

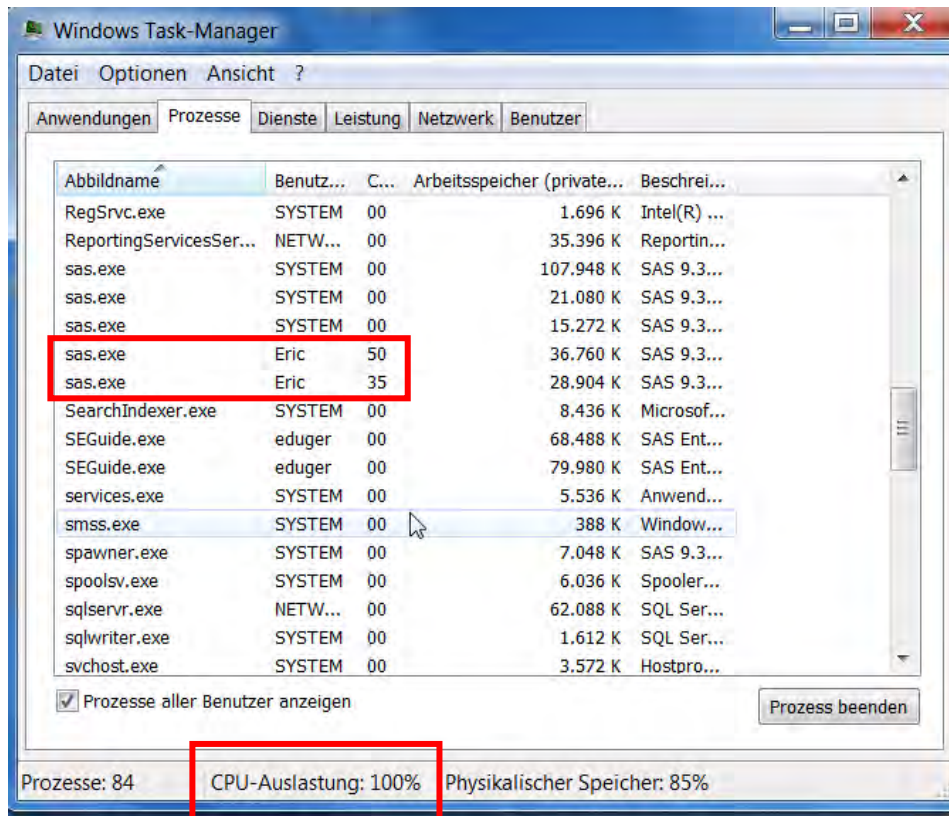
Don't touch this... oder Dinge, die man im SAS Enterprise Guide nicht tun sollte...

– Teil 1 - Performancekiller

Der SAS Enterprise Guide ist ein sehr intuitives und dabei trotzdem ausgesprochen mächtiges Graphical User Interface. Neben einer kompletten Programmieroberfläche bietet er eine Menge Point- and Click Anwendungsroutinen, mit denen man vom Einlesen von Excel-Dokumenten über SQL-Abfragen, bis hin zu komplexen Graphiken und Vorhersagemodellierungen mit ein paar Klicks alles erreichen kann. Aber Point- and Click bedeutet immer auch die Gefahr, dass man in Performanceengpässe läuft, weil man nicht genügend über die Funktionsweise und Hintergrundprozesse der Anwendung weiß. Hierfür habe ich im Teil 1 meiner Rubrik mal zwei klassische Beispiele ausgewählt.

