

## Desafío Técnico

El presente desafío consiste en construir un maestro de ratings utilizando Python y Spark SQL. Para ello, se cuenta con el excel **data.xlsx** con las siguientes 4 pestañas: **rating\_empresa** (clasificaciones de riesgo de algunas empresas), **rating\_soberano** (clasificaciones de riesgo de países), **homologacion\_rating** (homologación de las calificaciones de riesgo de las agencias) y **homologacion\_pais** (homologación de los nombres de los países).

La siguiente tabla corresponde al diccionario de datos de las columnas relevantes del conjunto de datos.

variable	descripcion
rut	rut de la empresa
dv	dígito verificador del rut
nombre	nombre de la empresa
pais_bbg	nombre país (no estandarizado)
pais	nombre país (estandarizado)
sp	clasificación de riesgo proveniente de la agencia S&P
fitch	clasificación de riesgo proveniente de la agencia FitchRatings
mdy	clasificación de riesgo proveniente de la agencia Moodys
agencia_homol	nombre corto/homologado de la agencia clasificadora
rating	rating de la agencia clasificadora (no estandarizado)
rating_norma	rating de la agencia clasificadora (estandarizado)
orden_norma	correlativo de los ratings homologados, un valor mayor indica un mayor riesgo o una peor clasificación

El desafío se divide en dos partes. En primer lugar, se debe crear un código en Python que lea el archivo Excel y exporte cada pestaña a un archivo .csv con separador de columnas "|". En segundo lugar, debe crear un código Python que use Spark SQL para generar el archivo master\_rating.csv con separador de columnas "|" con la estructura de la siguiente tabla:

rut	dv	nombre	pais	rating_empresa	rating_soberano
41203332	9	INDUSTRIAL AND COMERCIAL BANK OF CHINA	china	A+	A+
44000061	4	CITIBANK NA ESTADOS UNIDOS	estados unidos	AA-	AA+

A continuación se detalla la lógica de construcción del maestro de ratings:

- El maestro de ratings se construye a partir de la tabla rating\_empresa, donde los campos *rut*, *dv* y *nombre* se obtienen de forma directa, por otro lado, los campos *pais*, *rating\_empresa* y *rating\_soberano* son calculados.

rut	dv	nombre
41203332	9	INDUSTRIAL AND COMERCIAL BANK OF CHINA
40000177	4	ITAU UNIBANCO S-A
41001123	9	BCO BRADESCO SA BRASIL
42001513	5	BANCO SANTANDER BRASIL
41001310 K		BCO DO BRASIL S A-BRASIL
41202256	4	BANK OF CHINA HONG KONG BRA - CHINA
44000061	4	CHINA CONSTRUCTION BANK
41001391	6	CITIBANK NA ESTADOS UNIDOS
41001363	0	BCO SAFRA SA CAMPINAS
40000070	0	BBANK OF AMERICA N.A.
40000154	5	BNP PARIBAS S.A. FRANCIA
41202507	5	AGRICULTURAL BANK OF CHINA THE-CHINA
41202238	6	OVERSEA CHINESE BANKING CORPORATION LIMITED
41001270	7	BANK HAPOLIM B.M.
41202365 K		Coperative Rabobank
41203226	8	KOOKMIN BANK - COREA
41202513 K		CITIBANK REINO UNIDO
41001073	9	LANDESBANK HESSEN-THUERINGE - ALEMANIA
41001244	8	COMMERZBANK AG - ALEMANIA

- El campo *pais* es la homologación del campo *pais\_bbg* de la tabla **rating\_empresas** con la tabla **homologacion\_pais**.

