



News

Online publiziert: 25. Februar 2021

© Gesellschaft für Informatik e.V. and Springer-Verlag GmbH Germany, part of Springer Nature 2021

1 Auswirkungen von Covid-19 auf Veranstaltungen im Jahr 2021

Auch im Jahr 2021 beeinflusst Corona zahlreiche wissenschaftliche Veranstaltungen. Wie bereits in der letzten Ausgabe berichtet wird die BTW-Konferenz in den Herbst verlegt. Aus organisatorischen Gründen hat sich allerdings das Datum gegenüber den ersten Ankündigungen geändert. Die Konferenz wird jetzt vom *13. bis 17. September 2021* wie vorgesehen in Dresden stattfinden.

Zahlreiche andere für die DBIS-Community relevante Konferenzen finden ausschließlich digital statt, so zum Beispiel die EDBT (23.–26. März), die ECIR (28. März bis 1. April), die ICDE (19.–22. April) und die SIGIR (11.–15. Juli). Die SIGMOD wurde bei Redaktionsschluss noch als klassische Konferenz vor Ort in Xi'an (China) geplant. Für die VLDB ist eine hybride Veranstaltung geplant, die eine Veranstaltung in Kopenhagen mit einer starken Onlinekomponente kombiniert.

2 Workshop Grundlagen von Datenbanken 2021

Nachdem im letzten Jahr der Grundlagenworkshop Corona-bedingt ausfallen musste, plant der GI-Arbeitskreis „Grundlagen von Informationssystemen“ für dieses Jahr einen Workshop im Kloster Nimbschen in der Nähe von Leipzig. Die Veranstaltung wird von Andreas Thor (HTWK Leipzig) organisiert und soll vom 25. bis 28. Mai 2021 stattfinden.

Der Workshop soll die Kommunikation zwischen Wissenschaftlern/-innen im deutschsprachigen Raum fördern, die sich grundlagenorientiert mit Datenbanken und Informationssystemen beschäftigen. Er ist insbesondere als Forum für Nachwuchswissenschaftler/-innen gedacht, die ihre aktuellen Arbeiten in einem größeren Forum vorstellen wollen.

Auch hier wird man auf Basis der aktuellen Coronasituation entscheiden müssen, ob der Workshop in der ge-

planten Form stattfinden kann. Verfolgen Sie dazu bitte die Homepage der Veranstaltung: <https://fdit.htwk-leipzig.de/fakultaet/news-events/gvdb21/>

3 1. Workshop Conceptual Modeling for NoSQL Data Stores (CoMoNoS) auf der ER 2020

Auf der ER-Konferenz 2020 fand der 1. Workshop zum Thema *Conceptual Modeling for NoSQL Data Stores* (CoMoNoS) statt, organisiert von Meike Klettke (Universität Rostock), Stefanie Scherzinger (Universität Passau) und Uta Störl (Hochschule Darmstadt). Zunächst für Wien geplant, fanden die ER-Konferenz und alle assoziierten Workshops im November 2020 virtuell statt.

Thematischer Schwerpunkt des Workshops waren Methoden zur konzeptionellen Modellierung für NoSQL-Datenbanksysteme (wie MongoDB, Couchbase, Cassandra oder Neo4J). Ziel war es dabei, wissenschaftliche Arbeiten und den Blickwinkel der Industrie, Probleme aus der Entwicklung realer Anwendungen und den Stand der Technik zusammenzubringen.

Den Auftakt machte der eingeladene Vortrag von Pascal Desmarests (Gründer der Firma *Hackolade*) mit einem umfassenden und sehr interessanten Überblick zum Stand der Technik in kommerziellen Produkten sowie aktuellen Problemstellungen aus der Praxis. Thematische Schwerpunkte der wissenschaftlichen Beiträge waren Methoden des Schema-Reverse-Engineerings, sowie Ansätze zur modellbasierten NoSQL-Datengenerierung.

Ein weiteres Ziel des Workshops war es, den in diesem Themenfeld arbeitenden Vertreter:innen aus Forschung und Industrie ein gemeinsames Forum zu bieten. Den angeregten Diskussionen nach zu urteilen, ist dies den Organisatorinnen gelungen. Die aktive Beteiligung am Workshop lässt in diesem Themenbereich in den nächsten Jahren weitere Entwicklungen erwarten. Daher ist die Durchführung eines Folge-Workshops im Jahr 2021 bereits geplant.

