

Geschäftsprozess-Performance messen leicht gemacht

Alexander Voigt, areto consulting gmbh

Bei der Analyse komplexer Geschäftsprozesse oder Lebenszyklen ist die Zeitdauer zwischen Fortschritten von großem Interesse. Eine Möglichkeit, dem Endanwender diese Analysen zu erleichtern, ist das Aufbereiten der Daten in einem Accumulating Snapshot.

In diesem Snapshot sind alle Schritte eines Gesamtprozesses zu einem gemeinsamen Datensatz zusammengefasst, sodass Auswertungen über den Gesamtprozess stark vereinfacht werden. Dafür ein gutes Design zu finden, ist nicht trivial und fordert den Fachbereich gleichermaßen wie die IT. Als Lohn locken umfassendere Analysemöglichkeiten für den Endnutzer und damit das, was am wichtigsten ist: Kundenzufriedenheit.

Zum Hintergrund: Viele Geschäftsprozesse lassen sich als Serie von wohldefinierten Prozessschritten begreifen. Da werden Bestellungen abgewickelt, Ver-

sicherungsverträge verfolgt oder die Sales Opportunity analysiert. Die gängige Form, diese Daten im Enterprise Data Warehouse aufzubereiten, ist die transaktionale Faktentabelle. Jeder Prozessschritt wird als einzelner Datensatz gespeichert und enthält Fremdschlüssel zu den Dimensionen, die den jeweiligen Prozessfortschritt definieren. Ein Fremdschlüssel zu einer Status-Dimension ermöglicht die Zuordnung zu den einzelnen Prozessschritten. Damit sind die Informationen des Gesamtprozesses über mehrere Datensätze verteilt. Weitere Referenzen ermöglichen die Zuordnung des einzelnen

Prozessschritts zu einem Gesamtprozess. Im Folgenden wird als Beispiel ein Prozess für die Bestellabwicklung verwendet.

Diese Form der Datenmodellierung hat den Vorteil der Flexibilität. Sie kann ohne Probleme nicht-lineare Prozessabläufe abbilden und auch Prozess-Änderungen lassen sich ohne großen Aufwand integrieren. Ein weiterer Prozessschritt wird als neuer Eintrag in der Status-Dimension realisiert, auf den zukünftig in der Faktentabelle referenziert wird. Es können beliebige zeitliche Anordnungen und Wiederholungen von Prozessschritten abgebildet werden. Auswertungen auf Details einzel-

