



**ENGINEERED
FOR INNOVATION**

ORACLE®
DATA WAREHOUSE

ORACLE®

Datenqualitätsportal für das Data Warehouse mit APEX und Regelwerk

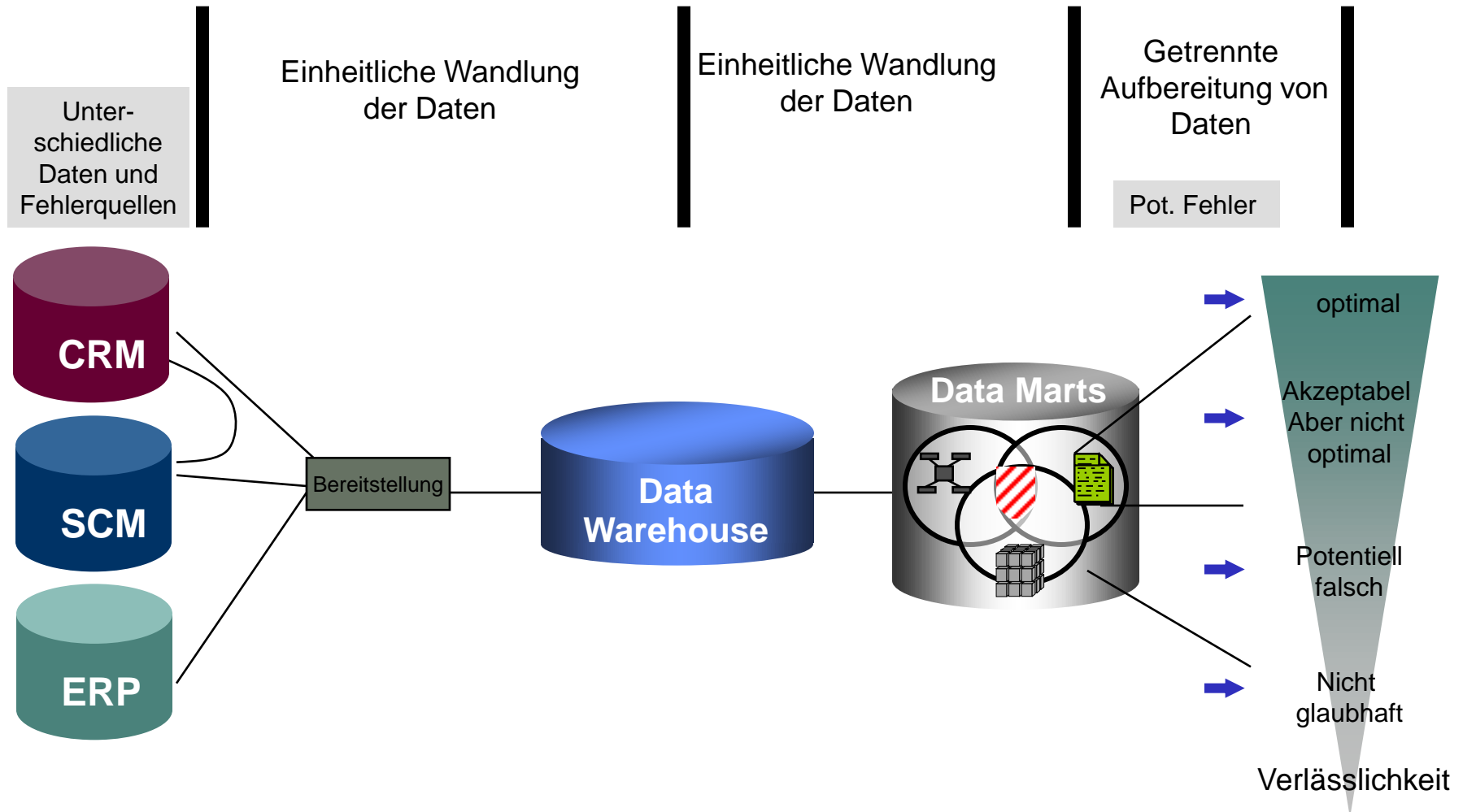
Alfred Schlaucher, Oracle Data Warehouse

