utilizados en el Workflow *Guantes Blancos* no fueron pensados para tener una buena ganancia, sino para que corra rápido y usted pueda conceptualizar las etapas del workflow.

2.17 Al terminar la virtual machine [vm-fantasmita](#) se auto.suicidará y desaparecerá. Esa será la forma de saber que terminó.

2.18 Los resultados de la corrida quedan en el bucket, dentro de la carpeta `exp`, en las siguientes carpetas

- DT0001
- CA0001
- DR0001
- FE0001
- TS0001
- TS0002
- HT0002
- HT0001
- ZZ0001
- ZZ0002

2.19 Analice lo que quedó en las carpetas de salida del workflow, las ZZ0001 y ZZ0002 y suba a Kaggle los archivos que encuentre por ahí que suponga que pueden ser alimento para Kaggle.

2.20 Analice con su grupo lo que quedó en la carpeta correspondiente a la etapa del Experimento Colaborativo que eligió

2.21 Finalmente, el paso más importante de todos, lea en grupo detalladamente el script correspondiente al experimento colaborativo elegido Experimento Colaborativo, el haber visto e interpretado las salidas les facilitará la tarea.

Anexo

El workflow se comunica con usted con archivos que deja en la carpeta del experimento.

La presencia de un archivo llamado `z-Rend.txt` indica que la corrida fue completamente exitosa.

La presencia de un archivo llamado `z-Rabort.txt` indica que el R ha abortado. Mirar el contenido del archivo `outfile` le podrá dar una pista del error.

Si una etapa del workflow aborta, entonces se detiene el avance del mismo. Usted podrá ver con los nombres de carpeta hasta donde llegó.

En el archivo `log.txt` queda la fecha y hora en que se inició el script, y en el caso que terminó exitosamente está la fecha de finalización

En el archivo `output.yml` quedan relevantes resultados intermedios de la corrida del script

El archivo `parametros.yml` son los parámetros que recibió el script, los que usted configuró en el workflow.

En cada carpeta de un experimento, hay al menos un soft links, que es una carpeta virtual que apunta a inputs que necesita el proceso para poder correr.

En cada carpeta de un experimento hay una copia del script `.r` que se corrió para ese experimento.

Va la descripción conceptual de lo que sucede con el siguiente código extraído del script `z501_workflow_guantesblancos.r`

```
TS_strategy_guantesblancos_202109( "TS0001", "FE0001" )  
HT_tuning_guantesblancos( "HT0001", "TS0001" )
```

La primer función `TS_strategy_guantesblancos_202109` es un experimento, que creará la carpeta `"TS0001"` donde dejará los resultados. Dicha función se alimentará de los resultados de un experimento previo llamado `"FE0001"`

Lo que genera el experimento anterior y que quedó en la carpeta `"TS0001"` ahora es tomado como input por el experimento `HT_tuning_guantesblancos` y dicho experimento guardará su salida en la carpeta `"HT0001"`

Usted es quien elige los nombres de los experimentos (la etiqueta que va en cada tubo de ensayo) , y debe ser muy, pero muy cuidadoso de asignar un nuevo nombre, jamás utilizado con anterioridad a cada nuevo experimento.

(Un sistema de workflow productivo, va a funcionar en forma sensiblemente distinta, pero en esta etapa usted debe conceptualizar los procesos)