



**PURPLE
STACK**



IT-P

Personenbezogene Datenanonymisierung

Theorie und Praxis

Bharat Ahuja



Wir sehen unseren Erfolg im Erfolg unserer Kunden!

Gemeinsam als Partner die Zukunft gestalten mit IT-Dienstleistungen, die begeistern.



› **Unsere Gründung:** 1. September 1996 › **Unser Umsatz:** Mehr als 30 Mio. € in 2017

› **Unsere Portfolioübersicht:**



Application Engineering



Cloud Solutions



Service Management



Enterprise Solutions



Project Management



Sie finden uns auf Stand 222

IT-P – Das macht uns zum richtigen Partner für unsere Kunden, Kollegen und Partnerunternehmen



- › IT-P – zwei Familien – zwei Generationen: Wir verbinden Erfahrung mit Innovation.



- › Gemeinsam die Zukunft gestalten mit Flexibilität und Persönlichkeit.



- › IT-P – vielseitige Expertise unter einem Dach!



- › IT(-P) Software & Services made in Germany – begeisterte Kunden & höchste Qualität!

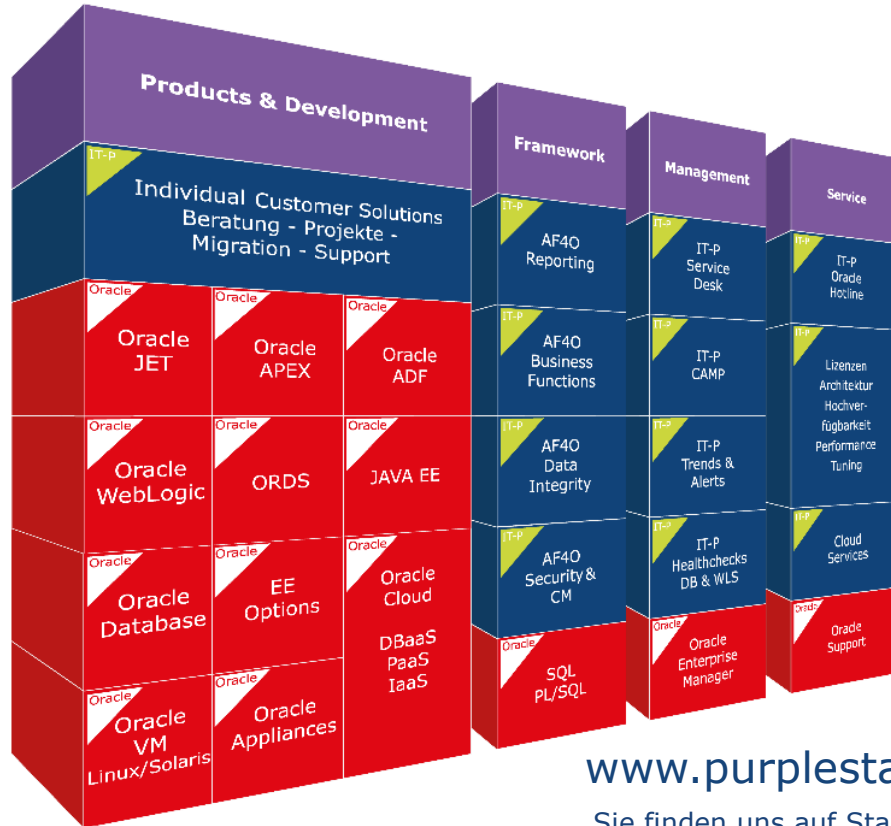


- › Unsere Kunden stehen im Mittelpunkt, denn ihr Erfolg ist unser Erfolg.



- › IT-P – regionale Kompetenz deutschlandweit!

