

# 1. Aufgabenblatt

## - Auswertung -

**Patrick Schäfer**

Berlin, 13. November 2017

[patrick.schaefer@hu-berlin.de](mailto:patrick.schaefer@hu-berlin.de)

# Aufgabenstellung

- Thema: Umgang mit großen Datenmengen und begrenzten Ressourcen
- Aufgabe: **Berechne Schnittmenge zweier Dateien**
  - Beide Dateien sind **1,2 GB groß** und bestehen aus jeweils 100.000.000 Einträgen
  - Das Ergebnis (**Schnittmenge**) soll in eine dritte Datei geschrieben werden
- Anforderung: Euer Programm **muss** mit **100 MB RAM** auskommen
- Wettbewerbsaufgabe: es gibt Punkte für die schnellste, korrekte Implementierung

# Schnittmenge

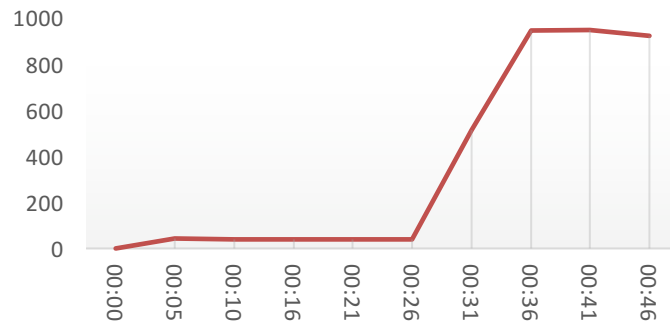
- Die Schnittmenge sollte 2097858 Ergebnisse haben.
- Messung der CPU- und Memory-Nutzung mit System Resource Usage Profiler:  
<https://github.com/jeetsukumaran/Syrupy>
- Zur Verfügung standen 40 Cores (80 mit Hyperthreading).

# 11 Abgaben

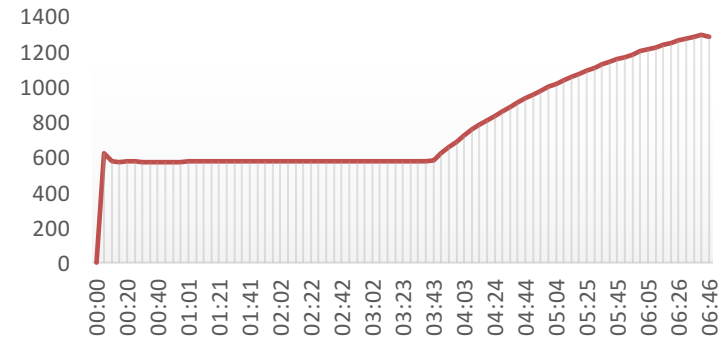
Gruppe	Ansatz
Blauer Vogel	Sorted Merge Join, Char-Arrays, Single-Threaded
CJLdwhdm	Sorted Merge Join, Single-Threaded
dsh	Naiver Join
DWHMiMoJoin2017	Hash Join, Schreiben Parallelisieren
Gruppe 1	Hash Join, Multi-Threading beim Read und Join
Join Venture	Sorted Merge Join, External Sort Multi-Threaded
MondayMorning	Hash Join, in Buckets: Sorted Merge Join. Integer Repräsentation, Join Multi-Threading, Direct Memory
SMH	Sorted Merge Join
SpeedyGonzales	Hash Join und Merge Join
TeamHA9876	Sorted Merge Join
wwgds	Sorted Merge Join (Sort Files, Merge, Remove Duplicates)

