DATENBANKGESTÜTZTE DATENVERWALTUNG IM WORKFLOW-TOOL JUBE

Entwurf, Implementierung und Analyse einer datenbankgestützten Datenverwaltung in der internen Ablaufsteuerung des Workflow-Tools JUBE

18. AUGUST 2023 I JULIA WELLMANN I JÜLICH SUPERCOMPUTING CENTRE

MOTIVATION

Workflow-Tool JUBE

- Konfigurieren, Ausführen, Verwalten, Optimieren von Workflows
- Beispielsweise Benchmarking oder Parameterstudien
- HPC- und lokale Systeme (Laptop)
- Konfiguration der Workflows durch Eingabedatei
- Speicherung und Weiterverarbeitung der Workflow-Konfigurationen

MOTIVATION

Forschungsfrage

- Workflow-Tool JUBE speichert Workflow-Konfigurationen
- Speichern von Informationen über Laufzeit hinaus
- Nachteile der XML-Datenhaltung: Speicher und Laufzeit
- Ersetzung durch schnellere, zuverlässigere und speicherorientiertere Datenhaltung

GRUNDLAGEN

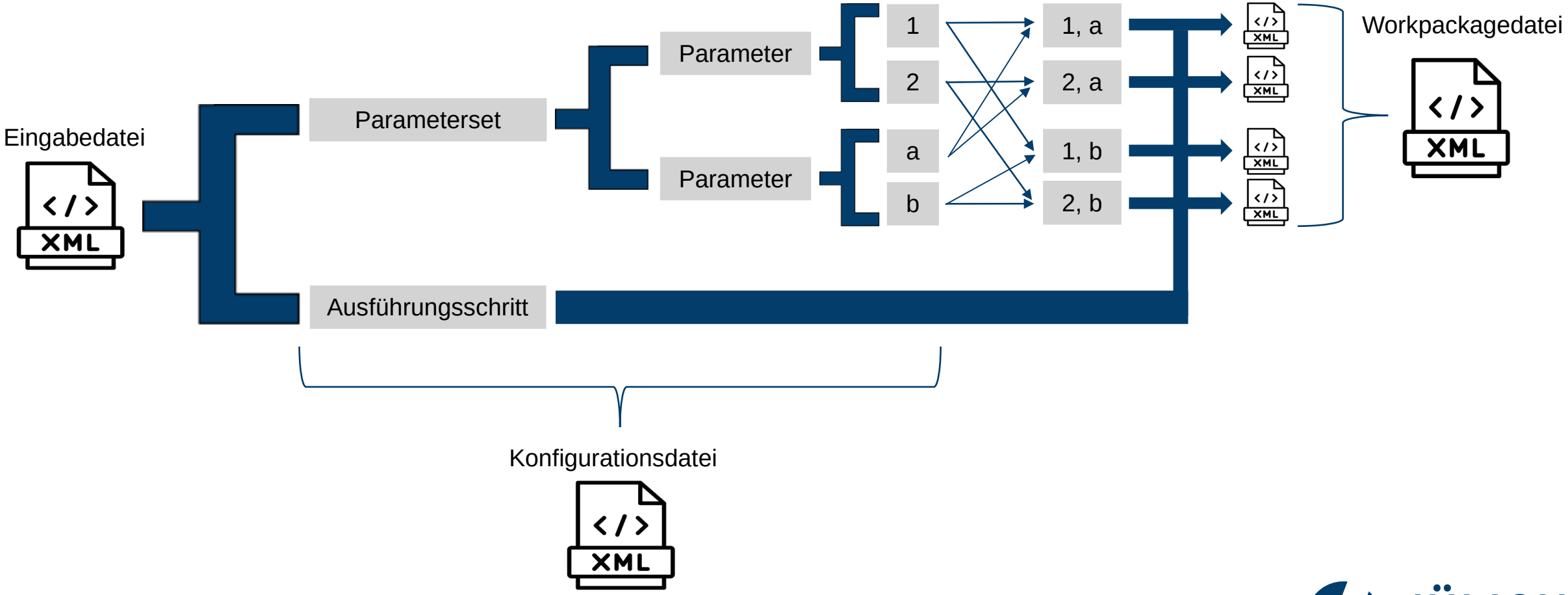
Datenverwaltung

- Datenhaltung: Speicherung in bestimmten Format/Struktur
- Datenverwaltung: Prozess der Speicherung

