

TECHNISCHE DOKUMENTATION

clinical staging Data Warehouse (csDWH)



Acronyms

AKTI	N Aktionsbündnis für Informations- und Kommunikationstechnologie in Intensiv- und Notfallmedizin
BfArN	Bundesinstitut für Arzneimittel und Medizinprodukte
DIZ	Datenintegrationszentrum6
csDW	H clinical staging Data Warehouse1
LTS	Long-term support7
DB	Datenbank6
ICD-1	0-GM International Statistical Classification of Diseases and Related Health Problems, 10. Revision, German Modification
IMBE	I Institut für Medizinische Biometrie; Epidemiologie und Informatik
IS-H	Industrial Solution - Healthcare
TDE	Transparent Data Encryption
ZIP	Zipper
MD5	Message-Digest Algorithm 5
\mathbf{ETL}	Extraction Tranformation Load
KIS	Krankenhausinformationssystem
GTDS	Gißener Tumordokumentationssystem
KHEn	tgG Krankenhausentgeltgesetz
OPS	Operationen- und Prozedurenschlüssel
PDMS	S Patientendatenmanagementsystem
\mathbf{SGB}	Sozialgesetzbuch
SC 6	Servicecenter 6 IT
GHH	Gutenberg Health Hub

Inhaltsverzeichnis

1	Dok	xumentationsinformationen	5
	1.1	Beteiligte Personen und Status	5
	1.2	Änderungshistorie	5
2	Ein	führung	6
3	Inst	allation und Konfiguration der Instanz des csDWH	7
	3.1	TDE-Installation	7
	3.2	TDE-Konfiguration	8
		3.2.1 TDE-Instanz-Dateien	8
		3.2.2 Konfigurationsdateien	9
4	Nut	zung	10
	4.1	Start die DB	10
	4.2	Arbeiten mit csDWH via ssh/psql	11
5	Stru	ıktur des csDWH	12
	5.1	Beschreibung	12
	5.2	Liste der vorhandenen Schemata	12
6	Bac	kup	14
	6.1	Konzept	14
	6.2	Technische Aspekte	14
7	Sch	emata	15
	7.1	p21	15
		7.1.1 Tabellen	15
		7.1.2 Views für Tools	16
	7.2	kis	16
		7.2.1 Tabellen	16
	7.3	Patientendatenmanagementsystem (PDMS)	17

8	Ben	utzer	22
		7.7.1 Tabellen	19
	7.7	metadata_repository	18
		7.6.1 Tabellen	18
	7.6	imagic	18
		7.5.1 Tabellen	18
	7.5	centrallab	18
		7.4.1 Tabellen	17
	7.4	gtds	17
		7.3.1 Tabellen	17

Tabellenverzeichnis

5.1	Schemata im csDWH	12
7.1	Tabellen im Schema p21	15
7.2	Views im Schema p21	16
7.3	Tabellen im Schema kis	16
7.4	Tabellen im Schema copra	17
7.5	Tabellen im Schema gtds	17
7.6	Tabellen im Schema centrallabor	18
7.7	Tabellen im Schema imagic	18
7.8	Tabellen in Schema metadata_repository	19
8.1	Benutzer im csDWH	23

Dokumentationsinformationen

1.1 Beteiligte Personen und Status

An Dokument beteiligte Person(en)	Status	Anmerkung zum Status
Abel Hodelin Hernandez, SC 6, GHH [Verfasser]	⊠ in Bearbeitung	
Sami Habib SC 6, GHH	☐ Aktiviert	
Daniel Schmitz, IMBEI	□ Deaktiviert	
	☐ Sonstig:	

1.2 Änderungshistorie

Wann?	Wer?	Was?
18. Februar 2022	Abel Hodelin Hernandez	Version 1.1

Einführung

Im Datenintegrationszentrum (DIZ) werden Daten aus verschiedenen Fachabteilungen und Systemen zusammengeführt. Ein zentrales Puzzleteil für die Zwischenspeicherung der Information dieser Systemen ist das clinical staging Data Warehouse (csDWH). In dieser Datenbank (DB) werden alle relevanten klinischen Systeme abgebildet. Diese Daten werden im Rahmen des Datenschutz sowie der Datenqualität aufbereitet und anschließend an weitere Komponenten des DIZ übertragen.

