Assessment of Disruption

Abandoning Enterprise Data Science Software Posit:

Impacts, Risks, and Mitigation Strategies

DATE: 8 August 2025

AUDIENCE: Luis Oquendo, Liz Conder

AUTHOR: Bill Prisbrey

VERSION: version A0

PURPOSE: The purpose of this report is to describe the impact and risks of abandoning the enterprise data science software Posit, and suggest mitigation strategies. This scope is limited to Bill Prisbrey’s work.

**EXECUTIVE SUMMARY:**

In order to avoid a renewal fee of $5,538 Posit Workbench is being abandoned. This will reduce Prisbrey’s work efficiency as the laptop is up to twenty times slower on many benchmarks. It risks disruption to current and historical projects due to things like moving them from Linux to Windows, and data loss due to no longer using the laptop as a thin client. It could be mitigated by establishing multiple VMWare sessions (which introduces its own risk due to being an untested alternative.)

Current customers and projects include studying retention and turnover (and identifying employees likely to separate) for Dr. Rothwell (Office of Sponsored Projects) and Ms. Topyrik (Human Resources Management.)

**IMPACTS:**

* *Reduced efficiency.* Not only is the laptop much slower on many benchmarks, but the laptop becomes unusable for all other work and applications when training a model.
* *Reduced ability to share work* via Shiny apps hosted on Connect (at a separate renewal of $19,309)

**RISKS:**

* *Disruption to current projects* (and re-visiting historical projects) takes time to repair
  + Linux to Windows
* *Loss of data*
  + Laptop is no longer used as a thin client but hosts critical files
* *Unknown quality of alternatives*

**MITIGATION STRATEGIES:**

* Establish multiple VMWare sessions to avoid loss of laptop ability while training large models and loss of data by resuming using the laptop as a thin client.

**CURRENT CUSTOMERS AND PROJECTS:**

Dr. Rothwell:

* HR metrics related to turnover and retention of principal investigators; identification of investigators likely to separate.

Ms. Topyrik:

* Expand the study of turnover and retention for Dr. Rothwell to all University employees.

**APPENDIX A: BENCHMARK**

**A.1: Posit workbench and laptop compared**

The laptop performs ten to twenty times slower on matrix calculations, five to twenty times slower on matrix functions, and thirty to forty percent faster on a programmation benchmark.

**A.2: Posit Workbench**

> source("~/Benchmark script vB0.R")

I. Matrix calculation

---------------------

Creation, transp., deformation of a 2500x2500 matrix (sec): 0.527

2400x2400 normal distributed random matrix ^1000 (sec): 0.968

Sorting of 7,000,000 random values (sec): 1.501

2800x2800 cross-product matrix (sec): 0.913

Linear regression over a 3000x3000 matrix (sec): 2.311

II. Matrix functions

---------------------

Determinant of a 2000x2000 matrix (sec): 0.375

Eigenvalues of a 640x640 matrix (sec): 0.548

FFT over 4,000,000 random values (sec): 1.376

Cholesky decomposition of a 3000x3000 matrix (sec): 1.126

Inverse of a 1600x1600 random matrix (sec): 0.329

III. Programmation

------------------

3,500,000 Fibonacci numbers calculation (sec): 3.437

Creation of a 3000x3000 Hilbert matrix (sec): 0.334

Grand common divisors of 400,000 pairs (sec): 1.758

Creation of a 500x500 Toeplitz matrix (sec): 0.009

Escoufier's method on a 45x200 matrix (sec): 0.015

**A.3: Laptop**

> source("~/Software and Systems/Posit/Renewal 2025/Benchmark script vB0.R")

I. Matrix calculation

---------------------

Creation, transp., deformation of a 2500x2500 matrix (sec): 0.54

2400x2400 normal distributed random matrix ^1000 (sec): 20.19

Sorting of 7,000,000 random values (sec): 0.83

2800x2800 cross-product matrix (sec): 12.33

Linear regression over a 3000x3000 matrix (sec): 22.3

II. Matrix functions

---------------------

Determinant of a 2000x2000 matrix (sec): 1.64

Eigenvalues of a 640x640 matrix (sec): 1.64

FFT over 4,000,000 random values (sec): 0.74

Cholesky decomposition of a 3000x3000 matrix (sec): 19.78

Inverse of a 1600x1600 random matrix (sec): 3.89

III. Programmation

------------------

3,500,000 Fibonacci numbers calculation (sec): 2.33

Creation of a 3000x3000 Hilbert matrix (sec): 0.13

Grand common divisors of 400,000 pairs (sec): 1.23

Creation of a 500x500 Toeplitz matrix (sec): 0

Escoufier's method on a 45x200 matrix (sec): 0

**A.4: Benchmark script**

# R-benchmark-25.R (Cleaned & Fixed)

# Based on Philippe Grosjean's benchmark, fixes Escoufier bug for modern R

library(Matrix)

cat("\nI. Matrix calculation\n---------------------\n")

# 1. Matrix creation, transpose, deformation

gc()

t1 <- system.time({

m <- matrix(runif(2500\*2500), nrow = 2500, ncol = 2500)

m <- t(m)

dim(m) <- c(1250, 5000)

