

The logo for SHS VIVEON AG, featuring the company name in a bold, white, sans-serif font against a green background of a leaf with water droplets.

SHS VIVEON AG

The Customer Management Company.

ETL mit PL/SQL Erfahrungen aus der Praxis

Nürnberg, 16.11.2011

Alexander Mendle

Professional Consultant

Alexander.Mendle@SHS-VIVEON.com

SHS VIVEON im Überblick

Facts & Figures

- Gründung: 1991, seit Mai 1999 börsennotiert, seit 12/2009 im m:access
- Langjährige Erfahrung im Customer Management:
 - Business Intelligence
 - Geschäfts- und Kundenprozesse
- Ganzheitlicher Beratungspartner Business Consulting – Implementierung
- Eigenes Produkthaus GUARDEAN
- 8 Standorte (Hamburg, Kastl, Köln, München, Stuttgart, Wien, Wiesbaden, Zürich)
- 25 Mio. Umsatz
> 200 Mitarbeiter
- Über 150 Kunden weltweit - davon 30 % aus der Liste der Fortune 500 International

Ausgewählte Kunden

Telekommunikation



Finanzdienstleistungen



Industrie & Handel



Die SHS VIVEON AG gehört als Beratungshaus zu den führenden Business- und IT-Lösungsanbietern im Customer Management

Inhalt

- 01** Abgrenzung des Themas
- 02** Tool & PL/SQL – die Unterschiede
- 03** ETL mit PL/SQL
 - Metadaten & Bibliothek**
 - Prozesse**
 - Hilfsmittel**
 - Organisation**
- 04** Zusammenfassung

Bildnachweise

01 ABGRENZUNG DES THEMAS

02 TOOL & PL/SQL – DIE UNTERSCHIEDE

03 ETL MIT PL/SQL

04 ZUSAMMENFASSUNG



ETL mit PL/SQL – Abgrenzung des Themas



01 ABGRENZUNG DES THEMAS

02 TOOL & PL/SQL – DIE UNTERSCHIEDE

03 ETL MIT PL/SQL

04 ZUSAMMENFASSUNG

Spezifische Eigenheiten bei...

...ETL mit Tool

- 🍃 Metadatenverwaltung
- 🍃 Releasemanagement
- 🍃 Systematisches Logging
- 🍃 Integrierte Ablaufplanung
- 🍃 Grafische Entwicklung
- 🍃 Lizenzkosten

...ETL mit PL/SQL

- 🍃 Individualisierbarkeit
- 🍃 Aktualisierung, Migrierbarkeit
- 🍃 Verfügbarkeit von Ressourcen
- 🍃 Performance
- 🍃 Flexibilität
- 🍃 Dynamisierung mit Code-Generatoren

Spezifische Eigenheiten bei...

...ETL mit Tool

- Metadatenverwaltung
- Releasemanagement
- Systematisches Logging
- Integrierte Ablaufplanung
- Grafische Entwicklung
- Lizenzkosten

- Informationen zu Ladeläufen
 - Mappings, Datenflüsse, Tabellen- und Spaltendefinitionen
 - Abhängigkeiten
 - Logging
 - Informationen zur Datenqualität
 - Repository
- Metadatenverwaltung ist Basis für einige wichtige Funktionalitäten!

Spezifische Eigenheiten bei...

...ETL mit Tool

- Metadatenverwaltung
- Releasemanagement
- Systematisches Logging
- Integrierte Ablaufplanung
- Grafische Entwicklung
- Lizenzkosten

- Releasemanagement und Repository
- Logging
- Ablaufsteuerung
- Datenqualität
- Produktivität

Comes-with-function

Spezifische Eigenheiten bei...

...ETL mit Tool

- Metadatenverwaltung
- Releasemanagement
- Systematisches Logging
- Integrierte Ablaufplanung
- Grafische Entwicklung
- Lizenzkosten

- Erweiterung auf Datenbankseite (Packages / Procedures) oder
- Programmierung von Erweiterungen aufwändig

...ETL mit PL/SQL

- Individualisierbarkeit
- Aktualisierung, Migrierbarkeit
- Verfügbarkeit von Ressourcen
- Performance
- Flexibilität

- Dynamisierung mit Code-Generatoren

- Beliebige Funktion umsetzbar
- Typische „Wildwuchsgefahren“
- Einsatz von Code-Generatoren!

Spezifische Eigenheiten bei...

...ETL mit Tool

- Erweiterung auf Datenbankseite (Packages / Procedures) oder
- Programmierung von Erweiterungen aufwändig

...ETL mit PL/SQL

- Individualisierbarkeit
- Aktualisierung, Migrierbarkeit
- Verfügbarkeit von Ressourcen
- Performance
- Flexibilität
- Dynamisierung mit Code-Generatoren

- Beliebige Funktion umsetzbar
- Typische „Wildwuchsgefahren“
- Einsatz von Code-Generatoren!

