



News

Online publiziert: 26. Oktober 2020

© Gesellschaft für Informatik e.V. and Springer-Verlag GmbH Germany, part of Springer Nature 2020

1 Verschiebung der BTW 2021

Die zu erwartende Situation rund um Corona/Covid-19 und damit verbunden das notwendige Hygienekonzept erlaubt es nicht, die BTW-Konferenz wie bisher vorgesehen als Präsenzveranstaltung im März 2021 in Dresden durchzuführen. Das Leitungsgremium des GI-Fachbereichs Datenbanken und Informationssysteme hat daher beschlossen, die BTW 2021 in die Woche vom *20. bis 24. September 2021* zu verlegen. Sie wird dann wie vorgesehen in Dresden stattfinden.

Die Verlegung hat keine Auswirkung auf den Publikationsprozess der wissenschaftlichen Beiträge, die im Frühjahr wie geplant als Konferenzband veröffentlicht werden – nur die Präsentation und Diskussion der Beiträge wird in den Herbst verschoben. Für alle weiteren Termine wird es entsprechende Verschiebungen geben, die auf der Website der BTW 2021 kommuniziert werden: <http://btw-konferenz.de/2021>.

2 Data-Science-Challenge bei der BTW 2021

Passend zum Austragungsort Dresden ist das Thema der Data-Science-Challenge bei der BTW 2021 die Vorhersage von Leistungsdaten im Kontext der Halbleiterproduktion. Der vorgegebene Trainingsdatensatz beinhaltet Leistungsdaten einzelner Messeinrichtungen an einer Produktionsanlage für ein Jahr sowie die dazugehörigen Werkstücke. Die Leistungsdaten werden unabhängig von den Anlagen gemessen und stehen entsprechend in einer anderen zeitlichen Auflösung zur Verfügung. Zuerst müssen die Leistungsdaten mit dem korrekten Werkstück verbunden werden. Auf Basis der kombinierten Daten soll ein Modell (oder ein Ensemble) erstellt werden, welches für gegebene Produktkonfigurationen die Leistungsdaten für alle Anlagen vorhersagt.

Die Bewertung erfolgt in zwei Phasen. In der ersten, öffentlichen Phase findet die Bewertung ausschließlich auf Basis eines Leaderboards statt. Dazu wird nur die Genauigkeit des Modells bewertet. Auf der Grundlage eines Testda-

tensatzes soll die benötigte Leistung vorhergesagt werden. Die Top-5 Teams zum Zeitpunkt des Leaderboard Freezes werden auf Basis einer schriftlichen Ausarbeitung durch eine Fachjury in den Punkten Erklärbarkeit und Anwendbarkeit des Modells in der Praxis bewertet. Für die drei besten Teams sind Geldpreise vorgesehen. Folgender Ablauf ist vorgesehen:

- 01.02.2021: Start der Data Science Challenge
- 28.06.2021: Leaderboard Freeze
- 26.07.2021: Einreichung der Beschreibung der Ansätze und des Codes der Top-5 Teams
- 21.09.2021: Präsentation der Ergebnisse auf der BTW 2021 in Dresden
- 22.09.2021: Bekanntgabe der Gewinner und Preisverleihung

Weitere Informationen zur Data-Science-Challenge finden Sie unter <https://sites.google.com/view/btw-2021-tud/data-science-challenge>

3 Virtuelle LWDA 2020

Aufgrund der Corona-Situation konnte die Konferenz „Lernen, Wissen, Daten, Adaption“ (LWDA) 2020 nicht wie geplant in Bonn stattfinden, sondern wurde als reine Onlineveranstaltung vom 9. bis 11. September 2020 durchgeführt. Das Organisationsteam um Daniel Trabold (Fraunhofer-Institut für Intelligente Analyse- und Informationssysteme) hatte zahlreiche Präsentations- und Interaktionstools zusammengestellt, so dass die LWDA trotz der schwierigen Bedingungen ihrer Rolle als Kommunikationsplattform für die teilnehmenden Fachgruppen gerecht werden konnte.