Themen

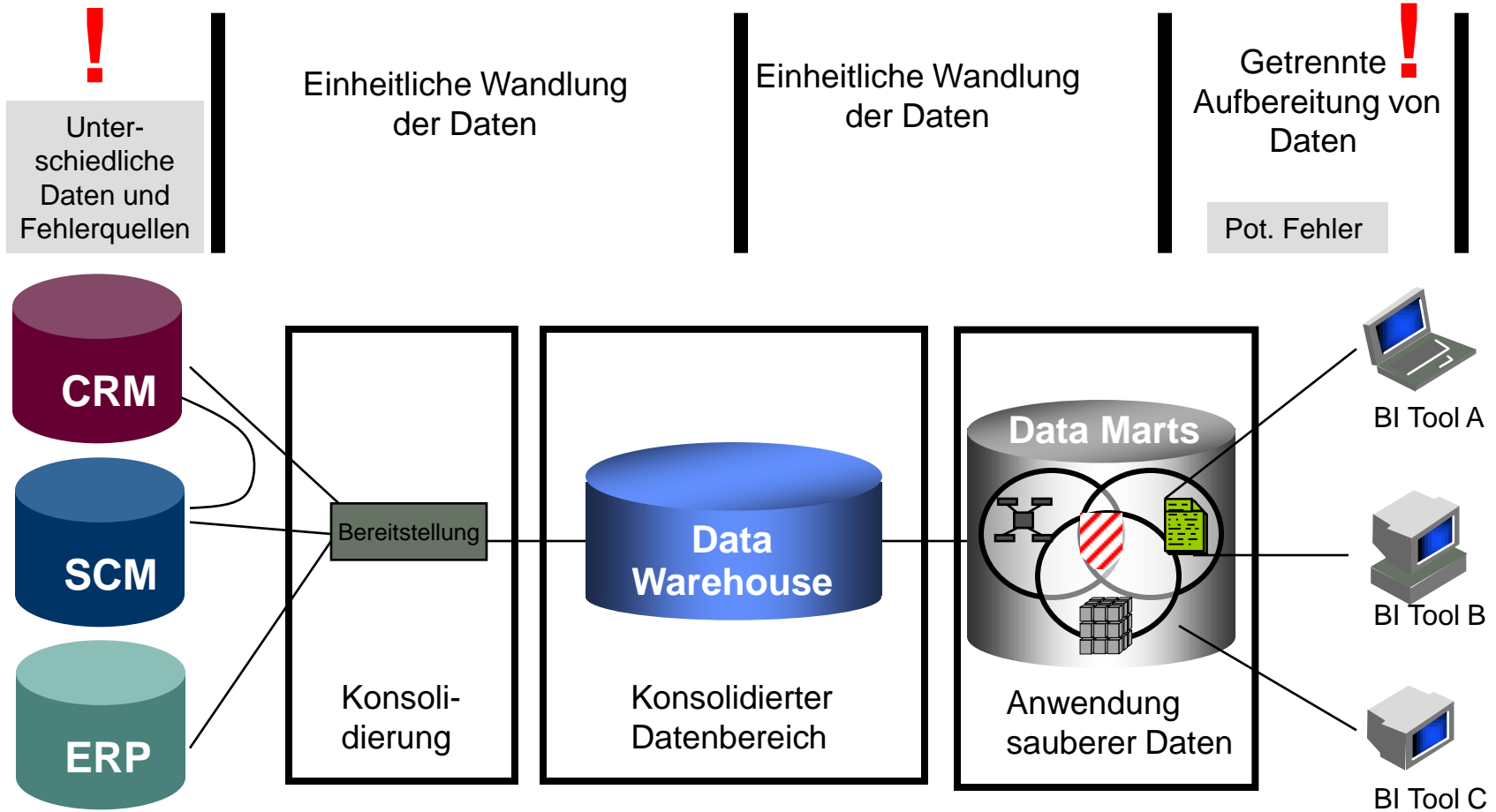
- Wo wird geprüft
- Was wird geprüft
- Dokumentation und Steuerung der Prüfungen über Metadaten
- Das System



Wo finden Profiling / Fehlersuchen statt?