Spezifische Eigenheiten bei...

- Viele Ressourcen am Markt verfügbar
- Die Frage ist lediglich, wie viel spezielles DWH Know-How verfügbar ist.
- Keine Einarbeitung in neue Technologie notwendig

...ETL mit PL/SQL

- Individualisierbarkeit
- Aktualisierung, Migrierbarkeit
- Verfügbarkeit von Ressourcen
- Performance
- Flexibilität
- Dynamisierung mit Code-Generatoren

Weitere Eigenschaften...

...ETL mit Tool

- 🍃 Grafische Entwicklung
- 🍃 Lizenzkosten

...ETL mit PL/SQL

- 🍃 Aktualisierung, Migrierbarkeit
- 🍃 Performance
- 🍃 Flexibilität

01 ABGRENZUNG DES THEMAS

02 TOOL & PL/SQL – DIE UNTERSCHIEDE

03 ETL MIT PL/SQL

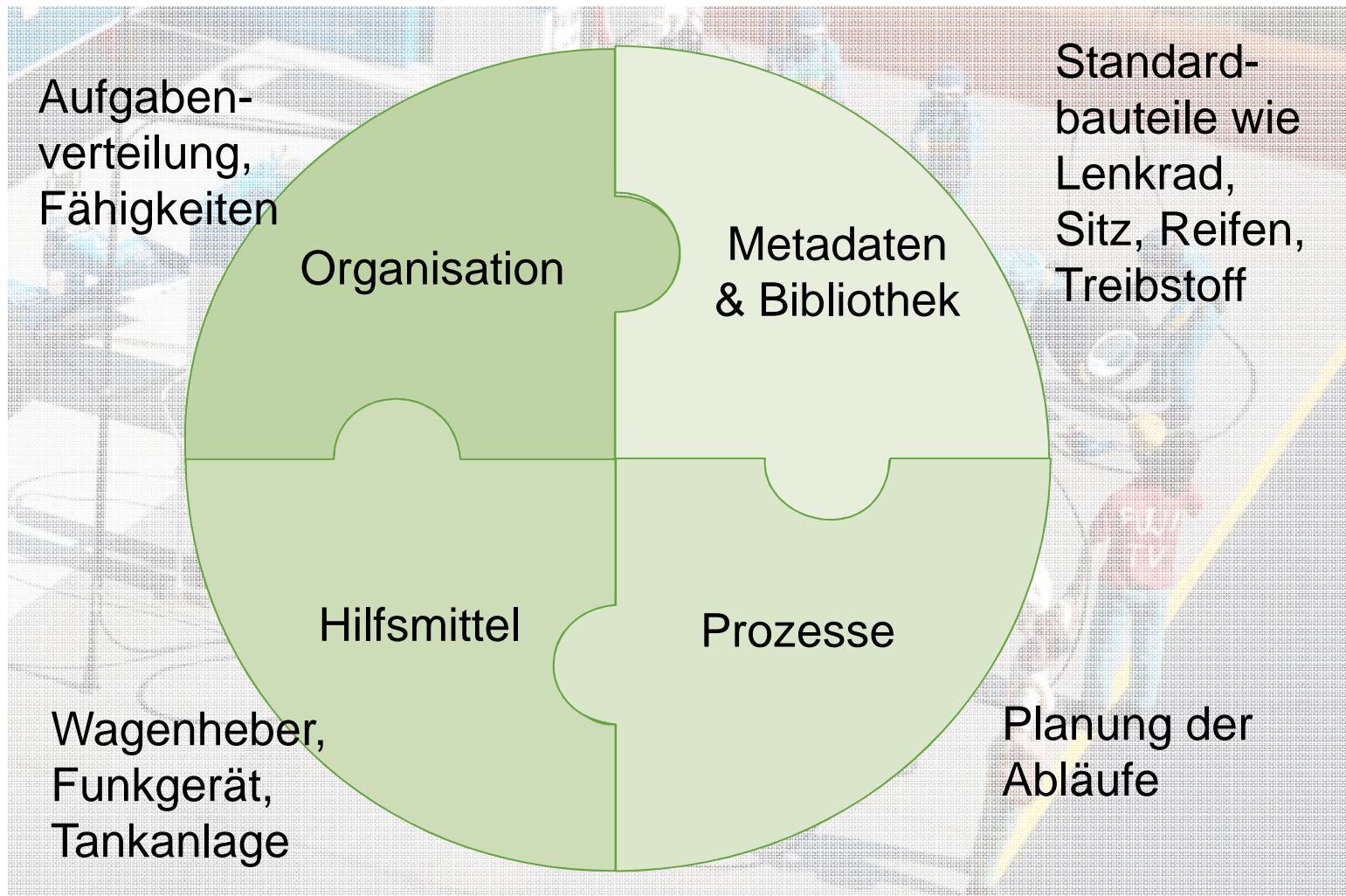
04 ZUSAMMENFASSUNG

03 ETL mit PL/SQL ETLs mit PL/SQL professionell umsetzen



03 ETL mit PL/SQL

ETLs mit PL/SQL professionell umsetzen



03 ETL mit PL/SQL

1 Metadaten & Bibliothek

Metadaten gibt es, wenn ETL-spezifische Funktionen und Informationen in bekannter, gemeinsamer und erweiterbarer Struktur bereitgestellt werden!



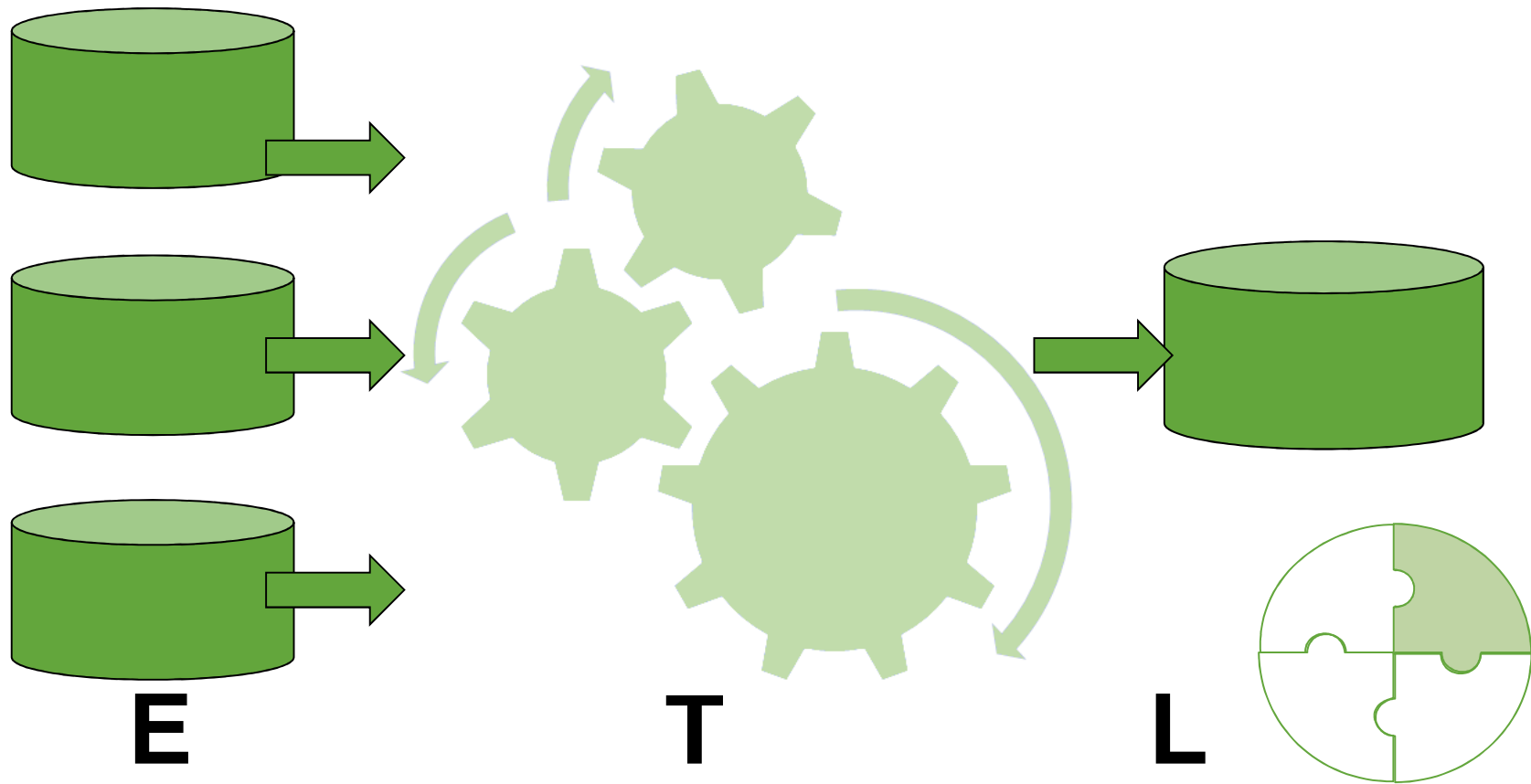
Welche Funktionalitäten werden für ETL benötigt?

