

UNIVERSITÄT SIEGEN

BACHELORARBEIT
IM FACH INFORMATIK

Transformation und Analyse von parallelen Datenbankprozessen

vorgelegt von

Marin BRESSEL

Betreuer:

Prof. Dr. M. LOCHAU,
Professur für Modellbasierte Entwicklung,
Universität Siegen

Eidesstattliche Erklärung

Ich versichere nach § 39 (1) der Einheitlichen Regelungen, meine Arbeit (bei einer Gruppenarbeit meinen entsprechend gekennzeichneten Anteil der Arbeit) selbstständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt sowie Zitate kenntlich gemacht zu haben.

Ort, Datum

Unterschrift

UNIVERSITÄT SIEGEN

Zusammenfassung

Transformation und Analyse von parallelen Datenbankprozessen

von Marin BRESSEL

[TODO: Deutsche Zusammenfassung]

UNIVERSITY OF SIEGEN

Abstract

Transformation and Analysis of Parallel Database Processes

by Marin BRESSEL

[TODO: English abstract]

Inhaltsverzeichnis

Abbildungsverzeichnis	xi
Tabellenverzeichnis	xiii
Listings	xv
Abkürzungsverzeichnis	xvii
1 Einleitung	1
2 Grundlagen	3
2.1 BPMN	3
2.1.1 Erste Schritte - Events und Aktivitäten	3
2.1.2 Mögliche Verzweigungen	5

Abbildungsverzeichnis

2.1	Einfache Task	4
2.2	Prozess Block	5
2.3	Intermediate Events	5
2.4	XOR-Gateway	6
2.5	Paralell-Gateway	6

Tabellenverzeichnis

Listings

Abkürzungsverzeichnis

LAH List Abbreviations Here
WSF What (it) Stands For

Kapitel 1

Einleitung

Im Rahmen dieser Arbeit wollen wir ermöglichen parallele Datenbankprozesse sowohl zu analysieren als auch zu Transformieren. Um zu verstehen was genau hier passieren soll und wie eine Mögliche Umsetzung aussehen könnte, müssen wir zunächst einige Begriffe definieren und deren Bedeutung verstehen. Da wir uns im generellen Bereich von Prozessen befinden, muss zunächst erläutert werden was unter einem Prozess zu verstehen ist und auf welche Art von Prozessen wir uns beschränken wollen. Bei Prozessen handelt es sich um sogenannte reaktive Systeme. Es sind also Systeme, welche auf eine Eingabe warten und je nach Zustand entsprechend reagieren. Hierfür sind in jedem erdenklichen Teil des Alltags Beispiele zu finden. Es kann sich um Komplexe Prozesse in Computerspielen oder Betriebssystemen handeln, oder auch um ganz einfache Abläufe im Alltag. Ein einfaches Beispiel würde zum Beispiel der Tagesablauf eines gewöhnlichen Angestellten in einer Firma bieten (Siehe Abbildung 1). Wir wollen uns in dieser Arbeit zum größten Teil mit Abläufen von Geschäftsprozessen auseinandersetzen. Damit sind unter anderem jene Abläufe gemeint, die innerhalb von beispielsweise Firmen passieren. Es kann sich hier um viele Verschiedene Abläufe handeln. Ein passendes Beispiel ist auf vielen erdenklichen Internetseiten, wie beispielsweise online-shopping Seiten zu finden. Wir beschränken uns hier der Einfachheit halber auf den Prozess des Erstellens eines Accounts. Also auf den Registrierungsprozess. In unserem Beispiel muss der Nutzer auf der Homepage zunächst den Button „Register“ drücken. Er wird im Anschluss auf die korrekte Internetseite weitergeleitet und hat da die Möglichkeit seine Anmeldedaten anzugeben. Nun muss die Korrektheit der Daten geprüft werden und der Nutzer wird entweder auf eine Bestätigungsseite oder auf eine Fehlerseite weitergeleitet. Sollte alles korrekt sein wird nun gleichzeitig der Nutzer angelegt und das Verfahren eingeleitet diesen Nutzer für beispielsweise einen Newsletter anzumelden. Im Anschluss wird der Nutzer dann angemeldet auf die Homepage zurückgeleitet (Siehe Abbildung 2). Solche und ähnliche Prozesse laufen in jeder erdenklichen Firma ab und unterscheiden sich sehr stark voneinander in Komplexität und Inhalt.