# 1.-4. Platz: CPU-Auslastung

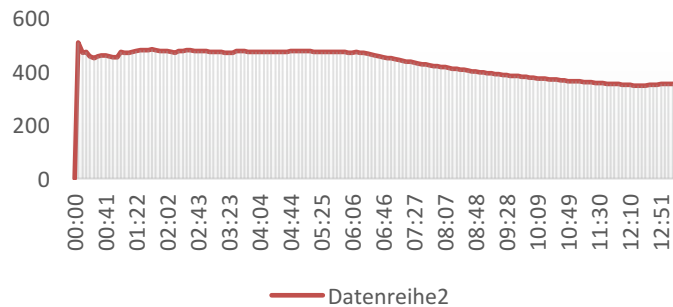
CPU



CPU

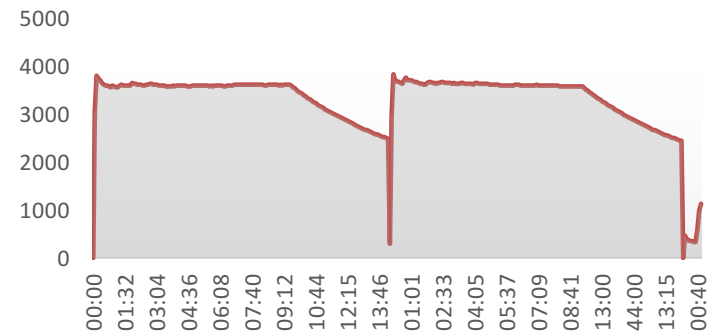


CPU

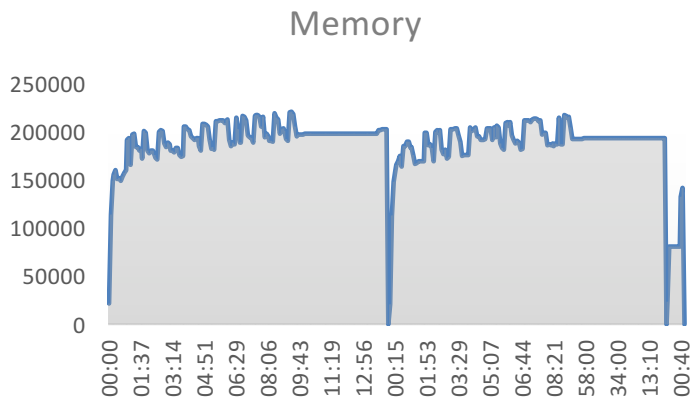
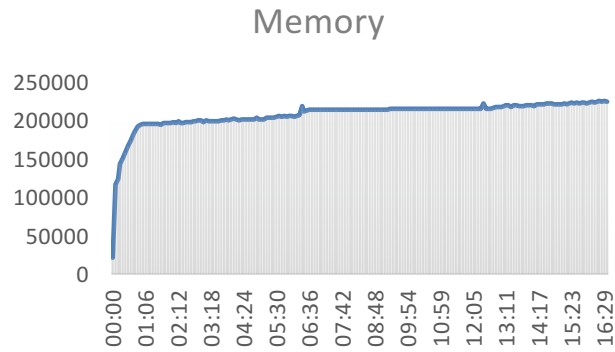
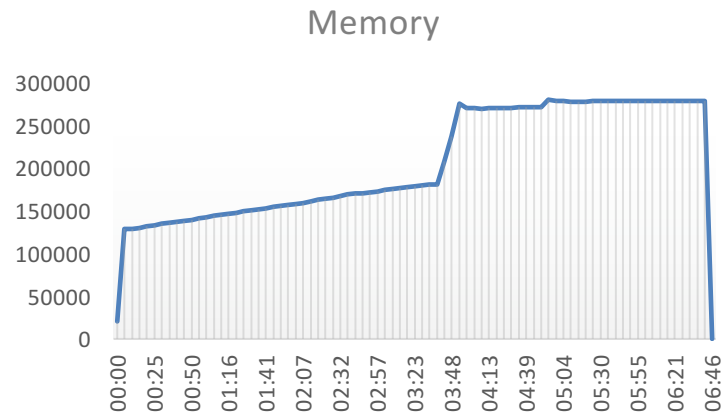
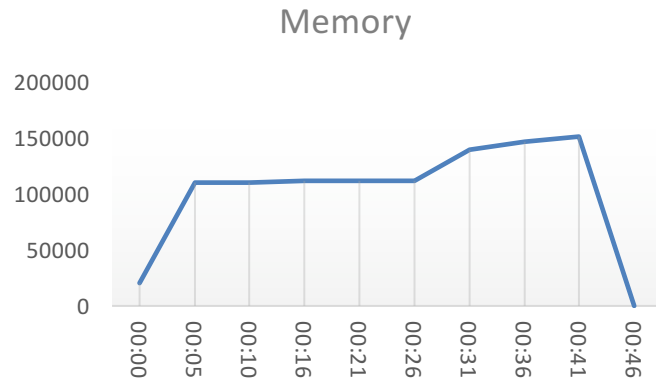


— Datenreihe2

CPU



# 1.-4. Platz: Hauptspeicherverbrauch



# 11 Abgaben

Gruppe	Gesamtdauer
Blauer Vogel	37:35 minute(s)
CJLdwhdm	26:37 minute(s), 26:22 minute(s)
dsh	pending (80 hour(s) ...)
DWHMiMoJoin2017	05:15 minute(s), 06:51 minute(s)
Gruppe 1	23:33 minute(s), 23:19 minute(s)
Join Venture	19:50 minute(s), 19:54 minute(s)
MondayMorning	00:51 minute(s), 00:51 minute(s)
SMH	42:24 minute(s)
SpeedyGonzales	?
TeamHA9876	30:16 minute(s), 29:30 minute(s)
wwgds	29:29 minute(s), 29:17 minute(s)

# 11 Abgaben

Gruppe	Ergebnis
Blauer Vogel	ok
CJLdwhdm	ok
dsh	pending
DWHMiMoJoin2017	ok
Gruppe 1	Falsche Ergebnisanzahl
Join Venture	Falsche Ergebnisanzahl (56)
MondayMorning	ok
SMH	Falsche Anzahl Ergebnisse
SpeedyGonzales	3 Ansätze? OutOfMemory und falsche Anzahl
TeamHA9876	ok
wwgds	ok



# Wettbewerbspunkte

Platz	Gruppe	Gesamtdauer	Punkte
1	MondayMorning	00:51 minutes	5
2	DWHMiMoJoin2017	05:15 minutes	3
3	CJLdwhdm	26:37 minutes	2
4	wwgds	29:29 minutes	
5	TeamHA9876	30:16 minutes	
6	Blauer Vogel	37:35 minutes	
7	SMH	42:24 minutes	

Fragen?

# Anmerkungen

- Gruppe: wwgd  
– Mail schicken für Oracle Account