Textdatei	Relationale Datenbanken	NoSQL-Datenbanken
		

AUSGANGSSITUATION

XML-Datenhaltung



AUSGANGSSITUATION

Problemanalyse: Ursachen

Speicherplatzbedarf	Leistung
Redundante Speicherung	Redundante Speicherung
XML-Formatierung	Keine Datentypnutzung
	Aufwendige Erzeugung
	Neuschreiben statt Aktualisieren

- Zudem: Persistenz und Konsistenz

ANFORDERUNGEN

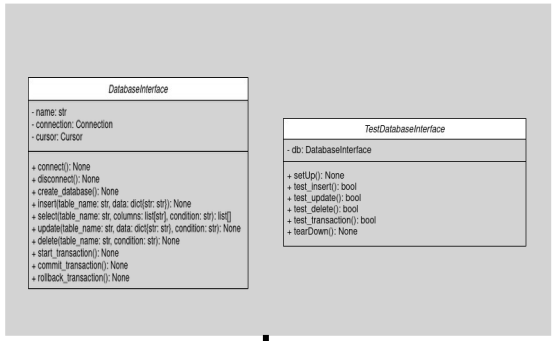
Nach ISO-Norm 25010

Funktionale Anforderungen	Nicht-funktionale Anforderungen
Informationen	Leistung
Dateneingabe	Sicherheit
Datenverarbeitung	Kompatibilität
Datenintegrität	Zuverlässigkeit
	Benutzerfreundlichkeit
	Wartbarkeit
	Übertragbarkeit

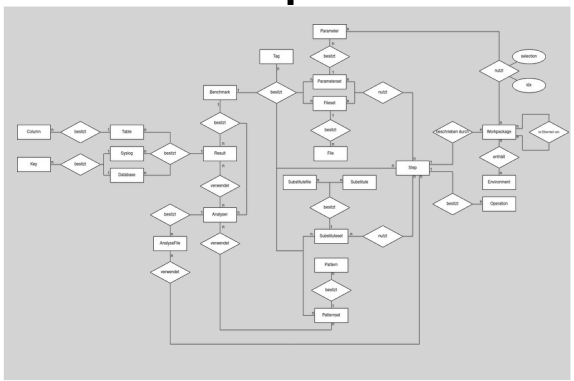
➡ Lokale Datenbank, Schnittstelle, Tests und Transaktionen

