Projekt 1: Semantische Datenanalyse

Programmiersprachen nach Wunsch

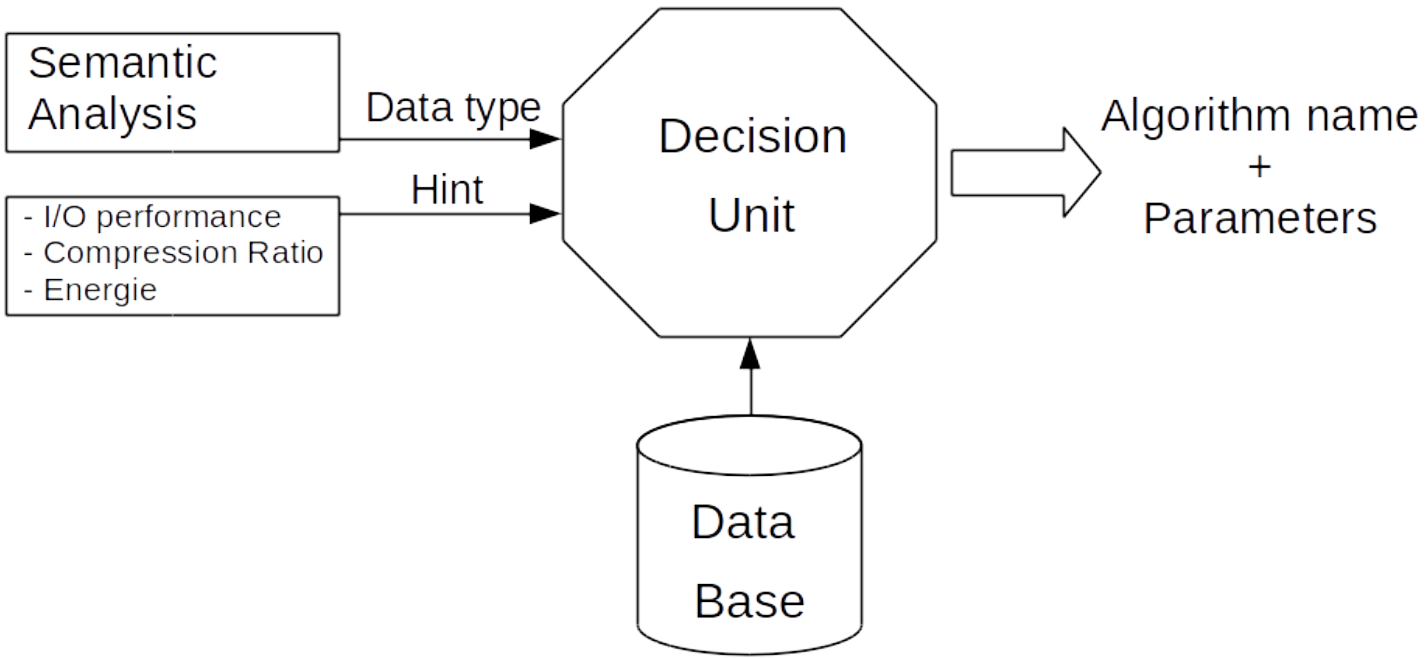
Schritt für Schritt

1. Daten aus verschiedenen Forschungsbereichen auswählen
   - Gendaten, Klimadaten, aus physikalischen Experimenten, ...
2. Verschiedene Kompressionsalgorithmen aussuchen
3. Kompressionsalgorithmen auf Daten anwenden
4. Die Resultate mit bestimmten Metriken in eine DB packen
5. Datentypen identifizieren
   - Z.B. random, spektrale, stetige Daten, ...
6. Bonus: Prototyp einer “Decision-Unit” erstellen

Projektziele

1. Häufig benutzte Datentypen in der Wissenschaft finden
2. Für jeden Datentyp den besten Kompressionsalgorithmus (inkl. Parameters) herausfinden
   - `CompAlg get_comp_alg(Datentyp);`

Bonus


Semantic Analysis
Data type
Hint
- I/O performance
- Compression Ratio
- Energie
Decision Unit
Algorithm name
+
Parameters
Data Base

Projekt 2: Paralle HDF5 Compression

Sprache: Python/Bash, C

1. NetCDF-Bench
   - Parallel I/O Benchmark
   - Geschieben in C-Programmierspache
   - `git clone git@github.com:joobog/netcdf-bench.git`
2. Parallele Kompression in HDF5
   - Beta-Feature in HDF5
   - `ftp://gamma.hdfgroup.org/pub/outgoing/jhenderson/releases/parallel_compression_hdf5/`
3. SCIL
   - Metakompressor
   - HDF5-SCIL-Filter bereits implementiert
   - `git clone git@github.com:JulianKunkel/scil.git`

Projektziele

Schritt für Schritt

1. Notwendige Erweiterung von Netcdf-bench
   - Einlesen von Eingangsdaten
2. Benchmarking durchführen und die E/A Leistung messen
3. Verschiedene Varianten vergleichen (ohne Kompression, mit deflate Kompression, SCIL)