



Hochschule Neubrandenburg
University of Applied Sciences

HOCHSCHULE NEUBRANDENBURG

STUDIENGANG GEOINFORMATIK

BACHELORARBEIT

Evaluierung von PostgreSQL/PostGIS für die Verwendung als alternatives DBMS zu Oracle GeoRaster in der Anwendung *profil rdv*

Florian Kaselow

Betreuer:

Prof. Dr.-Ing. Andreas WEHRENPFENNIG

Dipl.-Inform. Jörg SCHÄFER

Dipl. Ing. Frank CZARNIKOW

urn:nbn:de:gbv:519-thesis2019-0054-6

21.03.2019



Danksagung

Ich möchte mich bei meinen Betreuern Prof. Dr.-Ing. Andreas Wehrenpfennig und Dipl.-Inform. Jörg Schäfer von der Hochschule Neubrandenburg, sowie Dipl.-Ing. Frank Czarnekow von der Firma **data experts** für die Betreuung und Unterstützung bei dieser Arbeit bedanken.

Des weiteren möchte ich mich bei den Mitarbeiter der **data experts** bedanken, die mich zu jeder Zeit fachlich unterstützten.

Abschließend möchte ich mich bei meiner Familie und ganz besonders bei meiner Mutter bedanken, ohne die das Studium und diese Arbeit nicht möglich gewesen wäre.

Kurzfassung

Im Rahmen dieser Arbeit werden die Datenbankmanagementsysteme Oracle und PostgreSQL/PostGIS hinsichtlich der Funktionalitäten mit Rasterdaten analysiert. Dabei wird Bezug auf die Anwendung ***profil rdv*** der Firma **data experts** genommen.

Es wird die Datenstruktur von Rasterdaten und Kostenmodelle des jeweiligen DBMS dargestellt. Des weiteren werden Performanceanalysen im Bereich der Importzeit, des Speicherplatzverbrauches und der Anfragezeiten über einen Web Map Service durchgeführt.

Abschließend wird eine Nutzwertanalyse anhand der Analysen für den Fall der ***profil rdv*** erstellt.

Abstract

This thesis will evaluate the database management systems Oracle and PostgreSQL/PostGIS based on their functionalities with raster data. Reference is made to the application ***profil rdv*** of the firm **data experts**.

The data structure of raster data and cost models of the respective DBMS is presented. Furthermore, performance analyzes are carried out in the area of import time, storage space usage and request times via a web map service.

Finally, a utility value analysis is created on the basis of the analyses for the case of ***profil rdv***.

Inhaltsverzeichnis

1 Einleitung	1
2 <i>profil rdv</i>	2
2.1 Was sind Rasterdaten?	2
2.2 Fachliches Modell	3
2.3 Ausgewählte Werkzeuge der <i>profil rdv</i>	3
2.4 Systemarchitektur	6
2.5 Problemstellung: WCS-Anfrage	7
3 Rasterdaten in objektrelationalen Datenbanken	9
3.1 Oracle GeoRaster	9
3.2 PostgreSQL/PostGIS raster	14
4 Mögliche Umsetzung der DBMS in der <i>profil rdv</i>	18
4.1 Oracle GeoRaster	18
4.2 PostGIS raster	19
4.3 Gegenüberstellung	21
5 Kostenmodelle von Oracle und PostgreSQL/PostGIS	22
5.1 Oracle	22
5.2 PostgreSQL/PostGIS	25
5.3 Gegenüberstellung	26
6 Performanceanalyse	27
6.1 Rahmenbedingungen der Performanceanalyse	27
6.1.1 Was ist eine Performanceanalyse?	27
6.1.2 Beschreibung der Rahmenbedingungen	27
6.2 Oracle GeoRaster	30
6.3 PostGIS raster	39
6.4 Gegenüberstellung	45
7 Bewertung der rasterbasierten DBMS	48
7.1 Erklärung Bewertungssystem	48
7.2 Kosten	50
7.3 Performance	51
7.4 Abschlussbewertung	53



8 Zusammenfassung und Ausblick	54
8.1 Zusammenfassung	54
8.2 Ausblick	55
Glossar und Abkürzungsverzeichnis	56
Literaturverzeichnis	58
Abbildungsverzeichnis	59
Tabellenverzeichnis	60
Anhang A	61

1 Einleitung

Motivation

Die **data experts** haben im Rahmen des hauseigenen Softwarepaketes das **Programm-**
system zur rechnergestützten Verwaltung von Fördermitteln in der Landwirtschaft (*profil*), welches in mehreren Bundesländern Deutschlands eingesetzt wird, mehrere einzelne Programme entwickelt. Diese dienen dem Zweck der Umsetzung vom **integrierten Verwaltungs- und Kontrollsyste**m (InVeKoS) und angrenzenden Fördermaßnahmen [DEP19]. Diese Bachelorarbeit beschränkt sich auf die ***profil* Rasterdaten-Verwaltung (*profil rdv*)**. Dieses Programm ist spezialisiert auf die Speicherung und Bereitstellung von Rasterdaten.

Die ***profil rdv*** basiert aktuell auf dem Datenbankmanagementsystem (DBMS) Oracle (in der Version 12.1) mit der Erweiterung Oracle 'Spatial and Graph'. In den Jahren der Nutzung wurden einige Kritikpunkte am aktuellen DBMS identifiziert. Zum einen fallen durch das Lizenz- und Supportmodell von Oracle sehr hohe Kosten an, obwohl nur wenige Funktionalitäten in diesem speziellen Fall genutzt werden. Ein weiterer Grund ist die nicht zufriedenstellende Antwortzeit/Performance bei Bildanfragen an die Datenbank. Eine mögliche Ursache ist die Dekompression der Daten, bevor diese zur Verfügung gestellt werden können. In diesem Bereich gibt es Optimierungsbedarf. Der letzte Grund ist der Speicherplatzverbrauch durch die Menge an Daten, die dort abgespeichert werden. Im Fall der ***profil rdv*** werden Orthophotos für das gesamte Bundesland Schleswig-Holstein abgespeichert und zur Verfügung gestellt.

Ziel der Bachelorarbeit

Mit Hilfe dieser Arbeit wird im speziellen Fall der ***profil rdv*** die Anwendung von Rasterdaten in Datenbanken zwischen Oracle und PostGIS evaluiert. Dabei ist nicht direkt eine Umsetzung oder prototypische Implementierung angestrebt, sondern nur ein Vergleich der DBMS in ihrer Struktur sowie Performanceanalysen im Bereich der Anfragezeiten, aber auch Importzeit und Speicherplatzverbrauch sind relevant. Abseits der Funktionalitäten für Rasterdaten werden auch einmalige und Folgekosten verglichen.

Abschließend erfolgt eine Bewertung anhand der Kriterien und eine Einschätzung, ob PostgreSQL/PostGIS in diesem Fall eine Alternative zu Oracle GeoRaster darstellt oder eine Optimierung mit einem Datenbankupgrade auf 12.2 ausreicht.

2 *profil rdv*

Die ***profil rdv*** ist ein Teil der Softwarelandschaft der **data experts** und dient zur übersichtlichen Darstellung und zentralen Verwaltung von Rasterdaten inklusive deren Sach- und Metadaten. Sachdaten sind zusätzliche Informationen, die nicht zu den Metadaten des Bildes direkt zugeordnet sind, z.B. Bildname und Bildflugdatum. Die Rasterdaten werden über einen Web Coverage Service (WCS) und deren Sach- und Metadaten über einen Web Feature Service (WFS) für andere ***profil*** Anwendungen zur Verfügung gestellt.

2.1 Was sind Rasterdaten?

Rasterdaten sind als ein zweidimensionales Array angeordnet. Jedes Feld des Arrays wird als Pixel bezeichnet und enthält unterschiedlich viele Informationen. Die Anzahl an Informationen ist von der Anzahl der enthaltenen Bänder im Bild abhängig. Ein Beispiel für ein Bild mit drei Bändern ist ein RGB-Bild, welches pro Pixel jeweils 3 unterschiedliche Informationen enthält: Rot-Wert, Grün-Wert, Blau-Wert.

Rasterdaten können unterschiedliche Quellen haben, z.B. einfache Fotografien oder Fernerkundungsbilder, erstellt durch Satelliten, oder Orthophotos, die von Flugzeugen aufgenommen werden. Im Fall der ***profil rdv*** werden digitale Orthopotos (DOP's) verwaltet. Diese stellen eine verzerrungsfreie und maßstabstreue Abbildung der Erdoberfläche dar (siehe Abbildung 1).



Abbildung 1: Beispiel eines DOP's [LVG19]

2.2 Fachliches Modell

In der **profil rdv** sind die Rasterdaten in Bestände & Versionen eingeteilt. Bestände bedeutet hier, dass alle Rasterdaten in einem Bestand die selbe Herkunft bzw. den selben Hersteller haben, z.B. das Landesamt für Vermessung und Geoinformation Schleswig-Holstein (LVermGeo SH). Eine Version besteht aus einer Menge von Bildern, die zu einem angegebenen Zeitraum erstellt wurden. Eine Kachel ist ein georeferenziertes Bild.



Abbildung 2: UML Diagramm der fachlichen Bezeichnungen der **profil rdv**

2.3 Ausgewählte Werkzeuge der **profil rdv**

Die **data experts** entwickelt ihre Software nach dem WAM-Ansatz (Werkzeug, Automat, Material). Die nachfolgenden Werkzeuge beschreiben Funktionalitäten der **profil rdv**, mit denen das entsprechende Material (hier Rasterdaten) bearbeitet wird. Weitere Informationen zum WAM-Ansatz sind zu finden unter:

Züllighoven, H. (1998). *Das objektorientierte Konstruktionshandbuch nach dem Werkzeug & Material-Ansatz. dpunkt.* [WAM98]

Rasterdaten-Importeur

Der Import von neuen Daten beinhaltet folgende Schritte, die in Abbildung 3 dargestellt und darunter näher erläutert sind.

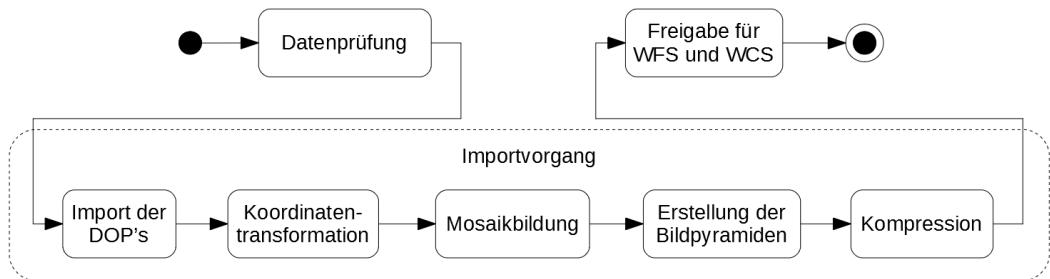


Abbildung 3: Aktivitätsdiagramm des Rasterdaten-Importeurs

Datenprüfung

Es wird eine automatische Prüfung der Sach- und Metadaten durchgeführt, zusätzlich ist eine visuelle Prüfung der DOP's durch den Nutzer möglich.

Import der DOP's

Die einzelnen Kacheln werden in die Datenbank geladen.

Koordinatentransformation

Die DOP's müssen in das genutzte Koordinatenreferenzsystem transformiert werden, damit die DOP's in der Datenbank korrekt abgespeichert werden.

Mosaikbildung

Aus den einzelnen Kacheln wird ein Mosaik gebildet. Oracle benötigt diesen Schritt, um eine topologische Beziehung zwischen den einzelnen Kacheln zu erstellen.

Erstellung der Bildpyramiden

Bildpyramiden ermöglichen Performanceverbesserungen bei Bildbetrachtungen in unterschiedlichen Auflösungsstufen. Für jedes Pyramidenlevel wird zusätzlich Speicherplatz benötigt, da neue Rasterdaten mit einer geringeren Auflösung erstellt werden.

Kompression

Um eine platzsparende Datenspeicherung zu gewährleisten, müssen die Daten abschließend komprimiert werden.

Freigabe für WFS und WCS

Nach erfolgreichem Import werden die DOP's für den WFS und WCS bereitgestellt um die neuen Daten zu nutzen.

Rasterbild-Explorer und -Viewer

Der Rasterbild-Explorer ermöglicht das gezielte Suchen von Bildern anhand unterschiedlicher Suchkriterien (siehe Abbildung 4).

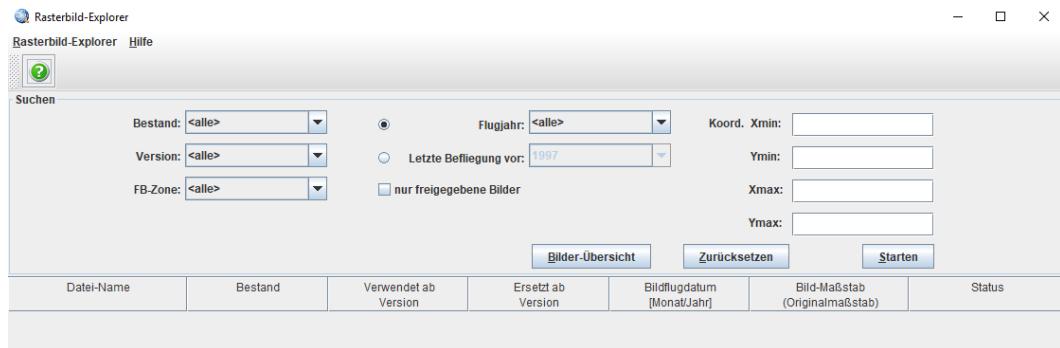


Abbildung 4: Darstellung der Suchkriterien des Rasterbild-Explorers [DEG13]

Bilder, die mit Hilfe des Rasterbild-Explorers gesucht wurden, werden mit dem Rasterbild-Viewer präsentiert. Im Viewer wird zum einen das Bild und zum anderen die dazugehörigen Sachdaten dargestellt (siehe Abbildung 5). Somit können die Rasterdaten aus der Datenbank betrachtet und stichpunktartig kontrolliert werden.

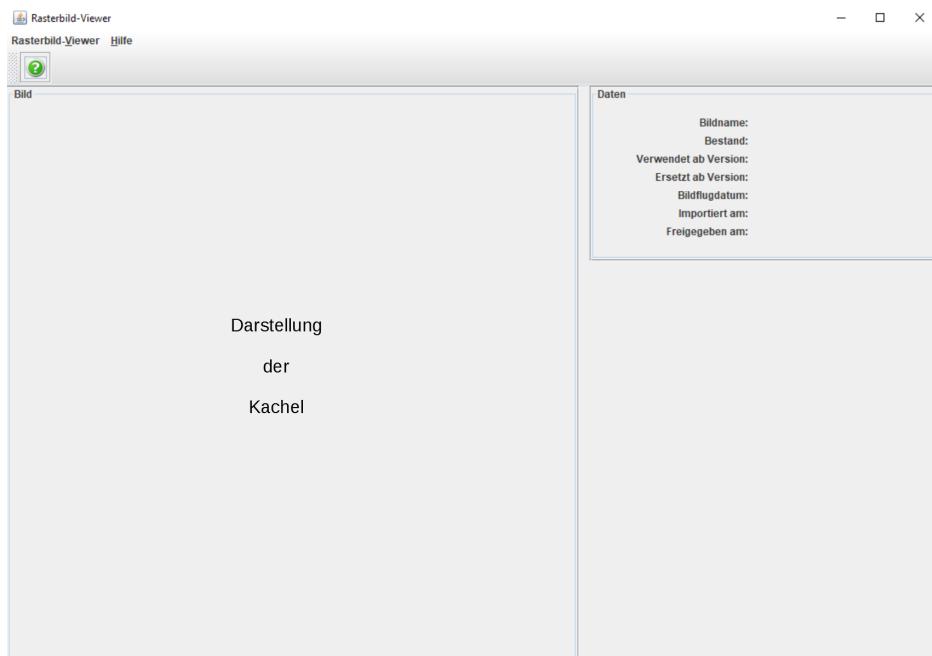


Abbildung 5: Beispielhafte Darstellung des Rasterbild-Viewer [DEG13]

2.4 Systemarchitektur

Die **profil rdv** besteht aus den folgenden Komponenten: Client, Applikation-Server, Oracle DBMS, WFS und WCS. Die Kommunikation mit anderen **profil** Anwendungen erfolgt ausschließlich über Open Geospatial Consortium (OGC)-konformen Webservices. In diesem Fall handelt es sich um eine WFS und WCS Schnittstelle. In Abbildung 6 sind **profil rdv** und **profil c/s** im 3 Schichten Modell dargestellt. Zusätzlich sind die Kommuni-

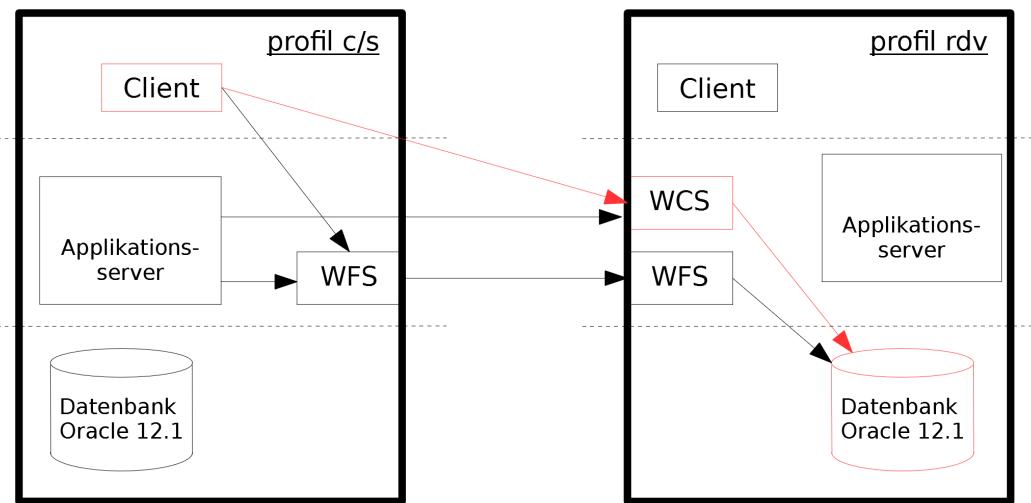


Abbildung 6: 3 Schichten Modell, **profil rdv** und **profil c/s**

Kommunikationswege mit der **profil rdv** über WFS und WCS dargestellt. Der rot hervorgehobene Weg beschreibt die Problemstellung dieser Arbeit und wird im Folgeabschnitt 2.5 näher erläutert.