A	B	C	D
rut	dv	nombre	pais_bbg
41203332	9	INDUSTRIAL AND COMERCIAL BANK OF CHINA	CHINA
40000177	4	ITAU UNIBANCO S-A	BRAZIL
41001123	9	BCO BRADESCO SA BRASIL	BRAZIL
42001513	5	BANCO SANTANDER BRASIL	BRAZIL
41001310 K		BCO DO BRASIL S A-BRASIL	BRAZIL
41202256	4	BANK OF CHINA HONG KONG BRA - CHINA	CHINA
44000061	4	CHINA CONSTRUCTION BANK	CHINA
41001391	6	CITIBANK NA ESTADOS UNIDOS	UNITED STATES OF AMERICA
41001363	0	BCO SAFRA SA CAMPINAS	BRAZIL
40000070	0	BBANK OF AMERICA N.A.	UNITED STATES OF AMERICA
40000154	5	BNP PARIBAS S.A. FRANCIA	FRANCE
41202507	5	AGRICULTURAL BANK OF CHINA THE-CHINA	CHINA
41202238	6	OVERSEA CHINESE BANKING CORPORATION LIMITED	SINGAPORE
41001270	7	BANK HAPOLIM B.M.	ISRAEL
41202365 K		Coperative Rabobank	NETHERLANDS
41203226	8	KOOKMIN BANK - COREA	KOREA, REPUBLIC OF
41202513 K		CITIBANK REINO UNIDO	UNITED KINGDOM
41001073	9	LANDESBANK HESSEN-THUERINGE - ALEMANIA	GERMANY
41001244	8	COMMERZBANK AG - ALEMANIA	GERMANY
41001850	0	BANCO SANTANDER MADRID ESPANA	SPAIN
41001007	0	Mizuho Bank Ltd	JAPAN
44000186	6	NONGHYUP BANK	KOREA, REPUBLIC OF
42001284	5	COMMERZBANK NEW YORK BRANCH	UNITED STATES OF AMERICA
44000158	0	BANK OF CHINA LIMITED - HONG KONG	HONG KONG
41000406	0	WUHAN DEVELOPMENT BANK	HONG KONG

A	B	C
pais	cod_pais	pais_bbg
abu dhabi	-1	abu dhabi
aruba	-1	aruba
bailea de guernsey	-1	balliwick of guernsey
islas cook	-1	cook islands
curazao	-1	curacao
dubai	-1	dubai
gambia	-1	gambia
gibraltar	-1	gibraltar
granada	-1	grenada
isla de man	-1	isle of man
serbia	-1	serbia
montenegro	-1	montenegro
montserrat	-1	montserrat
antillas neerlandesas	-1	neth. antilles
no informado	-1	uninformed
no informado	0	uninformed
senegal	101	senegal
guinea bissau	103	guinea-bissau
guinea	104	guinea

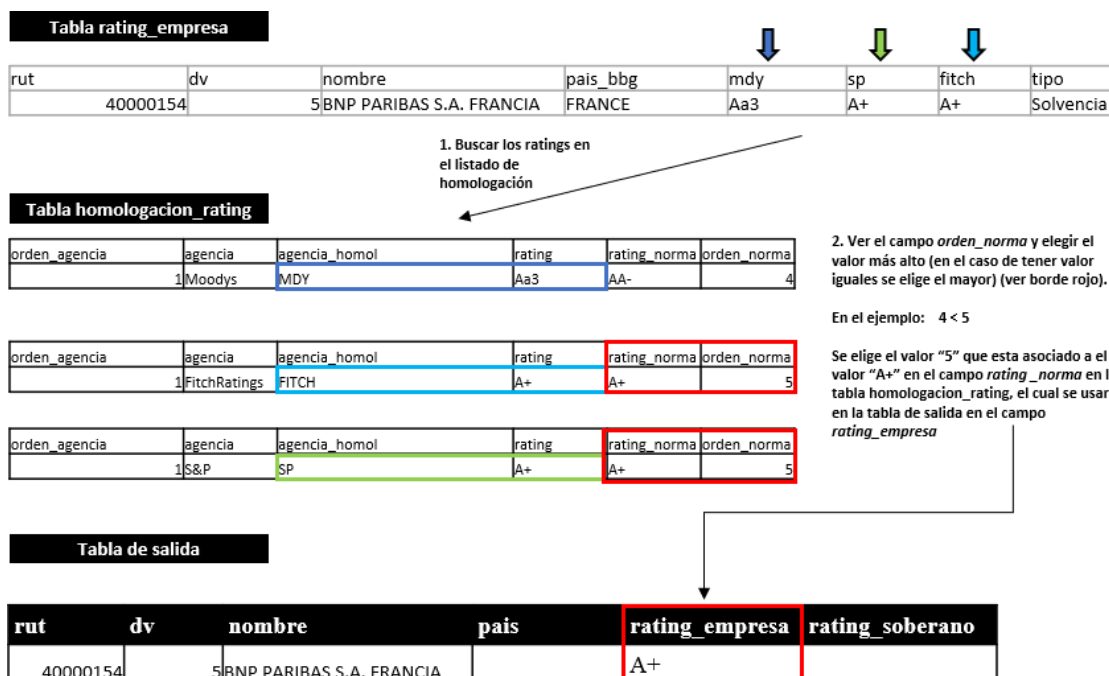
- El campo *rating\_empresa* de la tabla de salida es el peor entre los ratings *sp*, *fitch* y *mdy*. Para el cálculo se requiere homologar los campos *sp*, *fitch* y *mdy* de la tabla **rating\_empresas** por el campo *rating\_norma* de la interfaz **homologacion\_rating**. Luego, en base al valor del campo *orden\_norma* de la tabla **homologacion\_rating** se escoge el rating homologado con valor más alto.