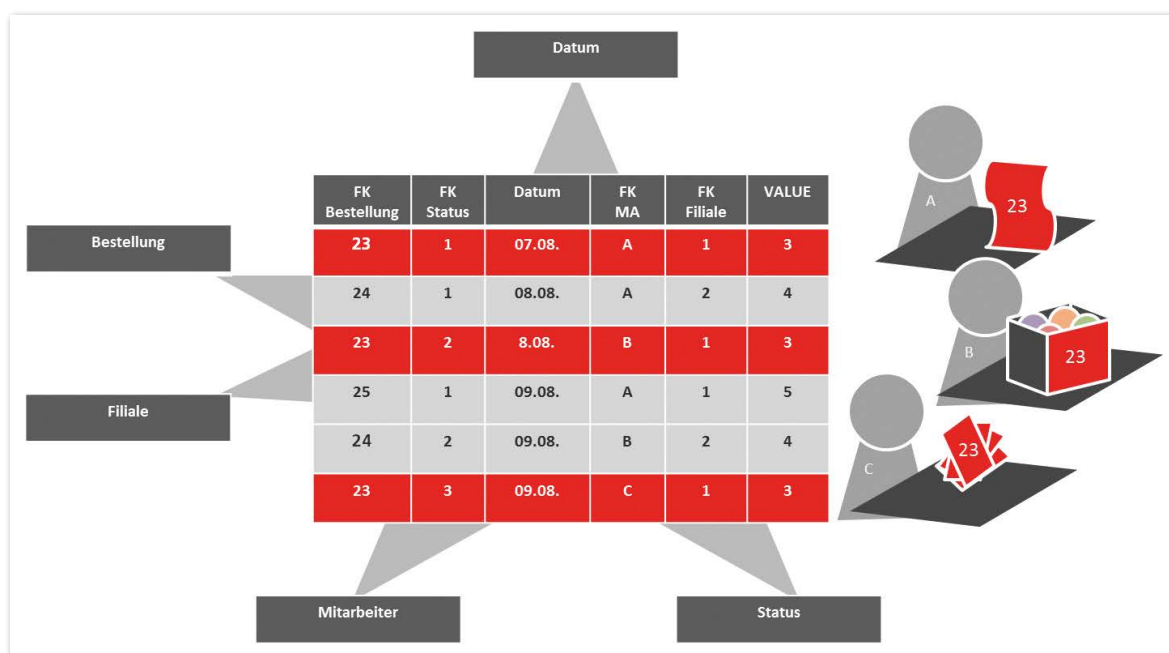


Abbildung 1: Transactional Star der Bestellabwicklung

ner Prozessschritte, wie zum Beispiel der Gesamtwert der Lieferungen in einem Zeitraum, lassen sich einfach umsetzen (siehe Abbildung 1).

Möchte man hingegen Fragen beantworten, wie zum Beispiel die Bestimmung der durchschnittlichen Zeitdauer zwischen Bestelleingang und Auslieferung, so werden die nötigen Abfragen schnell komplex. Allein aus der Information, dass eine Bestellung ausgelöst worden ist, lässt sich nicht ableiten, ob sie noch offen oder schon geschlossen ist.

Um die Frage zu beantworten, sind die Datensätze der einzelnen Prozessschritte zu verbinden. Die Zuordnung wird über den Fremdschlüssel der zu verarbeitenden Einheit, hier der Bestellung, vorgenommen. Die Auswertung der Zeitdauer ist abhängig vom Prozessschritt. Die nötige Logik ist zum Beispiel in Form von Case-Konstrukten hinzuzufügen. Ohne Wissen über sämtliche den Geschäftsprozess definierenden Regeln lässt sich das nicht durchführen (siehe Listing 1).

Das Aufbereiten der einzelnen Schritte zu einer fertigen Gesamtübersicht ist daher naheliegend. Eine erprobte und nützliche Form, dies abzubilden, ist der Accumulating Snapshot. Diese im Vergleich zum Transaction Star in der Praxis seltene Form der Faktentabelle bildet den Lebenszyklus eines Prozesses ab (siehe Abbildung 2).

Im Gegensatz zur transaktionalen Faktentabelle besitzt der Accumulating Snapshot mehrere Zeitbezüge. Dabei steht der einzelne Zeitbezug für das Erreichen eines neuen Prozessschritts. Je nach Geschäftsprozess können aber auch andere Dimensionen in verschiedenen Rollen referenziert werden, wie zum Beispiel der Einkäufer, der die Bestellung ausgelöst hat. Auf Basis mehrfacher Zeitbezüge lassen sich Kennzahlen in Form von Zeitdifferenzen zwischen Schritten berechnen. Sie werden häufig „Lags“ genannt. Darüber hinaus sind Vergleiche zwischen Soll- und Ist-Werten möglich, wie zum Beispiel Differenzen zwischen der Bestell- und der Liefermenge (siehe Abbildung 3).