Die Beiträge des 1. CoMoNoS Workshops finden sich auf der Webseite: <https://sites.google.com/view/comonos20/>. Der Workshop wurde durch die Deutsche Forschungsgemeinschaft (DFG) – 385808805 gefördert.

4 Produkt-News

Uta Störl

4.1 Neo4j 4.2

Seit November 2020 steht *Neo4j 4.2* zur Verfügung. Das neue Release bringt Verbesserungen rund um Performance und Bedienbarkeit. So wurde die durchsatzoptimierte Abfrageausführung auf alle Leseoperationen erweitert, die Abfrageplanung für komplexe Cypher-Anfragen mit mehreren MATCH-Klauseln verbessert und die Geschwindigkeit des Parsers um den Faktor 10 erhöht. Neo4j 4.2 vereinfacht feingranulares Backup und Recovery, die Portierbarkeit von Zugriffsrechten und die Lastenverteilung. Bei geshardeten Graphen erlaubt Neo4js Fabric-Architektur nun auch, Teilgraphen anhand von Knotenlabels und Kanteneigenschaften zwischen Maschinen zu migrieren.

Neo4j, <https://neo4j.com/blog/introducing-neo4j-4-2/>

4.2 Oracle Database 21c verfügbar

Oracle Database 21c, das neue Release der Oracle Datenbank, steht seit Dezember 2020 in der Oracle Cloud auf Always Free Autonomous Database und im Oracle Database Cloud Service zur Verfügung. Weitere Freigaben für Cloud und auch für On-Premises-Installationen sind 2021 eingeplant.

Bei Oracle Database 21c handelt es sich im Gegensatz zu dem Long-Term Release 19c um ein Release mit vielen neuen Funktionalitäten. Alle Neuigkeiten und Erweiterungen aus dem Oracle Database 20c Preview wie JSON-Datentyp, Native Blockchain Tables, SQL-Macros etc. sind auch in Database 21c eingeflossen. Darüber hinaus gibt es weitere Features und Funktionen im Umfeld von Converged Database, Multitenant, Multi Workflow, Security und Hochverfügbarkeits.

Oracle, <https://blogs.oracle.com/coretec/oracle-database-21c-ist-da>

4.3 IBM Db2 11.5.5

Die neue Version 11.5.5 von *IBM Db2* bietet Erweiterungen im Bereich Hochverfügbarkeit insbesondere durch die vermehrte Nutzung von Pacemaker sowie Verbesserungen bezüglich Datenkomprimierung, Datenföderation und Log-

ging. Die Verwendung von Machine Learning durch den Optimierer wurde weiter ausgebaut.

IBM, <https://www.ibm.com/de-de/products/db2-database>

4.4 Übernahme von erwin durch Quest

Quest Software hat im Januar 2021 die erwin, Inc. übernommen. Erwin bietet neben dem bekannten erwin Data Modeler auch eine Data Intelligence Suite für das Metadatenmanagement und mit der erwin Evolve Suite, Möglichkeiten zur Business-Process-Modellierung an. Der erwin Data Modeler bietet seit einiger Zeit nicht nur Unterstützung für relationale, sondern auch für NoSQL Datenbanksysteme.

Erwin, <https://erwin.com/>

4.5 SAP HANA Cloud

SAP hat sein Database as a Service (DBaaS) Angebot *SAP HANA Cloud* erweitert. Inzwischen sind maximal 4TB verfügbar, die der Kunde mit entsprechenden Self Services selbständig skalieren kann. Dazu kommt die erweiterte Unterstützung für die SAP HANA Native Storage Extension und der verknüpfte SAP HANA Data Lake. SAP HANA Cloud ermöglicht den Zugriff auf Parquet Files und die Nutzung des HANA DocStores. Die Anonymisierungsfunktion unterstützt nun auch Remote-Sourcen im Real-Time-Zugriff, und auch X.509-basiertes Zertifikatshandling ist nun verfügbar. Im Bereich View-Modellierung gibt es Unterstützung für non-equi-join-Ausdrücke innerhalb von Joinkriterien sowie verbesserte Performance durch erweiterte Caching-Funktionen. Insgesamt wurden die Analysefähigkeiten durch die Unterstützung der PAL/APL-Engine erweitert. Hier wurden auch neue Algorithmen im Bereich Time Series und ein Long-Short-Term-Memory zur Verfügung gestellt. Besonders interessant für Data-Scientisten ist die Einbettung von HANA-Spatial und HANA-Graphalgorithmen in den Python ML Client. Dazu gibt es jetzt auch die ESRI-Zertifizierung für die SAP HANA Cloud. Für Kunden, die von früheren Cloudangeboten wechseln wollen, gibt es nun erste Migrationsunterstützung.