Welche (Meta-) Daten werden benötigt?



03 ETL mit PL/SQL

1 Metadaten & Bibliothek

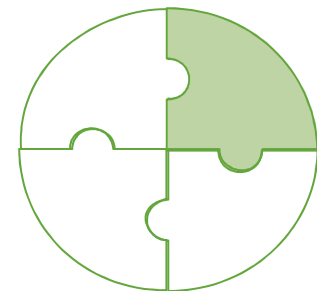


03 ETL mit PL/SQL – Gemeinsame Bibliothek

1 Metadaten & Bibliothek

Gemeinsame Bibliothek für Standardaufgaben – Beispiele

- 🍃 Logging
- 🍃 Fehlerbehandlung (Datenfehler & Ablauffehler)
- 🍃 Ladelaufkontrolle (Initialisierung, Beendigung etc.) / Wiederanlauf
- 🍃 Fertige Bausteine für
 - Datumsfunktionen (Feiertage, Betriebskalender,...)
 - Erstellung und Verwaltung von Partitionen, Indizes, ...
 - Ausführung von dynamischem SQL
 - Pivotierung
 - Lookups für Konstanten, unternehmensweite Informationen
 - u. a. nach Bedarf



03 ETL mit PL/SQL – Gemeinsame Bibliothek

1 Metadaten & Bibliothek

Bibliothek muss leben

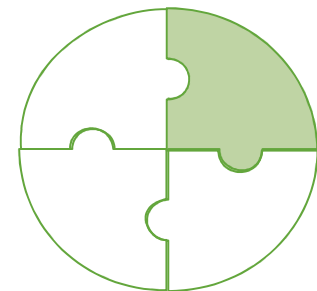
Metadaten gibt es nur bei Nutzung der Bibliothek.

Durchdachte Programmierhilfen und **zentrale** Datenspeicher!

Anreicherung von Metadaten nur bei zentralem Datenmodell.

Drei Paradigmen:

- 1.) Metadaten **zentral modellieren**.
- 2.) Klasse statt Masse
- 3.) Kontinuierliche **Weiterentwicklung**



03 ETL mit PL/SQL

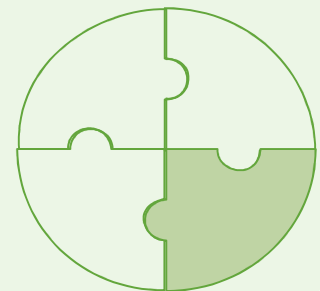
2 Prozesse

Entwicklungsprozesse

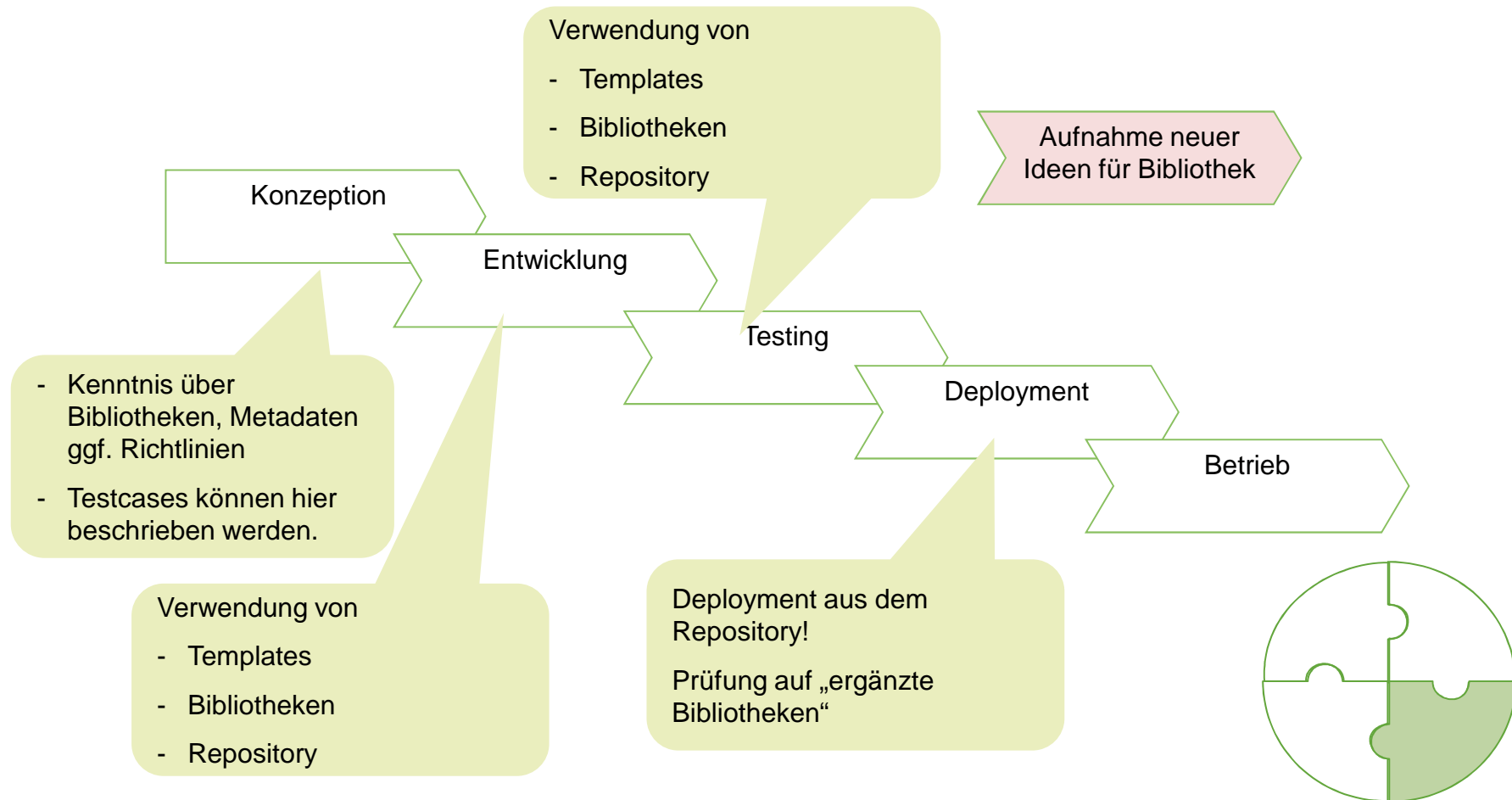
- Welche Schritte müssen beachtet werden?
- Wie wird die Bibliothek weiterentwickelt?