Der Nutzen ist, dass der Benutzer die fachlich sinnvolle Verbindung der Daten einzelner Prozessschritte bereits aufbereitet vorfindet. Ein weiterer Pluspunkt ist die Definition der daraus neu entstehenden Kennzahlen, auf die dann einheitlich

```
SELECT
    f.filiale,
    avg(CASE
        WHEN stat_3.datum IS NOT NULL THEN stat_3.datum - stat_1.datum
        WHEN stat_3.datum IS NULL THEN trunc(SYSDATE) - stat_1.datum
    END) lieferzeit
FROM
    bestellung b INNER JOIN
    (SELECT fk_bestellung, datum , fk_filiale
    FROM transact t
    WHERE t.fk_status = 1) stat_1
    ON b.pk_sik = stat_1.fk_bestellung
    LEFT JOIN
    (SELECT fk_bestellung, datum
    FROM transact t
    WHERE t.fk_status = 3) stat_3
    ON b.pk_sik = stat_3.fk_bestellung
    INNER JOIN filiale f
    ON stat_1.fk_filiale = f.pk_sik_filiale
GROUP BY
    f.filiale
```

Listing 1

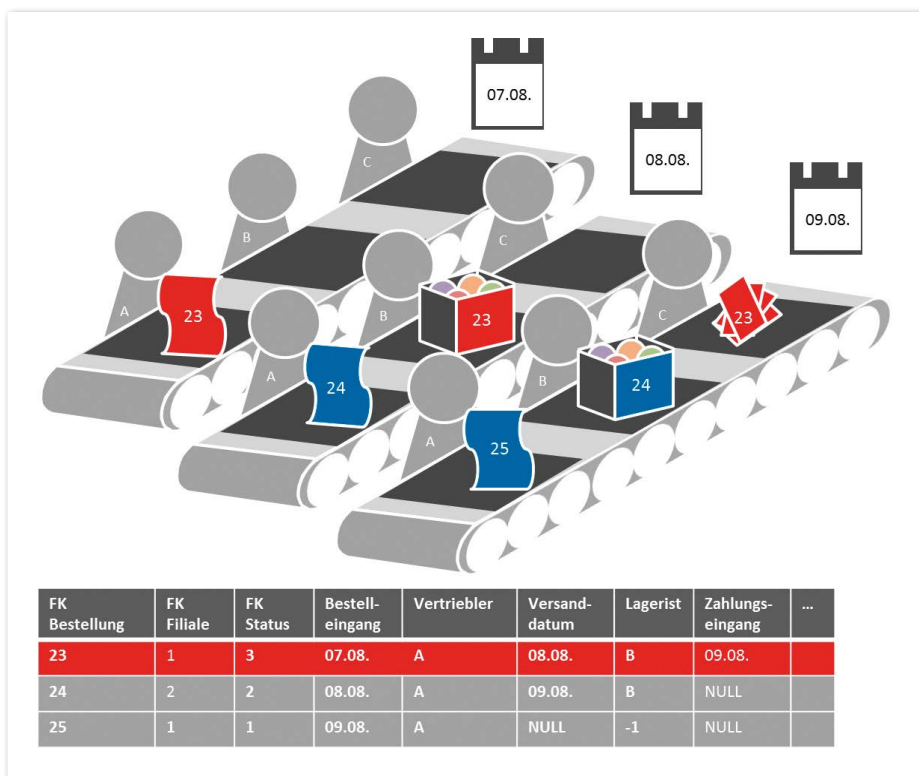


Abbildung 2: Datensatzentwicklung im Accumulating Snapshot

zugriffen werden kann. Der durch das Auswertungstool des Benutzers zu erzeugende Code ist im Vergleich zum Transaction Star simpel (siehe Listing 2).

Dynamisches Duo

Wichtig ist die Erkenntnis, dass ein Accumulating Snapshot im Gegensatz zum

dem Transaction Star ziemlich stark spezialisiert ist. Auswertungen auf einzelne Prozessschritte sind schwierig, insbesondere die historisch genaue Auswertung. Auch zwingt die notwendige Form der Aufbereitung zu vereinfachten Darstellungen und damit zu Detailverlust. Deshalb wird häufig sowohl Transaction