SAP, <https://www.sap.com/germany/products/hana/cloud.html>

4.6 Netezza auf Microsoft Azure

Die *Netezza-Technologie* für skalierbare Data Warehouses wird aus der Historie heraus vorwiegend mit On-Premises-Lösungen in Verbindung gebracht. IBM hat sich jedoch bereits seit geraumer Zeit als Ziel gesetzt, diese auch für den Cloudmarkt zur Verfügung zu stellen und hatte mit Netezza in der IBM- und AWS-Cloud den Anfang gemacht. Seit Ende letzten Jahres gibt es nun die nächste Form der

Erweiterung als *Netezza on Azure Cloud*. Als Basis hierfür dient wiederum Red Hat Openshift.

IBM, <https://www.ibm.com/>

4.7 Oracle Exadata Cloud at Customer

Exadata Cloud at Customer (ExaCC) ist die Bereitstellung von Exadata Infrastruktur als Cloud Infrastrukturservice im kundeneigenen Rechenzentrum. Im Announcement Anfang Juli wurden die Neuerungen und Erweiterungen der aktuellen ExaCC vorgestellt: Der mit der Exadata X8M-2 vorgestellte Architekturansatz basiert auf RDMA over Converged Ethernet (RoCE), Persistent Memory (PMEM) und KVM-Virtualisierung. Er steht nun auch für die Exadata Cloud at Customer zur Verfügung. Alle Exadata-Schlüsselfunktionen und die neuen Datenbankfunktionen können auf dem System mit dem aktuellen Softwarestand genutzt werden. Die Verwendung von KVM als Virtualisierungslayer erlaubt es, mehrere virtuelle Exadata Database Cluster parallel auf einer ExaCC X8M zu betreiben. Dadurch wird die Isolation unterschiedlicher Datenbanksysteme verbessert und ein Betrieb für unterschiedliche Mandanten ermöglicht.

Eine kürzlich abgeschlossene PCI-DSS-Compliance Zertifizierung der Exadata Cloud at Customer X8M erlaubt deren Einsatz für Unternehmen der Finanzbranche, die häufig eine solche Zertifizierung zwingend voraussetzen. Die Maximum Availability Architecture (MAA) ist ein seit Jahren bewährtes und stets an aktuelle Anforderungen angepasstes Konzept zur Realisierung von Hochverfügbarkeits-Architekturen. Eine wesentliche Komponente darin ist Oracle Data Guard, die Spiegelung einer Datenbank auf einen entfernten Standort. Mit Exadata Cloud at Customer kann Oracle Data Guard eingesetzt werden, um eine MAA zu betreiben

Oracle, <https://www.oracle.com/>

4.8 Cloud Pak for Data 3.5

Die Openshift-basierte Plattform *Enterprise Insights* von IBM – *Cloud Pak for Data* – ist jetzt in der Version 3.5 erschienen. Linux auf System z wird jetzt neben Linux auf Power- und Intel-Systemen auch als Plattform unterstützt. Es gibt sowohl viele Verbesserungen bei bestehenden Komponenten, etwa beim Watson Knowledge Catalog, bei Watson Machine Learning oder bei Watson Openscale, als auch neue Komponenten. Dazu gehört zum Beispiel die Db2 Data Management Console, die nicht nur zur Administration von Db2 (Warehouse), sondern auch zur Administration von Big SQL und Data Virtualization genutzt werden kann. Eine weitere neue Komponente ist der Watson Machine Learning Accelerator, der insbesondere bei großen Deep-Learn-

ing-Modellen hilft. Zu diesen neuen Komponenten gehören auch Open Pages und Open Data for Industries.