Potentielle Fehlerquellen

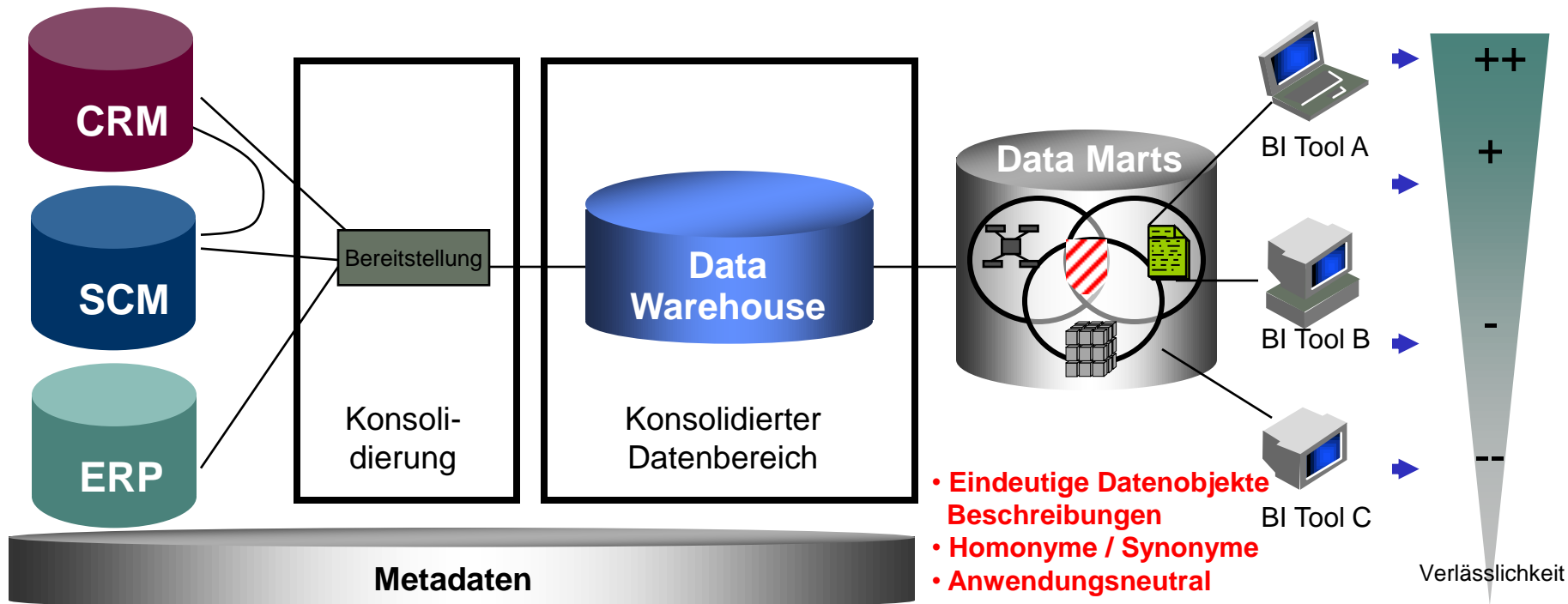


Datenqualität

Bezogen auf den Warehousing-Prozess

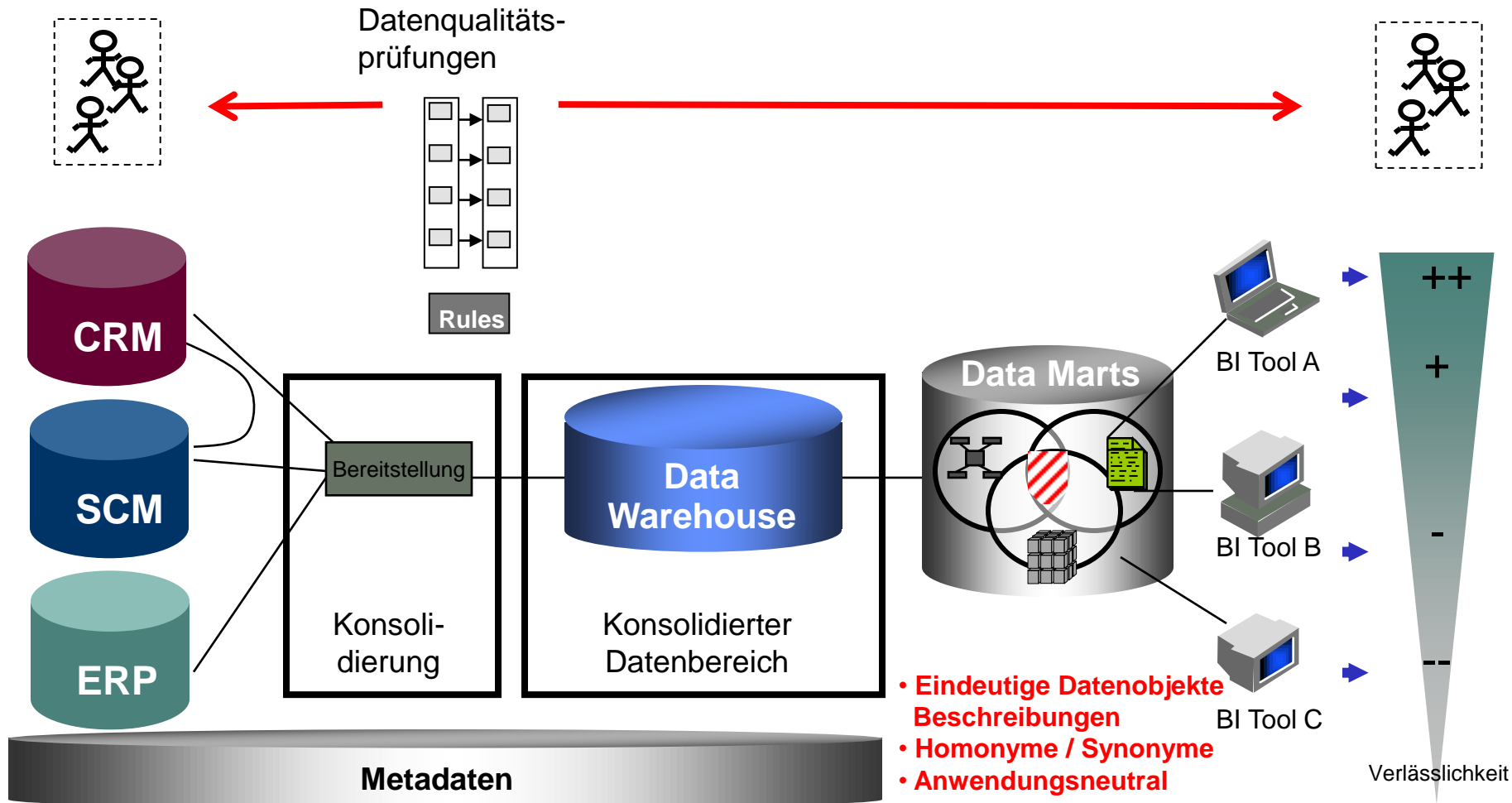
- Heterogene Datenmodelle / Konsistenz / Homonyme / Synonyme
- Kontinuität des Ladevorgangs / Vollständigkeit
- Widerspruchsfreiheit zwischen den Quellen

Unter-
schiedliche
Daten und
Fehlerquellen



Datenqualität

Bezogen auf den Warehousing-Prozess



Varianten von Prüfungen

Attribut-bezogen

- Not Null / Pflichtfelder
- Formatangaben
- Check Constraint
- Wertbereiche
 - Ober-/Untergrenzen / Wertelisten

Satz-bezogen (Tupel)

- Abhängigkeiten von Werten in anderen Attributen desselben Satzes

Satz-übergreifend (Relationen)

- Primary Key / Eindeutigkeit
- Aggregat – Bedingungen
 - Ober- Untergrenzen von Summen
 - Anzahl Sätze pro Intervall usw.
- Rekursive Zusammenhänge
 - Verweise auf andere Sätze derselben Tabelle (Relation)

Tabellen-übergreifend (interrelational)

- Foreign Key
- Aggregat – Bedingungen
 - Ober- Untergrenzen von Summen
 - Anzahl Sätze pro Intervall usw.
- Rekursive Zusammenhänge
 - Verweise auf Sätze einer anderen Tabelle (Relation)

Zeit-bezogen (Tupel)

- Zeitinvariante Inhalte
 - Anz. Bundesländer
- Zeitabhängige Veränderungen
- Über die Zeit mit anderen Daten korrelierende Feldinhalte

Verteilungs-bezogen

- Arithmetische Mittel
- Varianz / Standardabweichungen
- Qualitätsmerkmale und Mengen