Bei dem zweiten Begriff, der verstanden werden muss handelt es sich um BPMN (Business Process Modeling Notation). Hierbei handelt es sich um eine Standardisierte Modellierungssprache. Die in den Abbildungen 1 und 2 dargestellten Beispiele wurden bereits in BPMN dargestellt. Abbildung 2 zeigt einige der Features, die von uns genutzt werden sollen. Eine Detailliertere Beschreibung erfolgt im Abschnitt „Grundlagen-BPMN“ doch die wichtigsten Features in dieser Abbildung sind, neben den Aktivitäten welche durch abgerundete Rechtecke dargestellt werden und der Konkatenation dieser Aktivitäten durch die Pfeile, die sogenannten Gateways welche durch Rauten mit einem Entsprechenden Symbol dargestellt werden. Ist ein „X“ in die Raute geschrieben so handelt es sich um ein Exklusives OR-Gateway. Hier kann exakt eine der folgenden Pfade ausgeführt werden. Findet man ein „+“

vor, so liegt ein paralleles Gateway vor. Hier werden Alle Pfade gleichzeitig ausgeführt. Ein weites paralleles Gateway vereinigt die Pfade dann wieder. An dieser Stelle wird also gewartet, bis alle parallelen Aktivitäten durchgeführt worden sind. Diese Arbeit beschäftigt sich ausschließlich mit Prozessen in BPMN, allerdings wird nur eine Teilmenge benutzt. Hierbei handelt es sich um die selbe wie jene in der Arbeit „Delta-BPMN: a Concrete Language and Verifier for Data-aware BPMN“

Zuletzt muss nun noch erläutert werden zu was diese Prozesse transformiert werden sollen. Hierzu müssen wir den Begriff einer Prozessalgebra genauer erläutern. Bei einer Prozessalgebra handelt es sich nun um einen mathematischen Kernkalkül zur Darstellung von Prozessen. Es bietet also eine Möglichkeit die selben Prozesse durch eine Art Formeln darzustellen. Wir wollen in dieser Arbeit eine Prozessalgebra verwenden, welche auf ACP basiert. Zum besseren Verständnis möchte ich auch hier die wichtigsten Features von ACP erläutern. Ähnlich zu BPMN gibt es Möglichkeiten zur Darstellung von Konkatenation, oder-Zweigen und auch Parallelen Zweigen. Eine Aktivität wird durch einfach als Variable dargestellt. Wie diese genannt wird, ist zunächst irrelevant. Eine Konkatenation wird durch ein „ $*$ “ dargestellt. Ein Exklusiver oder-Zweig wird durch ein „ $+$ “ und ein paralleler Zweig durch „ $||$ “ dargestellt. Würde man das Beispiel in Abbildung 2 also nun nach ACP überführen, so würde ein Mögliches Ergebnis wie folgt aussehen:

$P := \text{Startseite} * \text{DatenÜberprüfen} * (P + (\text{NutzerAnlegen} * (\text{NutzerEinloggen} || \text{NewsletterAnmelden})) * F$

Das Ziel dieser Arbeit ist es nun Programm zu entwickeln, welches ein BPMN Diagramm einlesen und dieses zu einer Formel in der Prozessalgebra überführen kann. Es soll also ein Mapping von BPMN nach ACP erfolgen.

Der Zweck dieser Arbeit wird erkennbar, sobald Daten in form einer relationalen Datenbank für den Prozess relevant werden. BPMN bietet hier nur schwammige und unübersichtliche Methoden die Daten darzustellen. Es erweist sich also als schwierig zu Analysieren und zu verifizieren, was genau mit den Daten passiert. Zudem lässt BPMN zu viel Raum für Interpretation. Es wird viel mit Kommentaren gearbeitet und was genau in einem bestimmten Abschnitt des Prozesses passiert ist in vielerlei Hinsicht offen zur Interpretation. ACP ist etwas Strukturierter. Zudem bietet die Variante die von uns genutzt wird die Möglichkeit mit einer konkreten Datenbank zu kommunizieren und Anweisungen auf dieser auszuführen. Aus diesem Grund ist ein Nebenziel dieser Arbeit, die passenden Datenbankanweisungen zu dem jeweiligen Prozess auf einer relationalen Datenbank auszuführen.