2.5 Problemstellung: WCS-Anfrage

Der Ablauf einer WCS Anfrage ist in Abbildung 7 dargestellt und in Tabelle 1 sind die entsprechenden Parameter aufgelistet. Der größte Teil der Anfragezeit wird in der Datenbank verbraucht. Die Daten werden ermittelt, dekomprimiert und für den Export vorbereitet.

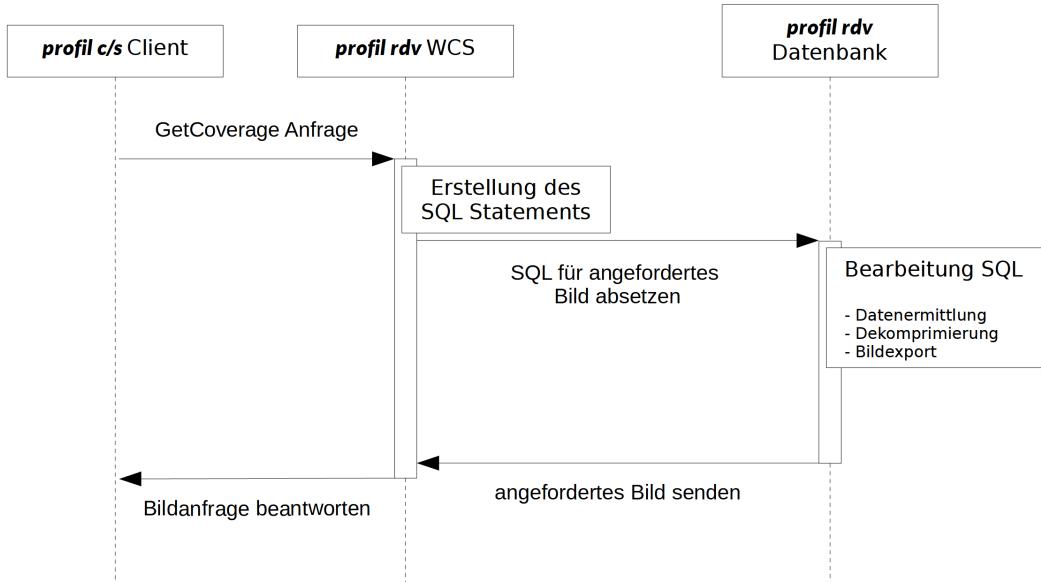


Abbildung 7: Sequenzdiagramm einer WCS Anfrage

Parameter	Beschreibung	Beispiel
service	Aufruf des abzufragenden Services	service=WCS
version	Versionsnummer des Services	version=1.0.0
request	Angabe der abzufragenden Informationsart	request=GetCoverage
coverage	Angabe der abzufragenden Daten	coverage=[LVerMGeo2016]
crs	Angabe des Koordinatenreferenzsystems	crs=EPSG:4647
BBOX	Angabe der Eckpunkte (unten links, oben rechts) des umgebenden Rechtecks der Rasterdaten	BBOX=[30,20,50,60] [x-min, y-min, x-max, y-max]
width	Anzahl Pixel in einer Reihe der Anfrageauflösung	width=[1000]
height	Anzahl Pixel in einer Spalte der Anfrageauflösung	height=[700]
format	Ausgabeformat der abzufragenden Daten	format=tif

Tabelle 1: Parameter einer einfachen WCS Anfrage

Weitere Informationen zu OGC WCS sind zu finden unter: [WCS19]

Bei dem Vergleich der Anfragezeiten an den WCS auf dem aktuellen Oracle DBMS mit Rasterdaten wird der Einfluss der Kompression deutlich. Die Rasterdaten sind sowohl in ihrer Struktur als auch in ihrer räumlichen Ausdehnung identisch und unterscheiden sich nur in der Kompressionsart. In Tabelle 2 sind die möglichen Kompressionsarten in Oracle 12.1 dargestellt, die in Kapitel 3.1 näher erläutert werden.

	NONE	DEFLATE	JPEG-F 90%
Anfragezeit (in Sekunden)	0,46	1,10	1,60

Tabelle 2: Anfragezeiten im aktuellen Oracle DBMS der profil rdv

Aufgrund der Kompression (JPEG-F) wird die Anfragezeit um mehr als eine zusätzliche Sekunde im Vergleich zum Unkomprimierten (NONE) verlängert. Die Speicherung der Daten mit einer Kompression wird benötigt. Der Betreiber der Datenbank begrenzt den Speicherplatz und dieser ist somit nicht flexibel erweiterbar.

3 Rasterdaten in objektrelationalen Datenbanken

'Objektrelationale Datenbanksysteme sind um objektorientierte Konzepte erweiterte relationale Datenbanksysteme.[TDB15]' Die objektorientierten Konzepte ermöglichen es, komplexe Datentypen für die Speicherung der Rasterdaten zu nutzen und in Relation mit anderen Daten (z.B. Sachdaten) zu bringen. Weitere Informationen zum Thema Datenbanken sind im folgendem Buch zu finden: [TDB15]

3.1 Oracle GeoRaster

Das Arbeiten mit Rasterdaten erfordert in Oracle die Lizenz für 'Spatial and Graph'. Die Lizenz enthält Funktionalitäten rund um Oracle GeoRaster. In diesem Kapitel wird auf die Oracle Version 12.2.0.1 eingegangen. Als Datengrundlage dient die Oracle 12.2 Spatial and Graph

Einige Funktionalitäten werden in diesem Kapitel nicht näher dargestellt, da sie für den Anwendungsfall der **profil rdv** nicht relevant sind. Darunter fallen Funktionen zum Abfragen und Bearbeiten von einzelnen Zellen einer Rasterdatei und dem gesamten Komplex der Rasterdatenanalyse. Weitere Informationen dazu sind in der entsprechenden Oracle Dokumentation zu finden: [ORA17]

Mit dem Oracle GeoRaster werden zwei Datentypen bereitgestellt, die benötigt werden, um Rasterdaten und deren Informationen in der Datenbank abzuspeichern, SDO_GEORASTER und SDO_RASTER. Die Attribute des jeweiligen Datentyps sind in Tabellen 3 und 4 aufgelistet.

Attributname	Datentyp	Beschreibung
rasterType	NUMBER	Fünfstellige Nummer zur Identifikation des Ausbaus, z.B. räumliche Dimension, ein oder mehrere Bänder
spatialExtent	SDO_GEOMETRY	Räumliche Ausdehnung des Bildes
rasterDataTable	VARCHAR2(32)	Name der Rasterdatentabelle
rasterID	NUMBER	Identifikation der zugehörigen SDO_RASTER Objekte
metadata	XMLTYPE	Metadaten des Bildes

Tabelle 3: Attribute des Oracle Datentyps SDO_GEORASTER [ORA17]

SDO_GEORASTER enthält die allgemeinen Informationen über die räumliche Ausdehnung des hinterlegten Bildes, als auch die entsprechenden Metadaten zum Bild selbst. Die Metadaten werden im binary XML Format abgespeichert. Zusätzlich zu den allgemeinen Informationen beinhaltet das SDO_GEORASTER den Verweis auf die eigentlichen Bildinformationen, die in einer anderen Tabelle in den zweiten Datentyp

SDO_RASTER abgespeichert sind.

Im SDO_RASTER werden die Bildinformationen in einem Binary Large Object (BLOB) abgespeichert. Abhängig von der Größe der Originaldatei wird diese in eines oder mehrere SDO_RASTER aufgeteilt und in Blöcke abgespeichert. Die von Oracle empfohlene Ausdehnung eines einzelnen Blocks liegt zwischen 512x512 und 2048x2048 Pixeln.

Attributname	Datentyp	Beschreibung
rasterID	NUMBER	Zuordnung zum SDO_GEORASTER Objekt
pyramidLevel	NUMBER	Pyramidenlevel
bandBlockNumber	NUMBER	Nummer des genutzten Bands
rowBlockNumber	NUMBER	Anzahl der Reihen des Blocks
columnBlockNumber	NUMBER	Anzahl der Spalten des Blocks
blockMBR	SDO_GEOMETRY	MBR des Blocks
rasterBlock	BLOB	Rasterdaten des Blocks

Tabelle 4: Attribute des Oracle Datentyps SDO_RASTER [ORA17]

In der Abbildung 8 ist die Verbindung zwischen SDO_GEORASTER und SDO_RASTER dargestellt. Deutlich wird, dass dem SDO_GEORASTER mit Hilfe des Namens der Rasterdatentabelle und der RasterID entsprechende SDO_RASTER zugeordnet werden.

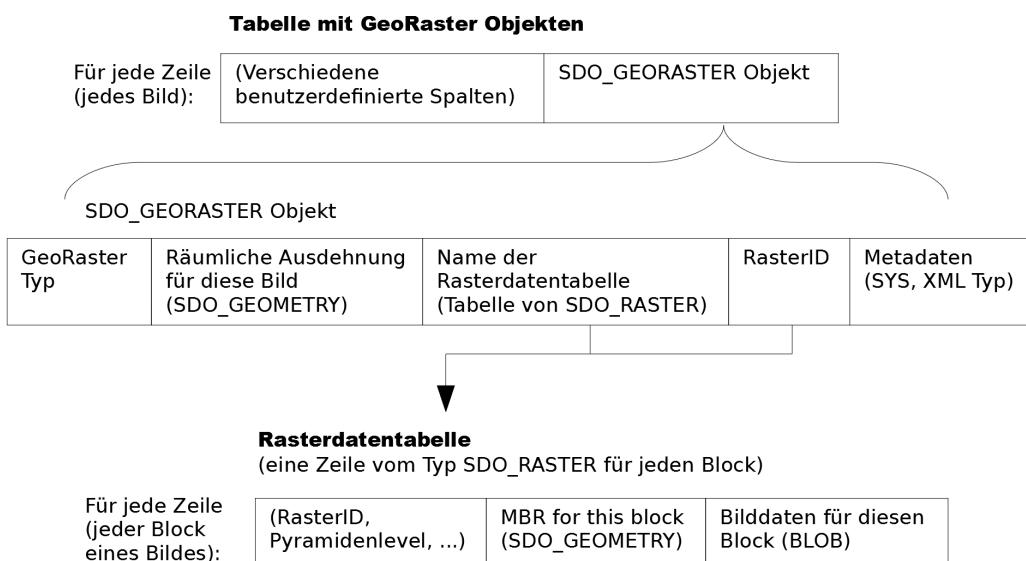


Abbildung 8: Aufbau der Datentypen SDO_GEORASTER und SDO_RASTER [ORA17]

Eine schemenhafte Gesamtdarstellung einer GeoRaster Datenbank ist in der Abbildung 9 zu sehen. Darin besteht die Datenbank aus den zwei Pflichttabellen, eine für die GeoRaster Objekte und eine Rasterdatentabelle für die Bildinformationen.

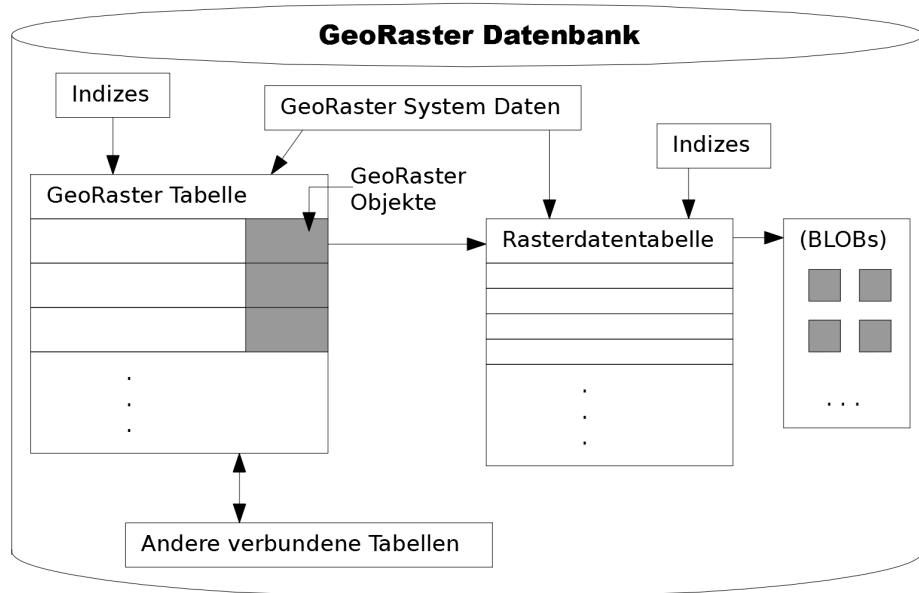


Abbildung 9: Darstellung der GeoRaster Datenbankstruktur [ORA17]

Arten der Datenkompression

Folgende Varianten der Kompression werden für Oracle GeoRaster angeboten:

NONE

unkomprimierte Datenspeicherung

DEFLATE

verlustfreie Datenkompression, Kompression nach RFC 1951 (Request for Comments) und Datenspeicherung nach RFC 1950 (ZLIB Format)

JPEG-F

verlustbehaftete Datenkompression, Grad des Datenverlustes wird mit dem Parameter 'quality' bestimmt, Oracle Standard beträgt 75%, auch 100% erzeugen Datenverlust aufgrund von Datenumstrukturierung

JPEG 2000 (neu ab Oracle 12.2)

Datenkompression nach ISO/IEC 15444-1, gesamtes Bild inkl. Pyramide in einem SDO_Raster

Alle Varianten haben unterschiedliche Auswirkungen auf Speicherplatzverbrauch und Performance.

Aufbau der Bildpyramide

Die erstellten Bildpyramiden werden ähnlich wie die Originaldaten in der Rasterdatentabelle in einen SDO_RASTER abgespeichert. Die Nummerierung der Pyramiden beginnt bei 0, dem Original. Ab Level 1 wird jeweils die halbierte Auflösung gespeichert. Die Formel für die Berechnung der Abstufungen der Pyramidenlevel ist in Abbildung 10 dargestellt.

$$\text{Pyramidenlevel} \quad \text{Anzahl Pixel}$$

$$2^{\text{Basis}} * 2^{\text{Basis}} = 16$$

Pixel pro Reihe Pixel pro Spalte

Beispiele:

- $2^1 * 2^1 = 4$
- $2^3 * 2^3 = 64$
- $2^4 * 2^4 = 256$
- $2^5 * 2^5 = 1024$

Abbildung 10: Formel zur Ermittlung der entsprechenden Pyramidenlevelauflösung

Das Level gibt immer den Exponenten zur Basis 2 an, jeweils für Reihe und Spalte der Pixel. Das Ergebnis der Berechnung ist die Gesamtanzahl der Pixel aus dem Original, die in dem angegebenen Pyramidenlevel zu einem Pixel zusammengefasst werden. In Abbildung 11 sind die einzelnen Pyramidenlevel visuell dargestellt.

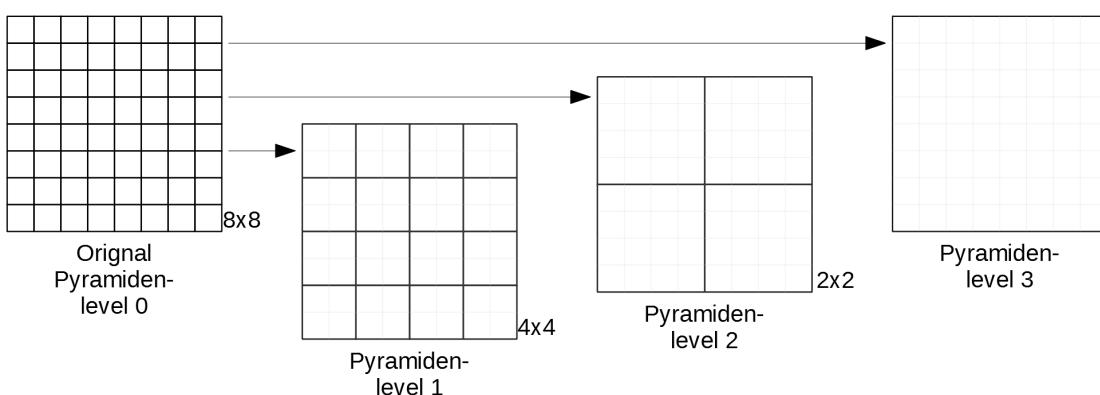


Abbildung 11: Darstellung der Pyramidenlevel in Oracle im Vergleich zum Original

Die Pyramiden eines Bildes werden nicht dynamisch generiert und müssen nach Veränderung des Originals neu erstellt werden [ORA17].

Erstellung von flächendeckenden Bilddaten (Mosaik)

Oracle GeoRaster benötigt die Funktionalität des Mosaiks, um aus mehreren einzelnen GeoRaster Objekten ein einzelnes GeoRaster zusammenzufügen.