Beschreibung von Regeln im Repository

| N° ▲ | Attribute | Value |
|------|----------------------------|--|
| 1 | KATEGORIE | COLUMN |
| 2 | KOMPLEXITAET | EINFACH |
| 3 | META_LANGUAGE | Column value between n and m |
| 4 | MASKE | NA |
| 5 | WERTEBEREICH | 100:400 |
| 6 | WERTE_LISTE | NA |
| 7 | LOESBAR_UEBER | Function_Call |
| 8 | FUNKTION | NR_between_Ck(scheme:table:column:from.t |
| 9 | ERWARTETER_ERFUELLUNGSGRAD | 100 |

Mögliche Einträge:
Column/Relationship/Data_Rule/
Process_Rule/Value_Rule

Mögliche Einträge:
einfach/komplex

Mögliche Einträge:
(Sprache mit der Regel programmiert wird)
Column value between n and m

Mögliche Einträge:
Feldstruktur, z. B. Regular Expressions

Mögliche Einträge:
Grenzen von Wertebereichen

Mögliche Einträge:
Liste mit gültigen Werten für eine Spalte

Mögliche Einträge:
Funktion/Sonst (Lösungsstrategie)

Mögliche Einträge:
Lösungs-Code (Pseudo-Code)

Mögliche Einträge:
Prozentwert

Auflistung von Regeln im Repository

Objects

+ Obj R% O'Tp RULE Scope all etc Act z/s 16

| Name | Version | Variant | Description | O'Type | Act. | Ref1 | Ref2 |
|------------------------------------|---------|---------|---|--------|------|------|------|
| R1_PROODUKTE_BESTAND | 1 | VAR1 | Produkte_Stamm: Bestand muss immer groe | RULE | ✓ | | |
| R3_KUNDENKARTE | 1 | VAR1 | Wenn Kundenkarte = 1 dann Status = ,P' | RULE | ✓ | | |
| R4_ZAHLUNGSZIEL | 1 | VAR1 | Wenn Zahlungsziel in der Vergangenheit u | RULE | ✓ | | |
| R5_FIRMENRABATT | 1 | VAR1 | Firmenrabatt darf nur gefuellt sein, wen | RULE | ✓ | | |
| R8_BESTELLPOSITION_OHNE_BESTELLUNG | 1 | VAR1 | Orphans: Bestellpositionen muessen einer | RULE | ✓ | | |
| R9_BESTELLUNG_OHNE_POSITION | 1 | VAR1 | Childless Parents: Bestellungen muessen | RULE | ✓ | | |
| R10_BESTELLWERT_GLEICH_LIEFERWERT | 1 | VAR1 | Bestellung (Wert, Produkte) = Lieferung | RULE | ✓ | | |
| R11_EINKAUFSMENG_ABSATZMENGE | 1 | VAR1 | Einkaufsmenge = Absatzmenge + Lagerbesta | RULE | ✓ | | |
| R7_GESAMTBESTELLWERT | 1 | VAR1 | Bestellung / Best_Position: Einzelnen Po | RULE | ✓ | | |
| R12_POSITION_MIT_STATUS_2_3 | 1 | VAR1 | Wenn Bestell_Position fü r Produkt = Stat | RULE | ✓ | | |
| R13_RABATT_P | 1 | VAR1 | Verkaufspreis-Rabatt } Einkaufspreis wen | RULE | ✓ | | |
| R6_ZAHLUNGSZIEL_LIEFERDATUM | 1 | VAR1 | Wenn Kundenkarte = 1 und Status = P dan | RULE | ✓ | | |
| R14_BESTELLUNG_KUNDE | 1 | VAR1 | Eine Bestellung muss immer eindeutig ein | RULE | ✓ | | |

1

Einbetten von Regeln in das System

Objects

+ Obj O'Tp Scope etc Act z/s

| ≡ ▲ | Name | Version | Variant | Description | O'Type | Act. | ≡ |
|-----|-----------------------------------|---------|---------|-------------------------------------|----------|------|---|
| | OU_SERVICE_GMBH | 1 | VAR1 | Beispielfirma | ORG_UNIT | ✓ | |
| | OU_CONTROLLING | 1 | VAR1 | Controlling Abteilung | ORG_UNIT | ✓ | |
| | OU_BUCHHALTUNG | 1 | VAR1 | Buchhaltungsabteilung | ORG_UNIT | ✓ | |
| | OU_DISPONIERUNG | 1 | VAR1 | Planung Material und Maschinen | ORG_UNIT | ✓ | |
| | OU_VERTRIEB_BAUMARKT_SERVICE_GMBH | 1 | VAR1 | Vertriebsabteilung der Service GMBH | ORG_UNIT | ✓ | |
| | OU_VERTRIEB_FINANZ_SERVICE_GMBH | 1 | VAR1 | Vertriebsabteilung der Service GMBH | ORG_UNIT | ✓ | |