ENTWURF

Vorgehen



LÖSUNGSENTWURF → **DATENBANKENTWURF** → **SCHNITTSTELLE** → **KONZEPT IN JUBE**

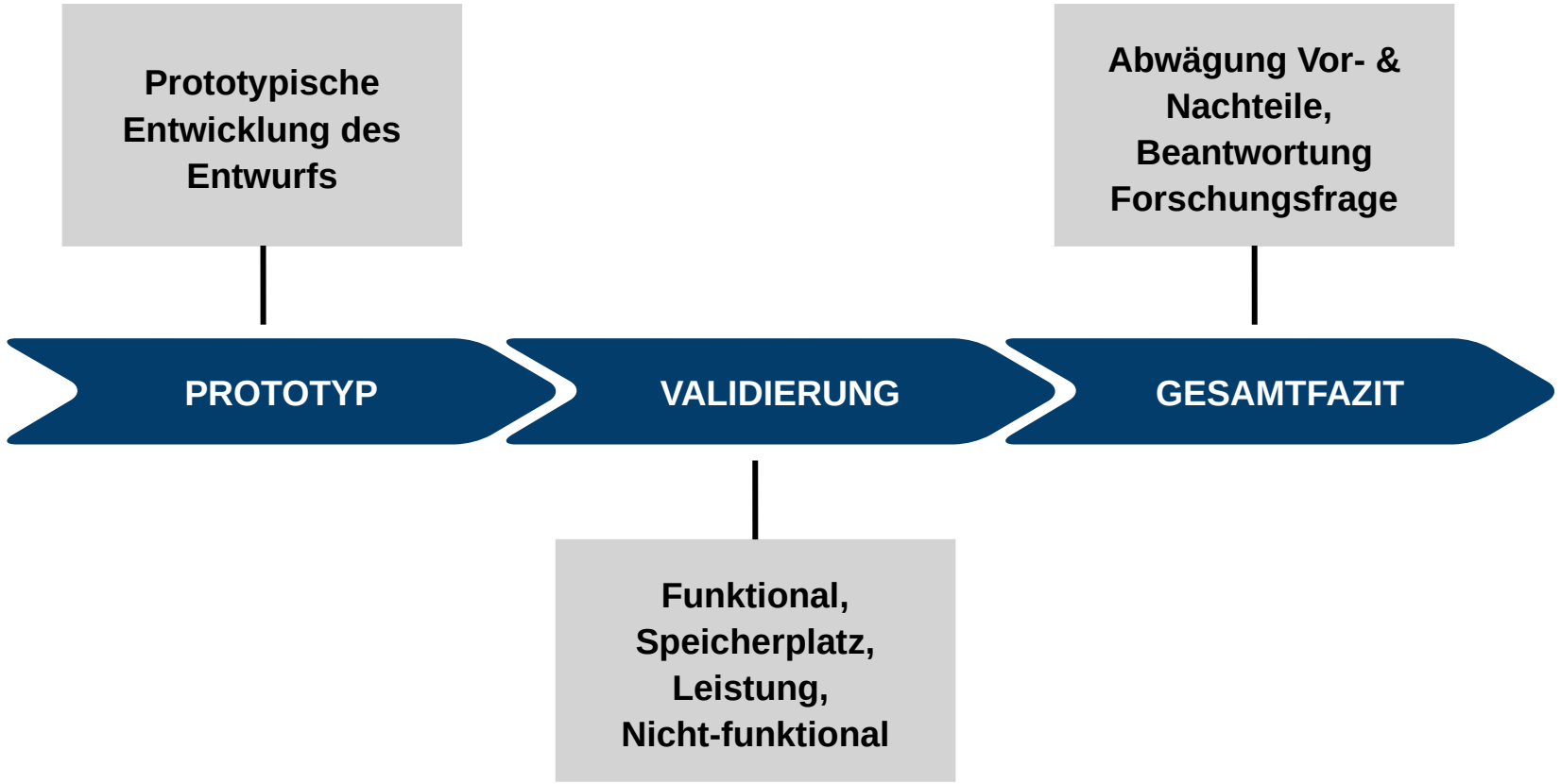


Anwendung der Schnittstellen-Methoden

JÜLICH Forschungszentrum

BEWERTUNG

Vorgehen



BEWERTUNG

Speicherplatzbedarf

Beispiel	Speicherplatz XML-Dateien	Speicherplatz Datenbank
Include	2 [kB]	254 [kB]
Scripting_parameter	7 [kB]	254 [kB]
Tagging	1 [kB]	254 [kB]
Hello_world_iter	128 [kB]	279 [kB]
Dependencies	185 [kB]	300 [kB]
Stream_Benchmark	2.406 [kB]	918 [kB]

synthetisch

anwendungsnah

➡ Verbesserung für anwendungsnahe Beispiele

BEWERTUNG

Leistung (*run*-Befehl)

Beispiel	Laufzeit XML-Dateien	Laufzeit Datenbank
Include	0,0050 [s]	0,0679 [s]
Scripting_parameter	0,0134 [s]	0,0763 [s]
Tagging	0,0019 [s]	0,0617 [s]
Hello_world_iter	16,9280 [s]	10,7989 [s]
Dependencies	50,6030 [s]	44,9279 [s]
Stream_Benchmark	1016,7354 [s]	385,6499 [s]

synthetisch

anwendungsnah

- Overhead Funktionsbeispiele: Erstellung
- Verbesserung für anwendungsnahe Beispiele

BEWERTUNG

Leistung (continue-Befehl)