})

cat("Creation, transp., deformation of a 2500x2500 matrix (sec):", t1[3], "\n")

# 2. Matrix exponentiation

gc()

t2 <- system.time({

m <- matrix(rnorm(2400\*2400), nrow = 2400, ncol = 2400)

m <- m %\*% m

m <- m %\*% m

})

cat("2400x2400 normal distributed random matrix ^1000 (sec):", t2[3], "\n")

# 3. Sorting

gc()

t3 <- system.time({

sort(runif(7000000))

})

cat("Sorting of 7,000,000 random values (sec):", t3[3], "\n")

# 4. Cross-product

gc()

t4 <- system.time({

crossprod(matrix(rnorm(2800\*2800), nrow = 2800))

})

cat("2800x2800 cross-product matrix (sec):", t4[3], "\n")

# 5. Linear regression

gc()

t5 <- system.time({

X <- matrix(rnorm(3000\*3000), 3000, 3000)

y <- rnorm(3000)

solve(t(X) %\*% X, t(X) %\*% y)

})

cat("Linear regression over a 3000x3000 matrix (sec):", t5[3], "\n")

cat("\nII. Matrix functions\n---------------------\n")

# 6. Determinant

gc()

t6 <- system.time({

det(matrix(rnorm(2000\*2000), nrow = 2000))

})

cat("Determinant of a 2000x2000 matrix (sec):", t6[3], "\n")

# 7. Eigenvalues

gc()

t7 <- system.time({

eigen(matrix(rnorm(640\*640), nrow = 640))

})

cat("Eigenvalues of a 640x640 matrix (sec):", t7[3], "\n")

# 8. FFT

gc()

t8 <- system.time({

fft(rnorm(4\*10^6))

})

cat("FFT over 4,000,000 random values (sec):", t8[3], "\n")

# 9. Cholesky

gc()

t9 <- system.time({

chol(crossprod(matrix(rnorm(3000\*3000), nrow = 3000)))

})

cat("Cholesky decomposition of a 3000x3000 matrix (sec):", t9[3], "\n")

# 10. Inverse

gc()

t10 <- system.time({

solve(matrix(rnorm(1600\*1600), nrow = 1600))

})

cat("Inverse of a 1600x1600 random matrix (sec):", t10[3], "\n")

cat("\nIII. Programmation\n------------------\n")

# 11. Fibonacci numbers

gc()

fib <- function(n) {

if (n < 2) return(n)

return(fib(n-1) + fib(n-2))

}

t11 <- system.time({

sapply(1:30, fib)

})

cat("3,500,000 Fibonacci numbers calculation (sec):", t11[3], "\n")

# 12. Hilbert matrix

gc()

t12 <- system.time({

1 / outer(1:3000, 1:3000, "+")

})

cat("Creation of a 3000x3000 Hilbert matrix (sec):", t12[3], "\n")

# 13. GCD

gc()

gcd <- function(a,b) { while(b != 0) { temp <- b; b <- a %% b; a <- temp }; a }

t13 <- system.time({

mapply(gcd, sample(1:10000, 4e5, replace=TRUE), sample(1:10000, 4e5, replace=TRUE))

})

cat("Grand common divisors of 400,000 pairs (sec):", t13[3], "\n")

# 14. Toeplitz matrix

gc()

t14 <- system.time({

toeplitz(1:500)

})

cat("Creation of a 500x500 Toeplitz matrix (sec):", t14[3], "\n")

# 15. Escoufier's method (FIXED)

gc()

t15 <- system.time({

library(stats)

library(utils)

library(Matrix)

library(graphics)

p <- 45

X <- matrix(rnorm(p\*200), ncol = p)

R <- cor(X)

vrt <- 1:p

vr <- rep(NA, p)

vr[1] <- which.max(colSums(R^2, na.rm=TRUE))

vrt <- setdiff(vrt, vr[1])

for (j in 2:p) {

cors <- colSums(R[vr[1:(j-1)], vrt, drop=FALSE]^2, na.rm=TRUE)

if (length(cors) == 0) break

vr[j] <- vrt[which.max(cors)]

vrt <- setdiff(vrt, vr[j])

}

})

cat("Escoufier's method on a 45x200 matrix (sec):", t15[3], "\n")

**APPENDIX B: EXPORTING FROM WORKBENCH**

This is a list of files and projects currently in Workbench that are variously ignored, exported, or copied or updated to GitHub. They are organized according to their original customer.

Generic

dotCalendar (IN GITHUB)

Draw Confusion Matrix (IN GITHUB)

Rothwell

Exploratory Analysis of OSP (6.2024) EXPORTED

Grants-Exploratory (IN GITHUB and EXPORTED)

Researcher-Turnover-in-Windows (IN GITHUB and EXPORTED)

St Andre

Course Feedback Jan 2023 EXPORTED

Course Feedback July 2024 EXPORTED

Course Feedback June 2024 EXPORTED

Demonstrations

First Round Instructor Feedback EXPORTED

GPA\_Comparison (IN GITHUB)

Three Year Graduation Aug 2022 EXPORTED and COPIED TO GITHUB

Topyrik

Headcount-App (IN GITHUB)

Individual\_Journey\_App (IN GITHUB)

Pay Equity Audit (Apr 2024) (EXPORTED)