1

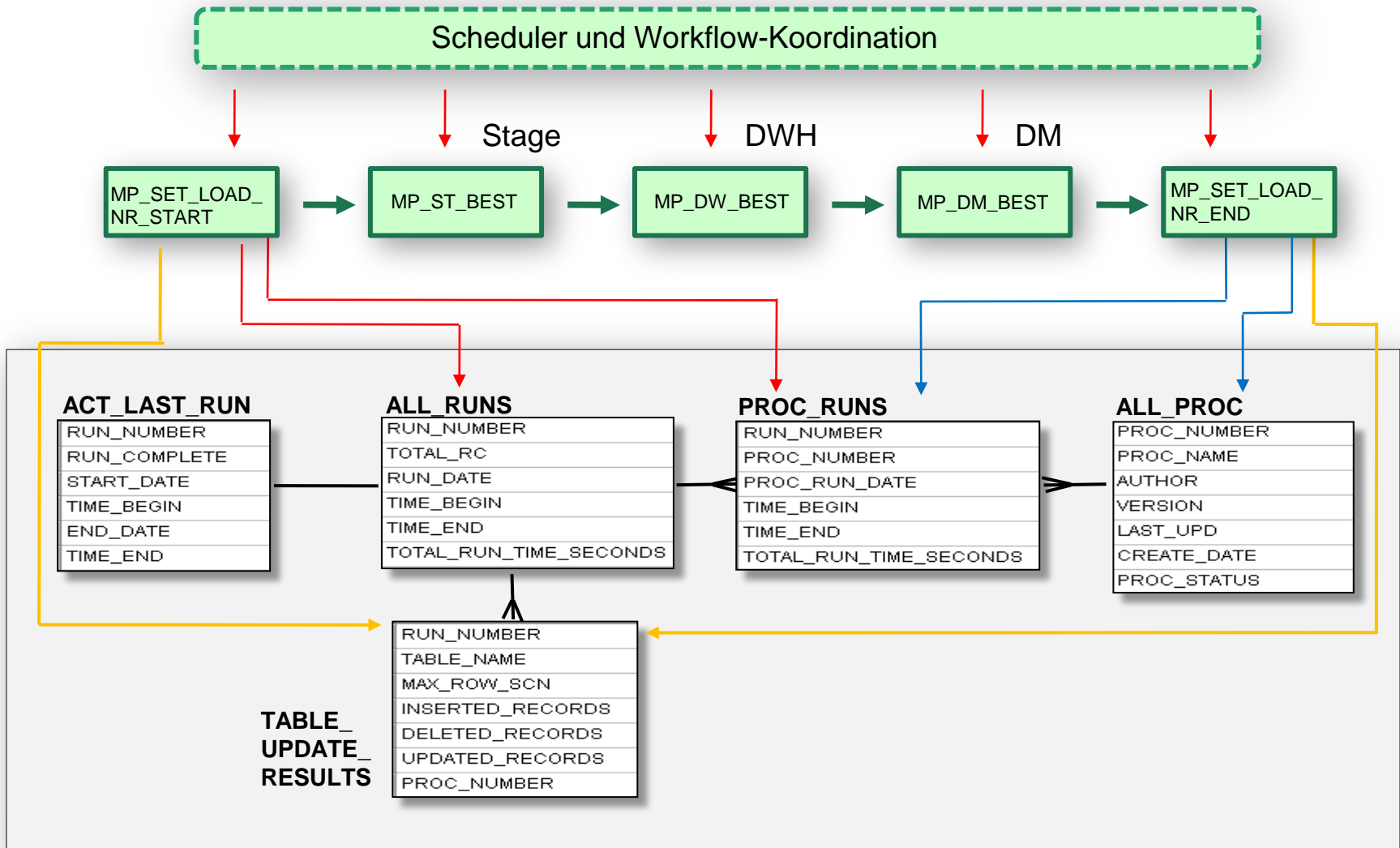
Object Structure Resolution

O'Type Obj #Lev. z/s

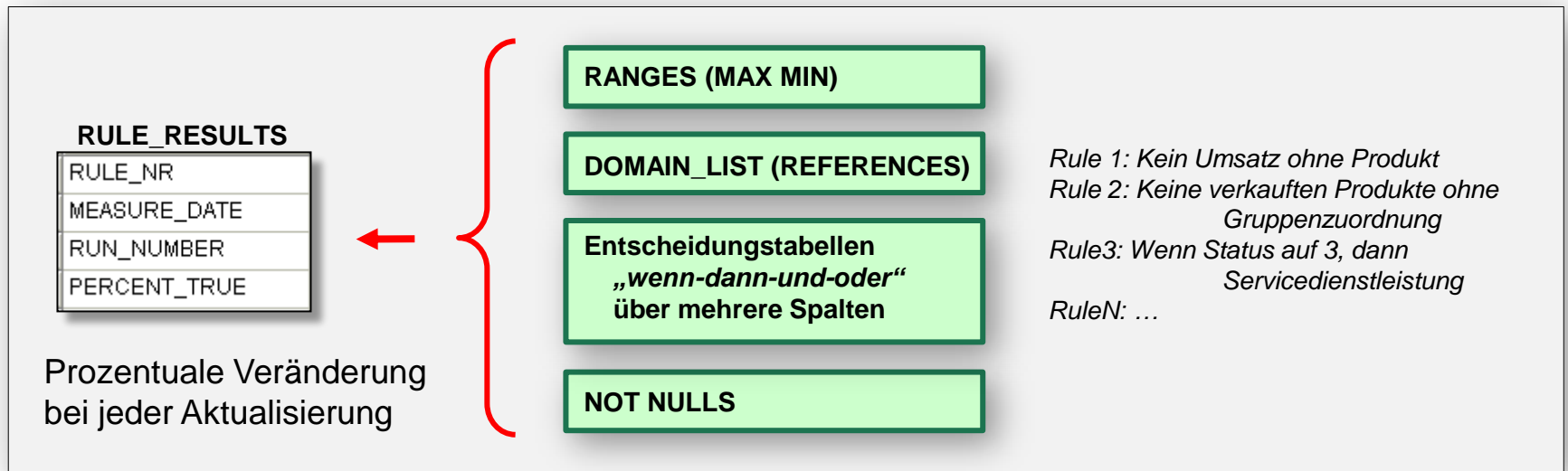
| Lv | ... | Relationship | Obj'Name | O'Type | ≡ |
|----|-------------------------------|--------------------|-----------------------------------|------------------|---|
| 0 | | | OU_VERTRIEB_BAUMARKT_SERVICE_GMBH | ORG_UNIT | |
| 1 | ---- | owns | GP_BESTELLUNGEN_KUNDEN | PROZESS | |
| 2 | ----- | is_controlled_by | BR_UMSATZ_PRODUKT_ZUORDNUNG | BUSINESS_RULE | |
| 3 | ----- ---- | is_covered_by | R43_UMSATZ_OHNE_PRODUKT | RULE | |
| 4 | ----- ----- | prueft | BEST_POSITION | TABELLE | |
| 5 | ----- ----- ----- | input_fuer | MP_ST_RULE_CHECK | MAPPING_PROGRAMM | |
| 6 | ----- ----- ----- ----- | is_used_by_process | PRSZ_SZ_1_TEIL1 | LOAD_PROCESS | |
| 7 | ----- ----- ----- ----- ----- | uses_check | CP_MP_ST_RULE_CHECK | CHECK_PROCESS | |