Beispiel	Laufzeit XML-Dateien	Laufzeit Datenbank
Include	0,0109 [s]	0,0104 [s]
Scripting_parameter	0,0272 [s]	0,0244 [s]
Tagging	0,0043 [s]	0,0040 [s]
Hello_world_iter	14,5909 [s]	11,6174 [s]
Dependencies	54,9885 [s]	47,0704 [s]
Stream_Benchmark	624,4697 [s]	145,1249 [s]

synthetisch

anwendungsnah

➡ Verbesserung für anwendungsnahe Beispiele

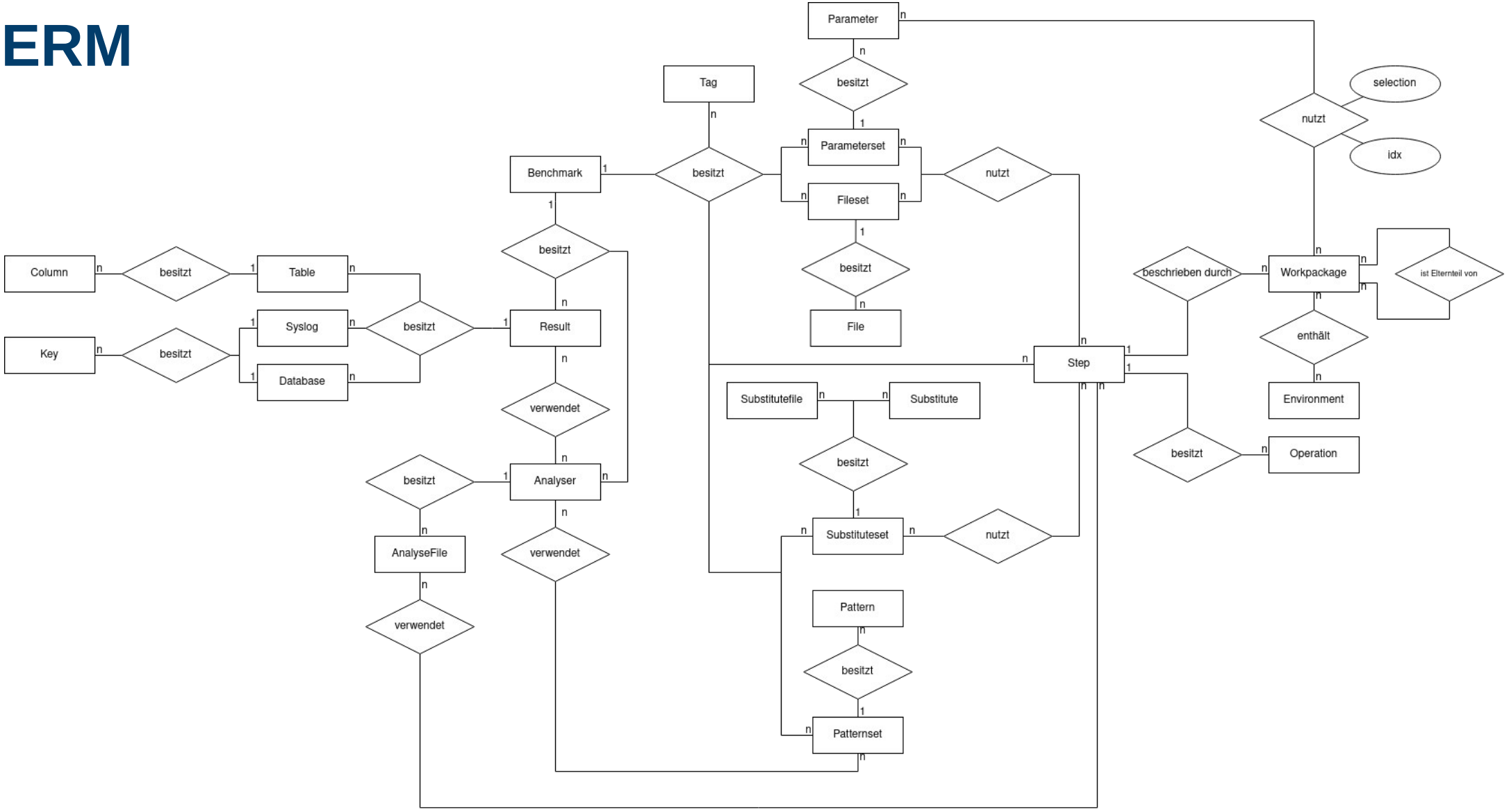
BEWERTUNG

Gesamtfazit

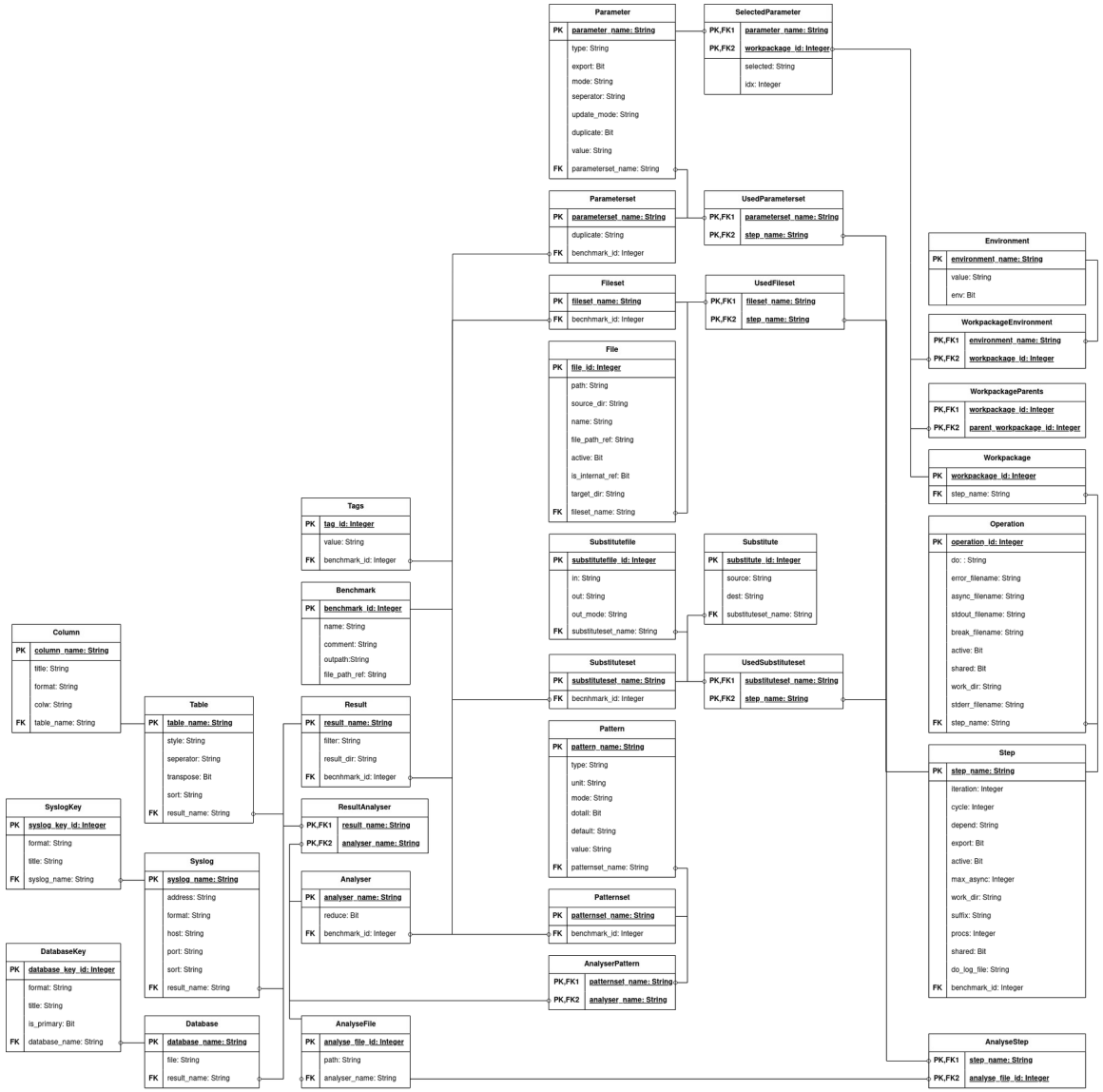
- Funktionale Gleichheit gegeben
- Leistungs- & Speicherbedarfsverbesserung für Anwendungsfälle
- Nicht-funktionale Anforderungen umgesetzt oder umsetzbar
- Ausblick:
  - Erstellung der Datenbank
  - Aktualisierung der Daten
  - Funktionale Erweiterbarkeit