Installation und Konfiguration der Instanz des csDWH

Das csDWH, welches die Forschungsdaten beinhaltet, befindet sich in einem Ubuntu Server mit der Version Ubuntu 18.04 Long-term support (LTS). Diese DB wurde in PostgreSQL mit Hilfe von PostgreSQL Transparent Data Encryption (TDE) implementiert und verschlüsselt. Somit sind alle Datensätze der Datenbank verschlüsselt auf der Festplatte gespeichert und werden erst bei Zugriff entschlüsselt.

3.1 TDE-Installation

Die Installation von PostgreSQL TDE Version postgresql-12.3_TDE_1.0 folgte dem Installation Guide Software unter dem Link (https://www.cybertec-postgresql.com/de/transparent-data-encryption-installation-guide/). Davor wurden die notwendigen Pakete und Abhängigkeiten auf dem Ubuntu-Server via apt installiert:

Schritte um Paketen in Ubuntu zu installieren:

- sudo apt update # liest alle eingetragenen Paketquellen neu ein
- sudo apt list –upgradable # List neuer Pakete
- sudo apt upgrade # bringt die installierten Pakete auf den neuesten in den Paketquellen verfügbaren Stand
- sudo apt install pakete_name # installiert pakete_name.

Liste der notwendigen Paketen und Abhängigkeiten

zlib1g-dev libssl-dev libldap2-dev libperl-dev python-dev libreadline-dev libxml2-dev

libxslt1-dev bison flex uuid-dev make make

gcc libsystemd-dev

libxml2-utils xsltproc

Das Install-Kommando für den Server lautet:

sudo./configure-prefix=/usr/local/pg14tde-with-openssl-with-perl-with-python-with-ldap-with-libxml-with-uuid=e2fs-with-systemd

Wobei /user/local/pg14tde ist den Ordner, wo die Befehlen zur Steuerung der Server sich befinden.

Das Start-Kommando für die Instanz lautet:

/usr/local/pg14tde/bin/initdb -D /media/db/cdw_database/clinic_instance

3.2 TDE-Konfiguration

3.2.1 TDE-Instanz-Dateien

Auf dem Betriebssystem wurde der Benutzer clinicuser angelegt, dieser ist für die Administration der DB-Instanz vorgesehen und besitzt keine administrative rechte auf dem Betriebssystem.

Die Dateien der TDE-Instanz befinden sich auf dem Server unter /media/db/cdw_database.

- clinic_instance Instanz der csDWH mit DB- und Konfigurationsdateien.
- sh_scripts Shell-Skript mit Schlüssel-Manager Datei
 - clinic_instance_key.sh Skript fürs Schlüssel-Manager. Der Schlüssel der Instanz wird in einem Message-Digest Algorithm 5 (MD5)-Hash umgewandelt.
- dbBack Täglicher Backup der ganzen Instanz. Hier werden die sieben letzten Backups der DB in verschlüsselten Zipper (ZIP)-Dateien aufbewahrt. Die Namenskonvention für die Backup Dateien ist staging_YYYY-MM-TT.all.zip.

3.2.2 Konfigurationsdateien

Die Datei postgresql.conf wurde wie folgt modifiziert:

- port = 5433 #Proxy der Instanz
- \bullet listen_addresses = '*' # Maschinen auf denen die Instanz abrufbar ist
- ullet password_encryption = scram-sha-256 # Kennwort-Verschlüsselung Protokoll
- encryption_key_command = '/media/db/cdw_database/sh_scripts/clinic_instance_key.sh' # Datei mit dem Schlüssel der Instanz

In der Datei pg_hba.conf wurden die Zugangsprotokoll der Instanz definiert.

