

Ophidia Analytics Framework

Nele Lips, Dominik Scherer
19.03.2018

Projekt „Big Data“

Betreuer: Julian Kunkel, Jakob Lüttgau



Universität Hamburg

Motivation

- Wissenschaftliche Arbeit produziert große Mengen an Daten
 - Schwierig effizient zu verarbeiten
- Ophidia soll Möglichkeit bieten, Daten parallel und damit effizient zu verarbeiten
- Ophidia auf Leistungsfähigkeit testen

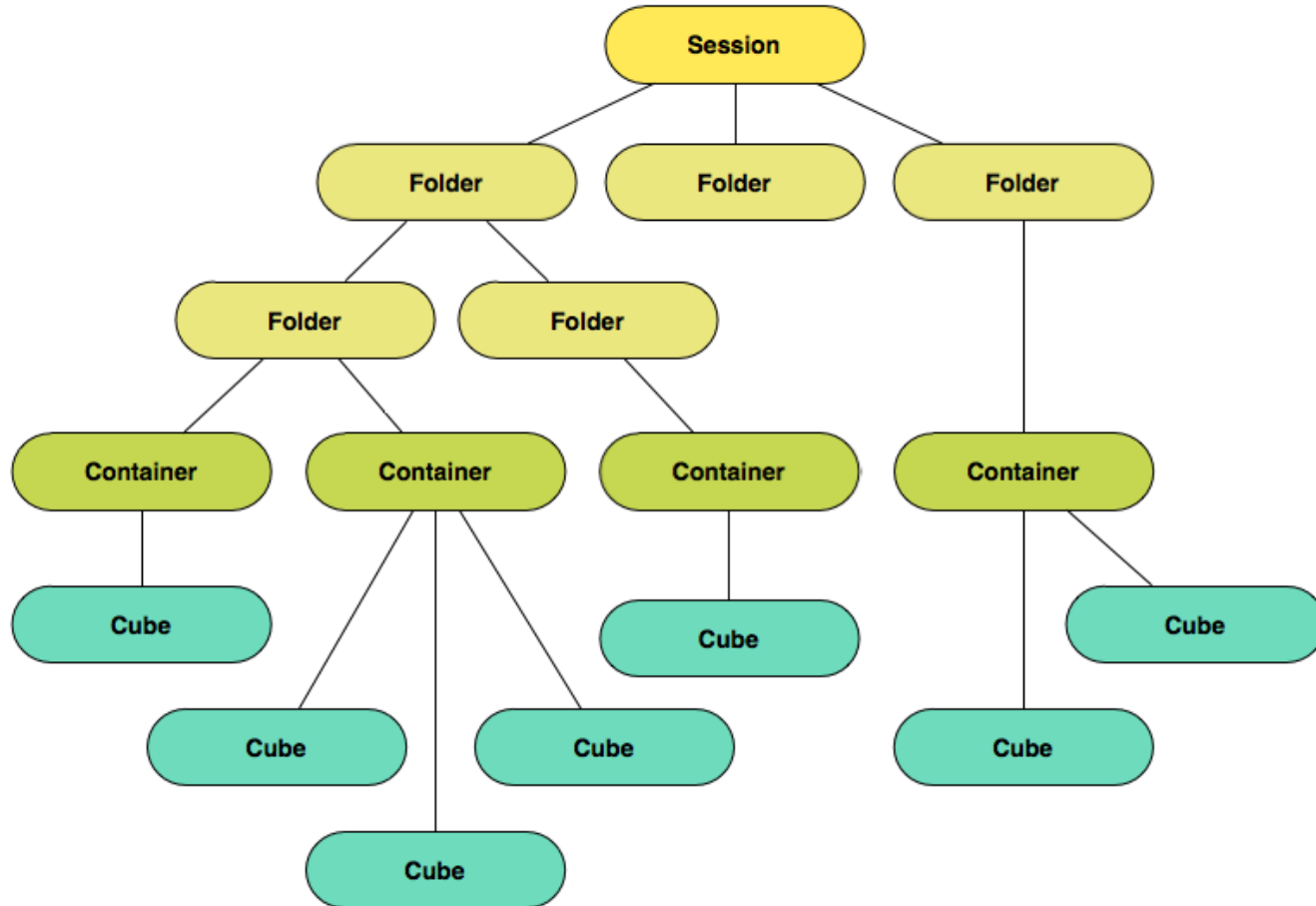
Agenda

- Einführung in Ophidia
- Probleme
- Benchmark Resultate
- Zusammenfassung

Einführung in Ophidia

- Framework zur parallelen Verwaltung, Analyse und Verarbeitung von Daten
- Bietet Funktionalitäten wie
 - Array-based Primitives
 - Datacube Operatoren
 - Single und Massive Operations
 - Workflows (JSON)
- Verarbeitet Daten lokal und minimiert Kommunikation zwischen Knoten

