

 PASS

SQLSATURDAY

SANTIAGO

◀ Sábado 26 de Mayo ▶

Conectando R con MS SQL Server

Nombre Speaker: Daniel Fischer

Email: dfischer@ug.uchile.cl

Blog: <http://geekosas.com/>

Twitter: @Geekosas_com (Artículos Blog)

GIT: <https://github.com/danielm123/sqlsaturday2018>

Gold Sponsors



Microsoft

METRICARTS

software**ONE**



punt**ticket.com**

Silver Sponsors

News**Report**^{beta}

cuatro**i**

Conectando R con MS SQL Server

R y MS SQL son el complemento perfecto.

- data.frame ~ table.
 - Tipos de datos equivalentes
 - Estructura columnar
 - Existen interfaces
- RAM y Disco Duro.
 - SQL permite preprocesar datasets grandes.
 - SQL permite almacenar información persistente.
 - R permite realizar transformaciones complejas.
 - R permite realizar modelos y reportes.
 - R permite controlar el flujo.

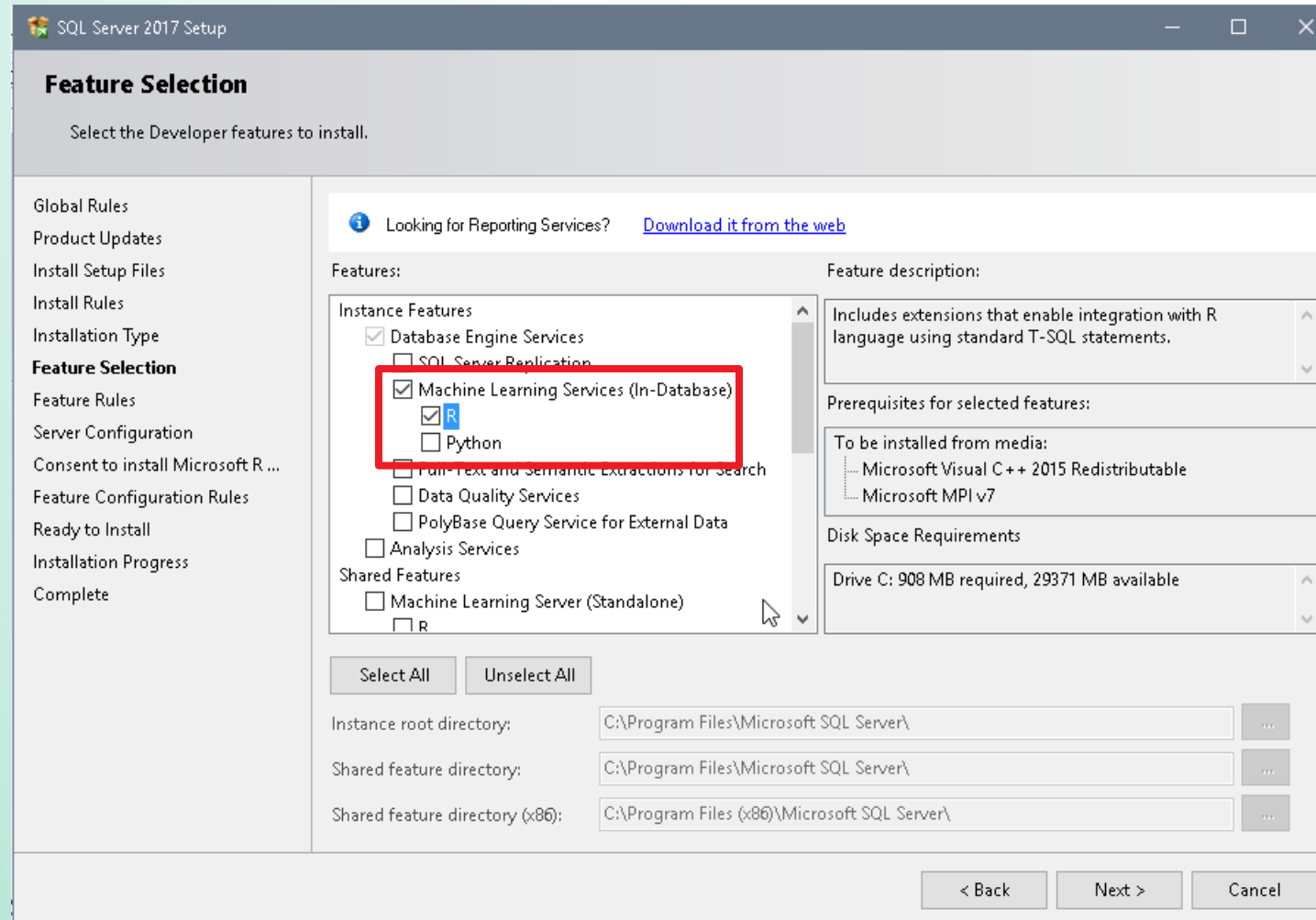
Formas de Conectarse

- R → SQL Server
 - Distintas Librerías.
 - Como extraer datos de SQL Server.
 - Como guardar datos en SQL Server.
- SQL Server → R
 - Ejecutar R desde SQL Server (Cualquier Versión).
 - Usar R desde SQL Server 2017.

R → SQL Server

Package	RSQLServer	RJDBC	RODBC	odbc
Descripción	Package con driver de JDBC.	Genérico para drivers JDBC	Usa framework ODBC	wrpapper ODBC a DBI
Framework R	DBI	DBI	RODBC	DBI
Tipo Driver	JDBC	JDBC	ODBC	ODBC
tidyverse	X	X		X
escalable			X	X
requiere driver		X	X	X
transacciones	X	X		X
linux setup			X	X
windows setup				

SQL Server 2017 → R (Machine Learning Services)