Zu den DQ-
Ergebnissen

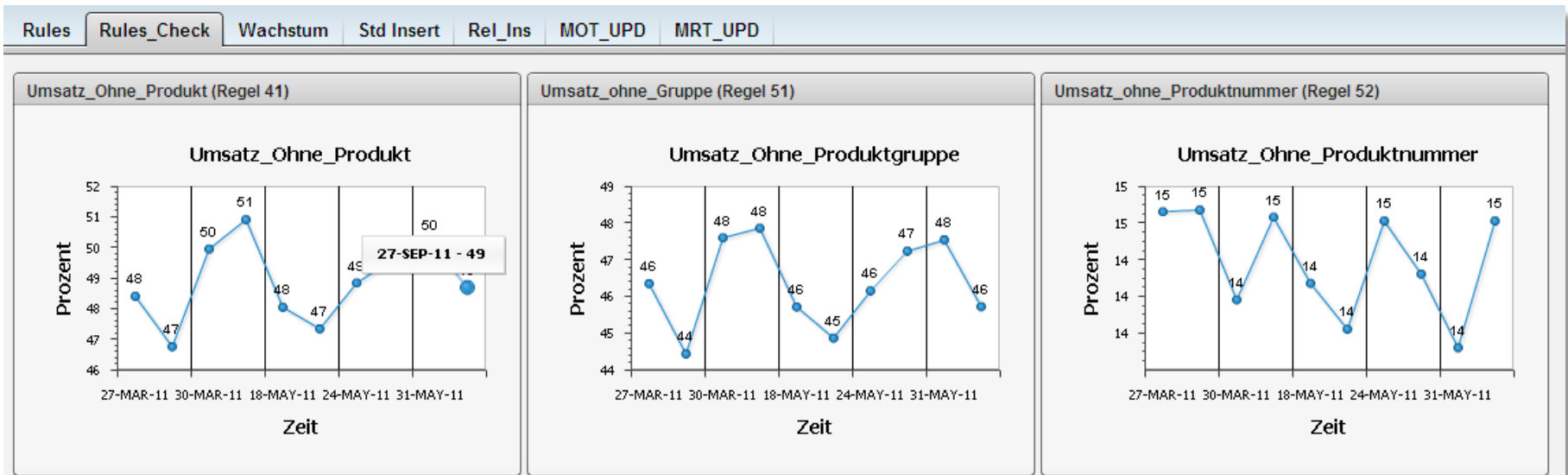
Laufzeit- und Zuwachsmonitoring



Parametrisierbare Rules-Bausteine



Visualisierung der Regelprüfung mit APEX



Visualisierung der Prozentwerte der Regelprüfung mit APEX

| Umsatz ohne Produkt (Regel 41) | | Umsatz ohne Gruppe (Regel 51) | | Umsatz ohne Produktnummer (Regel 52) | |
|--------------------------------|--------------|-------------------------------|--------------|--------------------------------------|--------------|
| Measure Date ▲ | Percent True | Measure Date ▲ | Percent True | MEASURE_DATE ▲ | PERCENT_TRUE |
| 27-SEP-11 | 48.68 | 27-SEP-11 | 45.72 | 27-SEP-11 | 14.61 |
| 31-MAY-11 | 50.16 | 31-MAY-11 | 47.53 | 31-MAY-11 | 13.92 |
| 25-MAY-11 | 49.61 | 25-MAY-11 | 47.23 | 25-MAY-11 | 14.32 |
| 24-MAY-11 | 48.83 | 24-MAY-11 | 46.14 | 24-MAY-11 | 14.61 |
| 19-MAY-11 | 47.33 | 19-MAY-11 | 44.86 | 19-MAY-11 | 14.02 |
| 18-MAY-11 | 48.04 | 18-MAY-11 | 45.71 | 18-MAY-11 | 14.27 |
| 17-MAY-11 | 48.75 | 17-MAY-11 | 46.31 | 17-MAY-11 | 13.74 |
| 17-MAY-11 | 50.9 | 17-MAY-11 | 47.85 | 17-MAY-11 | 14.63 |
| 30-MAR-11 | 49.94 | 30-MAR-11 | 47.59 | 30-MAR-11 | 14.18 |
| 29-MAR-11 | 46.75 | 29-MAR-11 | 44.43 | 29-MAR-11 | 14.67 |
| 27-MAR-11 | 48.54 | 27-MAR-11 | 46.03 | 27-MAR-11 | 13.94 |
| 27-MAR-11 | 50.18 | 27-MAR-11 | 47.77 | 27-MAR-11 | 13.88 |
| 27-MAR-11 | 48.4 | 27-MAR-11 | 46.34 | 27-MAR-11 | 14.66 |