Mehrere Enterprise Guides öffnen

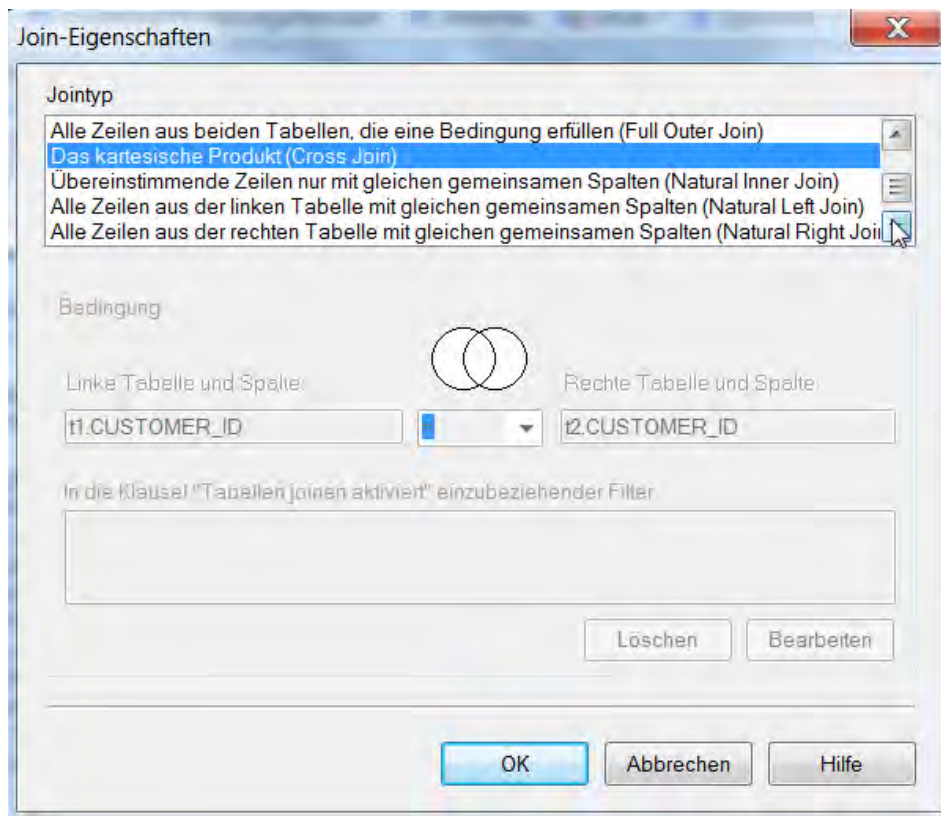
Ha! Erwischt! Sie haben eine riesige und lang dauernde Analyse abgeschickt und weil Ihnen das Warten aufs Ergebnis zu lange gedauert hat, haben Sie einen zweiten Enterprise Guide geöffnet und mit diesem eine zweite aufwändige Analyse losgeschickt und weil Sie jetzt gerade auf beide Ergebnisse warten, haben Sie Zeit, meinen Artikel zu lesen. Was ist passiert? Wenn Sie im Enterprise Guide auf Ausführen drücken, wird das von Ihnen erzeugte Programm auf einen externen Server geschickt. Dort wird für jeden geöffneten Enterprise Guide eine Workspace-Server Sitzung eröffnet. Das ist eine SAS-Sitzung (sas.exe), die die vom Enterprise Guide abgeschickten Programme abarbeitet. Wenn Sie mehrere SAS Enterprise Guides öffnen und Programme abschicken, gibt es auf dem Server mehrere konkurrierende SAS-Sitzungen. Bei älteren SAS-Versionen und Ein-Prozessor Computern musste noch die zweite SAS-Sitzung warten, bis die Erste fertig ist. Heute auf Mehr-Prozessor Maschinen, mit Load-Balancing Möglichkeiten, können die verschiedenen SAS-Sitzungen auf unterschiedliche Prozessorkerne verteilt werden. Zwei Enterprise Guides machen Ihre Arbeit also tatsächlich schneller - vorausgesetzt Sie sind der einzige Mitarbeiter Ihrer Firma. Wenn Sie sich jetzt vorstellen, dass jeder Ihrer Kollegen auch gerade an mehreren Enterprise Guides arbeitet und mehrere sas.exes für sich gestartet hat, dann geht die Gesamtperformance des Servers ziemlich in den Keller. Ich habe für mein Beispiel mal 2 Guides geöffnet und jeweils einen aufwändigen Prozess abgeschickt. Wie Sie in dem Screenshot unten sehen können, bringe ich damit meine CPU schon auf 100% Auslastung (Ich bin als Eric im Enterprise Guide angemeldet):



Cross Join

Der größte Performance- und Speicherkiller, den man momentan im SAS Enterprise Guide findet, ist der Cross Join. Falls Sie im Abfragefenster zufällig mal im „Tabellen joinen“...-Menü auf den Cross Join gestoßen sind und überlegt haben, drauf zu klicken: Tun Sie's besser nicht!

Der Cross Join macht ein sogenanntes Kartesisches Produkt. D.h. er berechnet alle erdenklichen Kombinationen aller Variablenausprägungen aus der linken Datei mit allen Variablenausprägungen aus der rechten Datei. Einen Cross Join sollte man nur machen, wenn die beiden Ausgangsdateien sehr klein sind. Sind die Ursprungsdateien groß, wird das Ergebnis riesig. Die Wahrscheinlichkeit, dass Sie den Ihnen zugewiesenen Speicherplatz komplett vollschreiben ist sehr hoch und natürlich dauert die Rechenoperation ausgesprochen lange.



Ich habe das mal für Sie ausprobiert:

Gejoint habe ich die beiden Dateien Order_Fact (Größe: 82 MB) und Customer_Dim (Größe: 18 MB). Nach ca. einer dreiviertel Stunde ist meine Crossjoin-Datei knappe 48 Gigabyte groß und es ist noch kein Ende in Sicht. Ich breche das Experiment jetzt ab.

crossjoin.sas7bdat.lck	13.06.2013 17:36	LCK-Datei	47.703.585 KB
------------------------	------------------	-----------	---------------