Zusätzlich ist auch die Funktionalität des virtuellen Mosaiks enthalten. Dabei werden mehrere GeoRaster Objekte virtuell zu einem zusammengefügt. Dadurch wird eine doppelte Abspeicherung von Daten vermieden. Die Erstellung erfolgt entweder über eine Menge von GeoRaster Tabellen, einer View mit GeoRaster Spalte oder einem SQL Statement [ORA17].

Indexierung von Rasterdaten

Die Indexbildung wird nicht direkt über die Rasterdaten erstellt, sondern über die Vektorgeometrien (kleinstes umschließendes Rechteck), welche die Ausdehnung der jeweiligen Rasterdaten darstellen. Daraus wird ein entsprechender R-Tree Index erzeugt. Zusätzlich zu diesem Index empfiehlt Oracle noch einen weiteren B-Tree über die Informationen zu der Geometrie. Bei beiden Indices handelt es sich um balancierte Indexstrukturen. Der B-Tree wird über die entsprechenden Primärschlüssel-Attribute des Typs SDO_RASTER generiert: rasterID, pyramidLevel, BandBlockNumber,

3.2 PostgreSQL/PostGIS raster

PostgreSQL ist eine OpenSource Datenbanksoftware. Für geometrische Funktionalitäten gibt es die Erweiterung PostGIS, die unter anderem auch das Speichern und Bearbeiten von Rasterdaten ermöglicht. In diesem Abschnitt werden PostgreSQL in der Version 10.7 und PostGIS in der Version 2.5.1 dargestellt. Analog zu Abschnitt 3.1 werden die **profil rdv** nicht relevante Funktionalitäten, wie die Verschneidung von Vector- & Rasterdaten oder Rasterdatenanalysen, nicht näher erläutert. Weitere Informationen sind zu finden unter [PGM18]

Attributname	Datentyp	Beschreibung
version	16 bit integer ohne Vorzeichen	WKB Format Version
number of bands	16 bit integer ohne Vorzeichen	Anzahl der Bänder im Raster
width	16 bit integer ohne Vorzeichen	Anzahl an Pixel in einer Reihe
height	16 bit integer ohne Vorzeichen	Anzahl an Pixel in einer Spalte
Informationen zur Georeferenzierung		
pixelsizex	64 bit double	Pixelgröße in x-Richtung
pixelsizey	64 bit double	Pixelgröße in y-Richtung
upperleftx	64 bit double	x-Koordinaten des Mittelpunkts des oberen linken Pixels
upperlefty	64 bit double	y-Koordinaten des Mittelpunkts des oberen linken Pixels
rotationx	64 bit double	Rotation der x-Achse
rotationy	64 bit double	Rotation der y-Achse
srid	32 bit double	ID des Georeferenzsystems
Bandinformation (ein Satz pro Band)		
isoffline	1 bit	gibt an, ob die Daten innerhalb oder außerhalb der Datenbank gespeichert sind
hasnodatavalue	1 bit	gibt an, ob der nodatavalue von Bedeutung ist
pixeltype	4 bits	Pixeltyp für das Band
nodatavalue	Abhängig von pixeltype	nodata Wert für das Band
Banddaten (ein Satz pro Band) bei in-db Speicherung		
values[]	Abhängig von pixeltype	Wert der einzelnen Pixel
Banddaten (ein Satz pro Band) bei out-db Speicherung		
bandnumber	8 bit integer ohne Vorzeichen	Nummer des Bandes
path	string	Pfad zur Rasterdatei außerhalb der Datenbank

Tabelle 5: Attribute von PostGIS raster [PRD12]

Der Datentyp 'raster' ermöglicht das Speichern von Rasterdaten in der PostGIS. Eine ausführliche Auflistung der Attribute des Datentyps 'raster' ist in der Tabelle 5 dargestellt. Die Funktionalitäten und Strukturen der Rasterdaten orientieren sich sehr stark

an die der Vektordaten, somit wird dort eine ähnliche Handhabung angestrebt. Die Datenhaltung der Bilddaten ist dabei einfach gehalten. Bild- und Metadaten sind im selben Datentyp 'raster' abgespeichert, d.h. pro Tabelleneintrag wird ein Bildblock (Tile) in ein 'raster' inklusive Metadaten gespeichert. Die empfohlene Ausdehnung eines Bildblocks liegt zwischen 50x50 und 500x500 Pixel, abhängig von Verwendungszweck der Daten [PGA15]. Die Datensätze innerhalb einer Tabelle können in ihrer Blockgröße und Georeferenzierung variieren.

Speicherort der Rasterdaten

PostGIS unterstützt sowohl die Rasterdatenspeicherung innerhalb (in-DB) als auch außerhalb (out-DB) der Datenbank. [PGA11]

Bei der in-DB Variante werden die gesamten Daten in der Datenbank abgespeichert.

Rasterdatenspeicherung innerhalb der Datenbank	
Vorteil	Nachteil
<ul style="list-style-type: none">Datensicherung durch Datenbank Backup'sschnelles Lesen der DatensätzeDatenintegrität bei Transaktionen	<ul style="list-style-type: none">hoher Speicherplatzverbrauch durch regelmäßige Datenbank Backup'sRasterdaten nur über Datenbank abrufbar

Tabelle 6: Vor- und Nachteile der in-DB Datenspeicherung [PRD12]

Als Alternative bietet PostGIS die out-DB Variante an. In diesem Fall sind nur die Metadaten und der jeweilige Pfad zur Bilddatei im Dateisystem in der Datenbank abgelegt. Alle Informationen über die Pixel-Werte sind als normales Bild im Dateisystem abgespeichert. Bei der Art der Bilddaten, die dort hinterlegt sind, werden nur Formate unterstützt, die auch durch Geospatial Data Abstraction Library (GDAL) unterstützt werden. In der Tabelle 7 sind die Vor- und Nachteile kurz dargestellt.

Rasterdatenspeicherung außerhalb der Datenbank	
Vorteil	Nachteil
<ul style="list-style-type: none">Einfache Strukturierung der Rasterdatendirekter Zugriff auf die Daten von anderen ProgrammenDatenbank Backup speichert nur Metadaten	<ul style="list-style-type: none">keine Sicherheit durch TransaktionenDaten im Dateisystem ungeschütztZuordnungsfehler durch Dateipfadefür Rasterdatenanalyse nicht geeignet

Tabelle 7: Vor- und Nachteile der out-DB Datenspeicherung [PRD12]

Aufbau der Overview's (Bildpyramiden)

Die Bildpyramiden werden in PostGIS als Overview's bezeichnet. Dabei wird für jede Overview-Ebene eine separate Tabelle erstellt und die Bilddaten entsprechend abgespeichert [PGM18]. In den Overview-Tabellen werden nur die Bildinformationen gespeichert. Die Metadaten werden nicht von den Originalen zusätzlich übernommen. Die Bezeichnung der Overview's beginnt mit Level 2. Die Formel zur Berechnung lautet wie folgt:

$$\text{Exponent} \quad \text{Anzahl Pixel}$$

Beispiele:

$$2^2 = 4$$
$$8^2 = 64$$
$$10^2 = 100$$
$$32^2 = 1024$$

Overview Level

$$4^2 = 16$$

Abbildung 12: Formel zur Berechnung der Overview's

Das Overview Level repräsentiert dabei die Basis, die den Exponenten vom Wert 2 besitzt. Daraus ergibt sich, dass die folgenden Overview Level jeweils die Hälfte des vorherigen darstellt: 2, 4, 8, 16, 32, 64, 128, ... (siehe Abbildung 13).

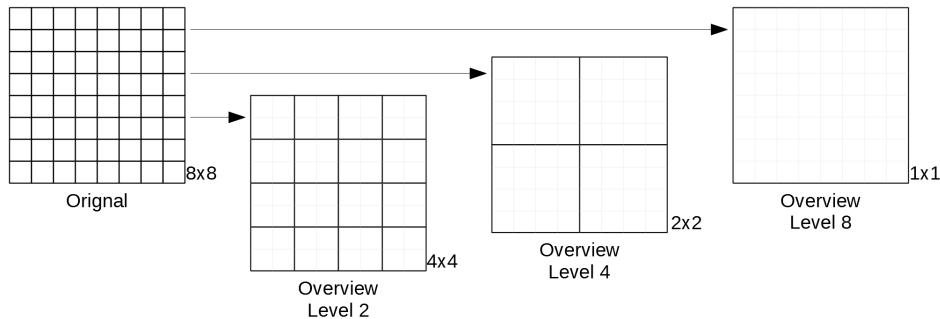


Abbildung 13: Darstellung der Overview's im Vergleich zum Original

Aufgrund dieser Funktionsweise kann die Bildpyramide sehr detailliert und individuell aufgebaut werden. Ein daraus folgender Nachteil ist, dass nicht sichergestellt wird, dass ein Pixel in der Overview immer eine passende Anzahl an Pixel im Original abdeckt.



Indexierung von Rasterdaten

In der PostGIS Umgebung werden folgende Indices für räumliche Datenbanken genutzt und in den anschließenden Abschnitten näher erläutert [PGM18]:

- Generalized Search Tree (GiST)
- Space-Partitioned Generalized Search Tree (SP-GiST)
- Block Range Index (BRIN)

Der GiST Index wird auch schon in PostgreSQL in einer einfachen Variante genutzt. In der PostGIS Umgebung wird in den GiST Index zusätzlich ein R-Tree implementiert.

Eine abgewandelte Form des GiST ist der SP-GiST, dieser arbeitet mit partitionierten Daten. Dieser Index wird auch teilweise für Daten genutzt, die keinen direkten räumlichen Bezug besitzen. Diese Variante des Index ist eine Alternative zum GiST, wenn es sich um Daten handelt, die sich häufig überlappen.

BRIN ist ein generischer Index, der seit der PostgreSQL 9.5 Version implementiert ist. Die Hauptnutzung des Indices ist hierbei ein Kompromiss zwischen Lese- und Schreibgeschwindigkeit. Es wird hierbei eine Verbindung zwischen natürlicher und physikalischer Lage ausgenutzt und somit Blockweise zusammengefasst. Dabei ist der Index nur effektiv, wenn die Menge an Daten sehr groß und die Verbindung zwischen Datenspeicher und natürlicher Lage gegeben ist. Im Index selbst werden nur das umgebende Quadrat gespeichert und nimmt somit selbst nur wenig Speicherplatz in der Datenbank in Anspruch.

Da es sich bei den hier angegebenen Indices um räumliche Varianten handelt, wird jeweils ein zweiter Index (B-Tree) für die anderen Attribute empfohlen.

4 Mögliche Umsetzung der DBMS in der *profil rdv*

In den nachfolgenden Kapiteln werden jeweils Strukturen der DBMS für eine Umsetzung im Falle der *profil rdv* dargestellt. Dabei werden dementsprechend Vor- und Nachteile in den Systemen aufgezeigt und abschließend gegenübergestellt. Es wird nur auf die Abspeicherung der Rasterdaten eingegangen, denn bei der Speicherung der zusätzlichen Sachdaten sind keine Unterschiede zu erwarten.

4.1 Oracle GeoRaster

Tabellenstruktur

In der *profil rdv* werden für das Abspeichern und Bereitstellen von Rasterdaten in einem Mosaik zwei GeoRaster Tabellen mit jeweils einer zugehörigen Rasterdatentabelle benötigt (siehe Abbildung 14). Zu Beginn eines Importvorgangs werden die einzelnen Bilder in die Import-Tabelle geladen. Für das zu bildende Mosaik werden alle GeoRaster Objekte aus der Import-Tabelle in ein GeoRaster Objekt der Mosaik-Tabelle zusammengeführt. Nach erfolgreicher Bildung des Mosaiks werden die entsprechend benötigten Pyramidenlevel gebildet. Abschließend wird eine Datenkompression durchgeführt, indem das aktuell unkomprimierte GeoRaster Objekt in ein neues GeoRaster Objekt mit angepassten Speicherparametern kopiert wird.

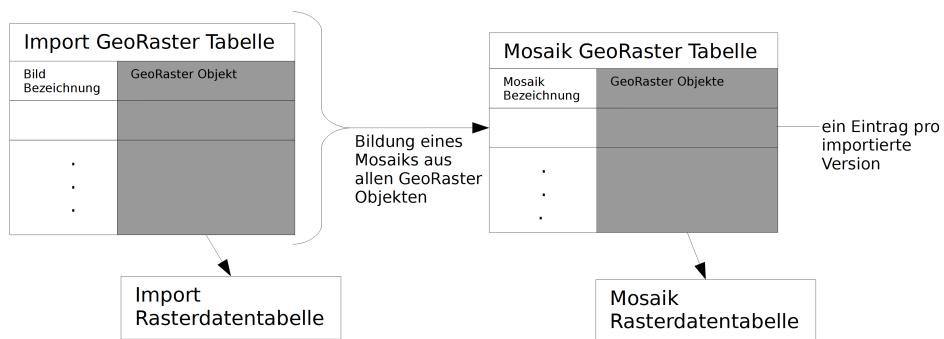


Abbildung 14: Darstellung der Tabellen in der *profil rdv* mit Hilfe von Oracle GeoRaster

Um die Speicherplatznutzung zu optimieren, werden nicht mehr benötigte Daten gelöscht. Zum einen handelt es sich dabei um die gesamte Import-Tabelle nach erfolgreichem Erstellen des Mosaiks. Zum anderen ist das unkomprimierte GeoRaster Objekt nach erfolgreicher Erstellung des finalen, komprimierten GeoRaster Objektes gemeint.

Der gesamte Importvorgang ist sehr speicherintensiv, aber das Ergebnis ist ein GeoRaster Objekt, das seine Bilddaten gebündelt in einer dazugehörigen Rasterdatentabelle abspeichert.

Bildpyramide

Die Abstufungen der Bildpyramide ist in Oracle klar vorgegeben und es stehen wenig Auswahlmöglichkeit zur Verfügung. Aufgrund von Kundendaten (Anhang A) werden nur Pyramidenlevel bis 8 benötigt, wobei Level 7 und 8 nur zu einem geringen Teil genutzt werden.

Datenkompression

In der **profil rdv** wird zur Datenspeicherung die JPEG-F Komprimierung mit einem Qualitätsparameter von 90% genutzt. Dies ist ein Kompromiss zwischen optimierten Speicherplatzverbrauch und Verlust von Bildinformationen.

Durch die Kompression ist ein zusätzlicher Schritt sowohl beim Import als auch beim Auslesen der Daten (Dekompression) erforderlich und führt somit zu längeren Anfragezeiten.

4.2 PostGIS raster

Tabellenstruktur

In PostGIS werden für jede Version eine Anzahl von Tabellen benötigt, eine für die Originaldaten und für jedes Overview Level eine zusätzliche Tabelle. Abbildung 15 ist eine beispielhafte Darstellung für diese Konstellation. Die einzelnen Tabellen benötigen eine zentrale Tabelle oder View (in Abbildung 15 Mosaik Tabelle), um Verbindungen zwischen diesen herzustellen. Somit werden die Overviews den entsprechenden Originaldaten zugeordnet.

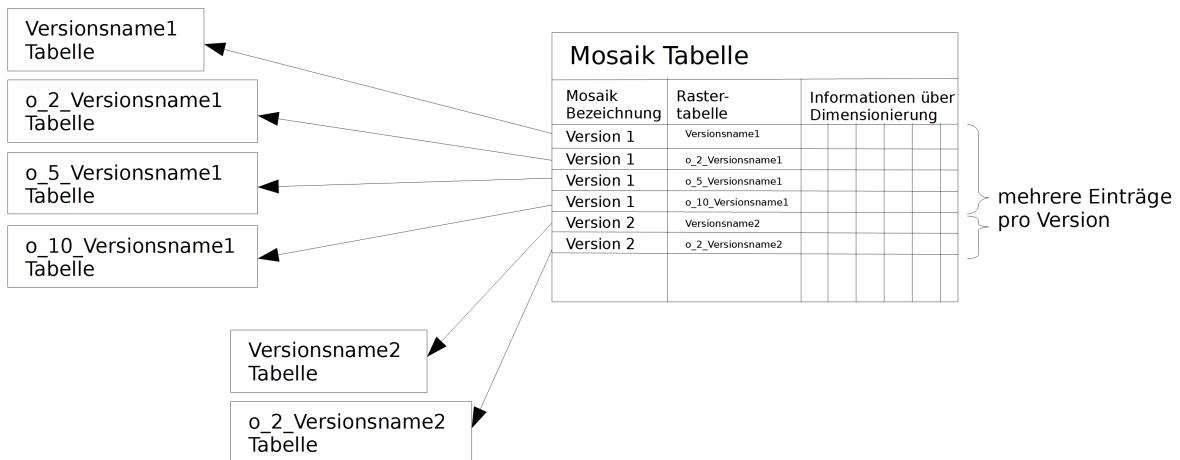


Abbildung 15: Darstellung der Tabellen in der **profil rdv** mit Hilfe von PostGIS raster

In der Mosaik Tabelle wird der Name des Mosaiks festgelegt, dieser wird aus den einzelnen Tabellen zusammengesetzt. Zusätzlich enthält die Tabelle auch Informationen über die räumliche Ausdehnung (BBOX), aber auch über die Pixelgröße der hinterlegten Tabelle.

Während des Imports werden keine zusätzlichen Tabellen erstellt. Es werden nur die Tabellen erstellt, in denen die Rasterdaten abgespeichert sind.

Die Struktur in der Datenbank unterscheidet sich nicht zwischen in-db und out-db Variante.