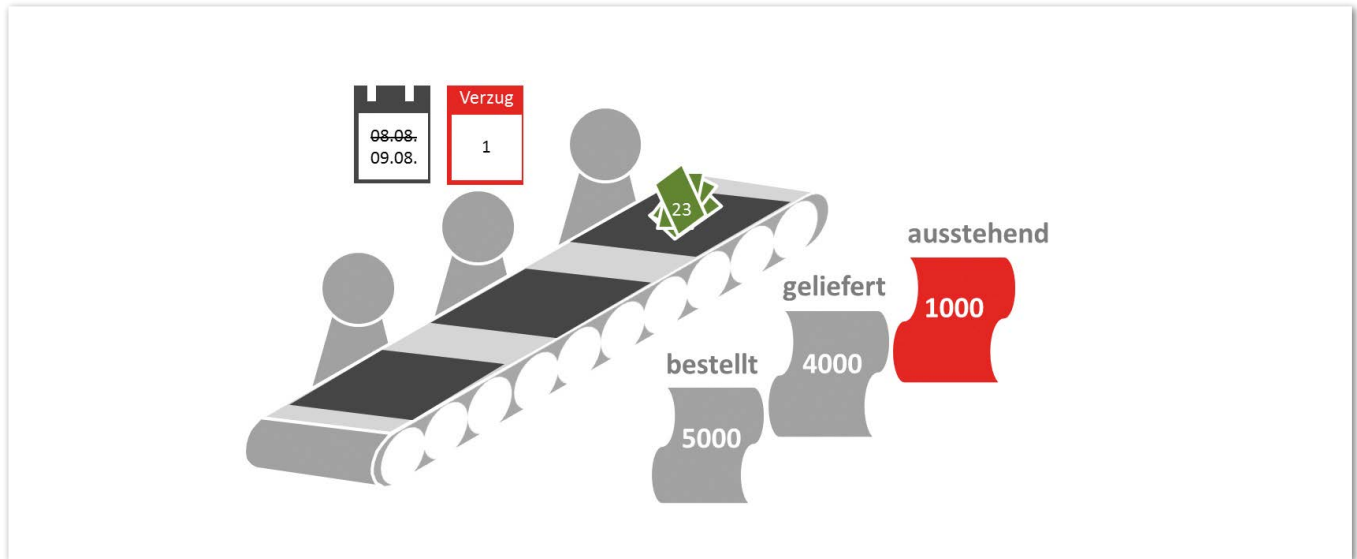


Abbildung 3: Neue Kennzahlen müssen sich keineswegs auf Lags beschränken

D Filiale						TF Bestellabwicklung				
SIK Filiale	ID	NAME	Adresse	VON	BIS	FK Bestellung	FK Status	Datum	FK Filiale	...
11	M	München	ABC 123	01.01.2011	06.08.2014	23	1	06.08.2014	11	...
12	M	München	XYZ 456	07.08.2014	31.12.9999	23	2	07.08.2014	12	...
						23	3	08.08.2014	12	...

AS Bestellabwicklung							
FK Bestellung	FK Filiale Vertrieb	FK Filiale Versand	FK Status	Bestelleingang	Versanddatum	Zahlungseingang	...
23	11	12	3	06.08.2014	07.08.2014	09.08.2014	...

Abbildung 4: Das Übernehmen der SIK aus der transaktionalen Faktentabelle erzeugt Fehler

Star als auch ein Accumulation Snapshot bereitgestellt.

Der Benutzer wertet auf dem Accumulating Snapshot dann zum Beispiel die Lieferzeit aus und analysiert bei Interesse die Details auf einem Transaction Star. Bei dieser Konstellation sollte der Accumulating Snapshot aus der transaktionalen Faktentabelle abgeleitet werden. Zum einen garantiert dies die Konsistenz beider Tabellen, zum anderen ist die Quelle gewöhnlich transaktional modelliert und

Doppelaufwände bei der Aufbereitung werden vermieden.

Beim Ableiten aus der transaktionalen Faktentabelle gilt es, den Geschäftsprozess derart zusammenzufassen, dass er in einer dem Endanwender verständlichen Form durch einen Datensatz repräsentiert werden kann und dabei analytischen Mehrwert generiert.