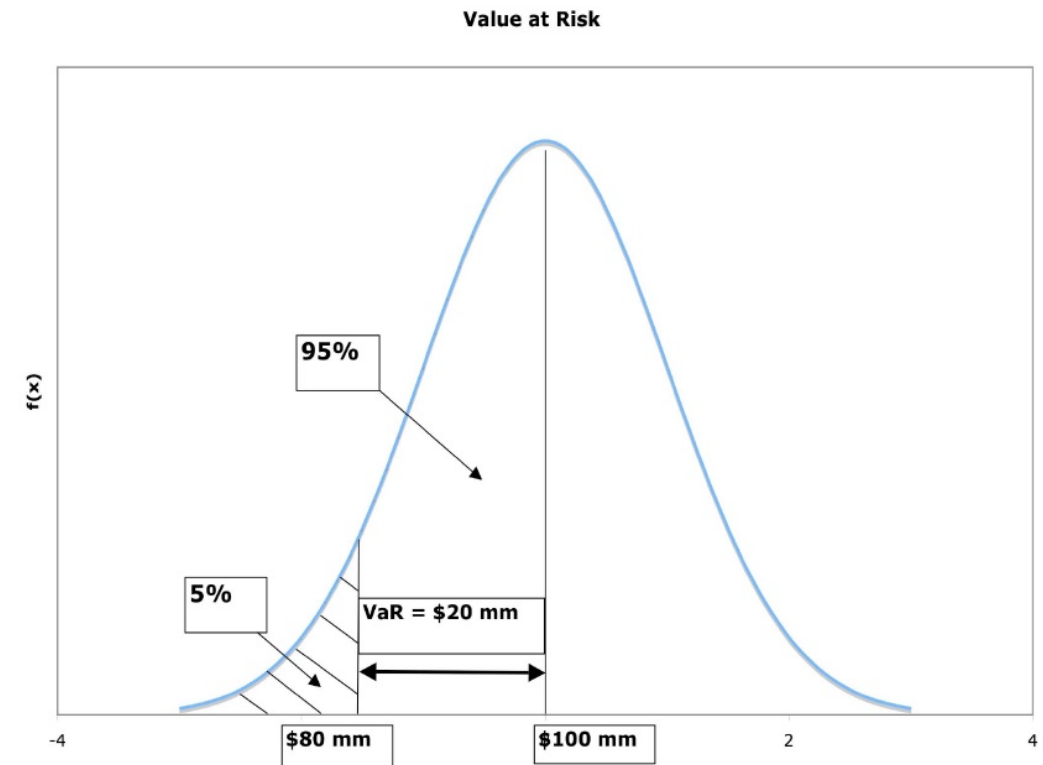


Caso VaR

Series de Retornos → Matriz Varianza Covarianza = MVC
Pesos Portafolio = W

$$\text{var} = W^t * MVC * W$$

$$\text{VaR} = 1.96 * \text{var}$$



VaR en SQL vs R

- SQL Server

```
1 --logaritmo de los valores
2 select fecha,
3     country,
4     log(value) log_value into #log_country
5 from country
6 order by country, fecha
7
8 -- retorno logartimico
9 select actual.fecha,
10     actual.country,
11     actual.log_value log_actual_value,
12     prev.log_value log_prev_value,
13     actual.log_value - prev.log_value log_proffit into #log_proffit
14 from #log_country actual
15 left join (select dateadd(MM,1,fecha) fecha, country, log_value from #log_country) as prev
16 on prev.fecha = actual.fecha and prev.country = actual.country
17 where prev.log_value is not null
18
19 -- retorno promedio
20 select country, avg(log_proffit) mean_log_proffit into #mean_log_proffit
21 from #log_proffit group by country
22
23 -- matriz varianza covarianza
24 select country_a.country country_a,
25     country_b.country country_b,
26     sum((country_a.log_proffit - mean_log_proffit_a.mean_log_proffit) * (country_b.log_proffit - mean_log_proffit_b.mean_log_proffit))/(count(*)-1) cov into #cov
27 from #log_proffit country_a left join #log_proffit country_b
28 on country_a.fecha = country_b.fecha
29 left join #mean_log_proffit as mean_log_proffit_a
30 on mean_log_proffit_a.country = country_a.country
31 left join #mean_log_proffit as mean_log_proffit_b
32 on mean_log_proffit_b.country = country_b.country
33 group by country_a.country, country_b.country
34 order by country_a.country, country_b.country
35
36 -- varianza portafolio
37 select sum( portafilo_b.peso*portafilo_a.peso*cov.cov) varianza
38 from #cov cov left join portafolio portafilo_a
39 on cov.country_a = portafilo_a.country
40 left join portafolio portafilo_b
41 on cov.country_b = portafilo_b.country
```

- R

```
1 library(odbc)
2 library(tidyverse)