Auf dem Programm der mittlerweile 22. Konferenz dieser Reihe standen parallele Workshops der GI-Fachgruppen Datenbanken (DB), Knowledge Discovery, Data Mining und Maschinelles Lernen (KDML), Wissensmanagement (WM) und Business Intelligence. Für die Teilnehmer bot sich so eine willkommene Gelegenheit, Arbeiten auch über die Grenzen der eigenen Fachgruppe hinaus zu prä-

sentieren. Um die Posterausstellung interaktiv gestalten zu können, war für jedes Poster ein eigener virtueller Diskussionsraum vorbereitet, was überraschend gut funktioniert hat.

Insgesamt drei Keynotes rundeten das Programm ab. Geoff Webb (Monash University) stellte skalierbare Ansätze zur Zeitreihenklassifikation vor. Kristian Kersting (TU Darmstadt) präsentierte eine neue Sicht auf maschinelles Lernen und Künstliche Intelligenz, die er „Systems AI“ nannte. Thomas Gärtner (TU Wien) sprach über interaktives maschinelles Lernen mit strukturierten Daten. Da die Veranstaltung in die „Intelligent Future Days“ eingebettet war, konnten die Teilnehmerinnen und Teilnehmer auch von Keynotes und Vorträgen des „KI-Forums NRW“ profitieren.

Die Proceedings der Gesamtagung finden Sie unter https://www.ifd2020.nrw/lwda2020_proceedings/.

Die LWDA wird im Jahr 2021 unter der Leitung von Thomas Seidl (LMU München) organisiert – wenn es die Coronasituation erlaubt, als Präsenzveranstaltung in München, sonst wieder als Onlineevent. Durch die zeitliche Nähe zur verschobenen BTW-Tagung ist noch unklar, in welcher Form die Fachgruppe Datenbanken teilnehmen wird. Der Arbeitskreis „Data Engineering for Data Science“ des GI-Fachbereichs DBIS plant, in geeigneter Form an der LWDA 2021 teilzunehmen (siehe dazu auch den Call für die Ausgabe 3/2021 im Editorial dieses Heftes; angenommene Beiträge können auf der LWDA präsentiert werden).

4 Produkt-News

Uta Störl

4.1 IBM Db2 11.5.4

Zu den Höhepunkten der neu erschienenen *IBM Db2 Version 11.5.4* zählt der weitere Ausbau von Db2 in Richtung Multi-Model Database: Neben dem relationalen und dem dokumentenorientierten Datenbankmodell bietet Db2 jetzt auch eine Graphdatenbank. Db2 definiert dazu eine Überlagerung der relationalen Tabellen mit einem Graphmodell, so dass die Daten nur einmal gespeichert werden, aber gleichzeitig mit SQL und Gremlin abgefragt werden können.

Darüber hinaus bietet Db2 eine REST-API, mit der nicht nur Abfragen, sondern auch INSERT-, UPDATE- und DELETE-Operationen durchgeführt werden können. Damit ist der Zugriff auf Db2 auch aus Programmiersprachen und -umgebungen möglich, für die keine Db2-Treiber existieren, sofern nur Unterstützung für REST-APIs vorhanden ist.

Neben weiteren Verbesserungen bezüglich Verfügbarkeit und Sicherheit, einer verbesserten Unterstützung für An-

wendungsentwicklung, analytische Anwendungen und die OpenShift-Plattform, ist ein weiterer Schwerpunkt die vermehrte Nutzung von Machine-Learning-Techniken durch Db2 selbst.