Kapitel 2

Grundlagen

Das folgende Kapitel dient dazu die in der Einleitung genannten Begriffe und Prinzipien genauer zu erläutern. Wir wollen also zunächst BPMN und die Bestandteile der Modellierungssprache genauer betrachten. Hierzu werden wir jedes Element einzeln betrachten dadurch stück für Stück ein passendes Beispiel erstellen. Dieses Beispiel wird auch im Rest der Arbeit Relevanz finden. Im Anschluss wird dann die verwendete Prozessalgebra genauer betrachtet. Auch hier wollen wir jeden Bestandteil des Modells im Detail erläutern.

2.1 BPMN

Wie in der Einleitung bereits erwähnt, dient BPMN der graphischen Darstellung von Business Prozessen. In BPMN werden diese Prozesse in einzelne Aktivitäten oder Aufgaben unterteilt und dann in der richtigen Reihenfolge aufgezeichnet. Es gibt unterschiedliche Möglichkeiten Verzweigungen, Abhängigkeiten und Ähnliches zu Modellieren doch zum größten Teil basiert alles auf der korrekten Aneinanderreihung dieser Aktivitäten.

2.1.1 Erste Schritte - Events und Aktivitäten

Im folgenden Abschnitt schauen wir uns also die von uns verwendete Teilmenge der Modellierungssprache BPMN an. Hierzu möchten wir zunächst einen einführenden Prozess als Beispiel betrachten. Für alle möglichen Homepages und Websites von unterschiedlichen Firmen und Anbietern können Konten erstellt werden. Diese dienen zur Wiedererkennung eines Kunden oder Mitarbeiters. In diesem Abschnitt wollen wir ein BPMN-Diagramm erstellen, welches den Registrierungsprozess auf einer solchen Website darstellen könnte. Wir werden die Modellierungssprache hierzu aufteilen und jeden Bestandteil der Sprache anhand des Beispiels einzeln erläutern so, dass wir am ende dieses Abschnittes ein erstes, vollständiges Diagramm vorfinden. Die von uns genutzte Teilmenge kann in einzelne Blöcke eingeteilt werden. Diese können in zwei Gruppen unterteilt werden. Die Basic Blocks und die Flow Blocks. Des Weiteren kann unterschieden werden zwischen „Leaf Blocks“ und „Nonleaf Blocks.“ Alle Nonleaf Blocks bestehen aus beliebig vielen Leaf Blocks. Es existieren genau zwei für uns relevante Leaf Blocks, welche wir hier zunächst erwähnen möchten, um deren Funktion genauer zu erläutern. Die unterschiedlichen Blöcke sind jeweils durch sogenannte „sequence flows“ verbunden. Diese werden dargestellt durch eine durchgezogene Linie mit einem ausgefüllten Pfeil am Ende. Sie verdeutlichen den Fluss des Diagrammes. Jeder Block hat einen eingehenden und einen ausgehenden Flow. Um zu verdeutlichen in welchem Zustand der Prozess sich zu einem bestimmten Zeitpunkt befindet, verwenden wir sogenannte Token. Diese beinhalten keine Daten, sondern stellen nur dar, welcher Teil des