Repository, Dokumentation, Ablage

- Strukturierungshilfen für den Entwicklungsprozess
- Versionsverwaltung

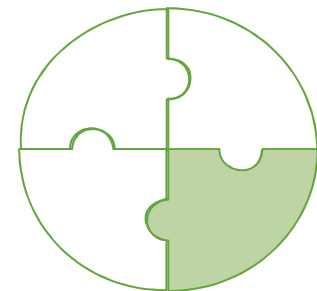
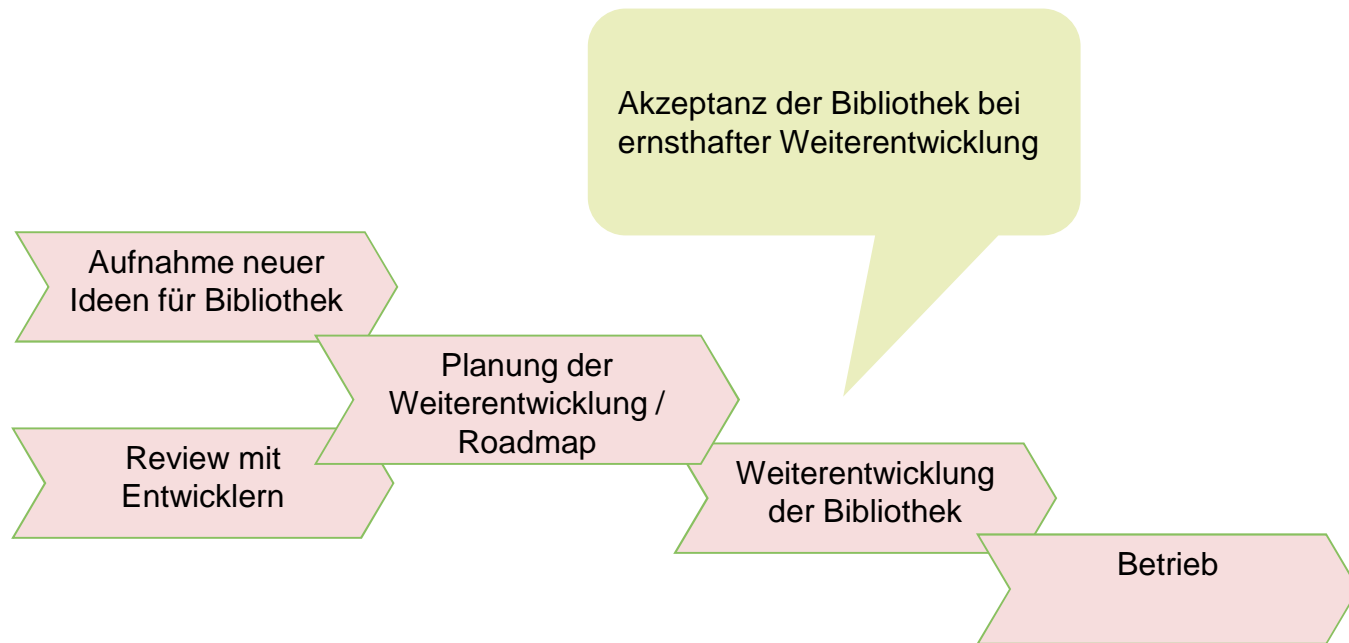