IBM, <https://www.ibm.com/analytics/de/de/technology/db2/>

4.2 Oracle Autonomous JSON Database

Oracle stellt einen neuen Cloud-Service auf der bewährten Autonomous-Datenbank-Infrastruktur zur Verfügung, die *Oracle Autonomous JSON Database (AJD)*. Sie richtet sich hauptsächlich an Entwickler, die JSON-basierte Applikationen mit einer einfachen API entwickeln und dabei nicht auf die Möglichkeiten von SQL verzichten wollen. Damit kann die Oracle-Datenbank auch als eine Dokumentdatenbank verwendet werden. AJD bietet dabei alle Automatisierungs- und Verfügbarkeits-Features, welche die Oracle Autonomous Database auszeichnet.

Bei den Schlüsselfunktionen handelt es sich um die Unterstützung einer einfachen Dokument-API (wie die Open Source API SODA), vollständige SQL-Unterstützung, aber auch die Möglichkeit der Verwendung von Python, REST oder Command-Line-Sprachen. Große binäre native JSON-Dateien wie auch unterschiedliche Indizierungen werden ohne Einschränkung unterstützt. Auch relationale (materialized) Views, Transaktionen, Stored Procedures usw. sind darüber hinaus vollumfänglich möglich und mit gleichen Security, Backup/ Recovery-Verhalten, Performance und Monitoring-Tools, wie sie standardmäßig in der Oracle Datenbank zur Verfügung stehen.

Bei Bedarf kann bei der Autonomous JSON Database auch ein Upgrade zu einer *Autonomous Transaction Processing Database (ATP)* erfolgen, wenn sehr viele relationale Daten und JSON gemeinsam verarbeitet werden sollen.

Oracle, <https://www.oracle.com/autonomous-database/autonomous-json-database/>

4.3 Informix 14.10.xC4

In Juni erschien die neue *Informix-Version 14.10.xC4*. Diese Version bietet viele Verbesserungen im Bereich der Anwendungsentwicklung, wie etwa einen neuen .NET Core Provider, Erweiterungen bei den REST- und MQTT-Listeners oder die Unterstützung für die *OpenGIS Simple Features Specification for SQL Revision 1.1* und für *ISO/IEC 13249-3 SQL/MM Part 3: Spatial* bei den spatio-temporal Datentypen von Informix. Ein Schwerpunkt dieser Version sind neben diesen Verbesserungen im Bereich Anwendungsentwicklung und solchen auf dem Gebiet der Administration (u. a. beim GUI Informix HQ) vor allem

Erweiterungen der Replikationsmöglichkeiten wie etwa mit dem Konzept der Smart Trigger.

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

4.4 Neo4j 4.1

Seit Juni 2020 steht *Neo4j 4.1* zur Verfügung. Das neue Release rundet die Neuerungen von Neo4j 4.0 ab und erweitert sie:

1. Verfeinerte Zugriffsrechte für Schreiboperationen: In Neo4j 4.1 können Rechte separat für das Erzeugen, Ersetzen und Löschen von Graph-Elementen sowie für das Hinzufügen, Entfernen und Modifizieren von Properties dieser Graph-Elemente vergeben werden.
2. Neo4j Causal Clustering: Ein Neo4j Causal Cluster wählt und verwaltet die Lead Role automatisch. In bestimmten Situationen ist es für die (Re-)Organisation eines Clusters aber hilfreich, die Leader-Auswahl explizit anzustoßen. Mit der *Lead Transfer Extension* ist dies nun möglich. Darüber hinaus steht Causal Clustering jetzt für Embedded Neo4j zur Verfügung.
3. Verbesserungen im Query Planner: Der Query Planner bietet Nutzern jetzt mehr Kontrollmöglichkeiten über das Replanning gecachter Abfragen. Die Planausgabe stellt zusätzlich Information wie I/O-Kosten, Speicherverbrauch und Cache-Effizienz bereit.
4. Speicherverwaltung: Eine neue Speicherbedarfsabschätzung für Abfragen und konfigurierbare Verbrauchsschranken erlauben bessere Kontrolle besonders speicherhungriger Abfragen.