ABBILDUNG 2.1: Einfache Task - Anmeldedaten eingeben

Prozesses gerade aufgeführt wird. Beim Ausführen des Prozesses wird ein Token immer in Richtung der Sequence Flows weitergegeben. Bei dem „Task Block“ und dem „Event Block“ handelt es sich um die beiden relevanten leaf-Blocks. Beide Blöcke werden durch den Namen bereits gut erklärt. Der Task Block, welcher durch ein abgerundetes Rechteck dargestellt wird, ist repräsentativ für eine beliebige Aufgabe. Diese Aufgaben benötigen in jedem Fall Zeit, um ausgeführt zu werden. Er kann durch einen Text in der Mitte des Rechtecks beliebig benannt werden. Diese Aufgaben können jede erdenkliche Form annehmen. Eine Mögliche Aufgabe ist in Abbildung 2.1 dargestellt. Es handelt sich um das Eingeben der Anmeldedaten. Es handelt sich hierbei um eine Aufgabe die Zeit beansprucht und vom Nutzer ausgeführt wird. Sie ist offensichtlich relevant für unser Beispiel.

Bei dem zweiten Leaf Block handelt es sich nun um sogenannte Events. Events werden durch einen Kreis dargestellt. Anders als die Aufgaben passieren Events sofort. Es gibt unterschiedliche Typen, welche durch unterschiedliche Variationen eines Kreises dargestellt werden. Für uns relevant sind allerdings nur sogenannte „catching Events.“ Erreicht der Token ein solches Event, wird gewartet bis das erwartete Event auftritt und erst dann wird der Token weitergeschickt. Auch unter den catching Events gibt es drei grundsätzliche Unterscheidungen. Die einfachsten Variationen sind sogenannte Startevents, welche einen Prozess Starten und durch einen einfachen dünn gezeichneten Kreis dargestellt werden und die Endevents, welche den Prozess terminieren und durch einen einfachen dick gezeichneten Kreis dargestellt werden. Durch die uns nun bekannten Bausteine, ist es uns möglich einen ersten Prozess aufzubauen und in BPMN zu Modellieren. In Abbildung 2.2 sehen wir ein Beispiel für den ersten und einfachsten „nonleaf-Block.“ Den „Prozess-Block.“ Dieser besteht aus einem Start- und einem Endevent. Zwischen den beiden Events liegt mindestens ein Task-Block und beliebig viele Event Blocks. Der Prozess in unserem Beispiel startet, indem ein Nutzer die Seite besucht und einen neuen Account anlegen möchte. Dieser wird zunächst auf eine Seite weitergeleitet, welche seine Anmeldedaten abfragt. Diese muss er dann in der ersten Aufgabe eingeben. Die zweite Aufgabe besteht nun darin den neuen Nutzer mit seinen eben angegebenen Daten in die vorhandene Datenbank einzutragen. Sobald diese Aufgabe erledigt ist, endet der Prozess und ein neuer Nutzer wurde erfolgreich angelegt.

Events können allerdings auch während eines Prozesses auftreten. Wollen wir in unserem Beispiel einführen, dass die Startseite besucht werden kann, ohne, dass direkt auf die Nutzer anlegen Seite weitergeleitet wird, so können wir eine neue Aufgabe und ein fangendes Event einführen. Abbildung 2.3 zeigt diesen Prozess. Er startet, indem ein Nutzer die von uns erstellte Website besucht. Im nächsten Schritt wartet das catching Event, bis der Nutzer sich dazu entscheidet einen neuen Account anzulegen. In Folge dessen verhält sich der Prozess so, wie im Beispiel aus Abbildung 2.2. Events welche während dem Prozess auftreten werden „Intermediate Events“ genannt.