Bildpyramide

Die Bildpyramide ist in PostGIS flexibel einsetzbar und sollte auf die gewählte Kachelgröße der Rasterdaten angepasst werden. Bei einer Kachelgröße von 1000x1000 Pixel sollten die Overview Level so gewählt werden, das diese immer Pixelgenau abschließen. Wenn dies nicht beachtet wird, kann es zu Überlappungen der einzelnen Bildern führen und das visuelle Ergebnis verfälschen. Aus diesem Grund werden folgende Overview Level genutzt:

- Overview Level 2
- Overview Level 5
- Overview Level 10
- Overview Level 25
- Overview Level 50
- Overview Level 100
- Overview Level 250

Datenkompression

Die Komprimierung der Daten ist in PostGIS nicht direkt vorhanden. Es ist möglich, die Rasterdaten außerhalb der Datenbank zu speichern und über einen Dateipfad auf diese zu verweisen. Die Rasterdaten können in jedes von GDAL unterstützten Dateiformat hinterlegt werden. Somit können komprimierte Dateiformate wie z.B. JPEG genutzt werden. Die Komprimierung ist aber ausschließlich für die Originaldaten möglich. Die daraus erstellten Overview Level werden in der Datenbank unkomprimiert abgespeichert.

4.3 Gegenüberstellung

In der Tabelle 8 werden die Kriterien aus den vorhergehenden Abschnitten gegenübergestellt. Oracle GeoRaster ermöglicht durch einen komplexeren Importvorgang eine klare Struktur der Enddaten, die zusätzlich komprimiert werden können. Im Gegenzug besitzt PostGIS raster einen einfacheren Importvorgang. PostGIS hinterlässt durch die Aufteilung der Rasterdaten über mehrere Tabellen einen unübersichtlichen Eindruck. Eine längere Nutzung in der **profil rdv** lässt somit die Menge an Tabellen rasch ansteigen. Zusätzlich wird aufgrund der fehlenden bzw. eingeschränkten Kompressionsmöglichkeiten der Speicherplatz ebenfalls deutlich schneller ansteigen im Vergleich zu Oracle GeoRaster.

Kriterium	Oracle GeoRaster	PostGIS raster
Import	- mehrstufiger Importvorgang über mehrere Tabellen hinweg	- einfacher Importvorgang in jeweils eine neue Tabelle
Pyramidenlevel / Overview Level	- Pyramidenlevel werden in das vorhandene GeoRaster Objekt abgespeichert - vorgegebene Abstufungen	- Overview Level werden in separaten Tabellen abgespeichert - individuell anpassbare Level
Datenfreigabe WCS, ...	- direkte Freigabe über das GeoRaster Objekt	- Freigabe erfolgt über eine zusätzliche Tabelle/View, in der die einzelnen Tabellen verknüpft werden
Kompression	- direkte Kompression der gesamten Datenmenge	- keine Kompression in der Datenbank - teilweise Kompression durch Verweis auf komprimierte Rasterdaten im Dateisystem

Tabelle 8: Auflistung der Kriterien und ihre Umsetzung in dem DBMS

5 Kostenmodelle von Oracle und PostgreSQL/PostGIS

In diesem Abschnitt der Arbeit werden die beiden zu untersuchenden DBMS hinsichtlich der Finanzierung, die auch eine Rolle bei der Entscheidung spielen, untersucht. Kriterien wie Serviceleistung und Kosten für Lizenzen werden im folgenden näher erläutert und gegenübergestellt.

5.1 Oracle

Oracle ist neben MySQL und Microsoft eines der führenden Unternehmen im Segment der relationalen Datenbanken. [DBE19]

Lizenzpolitik

Oracle nutzt zur Finanzierung ihrer Produkte eine modulare Lizenzierung der Produkte. Hier muss für jedes Funktionalitätspaket die entsprechende Lizenz gekauft werden. Im Beispiel der **profil rdv** wird neben der Grundlizenz 'Enterprise Edition' auch die Zusatzlizenz 'Spatial and Graph' benötigt. Die 'Spatial and Graph' Lizenz ermöglicht es mit geometrischen Daten wie Vector- und Rasterdaten zu arbeiten und diese in der Datenbank abzuspeichern. [ORE19]

Die relevanten Funktionen der 'Spatial and Gragh'-Lizenz wurden bereits im Kapitel 3.1 näher dargestellt. Weitere Informationen zu den einzelnen Oracle Lizenzen und ihren Inhalten gibt es unter [ORE19]

Support-Leistungen

Das Kostenmodell für den Support ist fest mit dem Lizenzmodell verbunden. Zum einen gibt es die Option, zu den einzelnen Lizenzen direkt Support zu buchen. Als Alternative sind im Rahmen eines Festpreises zehn Serviceanfragen im Jahr verfügbar. Des weiteren haben die einzelnen Produkte von Oracle eine Lebenszeit, in denen der Support gewährleistet wird. Im Regelfall ist diese auf 5 Jahre angesetzt, mit zusätzlicher Verlängerung um 3 Jahre. Die Lebenszeiten variieren von Produkt zu Produkt und sind unter folgendem Link aufgelistet: [OLS19]

Kostenrechnung für Lizenzen und Support

Mit Hilfe der von Oracle öffentlich zur Verfügung gestellten Preisliste [OPL19] wird die folgende Kostenrechnung beispielhaft für eine mögliche Konstellation für die **profil rdv** erstellt. Da die Preisliste nur Dollar(\$) Preise enthält, sind alle folgenden Rechenbeispiele auch in Dollar angegeben.

Die **data experts** setzt als Mindestanforderung für den Datenbankserver der **profil rdv** 4 Prozessoren voraus. In Tabelle 9 ist eine Auflistung der Kosten dargestellt, wobei in fünf unterschiedlichen Laufzeiten jeweils jährlich andere Kosten ermittelt werden.

Lizenz	Enterprise Edition	Spatial and Graph	
Kosten pro Prozessor	\$47.500,00	\$17.500,00	
genutzte Prozessoren	4	4	
Gesamtkosten	\$190.000,00	\$70.000,00	
Laufzeit in Jahren	Faktor für Jahreskosten	Jährliche Kosten	Summe
1	0,2	\$38.000,00	\$14.000,00
2	0,175	\$33.250,00	\$12.250,00
3	0,166	\$31.666,67	\$11.666,67
4	0,15	\$28.500,00	\$10.500,00
5	0,14	\$26.600,00	\$9.800,00

Tabelle 9: Beispielrechnungen für Lizenzkosten [OPL19]

Die Basis für die Berechnung der Kosten sind die benötigten Lizenzen mit ihren Kosten pro Prozessor und die Anzahl an genutzten Prozessoren. Daraus werden die Gesamtkosten für die entsprechende Lizenz ermittelt. Diese Gesamtkosten werden abhängig der gewählten Vertragslaufzeit mit einem Jahreskosten-Faktor multipliziert. Damit erhält man die jährlichen Kosten der einzelnen Lizenzen. Bei dem Vergleich der jährlichen Lizenzkosten in Abhängigkeit der Laufzeit wird deutlich, dass lange Verträge aufgrund von sinkenden Jahresbeiträgen angestrebt werden.

Die von Oracle angebotenen Supportleistungen sind immer direkt mit den Lizenzen verbunden. Der jeweilige Preis des Supports beträgt 22% der entsprechenden jährlichen Lizenzkosten.

Aus der Tabelle 10 wird deutlich, dass die Supportleistungen flexibel auf die einzelnen Lizenzen abgestimmt werden können.

Laufzeit in Jahren	Enterprise Edition		Spatial and Graph		Summe
	Lizenz	Support	Lizenz	Support	
1	\$38.000,00	\$8.360,00	\$14.000,00	\$3.080,00	\$63.440,00
2	\$33.250,00	\$7.315,00	\$12.250,00	\$2.695,00	\$55.510,00
3	\$31.666,67	\$6.966,67	\$11.666,67	\$2.566,67	\$52.866,67
4	\$28.500,00	\$6.270,00	\$10.500,00	\$2.310,00	\$47.580,00
5	\$26.600,00	\$5.852,00	\$9.800,00	\$2.156,00	\$44.408,00

Tabelle 10: Darstellung der einzelnen Supportkosten [OPL19]

Die Alternative zum klassischen Support mit zehn Fehleranfragen pro Jahr ist zum jährlichen Festpreis von \$2.300 verfügbar.

5.2 PostgreSQL/PostGIS

Bei PostgreSQL und PostGIS handelt es sich jeweils um eine OpenSource Software, somit ist sie für jeden kostenlos nutzbar.

Die Weiterentwicklung von PostgreSQL erfolgt durch die große Nutzerbasis und einigen Firmen, die Systeme entwickeln, die auf PostgreSQL aufbauen. Als namhafte Beispiele sind hier Amazon Web Services (AWS) oder Google zu nennen. Eine ausführliche Liste der Sponsoren sind auf der Homepage von PostgreSQL zu finden:

<https://www.postgresql.org/about/sponsors/>

PostGIS wird nach ähnlichem Schema weiterentwickelt. Die Gründer bzw. Hauptentwickler sind zum größten Teil in Firmen beschäftigt, die sich auf PostGIS spezialisiert haben und eine stetige Weiterentwicklung vorantreiben. Des weiteren ist es Teil der Open Source Geospatial Foundation (OSGeo) und wird durch regelmäßige Events und Konferenzen vorangetrieben. Weitere Informationen zu OSGeo sind zu finden unter <https://www.osgeo.org/>

Support-Leistungen

Einen Support oder einen Ansprechpartner gibt es bei PostgreSQL/PostGIS nicht direkt. Eine Möglichkeit der Problemlösung ist die Community mit Diskussionen in Foren oder Tickets die erstellt werden oder wurden. Als eine weitere Supportmöglichkeit gibt es kommerzielle Lösungen durch Unternehmen, die in den meisten Fällen PostGIS Produkte anbieten und unterstützen. Vereinzelt gibt es auch Support von Unternehmen für PostGIS Datenbanken, die nicht in direkter Verbindung mit einem ihrer Produkte stehen. Das Themengebiet der Rasterdaten grenzt zusätzlich das Angebot stark ein. Eine Liste an ausgewählten Unternehmen ist zu finden unter: <https://postgis.net/support/>

Aufgrund der eingeschränkten Nutzung der **profil rdv**, die sich hauptsächlich auf das Speichern von Rasterdaten spezialisiert haben, ist eine kommerzielle Lösung in diesem Umfeld nicht zu empfehlen. Die Unternehmen mit PostGIS Support sind meist auf Rasterdatenanalysen spezialisiert und sind somit sehr wahrscheinlich ungeeignet.

5.3 Gegenüberstellung

In der Tabelle 11 werden die Kriterien aus den vorhergehenden Abschnitten gegenübergestellt. Hier sind deutliche Unterschiede zu erkennen, da beide Systeme ein unterschiedliches Finanzierungsmodell als Grundlage besitzen.

Kriterium	Oracle	PostgreSQL/PostGIS
Einmalige Kosten	- benötigte Lizenzen und entsprechendes Supportangebot	- keine Kosten, da OpenSource Produkt
Folgekosten	- laufende Lizenz und Supportvertrag	- kommerzieller Support durch Unternehmen - kostenloser Support durch Communityforen

Tabelle 11: Auflistung der Kriterien und ihre Umsetzung in dem DBMS

6 Performanceanalyse

6.1 Rahmenbedingungen der Performanceanalyse

6.1.1 Was ist eine Performanceanalyse?

In den nachfolgenden Analysen werden insgesamt drei Kriterien ausgewertet: Importzeit, Speicherplatzverbrauch und Anfragezeit.

Die Importzeit umfasst nur die nötigen Tätigkeiten, die innerhalb der Datenbank durchgeführt werden. Die Vorbereitung der Daten wird nicht mit einbezogen. Der Zeitverlauf geht von 'Daten in die Datenbank laden' bis zum abschließenden Punkt 'Daten können nun abgefragt werden'.

Der Speicherplatzverbrauch beinhaltet die Analyse des Platzverbrauches der erstellten Daten und zeigt dabei auch die Verteilung innerhalb eines Datensatzes mit deren Pyramidenlevel bzw. Overview Level an.

Die Anfragezeit wird mittels Anfragen über einen WMS analysiert und ausgewertet. Zusätzlich werden einzelne WCS Anfragen getestet, um einen Vergleich zwischen WCS und WMS zu ermöglichen. Bei den Anfragen wird das gesamte Spektrum an Pyramidenlevel und räumlicher Ausdehnung abgedeckt und zusammengefasst.

Im Test werden im Gegensatz zur Problemstellung WMS statt WCS Anfragen durchgeführt aufgrund einer geplanten Umstellung der OGC Schnittstelle. Der Aufbau einer WMS-Anfrage ist ähnlich der WCS-Anfrage. Weitere Informationen zum WMS sind zu finden unter: [WMS19]

6.1.2 Beschreibung der Rahmenbedingungen

Die Messungen werden auf folgenden Datenbanken durchgeführt:

- Oracle 12.2 mit Spatial and Graph Lizenz
- PostgreSQL 10.7 mit PostGIS 2.5.1

Auf folgender Hardwareumgebung werden die Messungen durchgeführt:

- Intel(R) Xeon(R) CPU E5-1650 v4 mit 3,60GHz und 6 Kernen
- 64 GB RAM
- 2 TB SSD Festplatte

Die WMS und WCS Anfragen werden über einem Geoserver 2.14 und dem zusätzlichen MosaikJDBC Plugin realisiert[GEO19]. Die Anfragen werden mittels Apache JMeter durchgeführt [JME19]. Die SQL's auf den Datenbanken werden mittels Oracle SQL Developer Version 4.1.1.19 [SQL19] mit zusätzlichem PostgreSQL/PostGIS Treiber ausgeführt.

Die virtuellen Maschinen sind mit VMware Workstation Pro erstellt und das Betriebssystem ist CentOS 7 64 Bit.

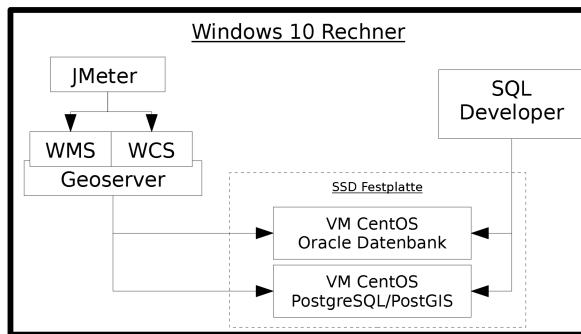


Abbildung 16: Struktur der Testumgebung

Aufgrund der regelmäßigen Anfragen (siehe Anhang A) in der **profil rdv** sind die Testdaten eingegrenzt. Es wird eine maximale räumliche Ausdehnung von 56 km x 36 km getestet. Die Daten bestehen aus insgesamt 504 DOP's, wobei jedes DOP eine räumliche Ausdehnung von 2km x 2km und einer Dimensionierung von 5000 x 5000 Pixel besitzt. Daraus ergibt sich eine Bodenauflösung von 0,4m pro Pixel. Zusätzlich wird eine Bildpyramide benötigt, die sich in Oracle bis Level 8 und in PostGIS bis Overview 250 erstreckt.

Die Messungen werden mit drei Datenkonstellationen durchgeführt. Die einzelnen Zusammenstellungen der Daten ist in Tabelle 12 näher dargestellt.

	Klein	Mittel	Groß
Anzahl Bilder	32	150	504
Bildverteilung	8 x 4	15 x 10	28 x 18
Räumliche Ausdehnung	16 km x 8 km	30 km x 16 km	56 km x 36 km
Anzahl Pixel	40.000 x 20.000	75.000 x 50.000	140.000 x 90.000

Tabelle 12: Datenkonstellationen mit den einzelnen Merkmalen

In Abbildung 17 sind die einzelnen Datenmengen visuell dargestellt. Die umschließenden Rechtecke stellen dabei sowohl räumliche Positionierung als auch Ausdehnung der Daten dar. Die Daten von 'Klein' sind ein Teil des unteren linken Bereiches der Daten von 'Mittel', der wiederum den unteren linken Bereich von 'Groß' darstellt.

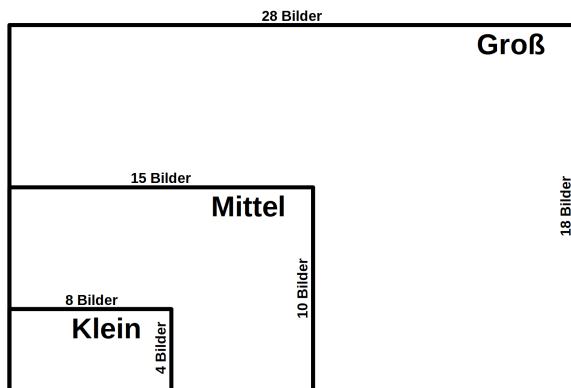


Abbildung 17: Räumliche Ausdehnung der Datenkonstellationen

Diese Konstellation ermöglicht es zusätzlich zu Import- und Speicherplatzvergleichen auch Anfragezeiten zu testen. Hier können Anfragen über dieselbe räumliche Ausdehnung abgefragt werden und dabei der Einfluss der Gesamtdaten berücksichtigt werden.

Die Anfragezeiten werden mit Hilfe von unterschiedlichen Bildauflösungen (3K, 1,5K, 0,7K, 0,3K) untersucht. Diese Auflösungen stellen mögliche Konstellationen in **profil c/s** dar, abhängig der Bildschirmauflösung des Nutzers: UHD, FullHD, HD und Fenstermodus. Die Stufen der Bildung sind etwas kleiner gewählt, da die Anzeigefläche der Karte nicht den gesamten Bildschirm belegt. Die restliche Bildfläche wird durch eine Kartenlegende und Sachdatentabelle abgedeckt. Die Auflösungen beeinflussen die Anzahl der zu ladenden Pixel und somit auch direkt die Anfragezeit.