➡ **Vielversprechender Ansatz zur Lösung der identifizierten Probleme**

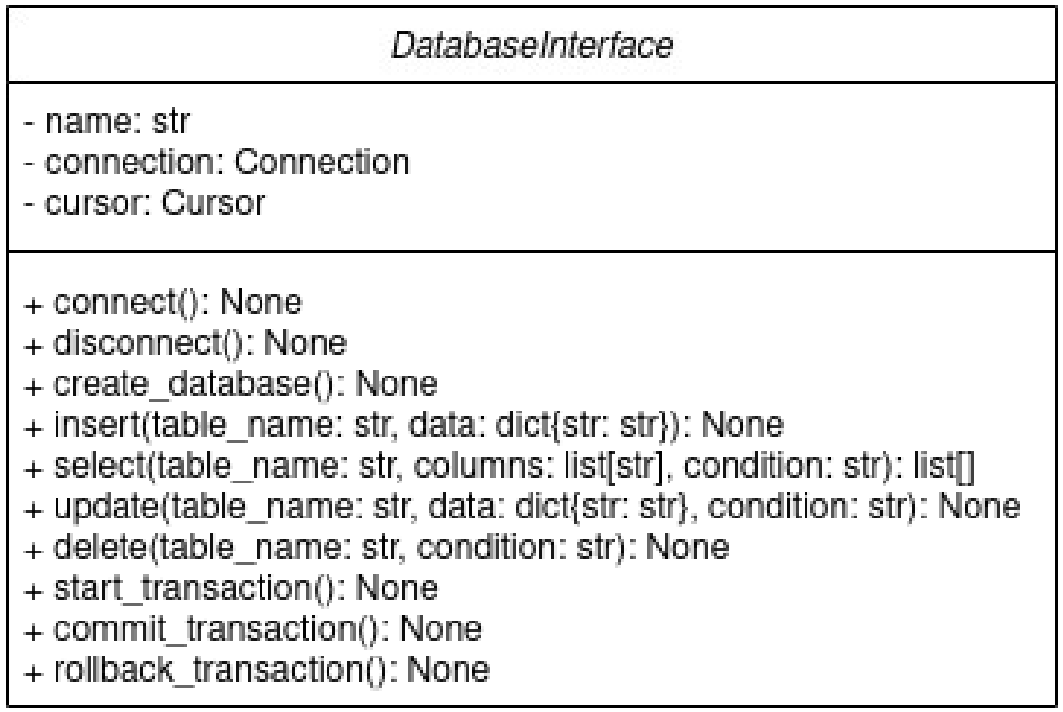
ERM

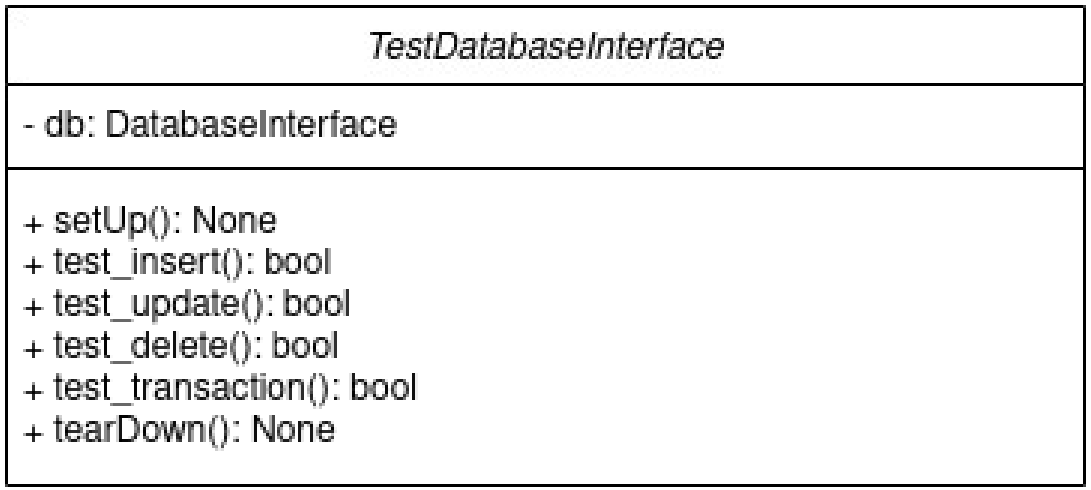

Parameter
Tag
besitzt
selection
nutzt
idx
Parameterset
Benchmark
besitzt
nutzt
Fileset
besitzt
File
Column
besitzt
Table
Key
besitzt
Syslog
Database
besitzt
Result
verwendet
Analyser
besitzt
AnalyseFile
verwendet
verwendet
Substitutefile
Substitute
besitzt
Substituteset
nutzt
Pattern
besitzt
Patternset
Step
beschrieben durch
Workpackage
ist Elternteil von
enthält
Environment
besitzt
Operation