03 ETL mit PL/SQL – Prozesse 2 Prozesse



03 ETL mit PL/SQL – Prozesse

2 Prozesse

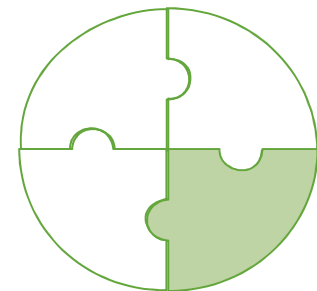


03 ETL mit PL/SQL – Prozesse

2 Prozesse

Berücksichtigung der Spezifika – insbesondere Metadaten und Bibliothek – in
Entwicklung und Deployment

- 🍃 Versionsverwaltung ist kritisch
- 🍃 Bereitstellung von Templates, insbesondere für Code-Templates für die Verwendung der Bibliothek (Aspekt „Bequemlichkeit“)
- 🍃 Releasewechsel auch für die **Bibliothek** und die **Metadaten** vorsehen.
- 🍃 Wartung der Bibliothek in einen **regulären Prozess** einbinden



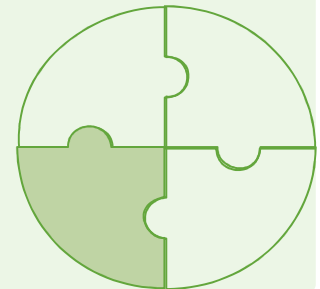
03 ETL mit PL/SQL

3 Werkzeuge & Hilfsmittel

Für die Unterstützung von

- 🍃 Dokumentation (hyperSQL)
- 🍃 Versionsverwaltung (SVN)
- 🍃 Testautomatisierung (FitNesse)
- 🍃 Abnahmen (Checklisten)
- 🍃 Deployment
- 🍃 Betrieb (Reporting von Ladeläufen und Qualität)
- 🍃 Issue-Tracking (JIRA)

Unterstützt das Werkzeug meinen Prozess?



03 ETL mit PL/SQL

3 Werkzeuge & Hilfsmittel

Automatisierung von Code/Package-Dokumentation

z. B. mit HyperSQL

schema_stats

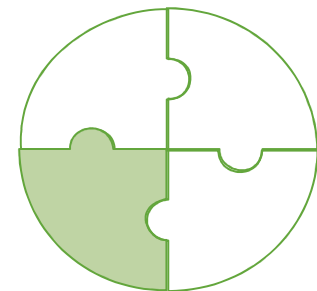
Generate statistics for all schemata we transfered data into

Syntax:
schema_stats (s_percent, s_degree)

Parameters:

Parameter	In/Out	Data Type	Description
s_percent	IN	NUMBER	
s_degree	IN	NUMBER	

Additional Info:
Utilizes the `dbms_stats` package. Walks over all schemata listed in `iz_owners`, drops its sta
On error for one schema, it walks down to handle each table of this owner separately.



TestCase01RegularDslCheck



Output
Captured

TEST RESULTS [\[history\]](#)

- Test
- Edit
- Properties
- Refactor
- Where Used
- Search
- Files
- Versions
- Recent Changes
- User Guide
- Test History

Assertions: 7 right, 0 wrong, 0 ignored, 0 exceptions

▶ *Set Up:* [.UnitTests.Vdc1wDw.SchemaDwCl.PackageS.SetUp \(edit\)](#) [Expand All](#) | [Collapse All](#)

Test Goal: B1, B2, B3

Description: [B1] Report only currently valid subscriptions. I.e. the latest tariff and options per subscriber without history. [B2] Report only the following latest options. [B3] Report just DSL customers.

