

Die InterSystems IRIS® Data Platform als Motor für künstliche Intelligenz

Technology Brief



Einführung

An der künstlichen Intelligenz (KI) ist nichts gekünstelt. Die Geschäftswelt wird schon bald in zwei Lager aufgeteilt sein: einerseits in Unternehmen, die auf KI setzen, um ihre Prozesse zu optimieren und agiler zu werden, und andererseits in Firmen, die den Sprung auf den KI-Zug verpassen und auf der Strecke bleiben.

Agilität beginnt mit dem Ergreifen der Initiative. Und jetzt ist die Zeit für Veränderung gekommen.

Die jüngsten Fortschritte im Bereich der künstlichen Intelligenz stellen die moderne Geschäftswelt auf den Kopf. Immer mehr Organisationen nutzen KI und Big Data, um innovative Dienstleistungen bereitzustellen, neue Einnahmequellen zu erschließen, Kundenerlebnisse zu verbessern und Geschäftsprozesse zu optimieren. Aber nicht nur etablierte Firmen implementieren KI-gestützte Initiativen. Insbesondere Startups nutzen KI und sorgen dadurch in nahezu jeder Branche für einen tiefgreifenden Umbruch und zunehmenden Druck auf die traditionellen Marktteilnehmer.

McKinsey hat die Vorteile des Einsatzes von KI in mehr als 400 Anwendungsfällen in 19 Branchen dokumentiert.¹ PwC prognostiziert, dass KI im Jahr 2030 15,7 Billionen US-Dollar zur globalen Wirtschaftsleistung beitragen wird.² Laut Gartner gaben zudem 77 Prozent aller Führungskräfte in einer Umfrage an, dass Data Science (Datenwissenschaft) bereits heute einen beträchtlichen Wert für ihr Unternehmen beiträgt.³

Um dieser Entwicklung in der Geschäftswelt Rechnung zu tragen, hat InterSystems eine umfassende Datenplattform entwickelt, welche die Entwicklung, Implementierung und Wartung von echtzeitbasierten, datenintensiven Analyseanwendungen für diverse Branchen beschleunigt und vereinfacht.

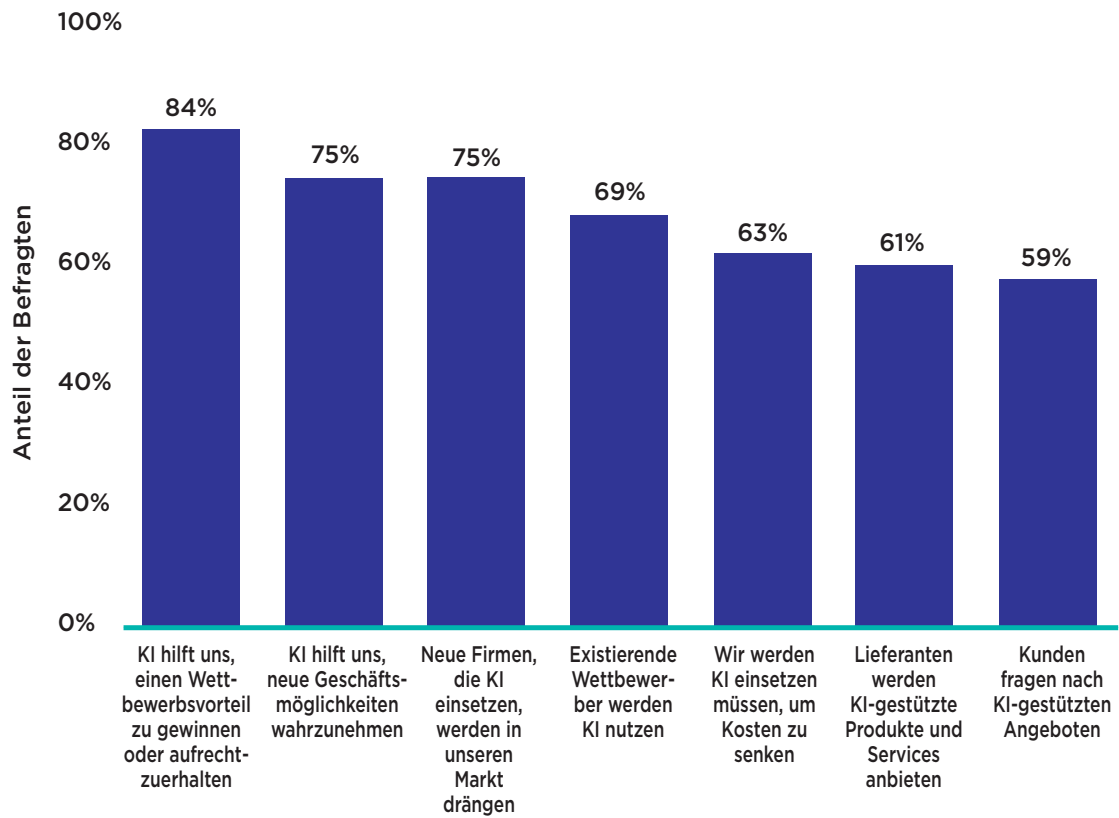
Dieses Dokument zeigt auf, wie die InterSystems IRIS® Data Platform die Wertschöpfung von Unternehmensinitiativen, die sich auf künstliche Intelligenz stützen, beschleunigen kann.