6.2 Oracle GeoRaster

Rasterdatenimport

Beim Import werden die drei Datenmengen gemessen. In der Tabelle 13 sind die einzelnen Schritte des Import chronologisch von oben nach unten sortiert und von der Initialisierung bis zur abschließenden Pyramidenlevelbildung dargestellt. An einigen Stellen des Imports ist auch eine Parallelisierung der Arbeitsschritte möglich, um die Zeiten zu verkürzen. Aus Gründen der Vergleichbarkeit wird auf die Parallelisierung nicht näher eingegangen.

Die Zeiten der darauf folgenden Komprimierung werden später in Tabelle 14 separat dargestellt.

Import Schritt	Klein		Mittel		Groß		Mittel-wert
	Zeit in Sekunden	Anteil in Prozent	Zeit in Sekunden	Anteil in Prozent	Zeit in Sekunden	Anteil in Prozent	
1. Initialisierung GeoRaster Objekte	0	0,00 %	1	0,10 %	4	0,11 %	0,07 %
2. Import der Rasterdaten	169	65,50 %	632	64,23 %	2217	63,05 %	64,26 %
3. Mosaikerstellung	34	13,18 %	132	13,41 %	434	12,34 %	12,98 %
4. Bildung Pyramidenlevel 1	41	15,89 %	165	16,77 %	642	18,26 %	16,97 %
5. Bildung Pyramidenlevel 2	10	3,88 %	40	4,07 %	163	4,64 %	4,19 %
6. Bildung Pyramidenlevel 3	3	1,16 %	10	1,02 %	41	1,17 %	1,12 %
7. Bildung Pyramidenlevel 4	1	0,39 %	3	0,30 %	10	0,28 %	0,33 %
8. Bildung Pyramidenlevel 5	0	0,00 %	1	0,10 %	3	0,09 %	0,06 %
9. Bildung Pyramidenlevel 6 - 8	0	0,00 %	0	0,00 %	2	0,06 %	0,02 %
Gesamt	258	100 %	984	100 %	3516	100 %	

Tabelle 13: Zeitliche Verteilung des Import von unkomprimierten Daten

Die Tabelle 13 verdeutlicht den großen Zeitverbrauch von Schritt 2 (Import der Rasterdaten) mit einem Anteil von 64% bis zur Bereitstellung der unkomprimierten Daten. Danach folgen 'Bildung Pyramidenlevel 1' mit 16,97% und 'Mosaikerstellung' mit 13,98%. Alle anderen Schritte fallen in geringe unter 5%-Bereiche. Visuell ist die Verteilung in Abbildung 18 als Kreisdiagramm dargestellt.

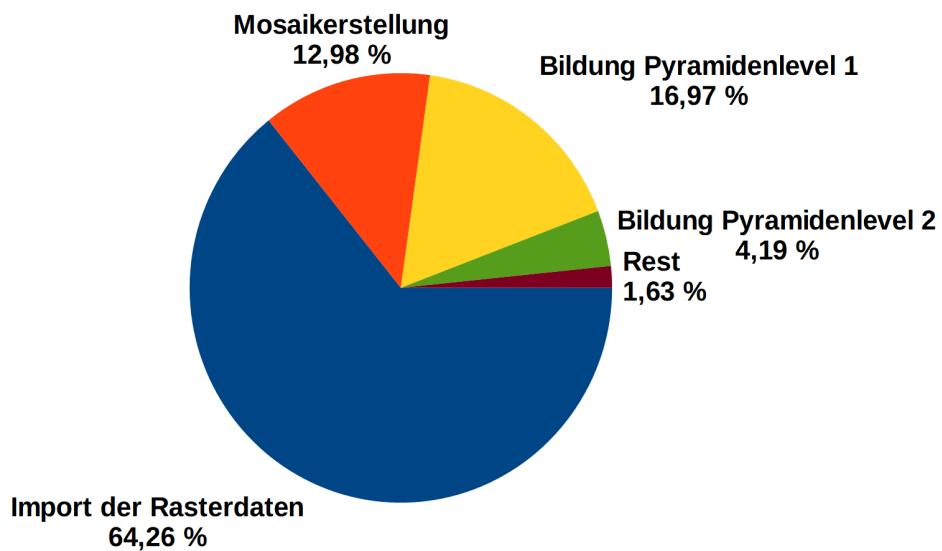


Abbildung 18: Prozentuale Gegenüberstellung der Zeit intensiven Importschritte

Nach Aufbereitung der unkomprimierten Daten erfolgt die Komprimierung mittels Komprimierungsvorgang des GeoRaster Objektes. In der Tabelle 14 sind die einzelnen Laufzeiten dargestellt. Es werden die zur Verfügung stehenden Komprimierungsarten getestet, wobei die JPEG-F Komprimierung wird sowohl in der Standardvariante (Qualität = 75%) als auch mit der Qualitätseinstellung von 90% getestet. Deutlich wird in der Tabelle 14, dass die DEFLATE-Kompression die schnellste von den aufgelisteten Varianten ist und die JPEG 2000 Kompression mit weitem Abstand die langsamste. Die beiden JPEG-F unterscheiden sich nur geringfügig. Die Kompressionen werden jeweils nur einmal durchgeführt und dienen somit nur als grobe Orientierung.

Komprimierungsart	Klein	Mittel	Groß
	Zeit in Sekunden		
DEFLATE	178	634	2487
JPEG-F 75%	209	733	3434
JPEG-F 90%	221	774	3526
JPEG 2000	514	1942	6109

Tabelle 14: Zeiten der einzelnen Komprimierungsvorgänge

In der folgenden Tabelle 15 ist zusammenfassend dargestellt, wie lange die jeweiligen Importvorgänge mit entsprechender Endkompression gedauert haben. Alle Angaben besitzen als Ausgangslage die Importzeit der NONE-Kompression, zusätzlich dazu werden dann die Zeiten der jeweiligen Kompression aus Tabelle 14 dazu addiert. Zur besseren zeitlichen Einordnung sind die jeweiligen gerundeten Minuten mit angegeben.

Endkompression	Klein	Mittel	Groß
	Zeit in Sekunden (Minuten gerundet)		
NONE	258 (4)	984 (16)	3516 (59)
DEFLATE	436 (7)	1618 (27)	6003 (100)
JPEG-F 75%	467 (8)	1717 (29)	6950 (116)
JPEG-F 90%	479 (8)	1758 (29)	7042 (117)
JPEG 2000	772 (13)	2926 (49)	9625 (160)

Tabelle 15: Zeiten eines gesamten Importvorgangs mit Endkomprimierung

Die Abbildung 19 verdeutlicht, wie zeitaufwändig Kompressionen sind. Die Importzeiten werden um mindestens 70% der NONE-Kompression erhöht und die JPEG 2000 Kompression liegt bei über 150% zusätzlich.

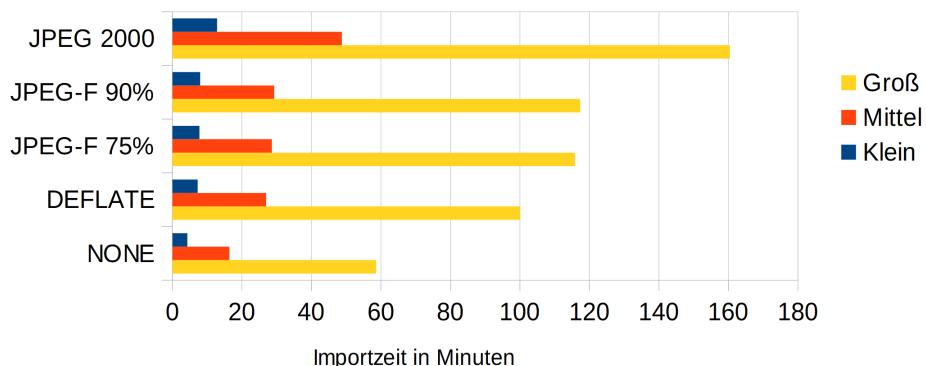


Abbildung 19: Importzeit abhängig von Kompression und Datenmenge

Speicherplatzverbrauch

Der Speicherplatzverbrauch basiert auf den im vorhergehenden Abschnitt importierten Daten und Kompressionsarten. In Abbildung 20 ist der jeweilige Speicherplatzverbrauch der unterschiedlichen Daten- und Kompressionskonstellationen dargestellt. Es wird deutlich, dass alle Kompressionsarten unterschiedliche Auswirkungen haben. Die nachfolgende Tabelle 16 beinhaltet die einzelnen Datengrößen aus Abbildung 20.

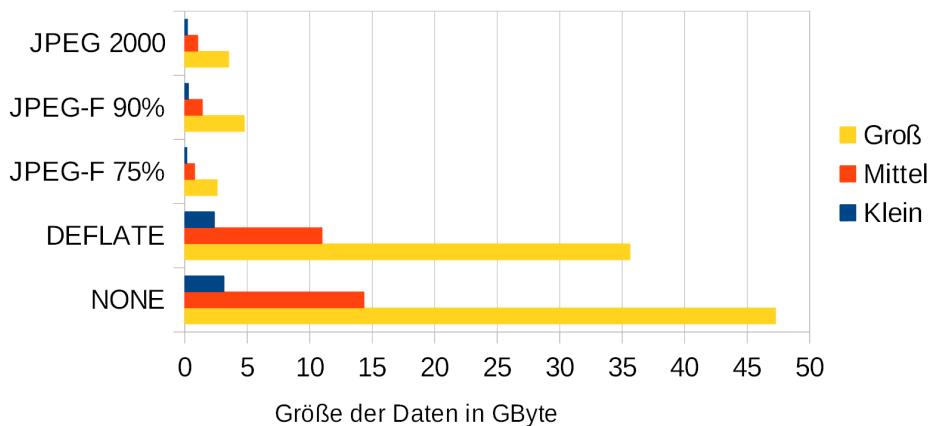


Abbildung 20: Speicherplatzverbrauch von Oracle GeoRaster

Die DEFLATE-Kompression ermöglicht nur eine Einsparung von ca. 25%. Im Gegensatz dazu sparen alle JPEG Kompressionen über 90% ein. Alle hier aufgelisteten JPEG-f bzw. JPEG 2000 Kompressionen sind verlustbehaftet und somit sind Bildinformationen verloren gegangen.

Endkompression	Klein	Mittel	Groß
	Speicherplatz in GByte		
NONE	3,15	14,35	47,28
DEFLATE	2,38	10,99	35,63
JPEG-F 75%	0,17	0,80	2,60
JPEG-F 90%	0,30	1,42	4,77
JPEG 2000	0,22	1,05	3,52

Tabelle 16: Zeiten der Komprimierung

Abschließend zur Analyse des Speicherplatzverbrauches wird ein Bezug zu den Pyramidenlevel hergestellt. In der Tabelle 17 werden die prozentualen Verteilungen des Speicherplatzes abhängig der Kompressionsart dargestellt. Dabei wurde die jeweilige Verteilung der drei Datenkonstellationen (Klein, Mittel, Groß) als Mittelwert zusammengefasst. Unabhängig der Datenkompression wird deutlich, dass die einzelnen Pyramidenlevel vom Speicherplatz ähnlich aufgeteilt sind. Die JPEG 2000 Kompression ist in der Tabelle nicht dargestellt, da sie keine von außen messbare Pyramidenlevel besitzt. Das gesamte Bild ist in einem SDO_RASTER enthalten.

Pyramiden-level	NONE	DEFLATE	JPEG-F 75%	JPEG-F 90%
	Anteil vom Gesamtspeicherplatz in Prozent			
0	74,607 %	73,749 %	69,637 %	69,455 %
1	18,739 %	19,460 %	21,387 %	21,830 %
2	4,740 %	5,050 %	6,380 %	6,292 %
3	1,301 %	1,298 %	1,854 %	1,755 %
4	0,419 %	0,331 %	0,533 %	0,486 %
5	0,128 %	0,084 %	0,151 %	0,133 %
6	0,054 %	0,021 %	0,043 %	0,037 %
7	0,009 %	0,005 %	0,012 %	0,010 %
8	0,004 %	0,001 %	0,003 %	0,003 %

Tabelle 17: Speicherplatzverbrauch der einzelnen Pyramidenlevel in unterschiedlichen Komprimierungsarten

Anfragezeiten

In der Tabelle 18 sind die jeweiligen Mittelwerte der Anfragezeiten über unterschiedlich große räumliche Ausdehnungen. Dadurch werden auch unterschiedliche Pyramidenlevel abgefragt und kein signifikanter Unterschied ist zwischen diesen Leveln erkennbar.

	NONE	DEFLATE	Anfragezeit in Sekunden	
			JPEG-F 75%	JPEG-F 90%
3 K	1546	1904	2223	2238
1,5 K	745	920	1108	1117
0,7 K	520	599	705	700
0,3 K	444	507	609	617

Tabelle 18: Anfragezeiten auf unterschiedlichen Bildgrößen und Kompressionsarten

Die Werte aus Tabelle 18 sind in der Abbildung 21 dargestellt. Die Ergebnisse zeigen deutlich, dass die Kompressionen längere Anfragezeiten hervorrufen. Die Anzahl der Pixel, abhängig von der Auflösung, haben auch einen großen Einfluss. So verdoppelt sich die Anfragezeit bei der Gegenüberstellung der 1,5 K und 3 K Auflösung. Im kleineren Auflösungsbereich von 0,3 K und 0,7 K ist die Pixelanzahl nicht so ausschlaggebend.

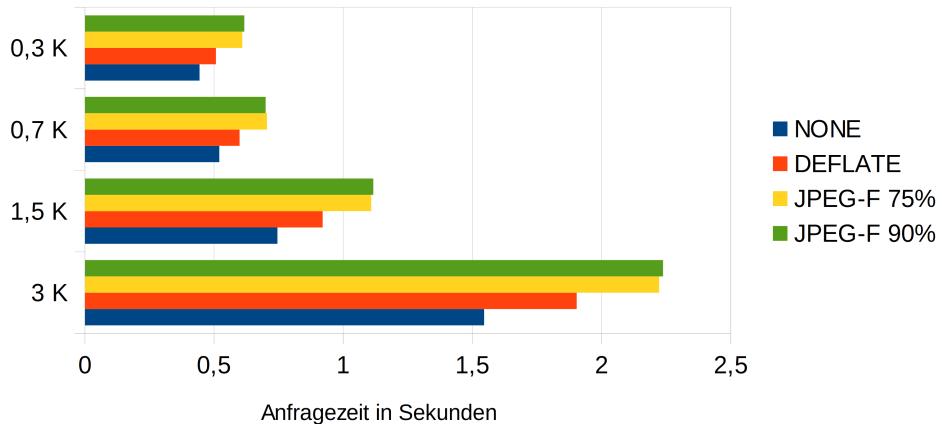


Abbildung 21: Anfragezeit von Oracle GeoRaster

Neben den eben dargestellten Messungen werden auch entsprechende Messungen auf die JPEG 2000 Daten durchgeführt. Aufgrund des sehr stark abweichenden Verhaltens sind diese separat in Abbildung 22 dargestellt. Hier hat die Bildauflösung keinerlei Auswirkung auf die Anfragezeit.

Auswirkung hingegen hat die Gesamtgröße des Bildes. Je kleiner das Bild ist, desto schneller wird die entsprechende Anfrage beantwortet. Unverhältnismäßig sind dabei die Anfragezeiten in allen Messungen. Aus diesem Grund werden die JPEG 2000 Daten in den nachfolgenden Abschnitten nicht weiter berücksichtigt.

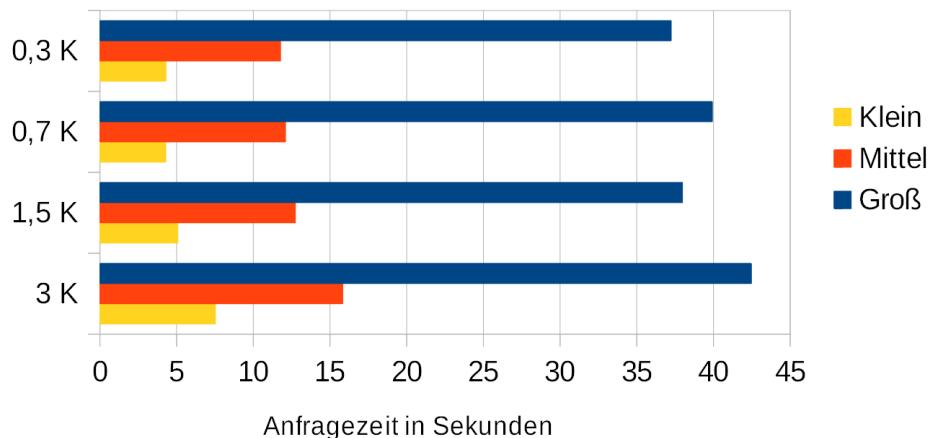


Abbildung 22: Anfragezeiten der JPEG 2000 Daten

Im direkten Verhältnis ergeben die gemessenen Anfragezeiten die in der Tabelle 19 dargestellten Werte. Dabei werden alle Auflösungen zu einem Wert zusammengefasst und ins Verhältnis gesetzt.

	NONE	DEFLATE	JPEG-F 75%	JPEG-F 90%
Anfragezeit	100,00 %	116,34 %	137,86 %	138,53 %

Tabelle 19: Vergleich der Anfragezeiten abhängig von der Kompressionsart

Neben den eigentlichen Performancemessungen werden auch Vergleichsmessungen durchgeführt. In Abbildung 23 werden gleiche Anfragen am WMS und WCS verglichen. Die Abbildung zeigt Schwankungen von bis zu 100 Millisekunden, wobei die WMS Anfragen in den meisten Fällen schneller sind als die vergleichbaren WCS Anfragen.