3 con = dbConnect(odbc(),.connection_string = "DRIVER=SQL Server;SERVER=192.168.56.101;PORT=1433;DATABASE=sqlsat;UID=dfischer;PWD=daniel;TDS_Version=8.0;")
4
5 paises = dbGetQuery(con,"select * from country")
6 mvc = paises %>%
7     spread(country,value) %>%
8     select(-fecha) %>%
9     mutate_all(function(x) c(NA,diff(log(x)))) %>%
10     var(na.rm = T, use = "pairwise.complete.obs")
11
12 portafolio = dbGetQuery(con,"select * from portafolio order by country")
13
14 #VAR
15 portafolio$peso %>% mvc %>% portafolio$peso
16
```

Calculos de VaR, formas de conectar R a SQL.

- VaR en SQL puro.
- VaR en R.
- VaR en R y guardado en SQL.
- VaR usando integración de R con SQL Server 2017 (sp_execute_external_script).
- VaR usando R desde SQL Server (xp_cmdshell).
- VaR modo batch desde linux.

Buenas Practicas.

- Usar un wrapper para las Querys.
- Integrar un log en el sql_wrapper.R e integrar ese log a errores.
- Limitar el uso de memoria RAM en SQL Server.
- Homologar nombres de columnas en query.
- Acelerar escritura usando paralelización en R.
- Manipulación de datos compleja en R.
- Si tabla es muy grande, usar data.table.

Próximas Presentaciones

01:00 PM - 02:00 PM	Almuerzo		
02:00 PM - 02:50 PM	Power BI in the New Age ! Cristobal Ibarra Level: Intermediate	La evolución de SQL Server - Azure SQL Database Managed Instance Javier Villegas Level: Intermediate	From Zero to Cosmos Patricio Cofre, Nestor Campos Level: Beginner
03:10 PM - 04:10 PM	Midiendo el éxito en una implementación Corporate de Power BI Gaston Cruz Level: Beginner	SQL Server Query Tuning and Optimization Kamal Valero Level: Beginner	SQL Server in Docker a new default Linux Carlos Aravena Level: Beginner
04:20 PM - 05:20 PM	AzureML Services, Workbench and Model Management Service. Adrian Fernandez Level: Intermediate	SQL Server 2017 Automatic Tuning Alejandro Cordero Level: Beginner	Encontrando el origen del problema PABLO JAVIER FERNANDEZ Level: Intermediate
05:20 PM - 06:00 PM	Rifa y Cierre		



◀ Gracias por vuestra asistencia ▶