D	E	F	G	H
pais_bbg	mdy	sp	fitch	tipo
CHINA	A1	A	A	Solencia
BRAZIL	Ba1	BB-	BB	Solencia
BRAZIL	Ba1	BB-	BB	Solencia
BRAZIL	Ba1	BB-	BB	Solencia
CHINA	A1	A	A	Solencia
CHINA	A1	A	A	Solencia
UNITED STATES OF AMERICA	Aa3	A+	A+	Solencia
BRAZIL	Ba1	BB-	BB	Solencia
UNITED STATES OF AMERICA	Aa2	A+	A+	Solencia
FRANCE	Aa3	A+	A+	Solencia
CHINA	A1	A	A	Solencia
SINGAPORE	Aa1	AA-	AA-	Solencia
ISRAEL	A1	A	A	Solencia
NETHERLANDS	Aa2	A+	A+	Solencia
KOREA, REPUBLIC OF	Aa3	A+	A+	Solencia
UNITED KINGDOM	-	-	-	Solencia
GERMANY	Aa3	A	A+	Solencia
GERMANY	A1	BBB+	BBB	Solencia
SPAIN	A2	A	A-	Solencia
JAPAN	A1	A	A-	Solencia
KOREA, REPUBLIC OF	Aa3	A+	A-	Solencia
UNITED STATES OF AMERICA	A1	BBB+	BBB	Solencia
HONG KONG	Aa3	A+	A	Solencia
KOREA, REPUBLIC OF	Aa3	A+	A	Solencia

A	B	C	D	E	F
orden_agencia	agencia	agencia_homol	rating	rating_norma	orden_norma
1	S&P	SP	AAA	AAA	1
2	S&P	SP	AA+	AA+	2
3	S&P	SP	AA	AA	3
4	S&P	SP	AA-	AA-	4
5	S&P	SP	A+	A+	5
6	S&P	SP	A	A	6
7	S&P	SP	A-	A-	7
8	S&P	SP	BBB+	BBB+	8
9	S&P	SP	BBB	BBB	9
10	S&P	SP	BBB-	BBB-	10
11	S&P	SP	BB+	BB+	11
12	S&P	SP	BB	BB	12
13	S&P	SP	BB-	BB-	13
14	S&P	SP	B+	B+	14
15	S&P	SP	B	B	15
16	S&P	SP	B-	B-	16
17	S&P	SP	CCC+	CCC+	17
18	S&P	SP	CCC	CCC	18
19	S&P	SP	CCC-	CCC-	19

Por ejemplo:



- El campo *rating\_soberano* de la tabla de salida es el peor entre los ratings *sp*, *fitch* y *mdy*. Para el cálculo se requiere homologar los campos *sp*, *fitch* y *mdy* de la tabla **rating soberano** por el campo *rating\_norma* de la interfaz **homologacion\_rating**. Luego, en base al valor del campo *orden\_norma* de la tabla **homologacion\_rating** se escoge el rating homologado con valor más alto. Similar al ejemplo anterior.

### Observaciones:

- Algunos ratings del campo *sp* de la interfaz **rating soberano** tienen la letra "u" al final. Para poder homologar dicho campo se debe eliminar dicho caracter con el código Python que genera el archivo *rating\_soberano.csv*.
- Los campos que no cruzan con las tablas de homologaciones se deben dejar nulos o en blanco.

### Entregables:

El postulante deberá enviar un archivo .zip con la respuesta del desafío a través del mismo mail que se enviaron las instrucciones ([ssosaq@deloitte.com](mailto:ssosaq@deloitte.com), [jdiazca@deloitte.com](mailto:jdiazca@deloitte.com)) a más tardar el día y hora indicado en el mail con los siguientes puntos:

- data/**: carpeta con los archivos .csv generados.
- process.py**: código que lee archivo *data.xlsx* y exporta un archivo .csv por pestaña.
- master\_rating.py**: código que genera archivo *master\_rating.csv*.