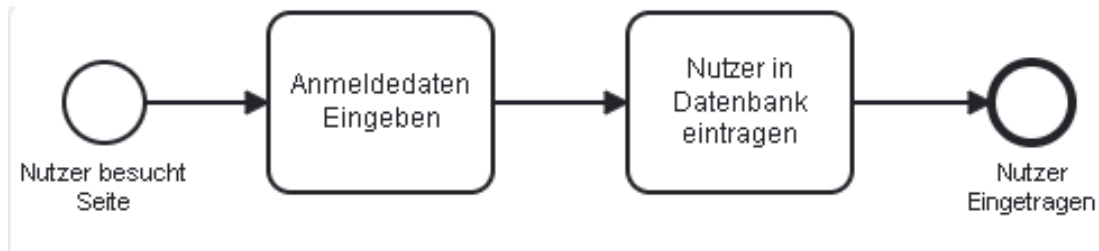


ABBILDUNG 2.2: Erster Vollständiger Prozessblock

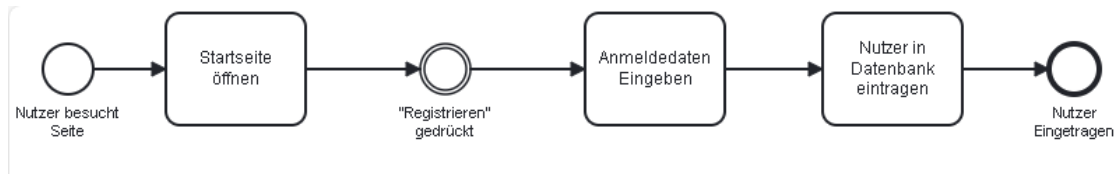


ABBILDUNG 2.3: Beispiel für ein intermediate Event

2.1.2 Mögliche Verzweigungen

In nun folgenden Abschnitt wollen wir uns mit Möglichkeiten beschäftigen den Sequenzfluss zu verzweigen. Hierzu bieten BPMN einige sogenannte „Gateways.“ Um Gateways darzustellen, werden Rauten verwendet. Unterschiedliche Gateways haben unterschiedliche Symbole in den Rauten eingezeichnet. Sie sind dafür da, den Verlauf des Prozesses aufzuteilen und wieder korrekt zusammenzufügen. Wenn ein Gateway den Verlauf des Prozesses aufspaltet, so muss immer ein zweites Gateway den Verlauf wieder zusammenfügen. Eine Ausnahme hierfür wäre, wenn der Prozess in einer Verzweigung durch ein Endevent terminiert. Um unser bislang erarbeitetes Beispiel etwas realitätsnaher zu gestalten, wollen wir eine neue Funktion einführen. Offensichtlich soll unsere Seite mehr Funktionen anbieten als neue Nutzer anzulegen. Die nächste logische Erweiterung ist eine Möglichkeit für bereits bestehende Kunden sich mit ihren vorhandenen Anmeldedaten einzuloggen. Hier spielt die erste Verzweigungsmöglichkeit eine Rolle. Da ein Nutzer immer entweder ein bestehendes Konto besitzt oder ein neues erstellen möchte, wird in jedem Durchlauf des Prozesses nur eine dieser Aktionen durchgeführt. Hier kann ein XOR-Gateway genutzt werden. Es ist zu erkennen durch ein X in der Raute. Es bietet die Möglichkeit zwischen unterschiedlichen Pfaden zu wählen. Im Beispiel aus Abbildung 2.4 sehen wir eine mögliche Verwendung für dieses Gateway. Nachdem der Nutzer auf der Startseite den entsprechenden Button betätigt, wird er entweder zur Anmeldung oder zur Registrierung weitergeleitet. In beiden Fällen muss er seine Daten angeben. Im Anschluss werden die zwei Äste wieder zusammengeführt und der Nutzer wird mit seinem zugehörigen Account auf der Seite angemeldet. Beim Zusammenführen der Äste, wartet das Gateway auf genau einen Token und gibt diesen dann weiter an das nächste Objekt. In diesem Beispiel ist der erste Flow-Block zu erkennen. Wird der Verlauf des Prozess durch ein XOR-Gateway aufgespalten und wieder zusammengeführt, so bilden alle Elemente innerhalb dieser Verzweigung im sogenannten Exclusive-choice-Block. Abbildung 2.4 zeigt außerdem eine mögliche Variante des XOR-Gateways. Wenn der Nutzer einen vorhandenen Account anmelden möchte, aber inkorrekte Anmeldedaten eingibt, so wird er auf die Startseite zurückgeleitet. Das XOR-Gateway kann also auch für Loops verwendet werden. Denselben Verlauf