Real laufen Geschäftsprozesse häufig nicht zwingend geradlinig ab. Da gibt es Wiederholungen und Rücksprünge, oder

ganze Schritte können übersprungen werden. Hier benötigt es Regeln, die das entsprechende Verhalten in solchen Fällen definieren.

Neu entstehende Kennzahlen müssen fachlich korrekt benannt werden und relevante Informationen eingegrenzt werden, um die Breite der Faktentabelle zu begrenzen. Deshalb sind für den Projekterfolg die Begriffe, die Prioritäten und die Regeln für Ausnahmen zusammen mit dem Fachbereich festzulegen.

```

SELECT
  f.filiale,
  avg(lieferzeit) lieferzeit
FROM accum_snap s INNER JOIN filiale f
  ON s.fk_filiale = f.pk_sik_filiale
GROUP BY
  filiale

```

Listing 2

Updates auf Faktentabellen

Updates auf Fakten sind atypisch für Data-Warehouse-Umgebungen. Beim Accumulating Snapshot verfolgen wir zeitliche Entwicklungen. Die Information darüber, wann eine Lieferung tatsächlich erfolgen wird, besteht bei Auslösen einer Bestellung noch nicht. Sie kann erst nachgereicht werden, wenn sie angefallen ist. In der Folge muss jeder Datensatz in seinem Lebenszyklus mehrfach aktualisiert werden. Das bedeutet hohe Anforderungen an die ETL-Prozesse. Auf der anderen Seite werden in einem Accumulating

Snapshot mehrere transaktionale Datensätze zu einem gemeinsamen zusammengefasst, sodass die Satzanzahl erheblich reduziert wird.

Ist es fachlich gefordert, Kennzahlen in Form von Zeitdauern, den „Lags“, zu bilden, kann das zusätzlichen Aufwand bedeuten, wenn diese auch im gerade aktiven Prozessschritt zu führen sind. In diesem Fall ist periodisch bei allen aktiven Prozessen der Lag des aktuellen Prozessschritts zu inkrementieren, unabhängig vom Statusfortschritt. Somit erhöht sich die Menge an durchzuführenden Aktuali-

sierungen und damit die der Anforderungen an die ETL-Prozesse.

Historisierung

Besondere Sorgfalt ist bei der Konstruktion von Accumulating Snapshots mit Fremdschlüsseln zu Slowly Changing Dimensions Type 2 (SCD2) aufzuwenden. Hier gilt es, einen einheitlichen Zeitbezug für den Gesamtprozess zu definieren. Ist dieser Zeitbezug der aktuelle Stand, so empfiehlt es sich, mit (abgeleiteten) SCD1-Dimensionen zu arbeiten. Verwendet die Architektur einen historischen Surrogate Integer Key (SIK), müssten andernfalls im Accumulating Snapshot alle existierenden Referenzen auf den neuen Schlüssel in der SCD2-Dimension aktualisiert werden. Auch kann der SIK nicht aus dem Transaction Star übernommen werden, da dieser zu unterschiedlichen Zeitbezügen ermittelt wurde (*siehe Abbildungen 4 und 5*).

Eine valide Form ist die Zuweisung des SIK über das Prozess-Startdatum. Auch in dieser Version repräsentieren alle Fremd-



FK Bestellung	FK Filiale Vertrieb	FK Filiale Versand	FK Status	Bestelleingang	Versanddatum	Zahlungseingang	...
23	11	11	3	06.08.2014	07.08.2014	09.08.2014	...

Abbildung 5: Der Prozess-Start als Bezugsdatum

schlüssel im Datensatz den Stand zu einem spezifischen Datum. Durch SCD2 ausgelöste Updates entfallen jedoch. Dieses Modell hat wiederum eigene Implikationen, da jeder Prozess historisch in Bezug zu dem Prozess-Start wiedergegeben wird und dies auch fachlich korrekt sein muss.