Einführung in Ophidia



Quelle: Ophidia.cmcc.it, CMCC Foundation, [1]

Einführung in Ophidia

- Beispiel Datacube
 - Messwert: Temperatur
 - Explizite Dimensionen: Breiten- und Längengrade
 - Implizite Dimension: Zeit

lat	lon	temperature
10.147903	46.190708	328.7889404297, 590.7707519531, 37.4648475647, 291.2965698242, 293.7053222656
10.147903	108.230979	883.2327880859, 822.5951538086, 901.1722412109, 840.0346069336, 502.5177001953
10.147903	143.593633	223.6679992676, 948.8746337891, 960.8160400391, 765.1795654297, 706.2957763672
10.147903	232.499386	317.4991455078, 185.2340240479, 51.6501235962, 974.2478637695, 329.8284912109
10.147903	236.835995	342.1051025391, 628.0217285156, 12.0141677856, 25.8362998962, 60.1650810242
10.147903	316.002566	559.1904296875, 982.9323730469, 578.8946533203, 332.0242919922, 867.5264892578
10.147903	374.621298	483.7551879883, 660.8132934570, 458.2972106934, 521.2200317383, 952.1098022461
10.147903	397.249060	752.0025634766, 404.4528198242, 774.7050170898, 653.1747436523, 244.4874420166
10.147903	463.691942	277.2226867676, 876.8427734375, 193.3621063232, 238.0387115479, 642.0223999023
10.147903	545.840266	899.6579369141, 555.5379028320, 827.2564086914, 951.3079833984, 529.7857666016

Probleme

■ Installation:

- Viele Abhängigkeiten
- Viele Konfigurationsdateien
- Dokumentation teilweise nicht mehr aktuell
- Meiste Probleme konnten gelöst werden, jedoch:

Internal server error: no response has been received from analytics framework

■ Begrenzungen unserer Cluster Accounts

- Rechenzeit
- Datacube Größe

Probleme

- Kommunikation mit Entwicklern vom CMCC eher schwergängig und langsam
- Rechenintensiver Benchmark schwer zu bewerkstelligen, meist eher speicherintensiv
- Zunächst keine Möglichkeit gefunden, Zeit technisch zu messen
 - Ophidia Terminal 1.2.0 Ende Februar released