▶ *Included page:* [DataSetupRegularDsl \(edit\)](#) [Expand All](#) | [Collapse All](#)

▶ *Included page:* [DataLoad \(edit\)](#) [Expand All](#) | [Collapse All](#)

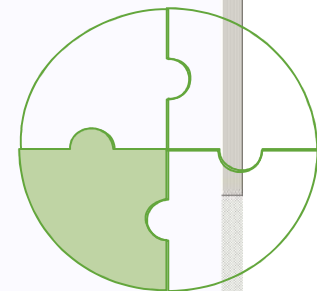
connect to

Expected result:

- * just subscriber -800 needs to be considered. -801 does not have a DSL order.
- * current tariff is DSL (60/60)
- * current speed: Speed 8MBit (the 4 MBit option is inactive and therefore not the latest)
- * currently booked options
- * current tech dsl info: tech_voice_03 and tech_data_03 (*_01 and *_02 are invalid, because of wrong subscriber and inactivity)

query	<pre>SELECT subscriberid , tariff , speed , tech_key_voice , tech_key_data , options_01 , options_02 FROM mv_dsl_alm_report</pre>					
SUBSCRIBERID	TARIFF	SPEED	TECH_KEY_VOICE	TECH_KEY_DATA	options_01	options_02
-800	DSL (60/60)	DSL 8 MBit/s	tech_voice_03	tech_data_03	Y	Y

close

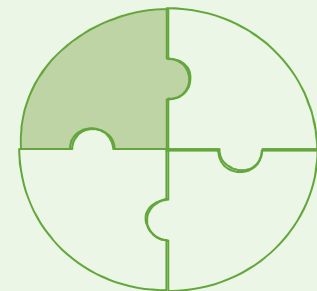


03 ETL mit PL/SQL

4 Organisation

Wer kümmert sich um

- 🍃 **Rollen- / Aufgabenverteilung**
- 🍃 **Entwicklungsrichtlinien**
- 🍃 **Bereitstellung von Templates**
- 🍃 **Weiterentwicklung und Planung der Bibliothek**
- 🍃 **Betrieb der Bibliothek**
- 🍃 **Beschaffung bzw. Implementierung und Betrieb von Hilfsmitteln / Werkzeugen**



03 ETL mit PL/SQL – Prozesse

4 Organisation

Beispiel: Wer übernimmt Betrieb und Weiterentwicklung der Bibliothek?

- **Variante: Kleine Entwicklungs- und Betriebstruppe in kleinem Umfeld**

Jeder macht alles

Risiko: Altlasten, Overdesign führen zu inkonsistenten / unvollständigen Metadaten

- **Variante: Betriebstruppe eines BICC kümmert sich auch um die Bibliothek**

Erkennen Entwicklungspotential schnell

Bei starkem Betriebsaufkommen Vernachlässigung

Risiko: Features mit Nachteilen zu Lasten der Entwickler entworfen → Akzeptanz schwindet.

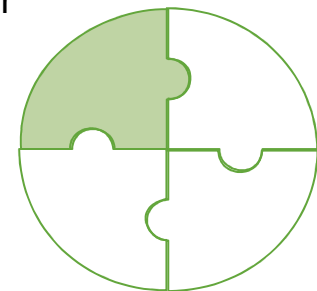
- **Variante: Eine zentrale Architektenstelle kümmert sich darum**

Weiteres Blickfeld als „nur“ ETL-Bibliotheken → Potential für Integration weiterer Metadatenquellen.

Kontinuierliche Arbeit an der Bibliothek möglich

Interessenausgleich zwischen verschiedenen Bereichen möglich

Risiko: Entwicklungen am Bedarf vorbei.



01 ABGRENZUNG DES THEMAS

02 TOOL & PL/SQL – DIE UNTERSCHIEDE

03 ETL MIT PL/SQL

04 ZUSAMMENFASSUNG

04 Zusammenfassung

Was ist wichtig für die professionelle ETL-Entwicklung mit PL/SQL?

- 🍃 Schaffung einer **zentralen Metadatenbasis**
- 🍃 Bereitstellung von **zentralen Funktionen**, Verzahnung mit Metadaten
- 🍃 KISS – einfach und bedarfsorientiert vorgehen
- 🍃 Offene Gestaltung der Metadaten
- 🍃 Einbindung in die **Prozesse**: Bibliothek und Metadaten weiterentwickeln!
- 🍃 Nutzung von Tools zur Unterstützung von Prozessen
- 🍃 Zuständigkeiten in der **Organisation** festlegen

04 ETL mit PL/SQL

Wie können diese Punkte nun bei der ETL Entwicklung helfen?

- 🍃 Besser: **Metadaten!**
- 🍃 Leichter: **Maintenance** durch Logging und Fehlerbehandlung.
- 🍃 Schneller: **Fehlerbehebung**
- 🍃 Strukturierter: Fertige Bausteine und **Templates**
- 🍃 Produktiver: Etablieren einer **DWH-Bibliothek für ETL-Prozesse** mit PL/SQL
- 🍃 So können Sie gezielt von den Vorteilen profitieren, die Ihnen PL/SQL bei der Entwicklung ETLs bietet.
- 🍃 Vorausgesetzt es wird richtig geplant und umgesetzt....