Oracle Technology meets IT-P Services



ORACLE Platinum Partner

Specialized
Oracle Database 11g
Performance Tuning

ORACLE Platinum Partner

Specialized
Oracle Database 12c

ORACLE Platinum Partner

Specialized
Oracle Linux 6

ORACLE Platinum Partner

Specialized
Java Platform Standard
Edition 8

www.purplestack.de

Sie finden uns auf Stand 222

Wissen = Macht

Der gefährliche Umgang mit Personendaten



Brüssel, wir haben ein Problem!

The background is a blurred photograph of a modern office hallway. In the foreground, a person's hand is visible on the left, holding a tablet. In the background, three people are walking away from the camera down a brightly lit hallway with a white floor and walls. The overall tone is professional and forward-looking.

Es gibt kein Zurück mehr.

Datenschutzausschuss zu Hilfe!

Datentransformationen, um Identitäten zu schützen

Pseudonymisierung und Anonymisierung

„Pseudonymisierung“
d.h. eine falsche Namensgebung

Name	Last_Login	Last_PW_Change	Wohnort
Alice K.	2018-09-22 10:10	2018-09-22 10:10	Hannover
Bob M.	2018-01-19 01:20	2016-08-11 23:12	Berlin
Steve C.	2016-09-22 10:10	(null)	Oldenburg

UID	Last_Login	Last_PW_Change	Wohnort
TSUID01	2018-09-22 10:10	2018-09-22 10:10	Hannover
TSUID02	2018-01-19 01:20	2016-08-11 23:12	Berlin
TSUID03	2016-09-22 10:10	(null)	Oldenburg

Name	UID
Alice K.	TSUID01
Bob M.	TSUID02
Steve C.	TSUID03

Verknüpfungstabelle

Name	Last_Login	Last_PW_Change	Wohnort
Alice K.	2018-09-22 10:10	2018-09-22 10:10	Hannover
Bob M.	2018-01-19 01:20	2016-08-11 23:12	Berlin
Steve C.	2016-09-22 10:10	(null)	Oldenburg

Name	Last_Login	Last_PW_Change	Wohnort
d932285...	2018-09-22 10:10	2018-09-22 10:10	Hannover
4278c5b...	2018-01-19 01:20	2016-08-11 23:12	Berlin
854b3f7...	2016-09-22 10:10	(null)	Oldenburg

„Pseudonymisierung“
d.h. eine falsche Namensgebung

Okay, ich halte meine Aliase geheim... Versprochen!

Oder besser, ich lösche die Verknüpfungstabelle.

Dann ist es nicht umkehrbar.

DSGVO: Spielt keine Rolle!

Die Daten identifizieren Individuen eindeutig.

Name
Maik D.
Tanja E.
Maria D.
Bob S.
...

UID
FJKLD879
OIUE890I
890JKJO8
890809FL
...

UID	SteuerID	...
FJKLD879	78979789	...
OIUE890I	80009087	...
890JKJO8	78594030	...
890809FL	78340084	...
...

UID
FJKLD879
OIUE890I
890JKJO8
890809FL
...


UID	PLZ	...
FJKLD879	30159	...
OIUE890I	30163	...
890JKJO8	30163	...
890809FL	30159	...
...

UID	GEB_DATUM	...
FJKLD879	01-01-1980	...
OIUE890I	05-02-1979	...
890JKJO8	01-01-1980	...
890809FL	28-02-1969	...
...

UID	PLZ	GEB_DATUM
FJKLD879	30159	01-01-1980
OIUE890I	30163	05-02-1979
890JKJO8	30163	01-01-1980
890809FL	30159	28-02-1969
...

Okay, ich halte meine Aliase geheim... Versprochen!
Oder besser, ich lösche die Verknüpfungstabelle.
Dann ist es nicht umkehrbar.

DSGVO: Spielt keine Rolle!



„Anonymisierung“
an-onym: namenslos

Wie kann ich aber meine Daten verarbeiten, so dass diese offiziell nach DSGVO anonym™ sind?

“Wahrscheinlich genutzte” Mittel?