- ullet local all all scram-sha-256 # lokale Verbindungen
- ullet host all all 0.0.0.0/0 scram-sha-256 # externe Verbindungen

Nutzung

4.1 Start die DB

Die Instanz startet automatisch nach jedem Reboot des Server. Wenn die Instanz auf diese Weise nicht startet, sollte man folgendes machen / überprüfen:

- ssh IP des Server -1 cdw # Login auf dem Server via ssh mit einem Benutzer mit sudo Rechten, unter Windows auch mit den Tools putty oder MobaX-term
- Die Partition der Instanz sollte automatisch gemountet werden, da es in fstab konfiguriert ist. Falls die Partition nicht gemountet ist, sollte man die folgende Schritte durchführen:
 - cdw\$ lsblk # Überprüfen ob die Partition /dev/sdb1 gemountet ist.
 - cdw\$ sudo mount /dev/sda1 /media/db #Falls die Partition /dev/sdb1 nicht gemountet ist
- Die PostgreSQL-Instanz startet automatisch nach 100 Sekunden nach jedem Neu Start des Servers, da das Script zum Starten via cron-daemon abgerufen wird:
 - Befehl in crontab: @reboot sleep 100 && /media/db/cdw_database/startDB.sh

Falls die Instanz nicht automatisch startet sollte man diese Befehlen verfolgen.

- cdw\$ sudo su clinicuser #Benutzer ändern
- clinicuser\$ cd /media/db/database # Gehe zum Ordner der Instanz

 clinicuser\$/usr/local/pg14tde/bin/pg_ctl -D clinic_instance restart #(Re)Start die Instanz

4.2 Arbeiten mit csDWH via ssh/psql

- ssh Server_IP -1 server_user # Login auf dem Server
- server_user\$/usr/local/pg14tde/bin/psql -p 5433 database_name -U database_user_name #Verbindung mit einer Datenbank der Instanz
- $\bullet \ database_name \# \ c \ another_database_name Verbindung \ mit \ anderer \ DB \\$

In der Datenbank ist ein Benutzer clinicuser angelegt. Dieser ist der Administrator der Datenbank, aber nicht mit dem Benutzer clinicuser des Systems verwechselt.

Struktur des csDWH

5.1 Beschreibung

Das csDWH besitzt zwei strukturell gleiche DB, staging für die Produktion und staging_test zum testen. Die DB sind in verschiedenen Schemata geteilt, jede davon entspricht eine Quelle oder Zusammenfassung von Systemen. Die Information der Schemata liegt in Kapitel 7.

5.2 Liste der vorhandenen Schemata

Tabelle 5.1: Schemata im csDWH

Schema	Information
centrallab	Information aus dem Zentral Labor
copra	Information aus COPRA-System (PDMS)
gtds	Information aus dem Gißener Tumordoku-
	mentationssystem (GTDS)
icd_metadatainfo	International Statistical Classification of Diseases and Related Health Problems, 10. Revision, German Modification (ICD-10-GM)
kis	Information aus dem Krankenhausinformationssystem (KIS)
metadata_repository	Metadata
ops_metadatainfo	Operationen- und Prozedurenschlüssel (OPS)
p21	Information aus §21
aktin	Information des AKTIN-Projekts
diz_intern	Administrative Information
imagic	Information aus iMagic

Backup

6.1 Konzept

Ein Dump der kompletten csDWH-Instanz wird täglich um 01:00 gemacht. Das sind zwei Prozeduren, erst verläuft dumpall der csDWH-Instanz und direkt danach werden die Backup-Dateien in einer ZIP-Datei verschlüssel komprimiert. Diese Datei wird auf dem Server und auf einer extra-VM gespeichert.

6.2 Technische Aspekte

Ein Shell-Script garantiert die Speicherung und Verschlüsselung der csDWH-Instanz sowie die lokale und ferne Speicherung. Dieses Skript wird jeden Tag um 01:00 via cron-daemon abgerufen.

• Shell-Script: backDB.sh

• Befehl in crontab: 0 1 * * * /media/db/cdw_database/backDB.sh

• Backup-Ordner: /media/db/cdw_database/dbBack

• Backup-Name-Format: staging_YYYY-MM-DD.all.zip

Schemata

Die Schemata speichern die "rohe" **pseudonymisierte** Information der ursprünglichen Systems oder die Metadaten. Diese Daten werden in Views analysiert oder weiter verarbeitet für andere Anwendungen oder Projekten. Wichtige Hinweis ist, dass die Daten in dem Data Warehouse unverändert bleiben sollen.