Neo4J, <https://neo4j.com/blog/introducing-neo4j-4-1/>

4.5 Graph Data Science 1.3

Seit Juli 2020 steht *Neo4j Graph Data Science (GDS)* in Version 1.3 zur Verfügung. GDS 1.3 ist voll kompatibel mit Neo4j 4.0 und 4.1 und verfügt über ein erweitertes Paket von Algorithmen. Insgesamt stehen 50 Algorithmen zur Verfügung, die in produktionsreif, Beta und Alpha kategorisiert sind. GDS 1.3 bietet Betweenness Centrality nunmehr in Produktqualität. In der Alpha-Kategorie sind drei Node Embeddings (Machine-Learning-Algorithmen – Node2Vec, Random Projection und GraphSAGE) hinzugekommen. Außerdem stellt GDS eine Pregel-API bereit, die es ermöglicht, eigene Algorithmen parallel auf GDS' effizienter Hauptspeicher-Graph-Repräsentation zu implementieren und auszuführen. GDS kann komfortabel über Neo4j Desktop installiert und aktualisiert werden.

Neo4J, <https://neo4j.com/graph-data-science-library/>

4.6 SAP HANA Cloud

Nach dem Launch Ende März gibt es weitere Neuerungen bei *SAP HANA Cloud*. Der SAP-eigene DBaaS bietet Echtzeit-Analyseverarbeitung mit deutlich gewachsenen Systemgrößen bei den meisten großen Cloud-Providern. SAP HANA Cloud bietet bedarfsorientiert flexible, mehrschichtige Datenspeicher, die Nutzung eines abfragefähigen Data Lakes für selten verwendete Daten und damit durch geeignete Speicherlösungen ein optimiertes Verhältnis zwischen Kosten und Leistung. Dazu kommt jetzt die einfache Möglichkeit, Daten von Amazon S3 und Azure Data Lake zu importieren und exportieren sowie den Zugriff auf Google BigQuery herzustellen.

Für die nachfolgende Datenverarbeitung wurden Neuerungen im Bereich Spatial und Graph freigegeben, bei letzterem mit deutlichen Performancegewinnen durch erweitertes Indexing. Spatial unterstützt jetzt GeoServer und DBEaver. Für den Zugriff auf SAP HANA gibt es jetzt zusätzliche .NET-Unterstützung für Linux und MacOS sowie node.js-Unterstützung für Alpine Linux. Darüber hinaus gibt es jetzt auch eine kostenlose Testversion.

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

4.7 Neo4j Aura Professional auf Google Cloud Platform

Neo4j Aura ist ein cloud-nativer Graphdatenbankservice basierend auf Neo4j, der sich auf einfache Weise für datenintensive Anwendungen einsetzen und skalieren lässt. Seit August 2020 ist *Neo4j Aura Professional* als nahtlos integrierter Graphdatenbank-Service im Google Cloud Marketplace verfügbar. Zielgruppe des automatisierten und vollständig verwalteten Service sind kleine und mittelständische Unternehmen (KMU) und Startups. Mit *Autopilot Graphdatenbank Management* kann Neo4j Aura direkt aus der Google Cloud Console gestartet werden. Automatisierte Updates garantieren, dass zu jeder Zeit die aktuellste Version von Neo4j verwendet wird. Zudem lässt sich die Graphdatenbank ohne Ausfallzeiten oder Wartungsfenster flexibel skalieren.

Die Nutzung von Neo4j Aura wird pauschal über einen Stundensatz für laufende Datenbanken verrechnet und beginnt bei 0,09 US-Dollar pro Stunde. Anwender bezahlen nur das, was sie brauchen, und erhalten eine einzige Rechnung für alle Google-Dienste einschließlich der Nutzung von Neo4j Aura. Neo4j Aura verfügt über eine selbstheilende Infrastruktur und automatisierte Backups. Die Daten sind durch eine End-to-End-Verschlüsselung geschützt und dank ACID-konformen Transaktionen sicher gespeichert. Das Visualisierungstool *Neo4j Bloom* ist in Neo4j Aura

integriert und erleichtert die Analyse von Mustern und die Darstellung von Ergebnissen von Algorithmen.