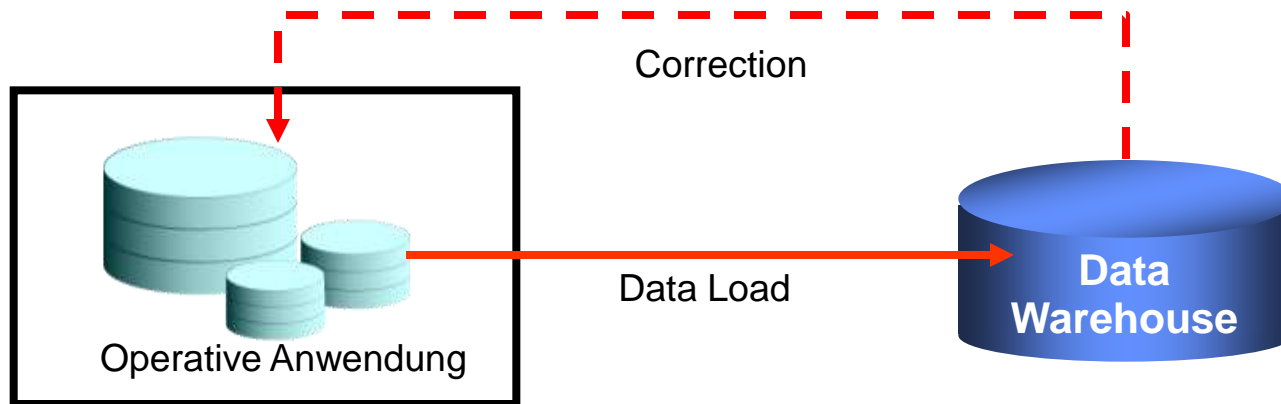
1 - 13

1 - 13

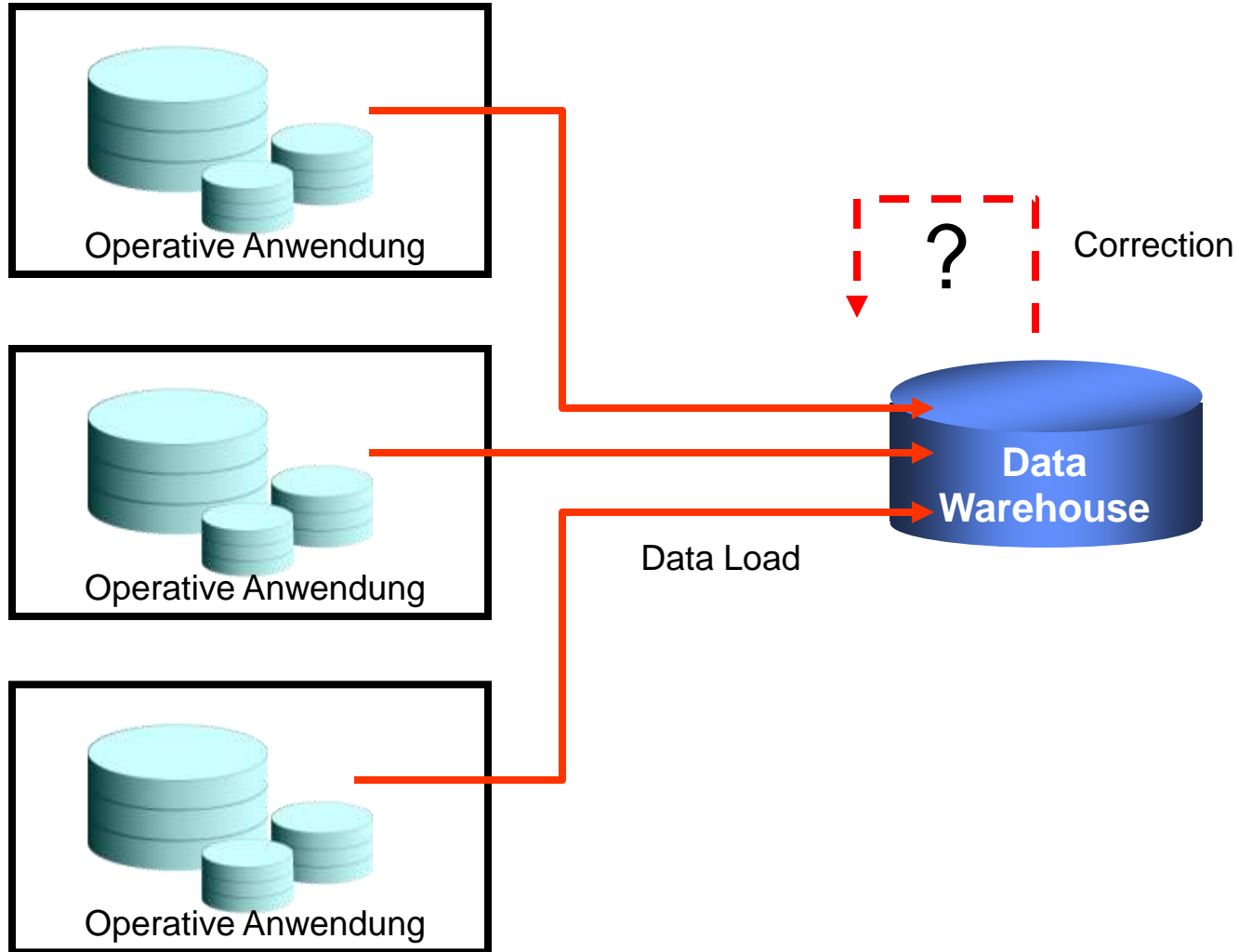
1 - 13

Wo sollten Korrekturen stattfinden?