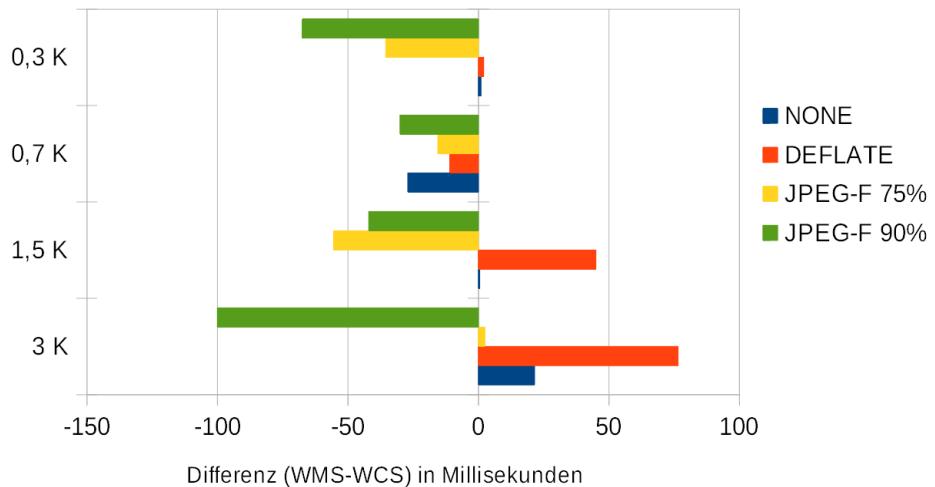


Abbildung 23: Performanceunterschiede zwischen WMS und WCS in Oracle

Eine weitere Messung wird anhand der unterschiedlichen Datengrößen durchgeführt. Damit kann die Frage beantwortet werden, ob die Größe der Datenmenge des GeoRaster Objekts Auswirkung auf die Anfragezeit hat. In Abbildung 24 wird jeweils die Datenmenge 'Groß' mit Datenmenge 'Klein' verglichen. Es sind Zeitunterschiede in beide Richtungen deutlich und deshalb kann keine Aussage zur Tendenz in diese Messrahmen getroffen werden.

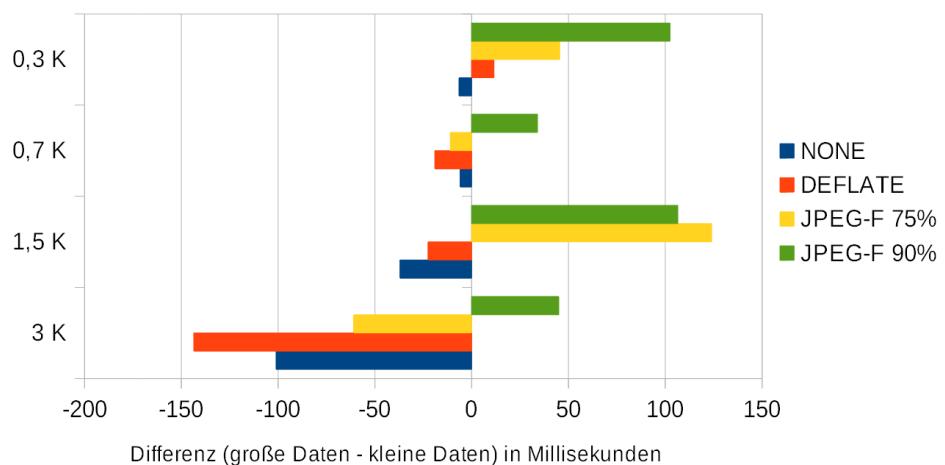


Abbildung 24: Vergleich Datenmenge Groß und Klein in Oracle

Zusammenfassung

Aus den vorhergehenden Abschnitten über die einzelnen Kriterien erfolgt nun ein Gesamtvergleich in der Tabelle 20. Als Vergleichsgrundlage dient die NONE-Kompression und ermöglicht somit eine Basis für die Bewertung der Performance von Oracle Geo-Raster mit den unterschiedlichen Kompressionsarten.

Die JPEG 2000 Kompression ist in der Tabelle auf Grund der unbefriedigenden und stark schwankenden Anfragezeiten nicht enthalten.

	NONE	DEFLATE	JPEG-F 75%	JPEG-F 90%
Importzeit	100 %	168 %	184 %	188 %
Speicherplatz	100 %	75 %	6 %	10 %
Anfragezeit	100 %	116 %	138 %	138 %
Bildqualität	100 %	100 %	75 %	90 %

Tabelle 20: Vergleich der Kompressionen in Bezug auf die analysierten Kriterien

6.3 PostGIS raster

Rasterdatenimport

Der Import der Daten erfolgt außerhalb der Datenbank mittels raster2pgsql Funktionalität von PostGIS. Die Overview Level werden anschließend einzeln erstellt. Die zeitliche Verteilung der einzelnen Schritte ist in Tabelle 21 zu sehen. Die gewählte Kachelgröße beträgt beim Import 1000x1000 Pixel.

Import Schritt	Klein		Mittel		Groß		Mittel-wert
	Zeit in Sekunden	Anteil in Prozent	Zeit in Sekunden	Anteil in Prozent	Zeit in Sekunden	Anteil in Prozent	
1. Import der Rasterdaten	238	37,84 %	1170	35,76 %	3932	36,94 %	36,85 %
2. Bildung Overview level 2	297	47,22 %	1517	46,36 %	4854	45,61 %	46,40 %
3. Bildung Overview Level 5	50	7,95 %	356	10,88 %	1000	9,40 %	9,41 %
4. Bildung Overview Level 10	23	3,66 %	117	3,58 %	400	3,76 %	3,66 %
5. Bildung Overview Level 25	7	1,11 %	43	1,31 %	156	1,47 %	1,30 %
6. Bildung Overview Level 50	5	0,79 %	28	0,86 %	116	1,09 %	0,91 %
7. Bildung Overview Level 100	5	0,79 %	21	0,64 %	101	0,95 %	0,80 %
8. Bildung Overview Level 250	4	0,64 %	20	0,61 %	84	0,79 %	0,68 %
Gesamt	629	100 %	3272	100 %	3516	100 %	

Tabelle 21: Zeitliche Verteilung des Imports bei in-db Speicherung

Schritt 1 und 2 belegen zusammen über 80% des gesamten Importvorgangs. Verdeutlicht wird diese Verteilung bei der in-db Speicherung in Abbildung 25. Die Schritte 5 bis 8 werden aus Gründen der Übersichtlichkeit in der Abbildung als Rest mit 3,69% zusammengefasst.

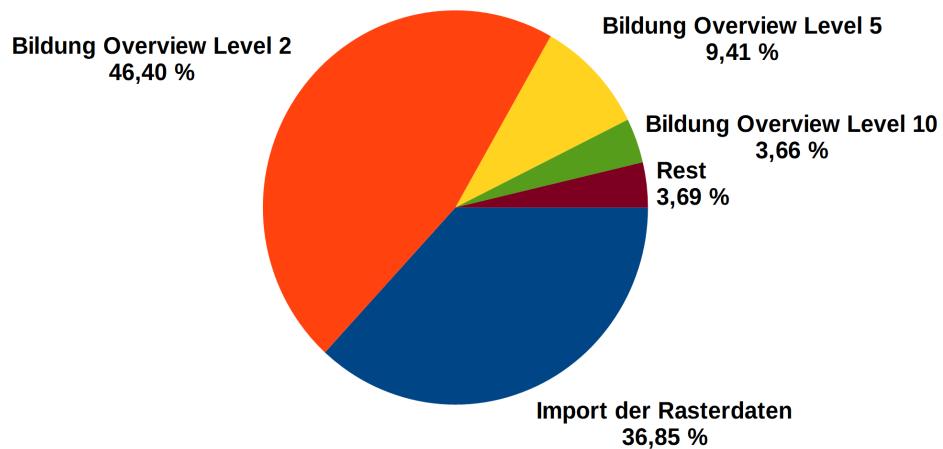


Abbildung 25: Prozentuale Verteilung der Importzeit in PostGIS in-db Variante

Bei der out-db Variante ist die zeitliche Verteilung des Imports deutlich verschoben. In Tabelle 22 wird deutlich, der Import der Rasterdaten ist deutlich kürzer, da nur die Metadaten in die Datenbank abgespeichert werden. Im Gegenzug werden die nachfolgenden Schritte der Bildung der Overview Level deutlich verlängert, da die Daten erst aus dem Dateisystem geladen werden.

Import Schritt	Klein		Mittel		Groß		Mittelwert
	Zeit in Sekunden	Anteil in Prozent	Zeit in Sekunden	Anteil in Prozent	Zeit in Sekunden	Anteil in Prozent	
1. Import der Rasterdaten	3	0,31 %	12	0,28 %	51	0,40 %	0,33 %
2. Bildung Overview level 2	556	57,44 %	2227	52,76 %	6949	54,61 %	54,94 %
3. Bildung Overview Level 5	142	14,67 %	699	16,56 %	1857	14,59 %	15,27 %
4. Bildung Overview Level 10	72	7,44 %	293	6,94 %	1016	7,98 %	7,45 %
5. Bildung Overview Level 25	50	5,17 %	250	5,92 %	897	7,05 %	6,05 %
6. Bildung Overview Level 50	50	5,17 %	237	5,61 %	650	5,11 %	5,30 %
7. Bildung Overview Level 100	51	5,27 %	252	5,97 %	611	4,80 %	5,35 %
8. Bildung Overview Level 250	44	4,55 %	251	5,95 %	693	5,45 %	5,31 %
Gesamt	968	100 %	4221	100 %	12724	100 %	

Tabelle 22: Zeitliche Verteilung des Imports bei out-db Speicherung

Die Gesamtimportzeiten aus in-db Variante (Tabelle 21) und out-db Variante (Tabelle 22) werden in der Abbildung 26 visuell gegenübergestellt.

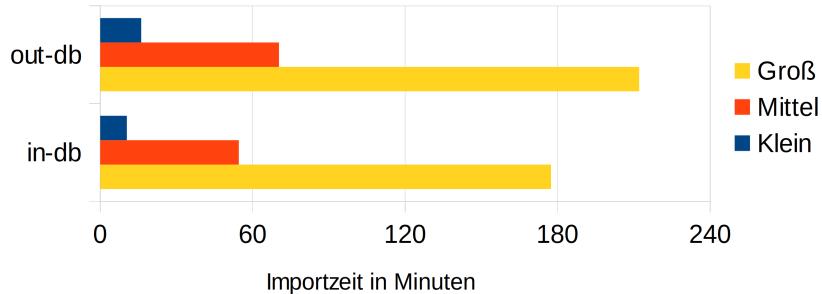


Abbildung 26: Importzeiten in PostGIS

Speicherplatzverbrauch

Die Speicherung von Rasterdaten in der Datenbank ist stark eingeschränkt. Der einzige mögliche Vergleich ist der Speicherplatzverbrauch zwischen der in-db und der out-db Speicherung. Bei der out-db Speicherung können komprimierte Daten hinterlegt werden, um Speicherplatz zu sparen. In Abbildung 27 sind die getesteten Varianten graphisch dargestellt.

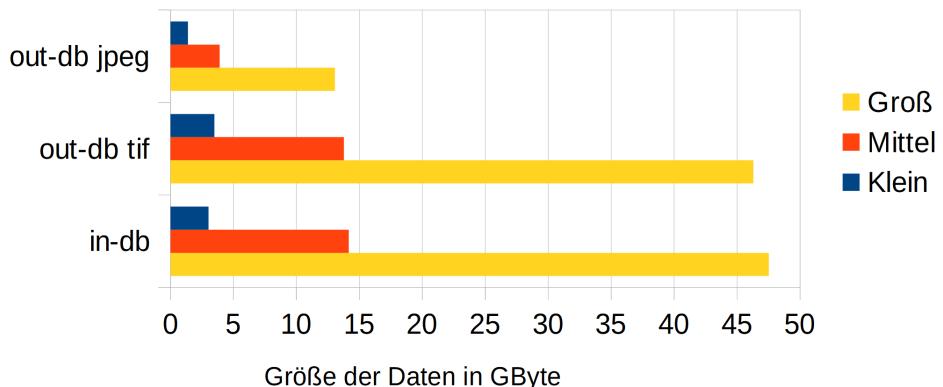


Abbildung 27: Speicherplatzverbrauch abhängig von Speicherort in PostGIS

In der nachfolgenden Tabelle 23 sind die jeweiligen Werte des Speicherplatzverbrauchs dargestellt, welche in Abbildung 27 dargestellt sind.

	Klein	Mittel	Groß
	Speicherplatz in GByte		
in-db	3,02	14,16	47,52
out-db tif	3,49	13,78	46,29
out-db jpeg	1,39	3,90	13,07

Tabelle 23: Speicherplatzverbrauch abhängig von Datenmenge und Speicherort

Der einzige Unterschied liegt bei den Originaldaten, die entweder direkt in der Datenbank liegen oder nur einen Verweis auf Bilddaten im Dateisystem aufweisen. In Tabelle 24 ist die prozentuale Verteilung in der jeweiligen Variante dargestellt. Es handelt sich dabei jeweils um die Mittelwerte über die 3 Testdatenmengen.

Overview Level	in-db	out-db tif	out-db jpeg
	Anteil vom Gesamtspeicherplatz in Prozent		
1	76,791 %	0,011 %	0,041 %
2	19,203 %	19,735 %	69,678 %
5	3,073 %	3,158 %	11,151 %
10	0,769 %	0,790 %	2,789 %
25	0,123 %	0,127 %	0,448 %
50	0,031 %	0,032 %	0,113 %
100	0,008 %	0,008 %	0,029 %
250	0,002 %	0,002 %	0,006 %
Dateisystem	0,000 %	76,137 %	15,746 %

Tabelle 24: Prozentualer Speicherplatzverbrauch der einzelnen Overview Level

In allen Varianten unterscheidet sich die absolute Datenmengen der Overview Level nicht. Diese sind exakt gleich groß, da es sich jeweils um dieselben Daten handelt.

Anfragezeiten

Die nachfolgenden Messungen beziehen sich nur auf die Daten der in-db Variante. Die out-db Variante kann aufgrund von Kommunikationsproblemen in der Testkonstellation zwischen Geoserver mit WMS und der PostGIS Datenbank mit out-db Speicherung nicht durchgeführt werden. Die Anfragezeiten der out-db Variante unterscheidet sich zu der in-db nur in den Anfragen auf die Originaldaten, da die einzelnen Overview Stufen sich in ihrer Struktur nicht unterscheiden. Es ist aber ein ähnliches Verhalten zu erwarten wie beim Import, d.h. längere Anfragezeiten in allen Konstellationen, nur konkrete Werte können hier nicht geliefert werden.

In Abbildung 28 sind die Anfragezeiten dargestellt, abhängig von der Auflösung. Deutlich wird der starke Anstieg der Anfragezeit bei höherer Pixelanzahl.

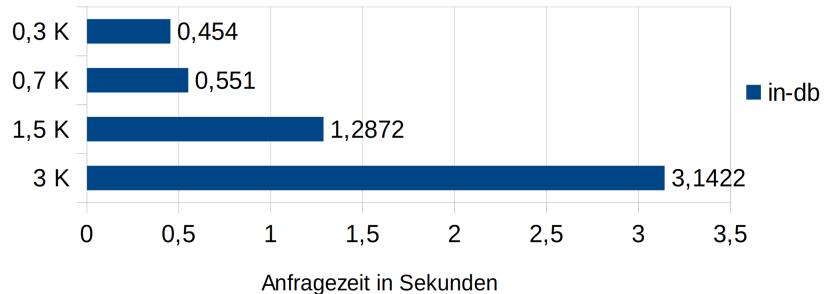


Abbildung 28: Anfragezeiten in PostgreSQL/PostGIS

Beim Vergleich zwischen WMS und WCS Anfragen wird in der Abbildung 29 verdeutlicht, dass diese kaum Unterschiede in der Performance erzeugen. Eine maximale Differenz von 37ms wurde dabei ermittelt.

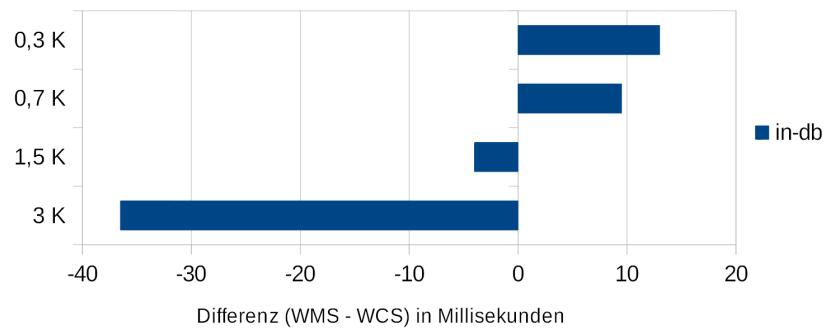


Abbildung 29: Performanceunterschiede zwischen WMS und WCS in PostGIS

In Abbildung 30 werden die Datenmengen gegenübergestellt und verglichen, ob die Größe der Datenmenge die Anfragezeiten beeinflussen. Es wird deutlich, dass die Größe in PostGIS eine entscheidende Rolle spielt. Bei kleinen Auflösungen von 0,3K und 0,7K werden die Anfragen um ca. 200 ms verzögert. In den größeren Auflösungen sind die Differenzen bei über einer halben Sekunde.