Neo4J, <https://neo4j.com/aura/>

4.8 Db2 Data Gate on Cloud Pak for Data

Das kürzlich erschienene *Db2 Data Gate on Cloud Pak for Data* erlaubt es, Daten aus einer Db2-Datenbank unter z/OS in eine Db2- oder Db2-Warehouse-Datenbank in einer Linux-Umgebung unter Openshift zu replizieren. Lesende analytische oder OLTP-Anwendungen können damit in Echtzeit auf die Daten der Mainframe-Datenbank zugreifen, ohne diese zu belasten.

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

4.9 Cloud Pak for Data als managed Service in der IBM Cloud

IBM Cloud Pak for Data as a Service ist ein Starter-Set für IBMs Enterprise Insights Platform Cloud Pak for Data in der IBM-Cloud. Durch die Integration von Daten und KI in einer einheitlichen Umgebung wird die Erstellung und das Management von Analysen und Machine-Learning-Modellen vereinfacht. IBM übernimmt die Bereitstellung, Verwaltung und Aktualisierung.

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

4.10 Netezza Performance Server in der Cloud

Nachdem letztes Jahr die Technologie der bewährten *Netezza Data Warehousing Appliance* als Teil von Cloud Pak for Data verfügbar wurde, kann man jetzt den Netezza Performance Server auch in der Public Cloud nutzen – sowohl in der IBM-Cloud als auch als Amazon Web Service. Dabei wird eine Datenbank-Engine genutzt, die zu 100 Prozent kompatibel zu den On-Premises-Systemen ist. Erweitert wird sie jedoch um die Möglichkeiten, wie sie die Cloud

zur Verfügung stellt, wie zum Beispiel die Nutzung von Cloud Object Storage als Teil der Datensicherungsstrategie. Für bestehende Appliance-Systeme wird ein einfacher Lift-and-Shift-Ansatz für die Migration in die Cloud unterstützt.

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

4.11 Oracle Academy Member Hub: Eigene Lern-Channels

Das kostenfreie *Oracle-Academy-Programm* (auf academy.oracle.com) wurde mit einem komplett neu gestalteten Mitgliederportal ergänzt. Über das sogenannte *Member Hub* erhalten registrierte Lehrkräfte jetzt schnelleren Zugriff auf alle Programminhalte. Konnten Lehrkräfte bisher ausschließlich auf vorgefertigte Academy-Lerninhalte für den Informatikunterricht zugreifen, so ist es jetzt auch möglich, eigene Lern-Channels individuell zu erstellen. In einem Channel kann man entweder diverse Academy-Lerninhalte hinzufügen oder aber auch einen kompletten Academy-Lernpfad einbinden. Alle derzeit verfügbaren Module/Lernpfade werden im Member Hub-Katalog aufgelistet. Möchte man eigenen Content in den Channel einbinden, so ist dies ebenfalls über eine Upload-Funktionalität problemlos möglich. Außerdem kann noch festgelegt werden, über welchen Zeitraum der jeweilige Channel verfügbar sein soll.

Nachdem der Channel erstellt wurde, legt man die Studenten-Accounts an und schaltet anschließend den Channel für seine Teilnehmer frei. Die Studierenden erhalten einen eigenen Member Hub-Zugang und können auf den erstellten Channel und dessen Inhalte zugreifen.

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

Dank an Jörg Latza (SAP), Frank Schneede, Ulrike Schwinn, Britta Wolf (alle Oracle), Hannes Voigt (Neo4j) sowie Wilfried Hoge, Stefan Lindenmeyer und Andreas Weinger (alle IBM) für ihren fachlichen Input.