RM


Parameter
PK parameter_name: String
type: String
export: Bit
mode: String
seperator: String
update_mode: String
duplicate: Bit
value: String
FK parameterset_name: String
SelectedParameter
PK,FK1 parameter_name: String
PK,FK2 workpackage_id: Integer
selected: String
idx: Integer
Parameterset
PK parameterset_name: String
duplicate: String
FK benchmark_id: Integer
UsedParameterset
PK,FK1 parameterset_name: String
PK,FK2 step_name: String
Environment
PK environment_name: String
value: String
env: Bit
Fileset
PK fileset_name: String
FK becnhmark_id: Integer
UsedFileset
PK,FK1 fileset_name: String
PK,FK2 step_name: String
WorkpackageEnvironment
PK,FK1 environment_name: String
PK,FK2 workpackage_id: Integer
File
PK file_id: Integer
path: String
source_dir: String
name: String
file_path_ref: String
active: Bit
is_internat_ref: Bit
target_dir: String
FK fileset_name: String
WorkpackageParents
PK,FK1 workpackage_id: Integer
PK,FK2 parent_workpackage_id: Integer
Workpackage
PK workpackage_id: Integer
FK step_name: String
Tags
PK tag_id: Integer
value: String
FK benchmark_id: Integer
Substitutefile
PK substitutefile_id: Integer
in: String
out: String
out_mode: String
FK substituteset_name: String
Substitute
PK substitute_id: Integer
source: String
dest: String
FK substituteset_name: String
Operation
PK operation_id: Integer
do: : String
error_filename: String
async_filename: String
stdout_filename: String
break_filename: String
active: Bit
shared: Bit
work_dir: String
stderr_filename: String
FK step_name: String
Benchmark
PK benchmark_id: Integer
name: String
comment: String
outpath:String
file_path_ref: String
Substituteset
PK substituteset_name: String
FK becnhmark_id: Integer
UsedSubstituteset
PK,FK1 substituteset_name: String
PK,FK2 step_name: String
Column
PK column_name: String
title: String
format: String
colw: String
FK table_name: String
Result
PK result_name: String
filter: String
result_dir: String
FK becnhmark_id: Integer
Pattern
PK pattern_name: String
type: String
unit: String
mode: String
dotall: Bit
default: String
value: String
FK patternset_name: String
Step
PK step_name: String
iteration: Integer
cycle: Integer
depend: String
export: Bit
active: Bit
max_async: Integer
work_dir: String
suffix: String
procs: Integer
shared: Bit
do_log_file: String
FK benchmark_id: Integer
Table
PK table_name: String
style: String
seperator: String
transpose: Bit
sort: String
FK result_name: String
ResultAnalyser
PK,FK1 result_name: String
PK,FK2 analyser_name: String
SyslogKey
PK syslog_key_id: Integer
format: String
title: String
FK syslog_name: String
Syslog
PK syslog_name: String
address: String
format: String
host: String
port: String
sort: String
FK result_name: String
Analyser
PK analyser_name: String
reduce: Bit
FK benchmark_id: Integer
Patternset
PK patternset_name: String
FK benchmark_id: Integer
AnalyserPattern
PK,FK1 patternset_name: String
PK,FK2 analyser_name: String
DatabaseKey
PK database_key_id: Integer
format: String
title: String
is_primary: Bit
FK database_name: String
Database
PK database_name: String
file: String
FK result_name: String
AnalyseFile
PK analyse_file_id: Integer
path: String
FK analyser_name: String
AnalyseStep
PK,FK1 step_name: String
PK,FK2 analyse_file_id: Integer

DATENBANKSCHNITTSTELLE





ORDNERSTRUKTUR