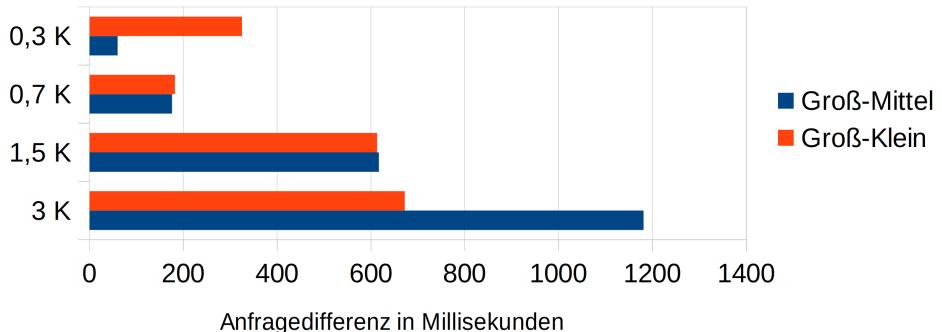


Abbildung 30: Anfragedifferenzen aufgrund unterschiedliche großer Datenmengen

Zusammenfassung

Die angeboten Importvarianten, in-db und out-db, unterscheiden sich in der Dauer des Imports deutlich. Das Lesen aus dem Dateisystem beeinträchtigt des Erstellen der einzelnen Overview Level deutlich um mindestens 20%.

Die Möglichkeiten zum Variieren des benötigten Speicherplatzes sind nur begrenzt vorhanden. Eine teilweise Komprimierung ist nur außerhalb der Datenbank möglich. Im getesteten Fall gab es eine Einsparung von ca. 70%.

Die Anfragezeit kann in PostGIS nicht gut vorausgesagt werden, da unterschiedliche Faktoren ineinander greifen und die Anfragezeit beeinflussen. Zum einen ist die Gesamtgröße der Daten von großer Relevanz, denn kleinere Datenmengen haben eine kürzere Antwortzeit als größere. Des weiteren ist die Anzahl der Pixel ein nicht zu unterschätzender Faktor. Dieser erhöht die Anfragezeiten deutlich.

6.4 Gegenüberstellung

Zeitverbrauch des Datenimports

In Abbildung 31 und Tabelle 25 sind die aus dem vorhergehenden Kapitel ermittelten Importzeiten dargestellt. Deutlich ist zu sehen, dass beide Importvarianten (in-db und out-db) von PostGIS höhere Importzeiten haben als alle Oracle Varianten. Die NONE-Variante von Oracle mit der Großen Datenmenge hat eine ähnliche Importzeit wie die mittlere Datenmenge in der in-db Variante.

Diese Gegenüberstellung macht deutlich, dass Oracle trotz der komplexen Rasterdatenstruktur mit zwei Datentypen und mehrstufigen Importvorgangs inklusive Mosaikbildung einen signifikanten Geschwindigkeitsvorteil gegenüber PostGIS besitzt.

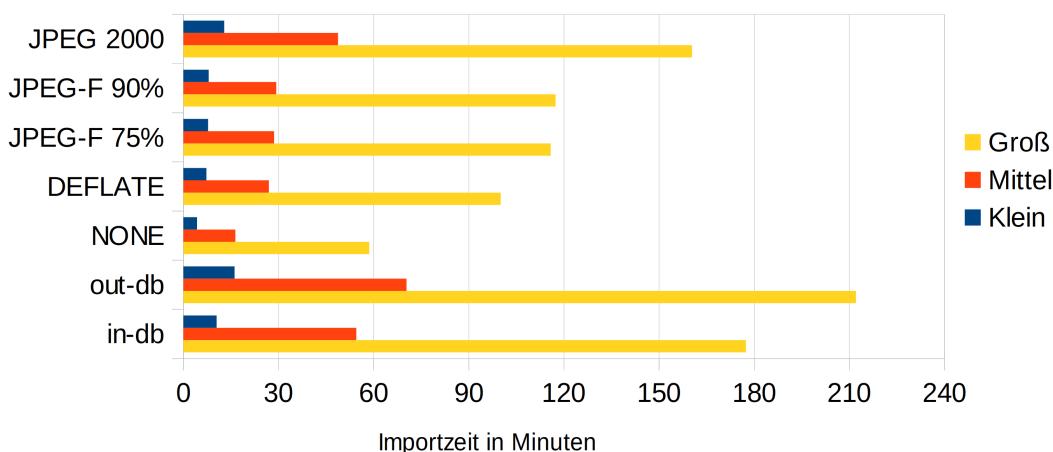


Abbildung 31: Gegenüberstellung der Importzeiten in Oracle und PostGIS

	Klein	Mittel	Groß
	Zeit in Minuten		
in-db	10	55	177
out-db	16	70	212
NONE	4	16	59
DEFLATE	7	27	100
JPEG-F 75%	8	29	116
JPEG-F 90%	8	29	117
JPEG 2000	13	49	160

Tabelle 25: Auflistung der Importzeiten in Oracle und PostGIS

Speicherplatzverbrauch

In Abbildung 32 und Tabelle 26 ist der aus dem vorhergehenden Kapitel jeweils ermittelte Speicherplatzverbrauch dargestellt. Die Ausgangssituationen beider DBMS sind ähnlich, sowohl die NONE-Variante, die in-db Variante und out-db tif Variante bewegen sich in ähnlichen Größenordnungen, mit minimalen Abweichungen. Die JPEG-F Kompressionsarten sind deutlich vor der out-db jpeg Variante aufgrund der unkomprimierten Overview Level in der Datenbank.

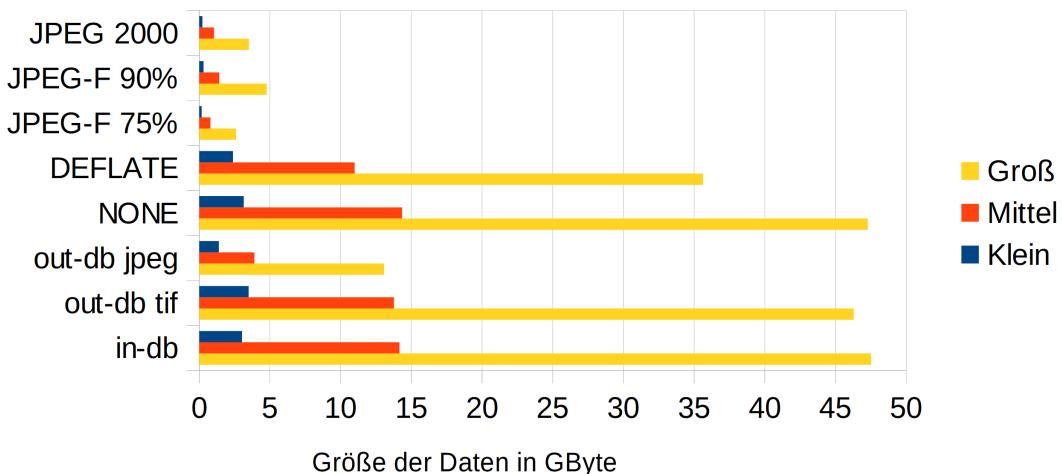


Abbildung 32: Gegenüberstellung des Speicherplatzverbrauchs in Oracle und PostGIS

	Klein	Mittel	Groß
	Speicherplatz in GByte		
in-db	3,02	14,16	47,52
out-db tif	3,49	13,78	46,29
out-db jpeg	1,39	3,90	13,07
NONE	3,15	14,35	47,28
DEFLATE	2,38	10,99	35,63
JPEG-F 75%	0,17	0,80	2,60
JPEG-F 90%	0,30	1,42	4,77
JPEG 2000	0,22	1,05	3,52

Tabelle 26: Auflistung des Speicherplatzverbrauches in Oracle und PostGIS

Analyse der Anfragezeiten

In Abbildung 33 und Tabelle 27 sind die aus dem vorhergehenden Kapitel ermittelten Anfragezeiten dargestellt. Bei kleinen Auflösungen ist PostGIS genau so schnell wie die NONE-Kompression von Oracle. Je größer die Auflösung wird, desto deutlicher wird der Unterschied zwischen Oracle GeoRaster und PostGIS raster. Ab 1,5K ist PostGIS schon langsamer als die JPEG-F Varianten und bei sehr großen Auflösungen wird der Unterschied noch deutlicher.

Bei den Ergebnisbildern ist auffällig, das vereinzelt Lücken zwischen den einzelnen Bildern trotz angepasster Overview Level entstehen. Dadurch wird in manchen Fällen ein unsauberes Ergebnis erzeugt.

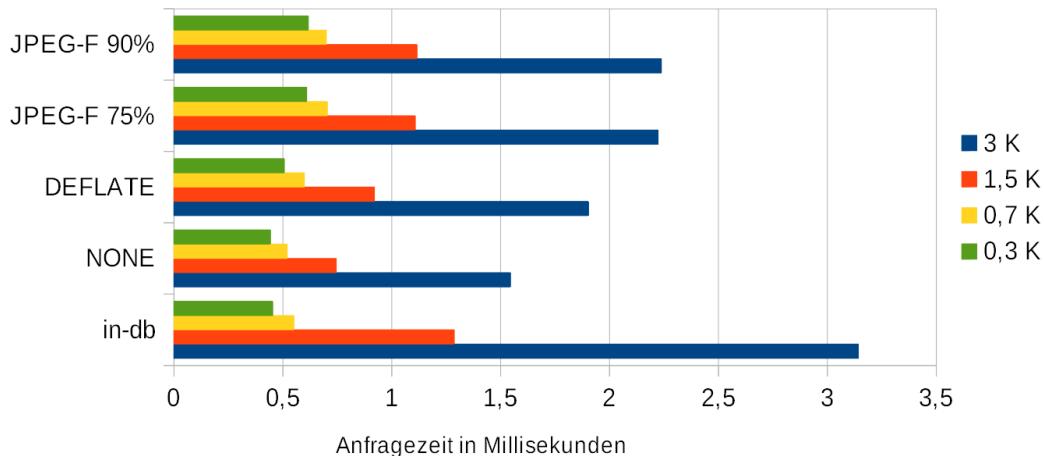


Abbildung 33: Gegenüberstellung der Anfragezeiten in Oracle und PostGIS

	in-db	NONE	DEFLATE	JPEG-F 75%	JPEG-F 90%
	Anfragezeit in Sekunden				
3 K	3,14	1,55	1,90	2,22	2,24
1,5 K	1,29	0,75	0,92	1,11	1,12
0,7 K	0,55	0,52	0,60	0,70	0,70
0,3 K	0,45	0,44	0,51	0,61	0,62

Tabelle 27: Auflistung der Anfragezeiten in Oracle und PostGIS

7 Bewertung der rasterbasierten DBMS

Die Nutzwertanalyse bezieht sich auf die vorhergehenden Kapitel 5 und 6. Aufbauend auf den erarbeiteten Informationen sind entsprechende Kriterien aufgestellt und gewichtet. In der folgenden Abbildung 34 sind die einzelnen Kriterien dargestellt.

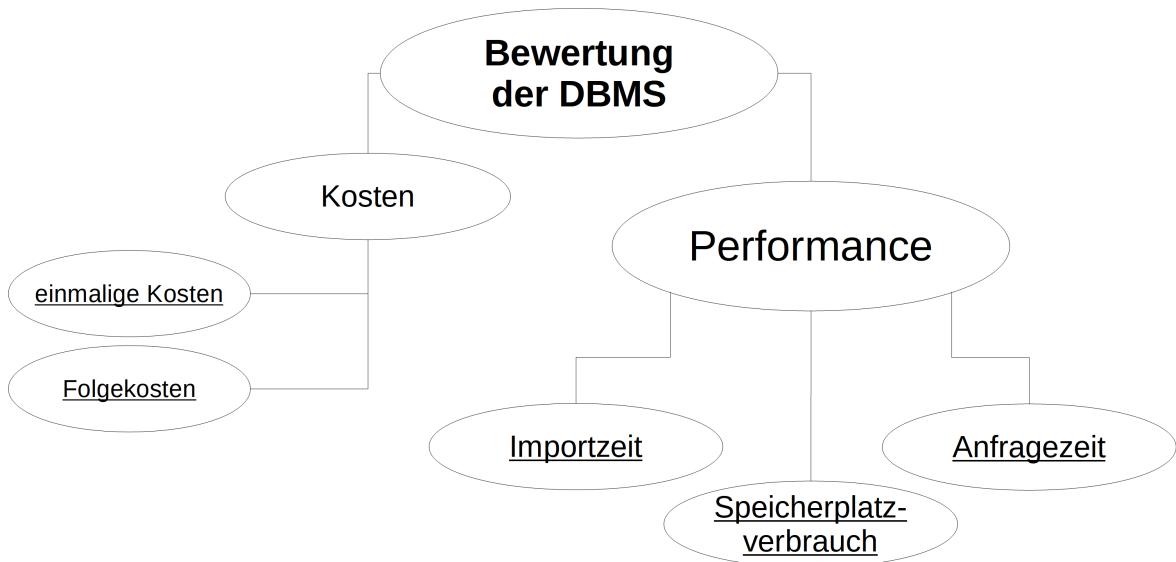


Abbildung 34: Darstellung der einzelnen Kriterien

7.1 Erklärung Bewertungssystem

Werteskala der Gewichtung

Mithilfe der Gewichtung wird in der nachfolgenden Nutzwertanalyse ein direkter Bezug zur Problemstellung der *profil rdv*. Der Wertebereich der Gewichtung lautet wie folgt:

1. kaum relevant
2. relevant
3. sehr relevant
4. besonders relevant

Die Gewichtung wird vor jeder aufgestellten Nutzwerttabelle für jedes Kriterium festgelegt und begründet.

Festlegung der Bewertungsskala

Die Bewertung der einzelnen Kriterien erfolgt Objektiv und unabhängig von der Problemstellung. Die Skala der Bewertung erstreckt sich von 1 (sehr schlecht) bis 5 (sehr gut). Die Bewertung wird vor jeder aufgestellten Nutzwerttabelle für jedes Kriterium festgelegt und begründet. Folgend sind die einzelnen Abstufungen dargestellt:

- Kosten: einmalige Kosten / Folgekosten
 1. sehr hohe
 2. hohe
 3. mittlere
 4. geringe
 5. keine
- Performance: Importzeit/Speicherplatzverbrauch/Anfragezeit
 1. sehr hohe(r)
 2. hohe(r)
 3. mittlere(r)
 4. geringe(r)
 5. sehr geringe(r)

7.2 Kosten

Gewichtung

1 einmalige Kosten

für langfristige Entscheidungen nicht unbedingt ausschlaggebend

3 Folgekosten

regelmäßige Ausgaben sind für längerfristige Entscheidungen ein wichtiger Faktor

Bewertung der Kriterien

- Oracle

2 einmalige Kosten

mehrere benötigte Lizenzen mit zusätzlichem Support

2 Folgekosten

aufgrund des Lizenz- und Supportmodell

- PostgreSQL/PostGIS

5 einmalige Kosten

Open Source Modell

4 Folgekosten

Kommerziellen Support oder eigener Mehraufwand bei Fehlerbehebung ohne Ansprechpartner

Nutzwerttabelle

Kriterium	Gewichtung	Oracle		PostgreSQL/PostGIS	
		Bewertung	Gewichtet	Bewertung	Gewichtet
einmalige Kosten	2	2	4	5	10
Folgekosten	3	2	6	4	12
Gesamt	5		10		22

Tabelle 28: Nutzwerttabelle der Kosten

7.3 Performance

Gewichtung

2 Importzeit

nur im gewissen Maße relevant, wird nicht oft in der **profil rdv** durchgeführt, soll trotzdem schnell sein

4 Speicherplatzverbrauch

aufgrund der Restriktionen der **profil rdv** des Festplattenspeichers hoch angesetzt

3 Anfragezeit

Kriterium war der Auslöser der Analyse und ist dementsprechend hoch gewichtet

Bewertung der Kriterien

Die Bewertung der Kriterien erfolgt anhand der Messwerte aus Kapitel 6. In Tabelle 29 sind diese jeweils anhand der Bewertungsskala umgewandelt. Die Zahlen in Klammern wurden nicht direkt gemessen, sind aufgrund der Vollständigkeit enthalten.

	Importzeit	Speicherplatz-verbrauch	Anfragezeit
in-db	2	1	2
out-db tif	1	1	(1)
out-db jpeg	1	3	(1)
NONE	5	1	5
DEFLATE	4	2	4
JPEG-F 75%	3	5	3
JPEG-F 90%	3	4	3

Tabelle 29: Bewertung der Performanceanalyse

Nutzwerttabelle

Aus Gründen der Übersicht ist die Nutzwerttabelle zweigeteilt, zum einen Oracle in Tabelle 30 und zum anderen PostGIS in Tabelle 31.

Kriterium	Gewichtung	NONE		DEFLATE		JPEG-F 75%		JPEG-F 90%	
		Bew.	Gew.	Bew.	Gew.	Bew.	Gew.	Bew.	Gew.
Importzeit	2	5	10	4	8	3	6	3	6
Speicherplatzverbrauch	4	1	4	2	8	5	20	4	16
Anfragezeit	3	5	15	4	12	3	9	3	9
Gesamt	9		29		28		35		31

Tabelle 30: Nutzwerttabelle der Performance von Oracle GeoRaster

Kriterium	Gewichtung	in-db		out-db tif		out-db jpeg	
		Bew.	Gew.	Bew.	Gew.	Bew.	Gew.
Importzeit	2	2	4	1	2	1	2
Speicherplatzverbrauch	4	1	4	1	4	3	12
Anfragezeit	3	2	6	(1)	(3)	1	(3)
Gesamt	9		14		(9)		(17)

Tabelle 31: Nutzwerttabelle der Performance von PostGIS raster

Die Bewertung macht deutlich, das Oracle in allen Punkten deutlich besser ist. In der Bewertung wird die Bildqualität nicht mit berücksichtigt. Demnach müsste die JPEG-F 75% noch etwas an Punkten verlieren.

7.4 Abschlussbewertung

In diesem Abschnitt werden die Kosten und die Performance nach gleichem Schema miteinander verknüpft und zusammengerechnet. Dabei ergeben die errechneten Werte die Bewertung.