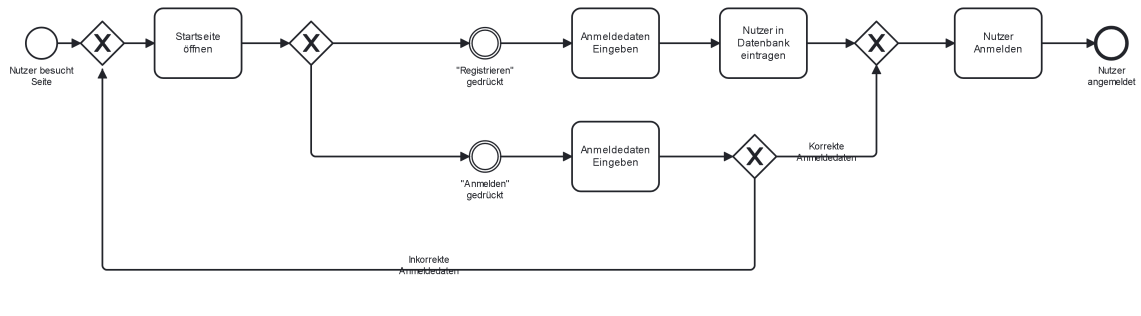


ABBILDUNG 2.4: Exclusive Entscheidung - Das XOR-Gateway

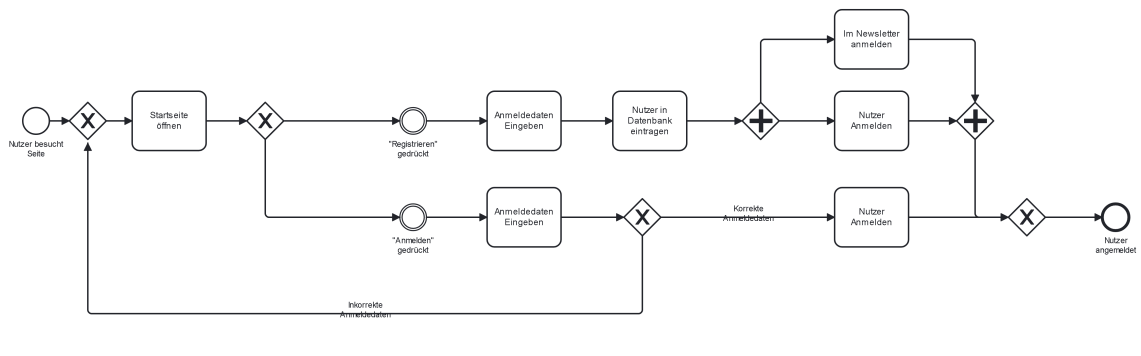


ABBILDUNG 2.5: Parallele Ausführung - Das Parallel-Gateway

sehen wir, falls der Nutzer bei der Registrierung invalide Daten angeben möchte. Auch hier wird der Nutzer auf die Startseite zurückgeleitet. Hier ist zu erkennen, dass das zusammenführende Gateway durchaus auch vor dem aufspaltenden Gateway liegen kann. Es ist zusätzlich möglich auch mehr als nur zwei Mögliche Ausgänge an das Gateway anzubinden.

Wir wollen nun ein weiteres Feature in unser Diagramm einfügen. Nachdem ein neuer Account erstellt wurde, soll dieser Nutzer nach wie vor angemeldet und auf die Homepage weitergeleitet werden. Zusätzlich soll er gleichzeitig auch für den Newsletter der Website eingetragen werden. Hierzu können wir ein weiteres Gateway nutzen. Das parallele Gateway wird in Abbildung 2.5 das erste Mal gezeigt. Es wird ebenfalls durch eine Raute dargestellt. Diese enthält allerdings ein + in ihrem Inneren. An diesem Gateway wird der Verlauf des Prozesses wieder aufgespalten. Anstelle von nur einem Strang werden hier aber alle gleichzeitig ausgeführt. Das zusammenführende Gateway muss deshalb nicht blind den ersten Token weiter-schicken, den es erhält, sondern wartet bis an allen Eingängen ein Token vorliegt, fügt diese wieder zusammen und gibt dann den Token weiter an das nächste Objekt. Hier können Schwierigkeiten auftreten, falls ein paralleler Zweig in einem Endevent endet. Bei der Modellierung eines Diagrammes muss dies also verhindert werden.