

Modulbezeichnung: Datenstromsysteme und Data Warehousing (DSSDW) **5 ECTS**
(Datastream Systems and Data Warehousing)

Modulverantwortliche/r: Klaus Meyer-Wegener

Lehrende: Klaus Meyer-Wegener, Thomas Ruf

Startsemester: SS 2018

Dauer: 1 Semester

Turnus: jährlich (SS)

Präsenzzeit: 60 Std.

Eigenstudium: 90 Std.

Sprache: Deutsch

Lehrveranstaltungen:

Ausschlussbedingung: Dieses Modul darf nur abgelegt werden, wenn keine der im Modul enthaltenen Lehrveranstaltungen auch noch in einem anderen Modul enthalten ist, das bereits abgelegt wurde.

Datenstromsysteme (SS 2018, Vorlesung, 2 SWS, Klaus Meyer-Wegener)

Data Warehousing (SS 2018, Vorlesung, 2 SWS, Thomas Ruf)

Inhalt:

DSS:

Datenströme gibt es schon lange. Sie fließen als Nachrichten durch Netze, werden in Protokolldateien abgelegt (und dort oft vergessen oder nur bei Problemen im nachhinein konsultiert), werden zur Anzeige gebracht, gefiltert, ausgewertet, transformiert usw. Beispiele sind die unterschiedlichsten Arten von Messwerten (Temperatur, Helligkeit, Feuchtigkeit, Druck, chemische Zusammensetzung), Börsen- und Wirtschaftsnachrichten (Preise), Gebote in Online-Auktionen, Aktionen in Rechnersystemen (Anmeldungen, Zugriffe, Programmausführungen), um nur ein paar zu nennen.

Bislang wurden die Systeme zur Verarbeitung dieser Datenströme als Individuallösungen erstellt, oft sogar auf sehr speziellen Plattformen (SPS). Um das Jahr 2000 herum begann sich auch die Datenbank-Forschung mit diesem Problem zu befassen, weil man erkannt hatte, dass sich eine Reihe von Techniken dieses Gebiets auch auf Datenströme anwenden lassen, vor allem die Anfrageverarbeitung und -optimierung, aber auch Verteilung und Datenunabhängigkeit. Damit lassen sich auch für die Verarbeitung von Datenströmen generische Lösungen erstellen - so wie Datenbanksysteme eine generische Lösung für die Auswertung gespeicherter Daten sind.

Die nun entstehenden Datenstromsysteme werden in dieser Vorlesung vorgestellt. Diskutiert werden dabei u.a.

- Anfragesprachen
- Schemastrukturen
- Sensornetze als eine der wichtigsten Datenquellen
- Anfrageauswertung und -optimierung
- Architekturen
- Anwendungsszenarien

Diese Lehrveranstaltung wurde von Prof. Meyer-Wegener und Dr. Daum erstellt und im Sommersemester 2007 erstmalig angeboten.

DW:

Das "Data Warehouse" ist sowohl im theoretischen Bereich der Datenbankforschung als auch in der praktischen Anwendung in der Wirtschaft ein wichtiges Thema. Die Hauptaspekte der Vorlesung liegen dabei in der Diskussion der unterschiedlichen Architekturansätze, der zugehörigen Datenmodelle und der verarbeitungstechnischen Grundlagen. Weiterhin wird auf aktuelle Techniken wie ROLAP und MOLAP oder Aggregatbildung, -verwendung und -haltung eingegangen. Die Vorlesung ist geprägt von einer Mischung aus theoretischen Grundlagen und Anwendungen neuer Forschungsergebnisse. Dabei berichten Praktiker aus Unternehmen zum einen von ihren Erfahrungen bei dem Aufbau eines "Warehouses" und stellen zum anderen den Stand der Technik in einschlägigen Data-Warehouse-Systemen dar.

Lernziele und Kompetenzen:

DSS:

Die Studierenden:

- beschreiben Datenstromsysteme als neues Systemkonzept;
- grenzen Datenstromsysteme von Datenbanksystemen ab;

- erklären ein Datenstrommanagementsystem;
- grenzen Datenstromverarbeitung von Complex-Event Processing ab;
- charakterisieren Datenstromquellen und ihre Anbindung an das System;
- unterscheiden Zeitmodelle für Datenströme;
- zählen die wichtigsten Datenstrom-Operatoren auf;
- trennen blockierende von nicht blockierenden Operatoren;
- unterscheiden die verschiedenen Arten von Fensterdefinitionen;
- benennen den Ansatz und die Vorgehensweise des DSAM-Projekts;
- charakterisieren kurz die Datenstromsysteme Aurora, STREAM, Cayuga, Esper und PIPES;
- erläutern den Umgang mit Abweichungen von der Ordnung in Aurora;
- erklären die Operatoren NEXT und FOLD in Cayuga;
- verstehen die drei Datenstrommodelle von PIPES;
- erklären den Begriff "Schnappschuss-Reduzierbarkeit";
- beschreiben die unterschiedlichen Semantiken von zeitbasierten Fenstern anhand des SECRET-Modells;
- können Techniken der lokalen und verteilten Optimierung von Datenstromanfragen nennen;
- erklären Eddies als Mittel der dynamischen Optimierung;
- zählen verschiedene Anwendungsszenarien von Datenstromsystemen auf;
- formulieren Anfragen in einer Anfragesprache für Datenstromsysteme;
- überprüfen Aufgabenstellungen bei der Auswertung von Daten daraufhin, ob sie mit Hilfe eines Datenstromsystems gelöst werden können.