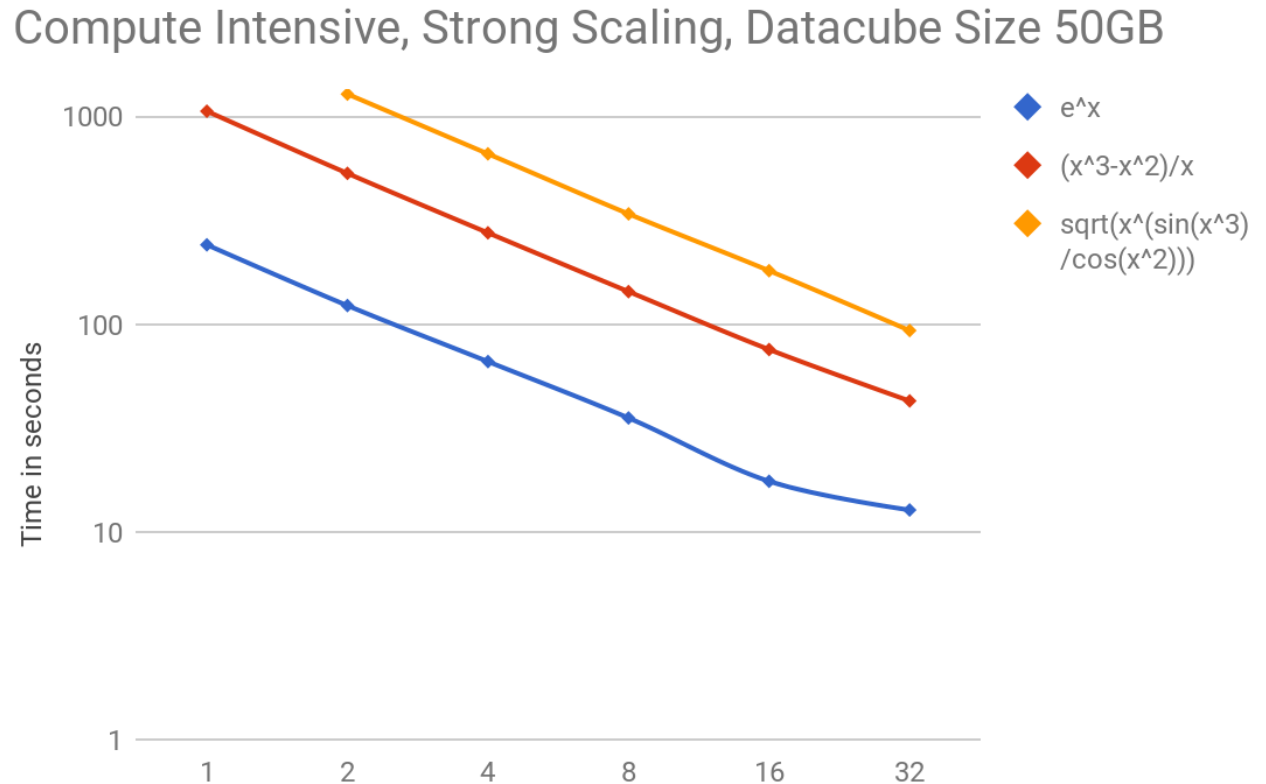
Benchmark

- Tests in 2 Kategorien: Compute intensive und I/O+Network
- Benchmarks mit:
 - Steigender Kern Anzahl bei gleicher Last
 - Steigender Last bei gleicher Kern Anzahl
- Zufällig erstellte Testdaten (OPH_RANDCUBE)

Benchmark Resultate

- Datacube Größe: 50GB
- Wende Formel $(x^3-x^2)/x$ auf alle Elemente an (rote Linie)

number of cores	time in seconds
1	1060.24
2	532.32
4	275.22
8	143.50
16	75.51
32	42.71

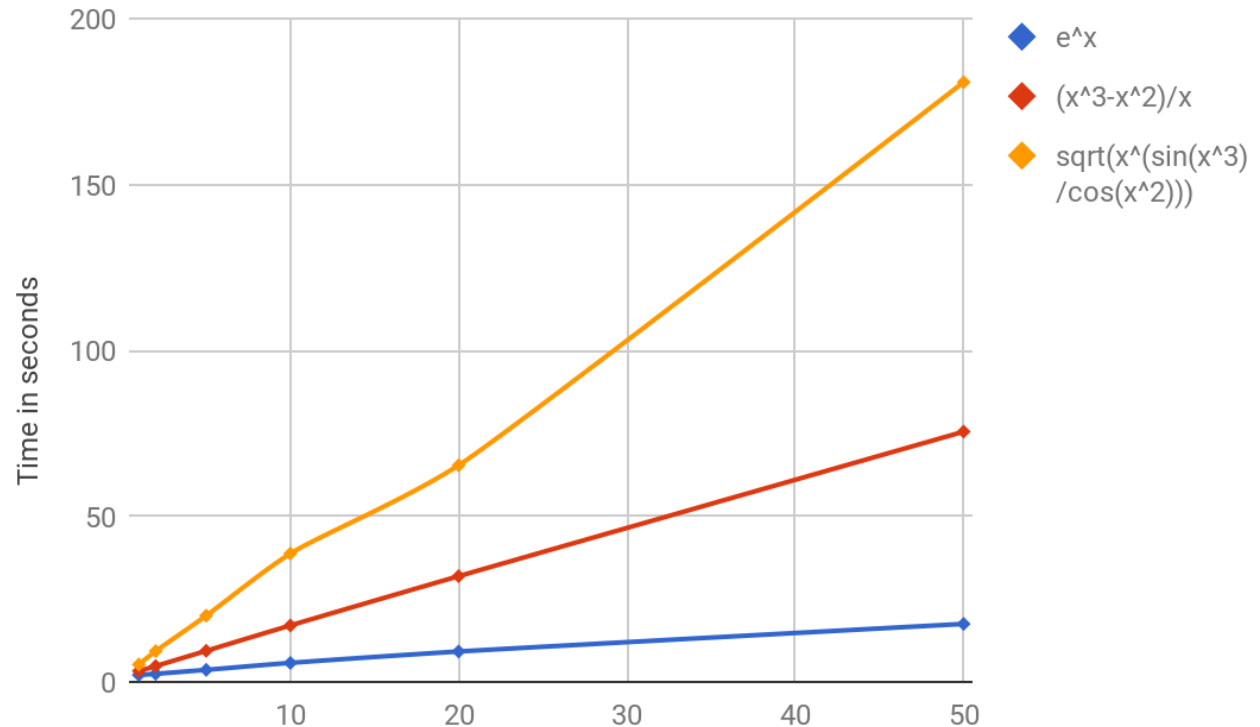


Benchmark Resultate

- Anzahl Cores: 16
- Wende Formel $(x^3-x^2)/x$ auf alle Elemente an (rote Linie)

