

Visual Analytics für Raumzeitdaten



Task 2

Markus Richter

Christopher Schiefer

Datentyp

- Relation
 - Siehe 1. Meilenstein: ausreichend komplex für Anwendung
- In-Memory
 - Daten sind ausreichend klein
 - Realtime-Verarbeitung notwendig
 - Verteiltes Berechnen/Bereitstellen der Daten unnötig

Implementierung

- ***Python***
 - Flexible, schöne Sprache mit sehr vielen guten Libraries:
- ***pandas:***
 - Sehr weit verbreitet für Datenanalyse, gut unterstützt
 - Schnelle und benutzerfreundliche Manipulation von Daten
- ***matplotlib:***
 - Einfache, aber mächtige Visualisierungs-library

Implementierung

- Daten in pandas-Datenframes:
 - 2-dimensionale Struktur für tabellarische Daten
 - Unterstützt schon sehr viele Operationen auf Daten
 - Nachteil: keine Sortierung/Baumstruktur, es muss bei jeder Operation die gesamte Liste durchgegangen werden
- DataSource als Dictionary
 - Neue Tabellen unter gegebenem Namen eingefügt

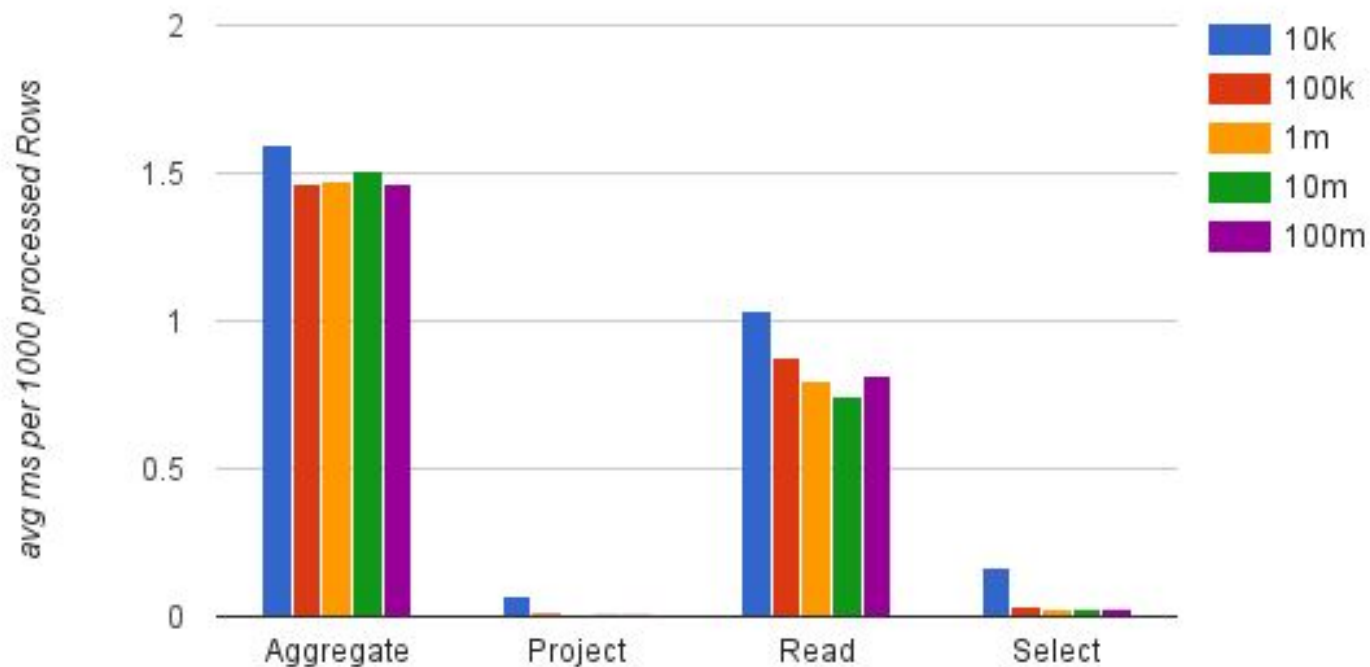
Features von Data-Source

- Select, Project, Aggregate
 - Mit entsprechenden Argumenten für Attribute usw.
 - Eingabetabelle spezifizierbar
- Read(filename)
- getdata(tablename)
- get_base_data()
- get_all_data()

Benchmark-Daten

- Zufällige Datenpunkte erzeugt
 - 4 Spalten mit gleichen Eigenschaften wie Originaldaten
 - ... gleicher Datentyp, gleicher Wertebereich
 - In Dateien abgespeichert
- 10 Durchläufe für Messungen
- Bedingungen zufällig gewählt
 - Attribute und Werte für Aufrufe unterschiedlich
 - Aber gleiche Bedingungen für alle Datensatzgrößen

Benchmark-Daten



Laufzeitverhalten

- Lineare Skalierung
- Project und Select vernachlässigbar
- Nur Aggregation wäre zu laufzeitkritisch
- Probleme bei überlaufendem RAM
 - ab 50M Einträgen auf Notebook

Beweisskizze

Identify: Simple Selektion mit entsprechender Bedingung, evtl. mehrmals angewendet für komplexere Bedingungen

Compare: ?

Determine: Selektionen für gegebene Bedingungen und anschließende Projektion für benötigte Attribute

Arbeitsaufteilung

- Gemeinsames Arbeiten, nur im Kleinen differenzierbar