Gewichtung

2 Kosten

wird nicht oft in der **profil rdv** durchgeführt aber trotzdem schnell sein

4 Performance

aufgrund der Restriktionen der **profil rdv** des Festplattenspeichers hoch angesetzt

Nutzwerttabelle

Aus Gründen der Übersicht ist die Nutzwerttabelle zweigeteilt, zum einen Oracle in Tabelle 32 und zum anderen PostGIS in Tabelle 33.

Kriterium	Gewichtung	NONE		DEFLATE		JPEG-F 75%		JPEG-F 90%	
		Bew.	Gew.	Bew.	Gew.	Bew.	Gew.	Bew.	Gew.
Kosten	2	10	20	10	20	10	20	10	20
Performance	4	29	116	28	112	35	140	31	124
Gesamt	9	136			132			160	

Tabelle 32: Nutzwerttabelle der Abschlussbewertung von Oracle GeoRaster

Kriterium	Gewichtung	in-db		out-db tif		out-db jpeg	
		Bew.	Gew.	Bew.	Gew.	Bew.	Gew.
Kosten	2	22	44	22	44	22	44
Performance	4	14	56	(9)	(36)	(17)	(68)
Gesamt	6	100		(80)		(112)	

Tabelle 33: Nutzwerttabelle der Abschlussbewertung von PostGIS raster

Trotz der finanziellen Vorteile die PostgreSQL/PostGIS mit sich bringt, ist Oracle aufgrund der besseren Performance im Fall der **profil rdv** das geeignete DBMS.

8 Zusammenfassung und Ausblick

8.1 Zusammenfassung

Ziel der Arbeit war die Analyse einer PostgreSQL/PostGIS Datenbank als mögliche Alternative zu einer Oracle Datenbank für die ***profil rdv***.

Eingangs wurden die grundlegende Struktur und Aufgabenbereiche der ***profil rdv*** dargestellt und der Auslöser einer Evaluierung der DBMS mit einer gegebenen Problemstellung beschrieben. Anschließend wurden die beide DBMS in ihrer Struktur in Bezug auf Rasterdaten näher untersucht und eine prototypische Umsetzung für die ***profil rdv*** grob aufgezeigt. Für die Bewertung der auftretenden Kosten beim Betreiben einer Datenbank wurden die Kostenmodelle der DBMS näher erläutert und gegenübergestellt. In der anschließenden Performanceanalyse wurde in einem klar definierten Testrahmen ***profil rdv*** relevante Kriterien genauer untersucht. Dabei stellte sich heraus, das PostGIS raster, aufgrund der unterschiedlichen Struktur, Nachteile in allen getesteten Kriterien aufweist. Zusätzlich korreliert die Anfragezeit mit der Größe der Datenmenge und da die Testdaten nur ein Teil der Daten in der ***profil rdv*** sind, wird dort eine größere Auswirkung des Verhaltens erwartet.

In der abschließenden Bewertung mittels Nutzwertanalyse stellte sich heraus, das PostgreSQL/PostGIS in den zur Bewertung aufgestellten Kriterien deutlich hinter Oracle liegt. Bei einer Bewertung mit anderen Schwerpunkten kann die Nutzwertanalyse ein anderes Ergebnis aufweisen.

8.2 Ausblick

Oracle GeoRaster

In Oracle 12.2 ist die JPEG2000 Kompression hinzugekommen, die in dieser Arbeit aufgrund von schlechten Anfragezeiten nur bis zu einem bestimmten Punkt untersucht wurde. Die eigentlich mögliche Performance bzw. passende Aufgabenbereiche konnte hier nicht aufgezeigt werden, aufgrund falscher Handhabung oder falscher Einsatzgebiete. Durch diese Neuerung wird deutlich, dass Oracle die Funktionalitäten des GeoRaster's stetig erweitert und in einigen Bereichen optimiert.

PostGIS raster

Aus den Unterlagen [PCB18] [PGA15] [PGM18] wird deutlich, dass PostGIS raster viele Funktionalitäten zur Bildanalyse enthält und somit einen anderen Fokus im Bereich der Rasterdaten hat. Bei neuen Anforderungen der **profil rdv** kann diese Analyse mit anderer Gewichtung als Grundlage für weitere Analysen genutzt werden.

Des weiteren wurde keine Erweiterungen durch Dritte mit einbezogen, die eine Performanceverbesserung oder Kompression innerhalb einer PostgreSQL/PostGIS Datenbank ermöglichen könnten.

Offene Punkte

Im Bereich der Importzeit von Oracle ist zusätzlich Optimierung durch Parallelisierung einiger Schritte möglich, z.B. bei der Bildung der Pyramidenlevel oder der Datenkomprimierung. Diese wurde hier nicht berücksichtigt.

Bei der Analyse der Anfragezeit wurde der Einflussfaktor 'Datenübertragung über ein Netzwerk' in der Testumgebung vollkommen ausgeschlossen. Deshalb spiegelt die Zeit nicht vollständig die Realität wieder und könnte somit zu etwas längeren Anfragezeiten führen.

Glossar und Abkürzungsverzeichnis

BLOB	Binary Large Object
BRIN	Block Range Index
CLOB	Character Large Object
DBMS	Datenbankmanagementsystem
DOP	digitales Orthophoto
GDAL	Geospatial Data Abstraction Library
GiST	Generalized Search Tree
InVeKoS	Integriertes Verwaltungs- und Kontrollsyste
in-DB	Rasterdatenspeicherung innerhalb der Datenbank
JAI	Java Advanced Imaging
LOB	Large Object
LVermGeo SH	Landesamt für Vermessung und Geoinformation Schleswig-Holstein
MELUND	Ministerium für Energiewende, Landwirtschaft, Umwelt, Natur und Digitalisierung
OGC	Open Geospatial Consortium
out-DB	Rasterdatenspeicherung außerhalb der Datenbank
profil	Programmsystem zur rechnergestützten Verwaltung von Fördermitteln in der Landwirtschaft
profil rdv	profil Rasterdaten-Verwaltung
SH	Schleswig-Holstein
SP-GiST	Space-Partitioned Generalized Search Tree
WCS	Web Coverage Service
WFS	Web Feature Service
WMS	Web Map Service

Literatur

- [DBE19] DB_Engines Ranking von Relational DBMS
<https://db-engines.com/de/ranking/relational+dbms> Zugriff 26.02.2019
- [DEG13] Detailkonzept: Verwaltung von Rasterdaten (ab 2013) Zugriff: 28.01.2019
- [DEP19] data experts ProfilGlossar
<http://profilglossar.data-experts.de/index.php/Profil> Zugriff: 5.03.2019
- [DVH19] DV-Handbuch für die Anwendung profil c/s, profident, rdv und aiv, Komplex Hard- und Softwareempfehlungen, gültig ab 3/2019 Zugriff: 29.01.2019
- [GEO19] Website GeoServer <http://geoserver.org/> Zugriff: 20.3.2019
- [JME19] Website JMeter <https://jmeter.apache.org/> Zugriff 13.2.2019
- [LVG19] LVerMGeo SH, Digitales Orthophoto
https://www.schleswig-holstein.de/DE/Landesregierung/LVERMGEOSH/Service/serviceGeobasisdaten/geodatenService_Geobasisdaten_DOP.html
Zugriff: 16.03.2019
- [NWA17] Macke S. (2017). Nutzwertanalyse in der Projektdokumentation
<https://fachinformatiker-anwendungsentwicklung.net/nutzwertanalyse-in-der-projektdokumentation/>
Zugriff: 05.02.2019
- [OLS19] <http://www.oracle.com/us/support/library/lsp-tech-chart-069290.pdf>
- [OPL19] Oracle Technology Global Price List January 17, 2019 Software Investment Guide <https://www.oracle.com/assets/technology-price-list-070617.pdf>
Zugriff: 28.01.2019
- [ORA17] Murray, C. (2017): Oracle(r) Spatial and Graph, GeoRaster Developer's Guide GeoRaster 12c Release 2 (12.2) E85863-01
<https://docs.oracle.com/en/database/oracle/oracle-database/12.2/geors/>
Zugriff: 07.03.2019
- [ORE19] Database Licensing Information User Manual 12.2
<https://docs.oracle.com/en/database/oracle/oracle-database/12.2/dblic/Licensing-Information.html>
Zugriff: 07.03.2019

-
- [PCB18] Zurbaran, M., Wightman, P. M., Corti, P., Mather, S. V., Kraft, T. J., & Park, B. (2018): PostGIS Cookbook-: Store, organize, manipulate, and analyze spatial data.
- [PGA11] Obe, R., & Hsu, L. (2011): PostGIS in Action
- [PGA15] Obe, R., & Hsu, L. (2015): PostGIS in Action: Second Edition
- [PGM18] PostGIS 2.5.1 Manual, Chapter 5 Raster Data Management, Queries, and Applications <https://download.osgeo.org/postgis/docs/postgis-2.5.1.pdf>
Zugriff: 20.03.2019
- [PRD12] PostGIS Raster Beta Documentation (2012)
<https://trac.osgeo.org/postgis/wiki/WKTRaster/Documentation01>
Zugriff: 29.01.2019
- [SQL19] Oracle SQL Developer
<https://www.oracle.com/database/technologies/appdev/sql-developer.html>
Zugriff: 14.2.2019
- [TDB15] Kudraß, T. (Ed.). (2015): Taschenbuch Datenbanken. Carl Hanser Verlag GmbH Co KG.
- [WAM98] Züllighoven, H. (1998): Das objektorientierte Konstruktionshandbuch nach dem Werkzeug & Material-Ansatz. dpunkt.
- [WCS19] OGC Web Coverage Service, Dokumentation
<http://www.opengeospatial.org/standards/wcs>
Zugriff: 14.03.2019
- [WMS19] OGC Web Map Service, Dokumentation
<http://www.opengeospatial.org/standards/wms>
Zugriff: 14.03.2019

Abbildungsverzeichnis

1	Beispiel eines DOP's [LVG19]	2
2	UML Diagramm der fachlichen Bezeichnungen der profil rdv	3
3	Aktivitätsdiagramm des Rasterdaten-Importeurs	3
4	Darstellung der Suchkriterien des Rasterbild-Explorers [DEG13]	5
5	Beispielhafte Darstellung des Rasterbild-Viewer [DEG13]	5
6	3 Schichten Modell, profil rdv und profil c/s	6
7	Sequenzdiagramm einer WCS Anfrage	7
8	Aufbau der Datentypen SDO_GEORASTER und SDO_RASTER [ORA17]	10
9	Darstellung der GeoRaster Datenbankstruktur [ORA17]	11
10	Formel zur Ermittlung der entsprechenden Pyramidenlevelauflösung	12
11	Darstellung der Pyramidenlevel in Oracle im Vergleich zum Original	12
12	Formel zur Berechnung der Overview's	16
13	Darstellung der Overview's im Vergleich zum Original	16
14	Darstellung der Tabellen in der profil rdv mit Hilfe von Oracle GeoRaster .	18
15	Darstellung der Tabellen in der profil rdv mit Hilfe von PostGIS raster . .	19
16	Struktur der Testumgebung	28
17	Räumliche Ausdehnung der Datenkonstellationen	29
18	Prozentuale Gegenüberstellung der Zeit intensiven Importschritte	31
19	Importzeit abhängig von Kompression und Datenmenge	32
20	Speicherplatzverbrauch von Oracle GeoRaster	33
21	Anfragezeit von Oracle GeoRaster	35
22	Anfragezeiten der JPEG 2000 Daten	36
23	Performanceunterschiede zwischen WMS und WCS in Oracle	37
24	Vergleich Datenmenge Groß und Klein in Oracle	37
25	Prozentuale Verteilung der Importzeit in PostGIS in-db Variante	40
26	Importzeiten in PostGIS	41
27	Speicherplatzverbrauch abhängig von Speicherort in PostGIS	41
28	Anfragezeiten in PostgreSQL/PostGIS	43
29	Performanceunterschiede zwischen WMS und WCS in PostGIS	43
30	Anfragedifferenzen aufgrund unterschiedliche großer Datenmengen	44
31	Gegenüberstellung der Importzeiten in Oracle und PostGIS	45
32	Gegenüberstellung des Speicherplatzverbrauchs in Oracle und PostGIS .	46
33	Gegenüberstellung der Anfragezeiten in Oracle und PostGIS	47
34	Darstellung der einzelnen Kriterien	48
35	Anfragenverteilung der Pyramidenlevel der profil rdv	61

Tabellenverzeichnis

1	Parameter einer einfachen WCS Anfrage	7
2	Anfragezeiten im aktuellen Oracle DBMS der profil rdv	8
3	Attribute des Oracle Datentyps SDO_GEORASTER [ORA17]	9
4	Attribute des Oracle Datentyps SDO_RASTER [ORA17]	10
5	Attribute von PostGIS raster [PRD12]	14
6	Vor- und Nachteile der in-DB Datenspeicherung [PRD12]	15
7	Vor- und Nachteile der out-DB Datenspeicherung [PRD12]	15
8	Auflistung der Kriterien und ihre Umsetzung in dem DBMS	21
9	Beispielrechnungen für Lizenzkosten [OPL19]	23
10	Darstellung der einzelnen Supportkosten [OPL19]	24
11	Auflistung der Kriterien und ihre Umsetzung in dem DBMS	26
12	Datenkonstellationen mit den einzelnen Merkmalen	28
13	Zeitliche Verteilung des Import von unkomprimierten Daten	30
14	Zeiten der einzelnen Komprimierungsvorgänge	31
15	Zeiten eines gesamten Importvorgangs mit Endkomprimierung	32
16	Zeiten der Komprimierung	33
17	Speicherplatzverbrauch der einzelnen Pyramidenlevel in unterschiedlichen Komprimierungsarten	34
18	Anfragezeiten auf unterschiedlichen Bildgrößen und Kompressionsarten .	35
19	Vergleich der Anfragezeiten abhängig von der Kompressionsart	36
20	Vergleich der Kompressionen in Bezug auf die analysierten Kriterien .	38
21	Zeitliche Verteilung des Imports bei in-db Speicherung	39
22	Zeitliche Verteilung des Imports bei out-db Speicherung	40
23	Speicherplatzverbrauch abhängig von Datenmenge und Speicherort .	42
24	Prozentualer Speicherplatzverbrauch der einzelnen Overview Level .	42
25	Auflistung der Importzeiten in Oracle und PostGIS	45
26	Auflistung des Speicherplatzverbrauches in Oracle und PostGIS	46
27	Auflistung der Anfragezeiten in Oracle und PostGIS	47
28	Nutzwerttabelle der Kosten	50
29	Bewertung der Performanceanalyse	51
30	Nutzwerttabelle der Performance von Oracle GeoRaster	52
31	Nutzwerttabelle der Performance von PostGIS raster	52
32	Nutzwerttabelle der Abschlussbewertung von Oracle GeoRaster .	53
33	Nutzwerttabelle der Abschlussbewertung von PostGIS raster	53

Die Auswertung der Anfragen an die ***profil rdv*** haben folgende Werte ergeben:

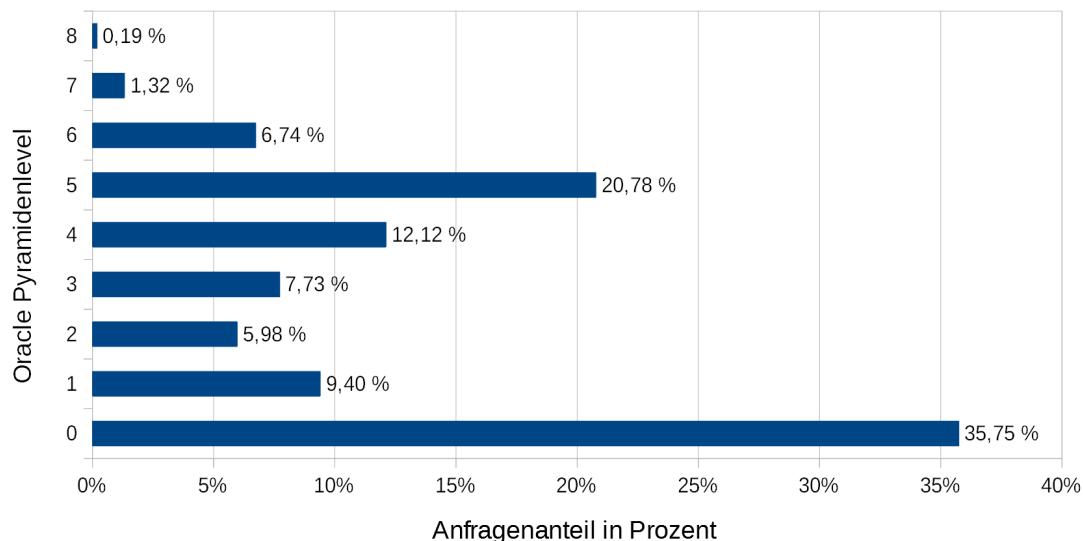


Abbildung 35: Anfragenverteilung der Pyramidenlevel der ***profil rdv***

Die maximal abgefragte Ausdehnung war in x-Richtung (Breite) 55539 Meter und in y-Richtung (Höhe) 33269 Meter.



Eidesstattliche Erklärung

Hiermit versichere ich, die vorliegende Bachelorarbeit ohne Hilfe Dritter und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe; die aus fremden Quellen direkt oder indirekt übernommenen Gedanken sind als solche kenntlich gemacht. Die Arbeit wurde bisher in gleicher oder ähnlicher Form in keiner anderen Prüfungsbehörde vorgelegt und auch noch nicht veröffentlicht.

Neubrandenburg, 21.03.2019

Unterschrift