Datcube size in GB	time in seconds
1	3.32
2	4.82
5	9.41
10	17.07
20	31.93
50	75.51

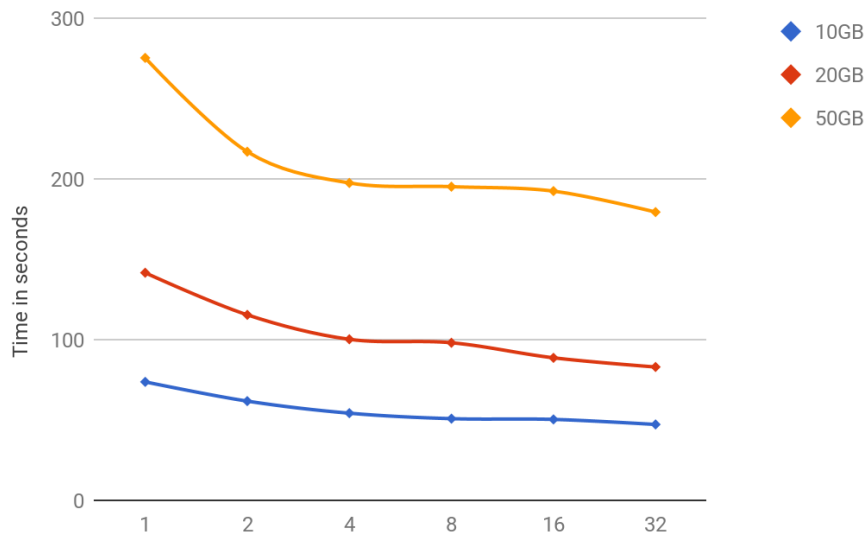
Compute Intensive, Workload Scaling, 16 Cores



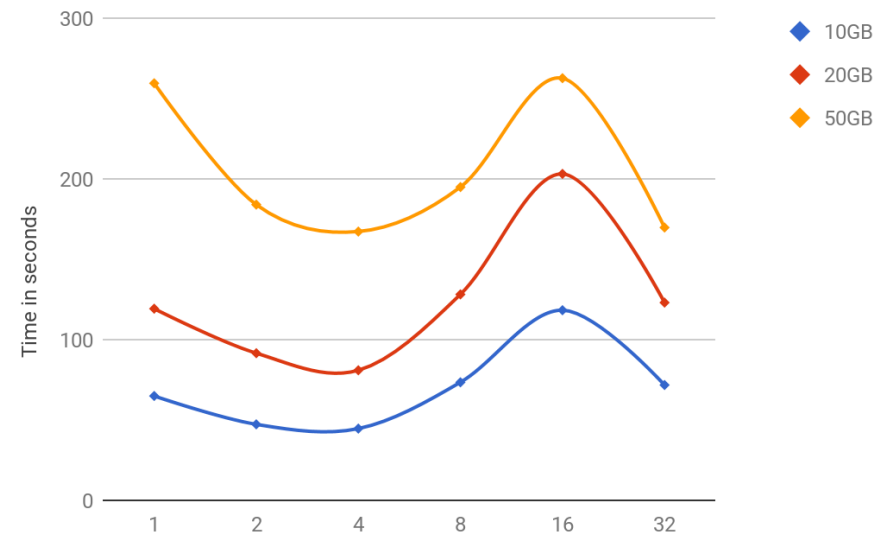
Benchmark Resultate

- EXPORT von Ophidia zu NetCDF
- IMPORT von NetCDF zu Ophidia

I/O & Network Intensive, Strong Scaling, EXPORT



I/O & Network Intensive, Strong Scaling, IMPORT



Zusammenfassung

- Ophidia am besten geeignet für Operationen auf Arrays und großen Datenmengen
 - Daten werden vorwiegend lokal verarbeitet
- Schwierige Installation und mittelmäßige Dokumentation
- Relativ unkomplizierte und intuitive Benutzung

- Quellen:

[1] <http://ophidia.cmcc.it/documentation/users/vfs/index.html>