IBM, <https://www.ibm.com/>

4.9 Graph Data Science 1.4.1

Seit Dezember 2020 steht *Neo4j Graph Data Science (GDS)* in Version 1.4.1 zur Verfügung. GDS 1.4.1 ermöglicht graph-natives Machine Learning (ML) direkt in Neo4j. Highlights des Release bilden die drei Graph-Embedding-Algorithmen Node2Vec, FastRP und GraphSAGE sowie ein ML-Modellkatalog. Damit lassen sich nun ML-Repräsentationen eines Neo4j-Graphen anlernen und verwalten. Graph Embeddings bilden die Topologie und Merkmale eines Graphen auf Vektoren fester Länge ab. Diese kompakte ML-Repräsentation eines Graphs lässt sich dann weiterführend für Analysen und Vorhersagen der Merkmale unbekannter oder neuer Teile des Graphs nutzen. GDS 1.4.1 ist voll kompatibel mit Neo4j 4.2.

Neo4j, <https://neo4j.com/aura/>

4.10 Neo4j Connector for Apache Spark

Seit November 2020 steht *Neo4j Connector for Apache Spark* zur Verfügung. Der Neo4j Connector for Apache Spark ist ein Integrationswerkzeug, das Neo4j mit Apache Spark verbindet und erlaubt, Daten bidirektional zwischen Neo4j und Apache Spark zu bewegen und zu transformieren. Der Connector implementiert die DataSource-API von Apache Spark und exponiert Neo4j als Lesequelle und Schreibziel für Apache Sparks DataFrames. Mit Hilfe des Connectors haben Neo4j-Nutzer nun Zugriff auf jegliche Daten und Systeme, die an das Spark-Ecosystem angebunden sind.

Neo4j, <https://neo4j.com/aura/>

4.11 Oracle: SQL-basierte Schnittstellen zum Contact Tracing

In Zeiten der COVID-19-Pandemie ist es essentiell zu verstehen, wann und wo Menschen sich durch zu enge und zu lange direkte Kontakte potentiell der Gefahr einer Ansteckung ausgesetzt haben. Um die betreffenden Personen auf das Risiko hinweisen zu können, bedarf es der raum-zeitlichen Analyse freiwillig geteilter (anonymisierter) Daten. Für eine solche Analyse werden die Ortsinformationen und Zeitstempel in einer Vielzahl von Datensätzen paarweise übereinandergelegt und verglichen. Treffer, also potentielle Risikokontakte, sind dann solche, bei denen sich 2 Personen eine definierte Mindestdauer ununterbrochen jeweils am gleichen Ort befunden haben.

Um diese Analysen auf der Basis vorhandener Daten zu vereinfachen und gleichzeitig die performante Analyse sehr großer Datenmengen zu ermöglichen, gibt es in der Oracle Datenbank zwei neue SQL-Schnittstellen: „Swipe IN/OUT Contact Tracing“ (API 1) und „GPS Tracking Contact Tracing“ (API 2). API 1 benötigt dabei für die Analyse keine direkte Ortsinformation in Form von GPS-Daten. Die Ortsinformation wird abgeleitet von einem konkreten Raum oder Gebäude, in dem sich eine Person mit anderen zur gleichen Zeit befunden hat.

Die Daten und insbesondere die Ergebnisse der SQL-basierten Analysen können extrem einfach über eine Webanwendung, z. B. mittels Application Express, oder auch einer Reporting-Umgebung zugänglich gemacht werden. Paarweise Kontakte lassen sich zudem zusätzlich in Form von Graphen abbilden. So können erweiterte Analysen poten-

tielle Infektionscluster oder besonders riskante Ereignisse aufdecken.

Auch wenn der sehr akute Anwendungsfall der COVID-19-Pandemie das auslösende Moment für diese Schnittstellen gewesen ist, so lassen sich doch damit Lösungen für eine Vielzahl anderer Anwendungsfälle mit geringem Aufwand bauen. Das sind all jene Anwendungsfälle, bei denen die raum-zeitliche Koexistenz von Personen und/oder Dingen eine wichtige Rolle spielt.

Oracle, <https://blogs.oracle.com/oraclespatial/contact-tracing-apis-in-oracle-database>

Dank an Jörg Latza (SAP), Karin Patenge, Ulrike Schwinn (beide Oracle), Hannes Voigt (Neo4j) sowie Andreas Christian, Stefan Lindenmeyer und Andreas Weininger (alle IBM) für ihren fachlichen Input.