Techniken:

- Generalisierung
 - k-Anonymity, l-Diversity, t-Closeness
- Randomisierung
 - Noise addition, Scrambling, usw.
 - Differential Privacy

Heuristiken:

- Singling Out
- Verlinkung
- Inferenz

Generalisierung:

München -> Bayern -> Deutschland

30159 -> 3015* ->... -> 3***** -> *****

2. Feb 1980 -> Feb 1980 -> Q11980 ... -> (1970,1990)

k-Anonymity

Jede Äquivalenzklasse $> k$ Datensätze

PLZ	GEB_DATUM	DIAGNOSE	...
30161	01.01.1980	Herzkrebs	...
30173	12.12.1980	Lungenkrebs	...
30185	03.04.1980	Hautkrebs	...
30332	16.05.1972	Schwangerschaft	...
30364	12.12.1972	Brustkrebs	...
...

PLZ	GEB_DATUM	DIAGNOSE	...
301**	1980	Herzkrebs	...
301**	1980	Lungenkrebs	...
301**	1980	Hautkrebs	...
303**	1972	Schwangerschaft	...
303**	1972	Brustkrebs	...
...

PLZ	GEB_DATUM	DIAGNOSE	...
301**	1980	Herzkrebs	...
301**	1980	Lungenkrebs	...
301**	1980	Hautkrebs	...
303**	1972	Schwangerschaft	...
303**	1972	Brustkrebs	...
...

Risikobewertung anhand DBSTAT Oracle Tool, um Identifier und Quasiidentifier zu erkennen.

```
[FULL_NAME]
```

```
COL_NAME_PATTERN = ^(PERSON|FULL).*NAME$
```

```
COL_COMMENT_PATTERN = (Full|Person).*Name
```

```
SENSITIVE_CATEGORY = PII
```

Risikobewertung anhand GROUP-BY Anfragen

Arbeitsblatt	Query Builder
1	<code>select count(*), plz, geburtsdatum</code>
2	<code>from personendaten</code>
3	<code>group by plz, geburtsdatum</code>
4	<code>order by 1 asc</code>

Risikobewertung anhand Histogrammen

Analyse der NDVs

- › Angenommen „PLZ, GEB_DAT“ -Werte sind nicht verzerrt, dann ist NDV hoch
- › D.h. jede einzelne „PLZ, GEB_DAT“ -Kombination deckt wenig ab. (Schubfachprinzip)
- › NDV über Quasiidentifizier hoch = Hohes Risiko

TopNFrequency Histogramm erstellen

Prüfen, ob das Histogramm besonders „schlecht“ ist.

Alternativ: Hybrid Histogramm erstellen und
„endpoint_repeat_count“ prüfen.

Datenverlust nicht unbedingt schlecht

Overfitting vermeiden

Analysen werden schneller (bessere Histogramme)

PLZ	GEB_DATUM	DIAGNOSE	...
30161	01.01.1980	Herzkrebs	...
30173	12.12.1980	Lungenkrebs	...
30185	03.04.1980	Hautkrebs	...
30332	16.05.1972	Schwangerschaft	...
30364	12.12.1972	Brustkrebs	...
...

if „PLZ“ ungerade: dann Krebs

if „PLZ“ like „301??“: dann Krebs

The background is a blurred image of a laptop keyboard in the foreground and several people in business attire walking away in the background, suggesting a professional or office environment.

Die ideale Lösung*

Sie finden uns auf Stand 222



Bharat Ahuja

IT-Consultant

E: b.ahuja@it-p.de

I: www.purplestack.de



IT-P Information Technology-Partner GmbH
Ein Unternehmen der IT-P Gruppe

Besuchereingang:
Seligmannallee 4
30173 Hannover
F: +49 511 61 68 04-17
I: www.it-p.de

Vielen Dank für Ihre Zeit

Wir freuen uns auf Ihren Besuch an unserem Stand 222