Die Dokumentation der Views für Datenqualität liegt in einem anderem Dokument.

7.1 p21

Dieses Schema speichert die jährliche Information der §21, die von Medizincontrolling in CSV-Dateien generiert wird.

Der jährliche Rhythmus ist zu groß, als dass die Daten bspw. zur Rekrutierung von Patienten für Studien aber auch zu Forschung genutzt werden können. Auf diesem Grund wird diese Information in der Zukunft nicht mehr aus CSV-Dateien genommen sondern direkt aus dem KIS.

7.1.1 Tabellen

Tabelle 7.1: Tabellen im Schema p21

Tabelle	Beschreibung
p21_encounter	Information der Datei FALL.csv: Fälle
p21_department	Inhalt der Datei FAB.csv: Fachabteilung
p21_operation	Information der Datei OPS.csv: Operationen
p21_diagnosis	Basiert auf der Datei ICD.csv: Diagnosen (ICD-10-GM)

7.1.2 Views für Tools

In diesem Schema befinden sich auch die Views für Extraction Tranformation Load (ETL)-Prozessen die, solche Information aus §21 benötigen. Der Inhalt dieser Views entspricht die Formatierung der Krankenhausentgeltgesetz (KHEntgG) §21 Übermittlung und Nutzung der Daten.

Tabelle 7.2: Views im Schema p21

View	Beschreibung
fall	Falldaten
fab	Fachabteilungsangaben
icd	Diagnosenangaben
ops	Prozedurenangaben

7.2 kis

Hier werden die tagesaktuellen extrahierten Daten zu Patienten, Fällen, Bewegungen, Diagnosen und Prozeduren direkt aus dem Quellsystem KIS gespeichert. Mit Hilfe diesem Schema lassen sich viele der Abbildungen für weitere Projekte realisieren.

7.2.1 Tabellen

In diesem Schema behalten die Tabellen denselben Namen wie in KIS.

Tabelle 7.3: Tabellen im Schema kis

View	Beschreibung
nbew	Bewegung
ndia	Diagnosen
nfal	Fälle
nicp	Prozeduren
npat	Patienten
norg	Organisationseinheiten

7.3 **PDMS**

Hier wird die tagesaktuelle Information aus dem COPRA-System gespeichert. Dieses Schema beinhaltet Befunde, ärztliche Anweisungen und Überblick über Behandlungsschritte.

7.3.1 Tabellen

Tabelle

co6_config_variables

co6_config_variable_types

In diesem Schema behalten die Tabellen denselben Namen wie im COPRA-System.

Beschreibung co6_data_decimal_6_3 Metadaten der nummerischen Messungen Metadaten der Messungen von Typ Objekt co6_data_object co6_medic_data_patient Demografische Information der Patienten co6_medic_pressure Daten der Herz-Messungen

Variables in copra

Art der Variables in copra

Tabelle 7.4: Tabellen im Schema copra

7.4 gtds

Dieses Schema speichert die Daten der mainzenen Instanz des GTDS und somit die Erfassung und Verarbeitung der Daten der revidierten Basisdokumentation klinischen Krebsregistern.

7.4.1Tabellen

Dieses Schema hat momentan nur eine Tabelle. Die ist auf eine View auf eine Auswertung auf die Daten des GTDS basiert.

Tabelle 7.5: Tabellen im Schema gtds

Tabelle	Beschreibung
auswertung_diz	Auswertung auf Daten auf GTDS

7.5 centrallab

Hier werden die Daten aus dem Zentrallabor (Institut für Klinische Chemie und Laboratoriumsmedizin) gespeichert.

7.5.1 Tabellen

Die Tabellen speichern die Messungen sowie Mapping zu LOINC-Code.

Tabelle 7.6: Tabellen im Schema centrallabor

Tabelle	Beschreibung
observation	Laborwerte der Patienten
observationreport	Verlinkung der Laborwerten mit Fälle und Patienten
loinc_mapping_central_lab	Mapping der LOINC-Code zu der Messungen und/Geräte

7.6 imagic

Hier wird die Information aus dem IMAGIC-System gespeichert. Dieses Schema beinhaltet Information aus der Hautklinik. Davon Metadaten der Bilder sowie Befunde anhand der Bilder.