Basiert die Architektur des Accumulating Snapshot hingegen auf dem Einsatz referenzierter Natural Keys, entfallen durch SCD2 ausgelöste Updates automatisch. Da pro Natural Key mehrere Datensätze in der Dimension vorhanden sein können, muss die Eindeutigkeit im Join mit dem Accumulating Snapshot über eine zusätzliche Datumsreferenz erreicht werden. Ansonsten besteht die Gefahr der Mehrfachzählung. Der Accumulating Snapshot besitzt mehr als eine Datumsreferenz. Das erlaubt mehrere Kombinationen technisch valider Umsetzungen. Das Festlegen erfolgt dann zusammen mit dem Fachbereich.

Für den Fall, dass die Prozess-Entwicklung selbst historisch nachvollziehbar sein muss, kann das bisher vorgestellte Design nicht die gewünschten Antworten liefern. Jede Zeile ist im Verlauf des Geschäftsprozesses mehrfach aktualisiert worden, die

Entwicklung des Datensatzes kann nicht mehr vollständig nachvollzogen werden. Ist der Accumulating Snapshot von einem Transaction Star abgeleitet, kann der historische Zustand jedoch aus diesem gebildet werden.

Soll die historische Entwicklung auch im Accumulating Snapshot ausgewertet werden, dann ist ein Ansatz die Anwendung von SCD2-Techniken auf die Faktentabelle – gleichbedeutend mit dem Hinzufügen von „gültig von“- und „gültig bis“-Informationen (Time Stamping). Statusänderungen führen zum Abschließen des alten Datensatzes und Einfügen des neuen Stands.

Offensichtlich enthält die Faktentabelle bei diesem Ansatz mehr Datensätze als ein Accumulating Snapshot, der auf Time Stamping verzichtet. Werten die meisten Analysen nur den aktuellen Zustand der Prozesse aus, dann ist die Partitionierung auf die „gültig bis“-Information eine zusätzlich zu evaluierende Möglichkeit. Damit befinden sich in der Default-Partition („MAXVALUE“ bei Partition by Range) die aktuell gültigen Stände, sodass durch den Optimizer bei entsprechenden Abfragen nur darauf zugegriffen wird.

Fazit

Transaktionale Faktentabellen komplexer Prozesse können häufig sinnvoll mit einem Accumulating Snapshot ergänzt werden. Bei Auswertungen sind sie eine große Hilfe, da komplexe SQL-Statements vermieden werden. Die notwendigen ETL-Prozesse sind allerdings sehr komplex. Mit ausreichender Abstimmung auf fachliche Fragestellungen kann jedoch ein hoher Nutzen generiert werden, der den Aufwand wert ist.



Alexander Voigt
alexander.voigt@areto-consulting.de

Zwischen 19. und 30. Januar 2015: Wahl zur DOAG-Delegiertenversammlung

Die Wahlen zur DOAG-Delegiertenversammlung stehen bevor. Vom 19. bis 30. Januar 2015 können wahlberechtigte DOAG-Mitglieder ihre Stimme abzugeben. Wie vor zwei Jahren finden die Wahlen elektronisch statt. Dies gab der DOAG-Wahl-ausschuss, vertreten durch Oliver Pyka, Andreas Ellerhoff und Reiner Bürger, am

16. Oktober 2014 bekannt. Wahlberechtigt sind nach der bestehenden Satzung natürliche Mitglieder, studentische Mitglieder, Ehrenmitglieder und korporative Mitglieder, vertreten durch ihre Repräsentanten. Alle Wahlberechtigten erhalten rechtzeitig eine entsprechende Wahlbenachrichtigung per E-Mail.

Für das zweijährige Amt des Delegierten kandidieren können nicht nur die Wahlberechtigten, sondern auch assoziierte Mitglieder. Damit sind alle Mitglieder der DOAG in der Delegiertenversammlung repräsentiert. Die Nominierung der Kandidaten fand vom 19. November bis 3. Dezember 2014 statt.