HERZLICHEN DANK FÜR IHRE AUFMERKSAMKEIT

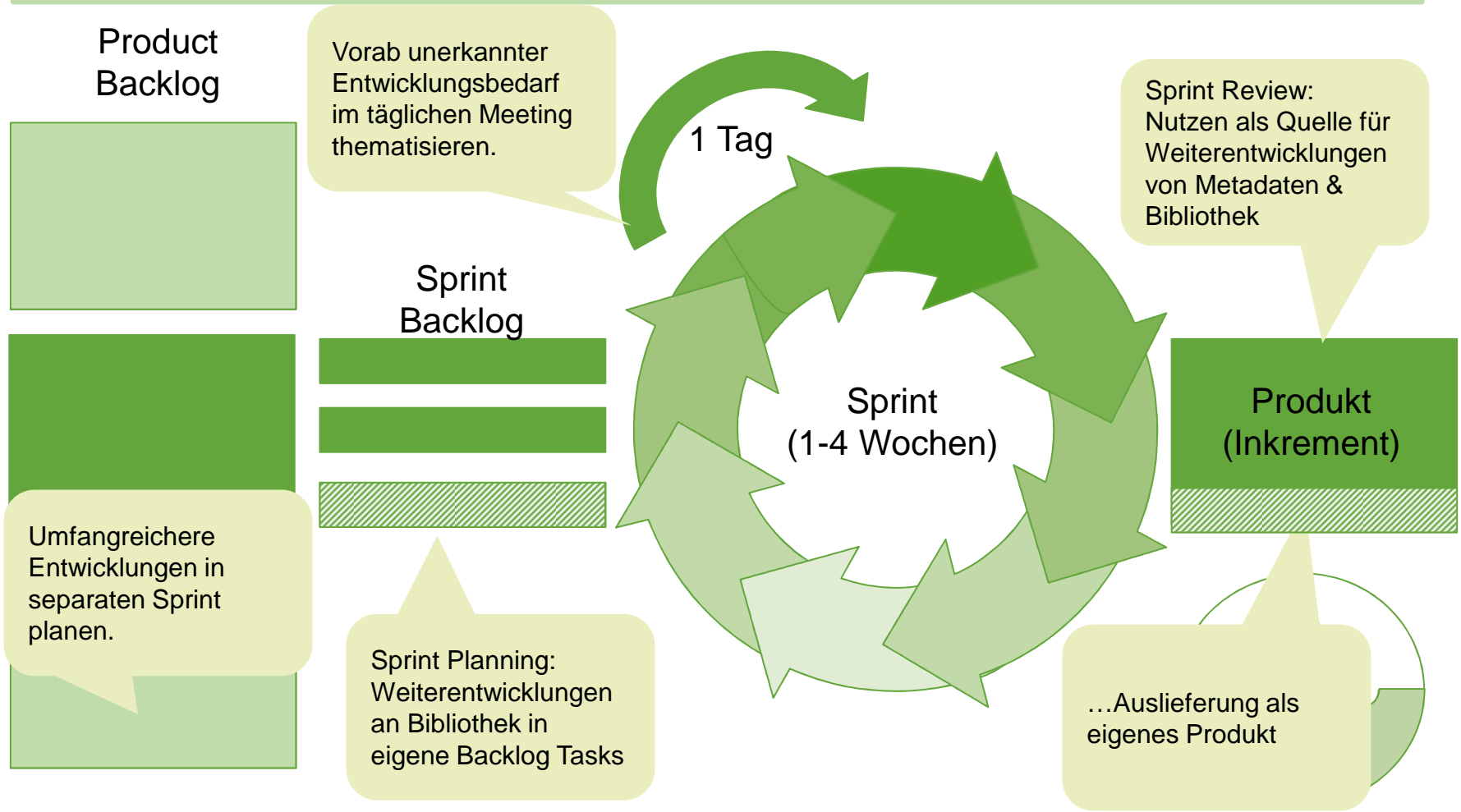
Alexander Mendle
Professional Consultant
Alexander.Mendle@SHS-VIVEON.com

SHS VIVEON AG
Clarita-Bernhard-Str. 27
81249 München
Germany
T +49 (0) 89 74 72 57 - 0
F +49 (0) 89 74 72 57 - 900

info@SHS-VIVEON.com
www.SHS-VIVEON.com



03 ETL mit PL/SQL – Prozesse 2 Prozesse



Bildnachweise

- **Boxenstopp:** Veröffentlicht auf flickr.com
<http://www.flickr.com/photos/aostamodel/5192207233/>
von „bs070180222“ (<http://www.aostamodel.net>) unter
Creative Commons Lizenz CC-BY-2.0
- **Seifenkiste:** Veröffentlicht auf flickr.com
<http://www.flickr.com/photos/coreforce/3586170715/> von
„CoreForce“ / Falk Lademann unter Creative Commons
Lizenz CC-BY-2.0