7.6.1 Tabellen

In diesem Schema behalten die Tabellen denselben Namen wie im IMAGIC-System.

Tabelle 7.7: Tabellen im Schema imagic

Tabelle	Beschreibung
image	Metadaten der Bilder
patient	Patienten Informationen
study	Information der Studien an der Hautklinik
visit	Besuch/Fall-Information

7.7 metadata_repository

Dieses Schema speichert die Information der Metadaten aller anderen Schemata.

7.7.1 Tabellen

Tabelle 7.8: Tabellen in Schema metadata_repository

Table	Beschreibung	Schemata/Quellen	
abschluss_grund	Auswahlliste - Abschluss Grund	GTDS	
ann_arbor_allgemein	Ann-Arbor Allgemeinsymptomatik	GTDS	
ann_arbor_extra	Ann-Arbor extralymphatischer Befall	GTDS	
ann_arbor_stadium	Auswahlliste - Ann-Arbor Stadium	GTDS	
applikationsart	Applikationsart der (Teil-)Bestrahlung	GTDS	
applikationstechnik	Applikationstechnik der (Teil-	GTDS	
)Bestrahlung		
arzt_anlass	Auswahlliste - Arzt Anlass	GTDS	
autopsie	Autopsie	GTDS	
behandlungsanlass	Auswahlliste - Tumor Ausprägung	GTDS	
behandlungskategorie	Behandlungskategorie	KIS	
bewegunsart	Bewegunsart	KIS	
bewegungstyp	Bewegungstyp	KIS	
body_localisation	Lokalisation im Körper	§21 KHEntgG	
complication_level	Komplication Ebene	KIS	
copra_biosignal	Biosignalen	PDMS	
copra_variables	Konfiguration Variablen	PDMS	
country	Länder	KIS	
todesursache	Todesursache	KIS	
fachabteilung	Fachabteilung	KIS, §301 Abs. 3 SGB	
		V, §21 KHEntgG	
fachabteilung_prefix	Präfix der Fachabteilung	§301 Abs. 3 SGB V,	
		§21 KHEntgG	
diagnosenzusatz	Diagnosenzusatz	KIS	
diag_gewissheit	Diagnostische Gewissheit	KIS	
diagnosis_type	Art der Diagnose	KIS, §21 KHEntgG	
diagnosis_type_icd	Art der ICD-10-GM-Diagnose	KIS, BfArM	
diagsich_hoechste	Höchste Diagnosesicherung	GTDS	
drg_prozedur	Kategorie einer DRG-Prozedur	KIS	
einrichtung	Einrichtung	KIS	
einweisungs_ueberweisungs	Einweisungs-, Überweisungs,- Nachbe-	KIS	
_nachbehandlungsart	handlungsart		
einwilligung	Einwilligung	GTDS	
erfassungsanlass	Erfassungsanlass	GTDS	

