

Saturday, April 19, 2008

About frogs, storks and Linux - or: The nature of pseudo correlations

This is really interesting. I've talked about this effect at my presentation on Friday. There is an effect in statistics called pseudo correlation. You have an pseudo correlation, when you can proof a correlation between two attributes, albeit the real relation lies somewhere else.

A good example: In years with many storks, more babies are born. And well, we were able to proof this relation in the past. But i hope all my reader are aware of the fact, that babies are not delivered by storks. The real correlation is somewhere else. Animals tend to reproduce them more effectively in years with good conditions. And albeit we try to hide it, humans are nothing than mammals. Okay, in former years wet years were good years for growing food. But wet years are good for frogs as well. And storks like frogs ... for lunch. More frogs, more storks. More water, more frogs. More water, more food. More food more babies. Albeit the amount of storks and the amount of babies seems to have a correlation, it's an indirect one.

Okay, getting from mammals, storks, babies and frogs back to the topic of this blog entry. I found an article in SearchEnterpriseLinux.com: "Move from Solaris to RHEL boosts performance for the Chicago Mercantile Exchange. In the first moment i thought: What the fsck ...? But then i've read the article and it's pseudo correlation time again.

The CME had Sun Server in the past. These Sun systems operated with Solaris. They purchased new servers. They use RedHat on them. Okay, that's bad for us, but hey ... that's life. The strange part are the conclusions in this articles.

What would you think, when you read the headline. "Solaris is so slow ... and Linux is much better". I want Linux. I want this boost, too.

When you read the article, you come to a different conclusion: Their baseline were Sun servers from 2000-2003. That's 8 years ago. In IT this is 32 hype cycles (a hype a quarter ... good rule of thumb) in the market and at least 3 real technology cycles (UltraSPARC III, UltraSPARC IV and IV+, Sparc64 VI and CMT) at Sun. They've redesigned their business logic. And obviously their application got faster by the help of Mr. Moore.

And now i come to the big pseudo correlation: The migration to Linux has paid off in improved performance, lower cost and greater stability, Kuttly said. Mr. Kuttly ... i don't want to insult you, but this is nonsense. Your performance boost didn't come from Linux, it came from using faster servers and redesigning your business logic. At least this is the real conclusion out of this article. I'm pretty sure, that with Solaris under the hood you would have yielded the same effect with new hardware and the rewritten application.

But wait ... when you really read the quote, he is pretty aware of the fact, that it wasn't Linux that gave him the boost and that it was an effect of the migration instead. But SearchEnterprise linux got this totally wrong. So, what's the sense behind this story? There is none ... i assume, at least in regard of the headline. Well ... pseudo correlations are still a standard tactic of marketing and to answer the "Let us know what you think about the story". Well, Mrs. Derringer ... i have to admit .. nothing good.

Posted by Joerg Moellenkamp in English, The IT Business at 22:05

Do you have any citation(s) to back up your stork/human correlation?
Anonymous on Apr 20 2008, 10:35

My professor in statistics used this example to explain pseudo correlations. But when you search for "Scheinkorrelation" and "Storch" you will find several examples.
Anonymous on Apr 20 2008, 10:54

Mr. Kuttly??? Isn't it Whoopi Goldberg???? g
Anonymous on Apr 21 2008, 10:17