DW:

Die Studierenden

- kennen verschiedene Anwendungsgebiete des Data Warehousing und charakterisieren diese
- modellieren multidimensionale Datenbanken
- entwickeln ein Grundverständnis für die Abbildung multidimensionaler Datenstrukturen in Datenbanken und bewerten die Vor- und Nachteile der verschiedenen Ansätze in konkreten Anwendungssituationen
- setzen grundlegende Verfahren der Datenauswertung in Data Warehouse-Systemen für konkrete Anwendungssituationen ein
- erläutern die Architektur und den Betrieb technischer Data Warehouse-Systeme und erläutern die sich hieraus ergebenden Implikationen im Betrieb solcher Systeme
- erklären die von Praxisvertretern vorgestellten Einsatzszenarien und Systemansätzen für Data Warehouse-Systeme
- schätzen aktuelle Entwicklungen im Themengebiet ein

Literatur:

DSS:

Die Literatur zu Datenstromsystemen ist derzeit noch sehr forschungsorientiert; es sind einzelne Aspekte in Artikeln dargestellt, aber es gibt nur wenige Übersichtsarbeiten und schon gar kein Standardwerk. Insofern ist die folgende Liste zwangsläufig unvollständig und laufenden Änderungen unterworfen.

- GOLAB, Lukasz ; ÖZSU, M. Tamer: Data Stream Management. Milton Keynes UK : Morgan & Claypool Publishers, 2010 (Synthesis Lectures on Data Management 5). - ISBN 9781608452729. - 14Gl mat 17.6.5-126
- Aggarwal, Charu C. (Hrsg.): *Data Streams : Models and Algorithms*. Boston/Dordrecht/London : Kluwer Academic Publishers, 2006
- Motwani, Rajeev ; Widom, Jennifer, u.a.: Query Processing, Resource Management, and Approximation in a Data Stream Management System. In: *Proc. 1st Biennial Conf. on Innovative Data Systems Research (CIDR, Asilomar, CA, USA, January 5-8, 2003)*. - Online Proceedings
- Golab, Lukasz ; Özsü, M. Tamer: Data Stream Management Issues - A Survey. Technical Report CS-2003-08, Waterloo 2003
- Golab, Lukasz ; Özsü, M. Tamer: Issues in Data Stream Management. *ACM SIGMOD Record* 32 (June 2003) issue 2, pp. 5-14
- Arasu, Arvind ; Babu, Shivnath ; Widom, Jennifer: The CQL Continuous Query Language : Semantic Foundations and Query Execution. *VLDB Journal* 15 (June 2006) no. 2, pp. 121-142

- Bry, Francois ; Furche, Tim ; Olteanu, Dan: Datenströme. Technical Report PMS-FB-2004-2, München 2004
- Koudas, Nick ; Srivastava, Divesh: Data Stream Query Processing. In: *Proc. VLDB 2003*

Verwendbarkeit des Moduls / Einpassung in den Musterstudienplan:

Das Modul ist im Kontext der folgenden Studienfächer/Vertiefungsrichtungen verwendbar:

[1] Informatik (Master of Science)

(Po-Vers. 2010 | TechFak | Informatik (Master of Science) | Wahlpflichtbereich | Säule der softwareorientierten Vertiefungsrichtungen | Vertiefungsrichtung Datenbanksysteme)

Dieses Modul ist daneben auch in den Studienfächern "Informatik (1. Staatsprüfung für das Lehramt an Gymnasien)", "Informations- und Kommunikationstechnik (Master of Science)", "Mathematik (Bachelor of Science)" verwendbar.

Studien-/Prüfungsleistungen:

Datenstromsysteme und Data Warehousing (Prüfungsnummer: 357826)

(englische Bezeichnung: Datastream Systems and Data Warehousing)

Prüfungsleistung, mehrteilige Prüfung

Anteil an der Berechnung der Modulnote: 100%

weitere Erläuterungen:

Die Bewertung der Prüfungsleistung setzt sich zusammen aus 1/2 Note der Klausur in Data Warehousing (von etwa 60 Minuten Dauer) und 1/2 Note der mündlichen Prüfung in DSS (von etwa 30 Minuten Dauer). Zum Bestehen der Modulprüfungen müssen beiden Teilprüfungen bestanden werden.

Erstablingung: SS 2018, 1. Wdh.: WS 2018/2019

1. Prüfer: Meyer-Wegener/Ruf (T10144)