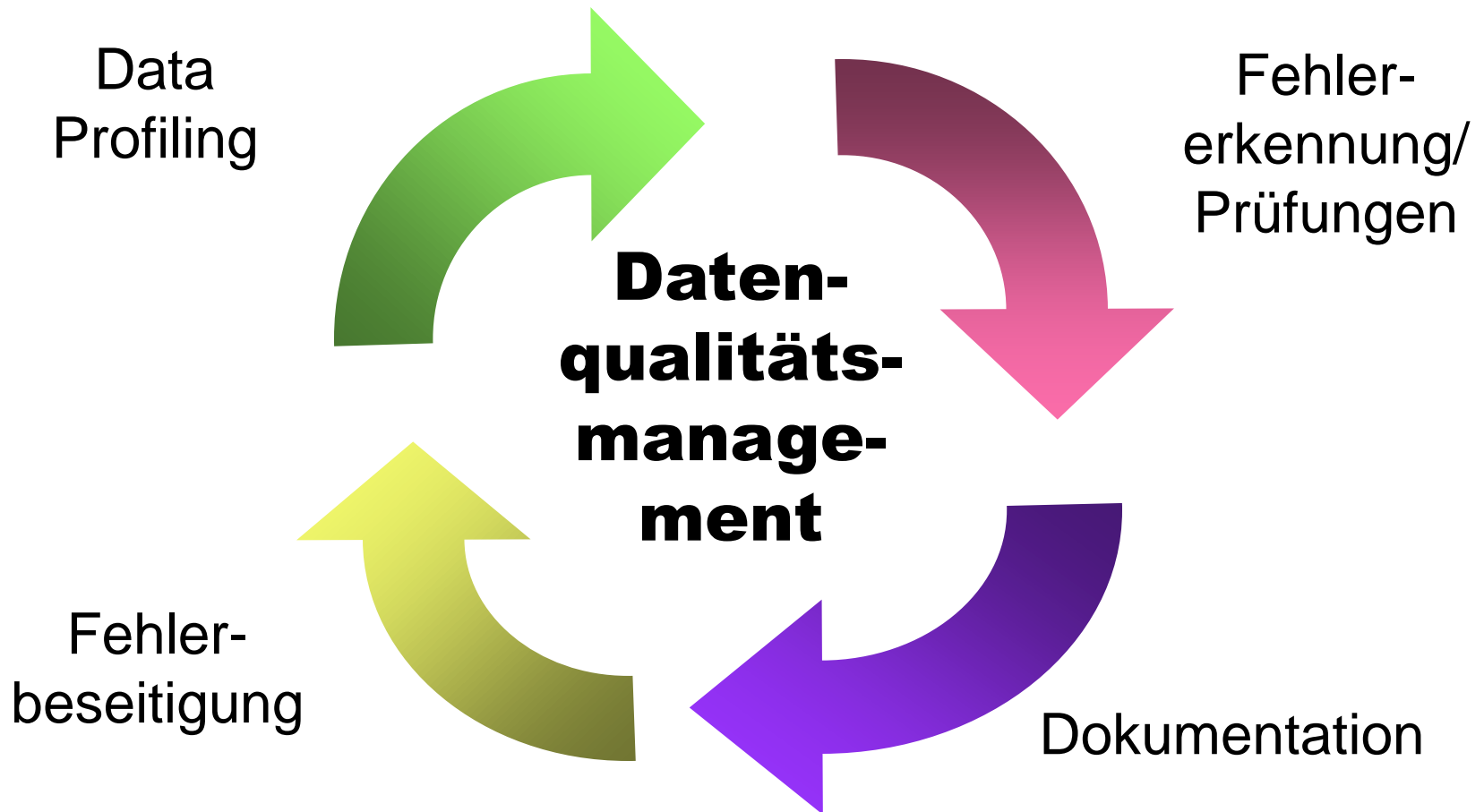
Vorsysteme bzw. Fachabteilungen sind in der Pflicht!



Wo sollten Korrekturen stattfinden?



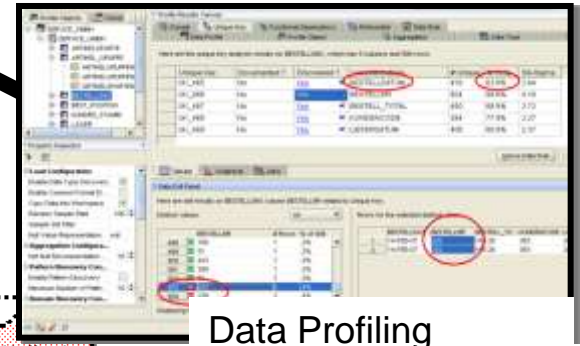
Lifecycle des Datenqualitäts-Management



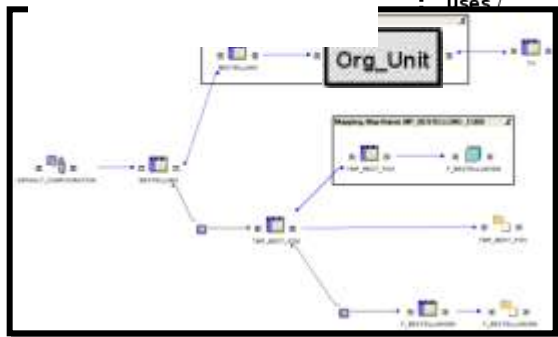
Data Quality Management Process



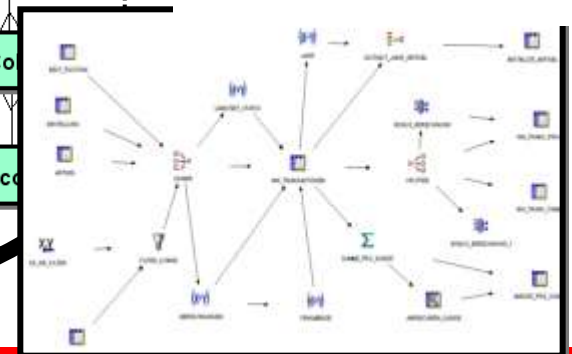
Data Quality Reporting



Data Profiling



Data Quality Auditing



Datenkorrektur