fachrichtung	Fachrichtung	KIS	
fallendes	Art des Fallendes	KIS	
fallstatus	Fallstatus	KIS	
falltyp	Falltyp	KIS	
gender	Geschlecht	KIS, GTDS, PDMS, §21	
		KHEntgG	
gtds_datenart	GTDS-Datenart	GTDS	
herkunft	Herkunft	GTDS	
histo_diagnose	Genkennzeichnung der Auswertungsre-	GTDS	
	levanz der Histologie		
histo_grading	Histologisches Grad	GTDS	
histo_haupt_neben	Haupt- oder Nebenhistologie	GTDS	
inter_status_amb_besu	Interner Status eines ambulanten Be-	KIS	
	suchs		
lok_haupt_neben	Kennzeichnung ob Haupt- oder Neben-	GTDS	
	lokalisation		
lok_rezidivart	Art des Rezidives	GTDS	
lok_seite	Seitenangabe zur Lokalisation	GTDS	
lokalisation	Lokalisationsschlüssel	GTDS	
op_intention	Intention der Operation	GTDS	
operation_katalog	Identifikation eines Operationsleis-	KIS	
	tungskataloges		
orgtype	Typ der Organisationseinheit	KIS	
p21_admission_cause	Aufnahmeanlass	§301 Abs. 3 SGB V,	
		§21 KHEntgG	
p21_admission_reason	Aufnahmegrund	§301 Abs. 3 SGB V,	
		§21 KHEntgG	
p21_admission_reason_1_2	Aufnahmegrund 1. und 2. Stellen	§301 Abs. 3 SGB V,	
		§21 KHEntgG	
p21_admission_reason_3	Aufnahmegrund 3. Stelle	§301 Abs. 3 SGB V,	
		§21 KHEntgG	
p21_admission_reason_4	Aufnahmegrund 4. Stelle	§301 Abs. 3 SGB V,	
		§21 KHEntgG	
p21_department	Fachabteilung	§301 Abs. 3 SGB V,	
		§21 KHEntgG	
p21_department_code_mean	Bedeutung der Ziffer im Code der Fach-	§301 Abs. 3 SGB V,	
	abteilungen	§21 KHEntgG	

p21_merging_reason	Fallzusammenführungsgrund	§301 Abs. 3 SGB V,	
		§21 KHEntgG	
p21_remission_reason	Entlassungs-/Verlegungsgrund	§301 Abs. 3 SGB V,	
		§21 KHEntgG	
p21_remission_reason_1_2	Entlassungs-/Verlegungsgrund 1. und	§301 Abs. 3 SGB V,	
	2. Stellen	§21 KHEntgG	
p21_remission_reason_3	Entlassungs-/Verlegungsgrund 3. Stelle	§301 Abs. 3 SGB V,	
		§21 KHEntgG	
p21_remuneration_area	Entgeltbereich	§301 Abs. 3 SGB V,	
		§21 KHEntgG	
protokoll_typ	Protokoll-Typ	GTDS	
prozedur_typ	Prozedurtyp	KIS	
pupille_reaction	Reaktion der Pupille	PDMS, AKTIN-Projekt	
pupille_width	Erweiterung der Pupille	PDMS, AKTIN-Projekt	
r_klassifikation	R Klassifikation	GTDS	
rezidiv	Rezidiv	GTDS	
sources	Information der System- Datenquellen	System- Datenquellen	
status_color	Status Farbe	PDMS, AKTIN-Projekt	
status_datenart	GTDS-Datenart zu einer Status	GTDS	
$sterbedatum_exakt$	sterbedatum_exakt		
stornierungsgrund	Stornierungsgrund	KIS	
strahlenart	Strahlenart	GTDS	
tn24	Tabelle TN24 in KIS (IS-H: Behand-	KIS	
	lungskategorien)		
transportart	Transportart	KIS	
tumorauspraegung	Tumorausprägung	GTDS	
tumortod	Tumortod	GTDS	
unfallart	Unfallart	KIS	
units	Einheiten	KIS	
zipcode	Postleihzahl	KIS	

Benutzer

Der Benutzer clinicuser ist der Administrator der Server. Jede Schema hat mindesten ein Benutzer, einer davon darf lesen und Operationen der Insertion, Modifikation, und Löschen/Truncate durchführen. Momentan hat das System ein Benutzer (onlyreader), der nur Lesen-Zugriff in allen Schemata hat.

Tabelle 8.1: Benutzer im csDWH

Benutzer	Schema	Berechtigungen	Anwendung
centrallabcdwuser	centrallab	select, insert, update, delete, truncate	-
clinicuser	all	administrator	-
grafana	all in staging	select	ja
kisvendor	kis	select, insert, update, delete, truncate	-
p21user	p21	select, insert, update, delete, truncate	-
gtdscdwuser	gtdscdwuser	select, insert, update, delete, truncate	-
onlyreader	all in staging_test	select	-
copracdwuser	copra	select, insert, update, delete, truncate	-
imagiccdwuser	imagic	select, insert, update, delete, truncate	-
emergencydepartmentuser	aktin	select, insert, update, delete, truncate	-