¹ <https://www.mckinsey.com/featured-insights/artificial-intelligence/notes-from-the-ai-frontierapplications-and-value-of-deep-learning>

² <https://www.pwc.com/gx/en/issues/data-and-analytics/publications/artificial-intelligence-study.html>

³ Gartner, 2018: „Magic Quadrant for Data Science and Machine Learning Platforms“

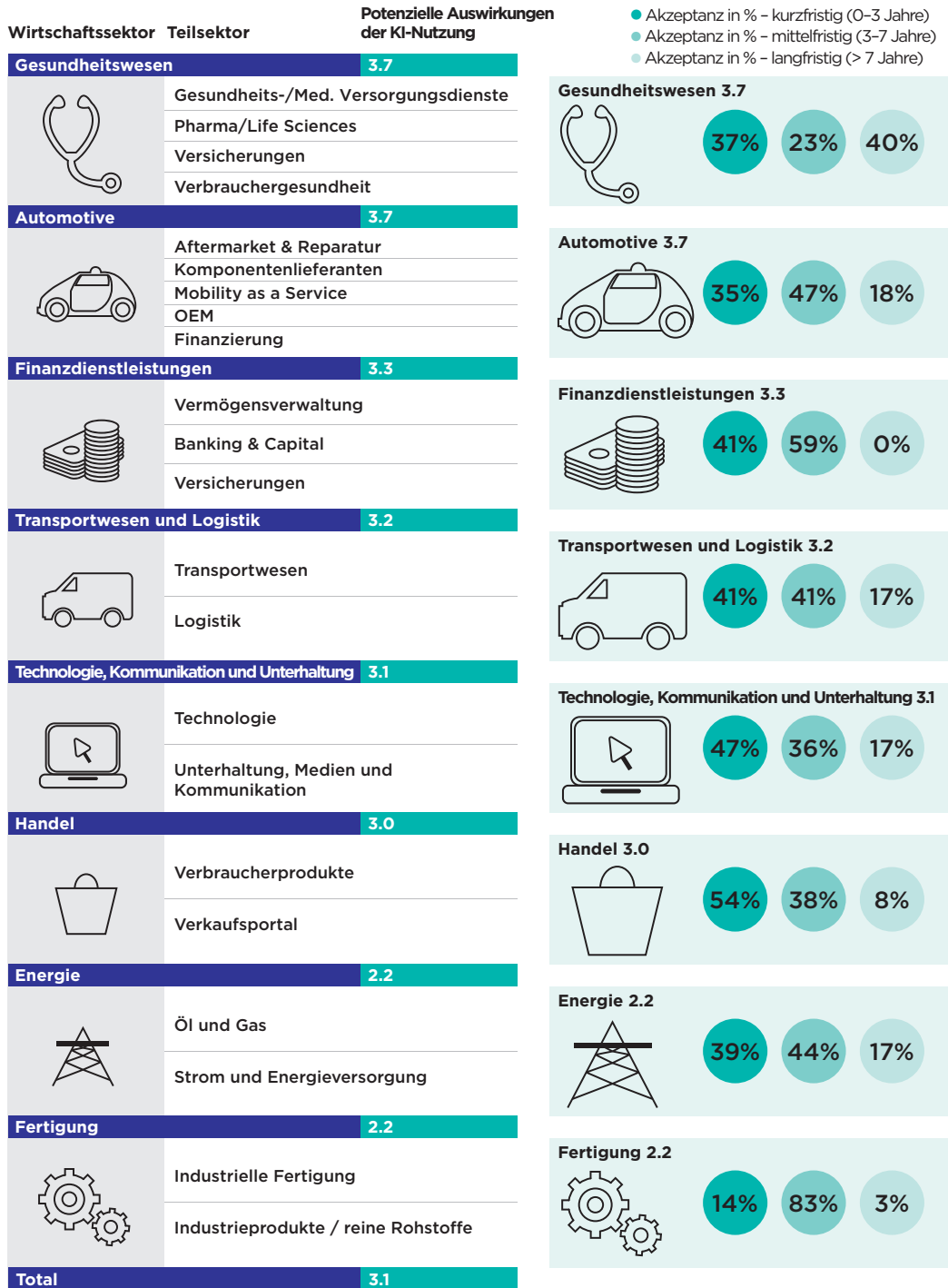
Warum Unternehmen KI implementieren



Quelle: Statista-Umfrage 2017 unter mehr als 3.000 Führungskräften weltweit⁴

⁴ <https://www.statista.com/statistics/747775/worldwide-reasons-for-adopting-ai/>

KI-Akzeptanz nach Branche



Quellen basierend auf der PwC-Indexbewertung des Einflusses von KI Bewertung des Potenzials mit 1-5, wobei 5 für das höchste und 1 für das geringste Einflusspotenzial von KI steht.

InterSystems-Technologie wurde bereits in Tausenden Unternehmensanwendungen in mehr als 80 Ländern implementiert. Mit ihrer Hilfe werden mehr als 500 Millionen elektronische Patientenakten und 15 Prozent der täglichen Aktiengeschäfte weltweit verarbeitet.

InterSystems stellt erstklassigen weltweiten Support für seine Kunden über das Worldwide Response Center bereit. Dieses ist rund um die Uhr mit kompetenten Software-Experten besetzt. Unser Ziel ist, das gesamte Unternehmen und nicht nur das akute Problem zu verstehen. Wir unterstützen Sie bei Ihrer Strategie, Planung, Performance-Analyse und Migration und verhelfen Ihnen so zum Erfolg.

KI vorantreiben

Obwohl es bereits eine Vielzahl von KI-Tools und -Frameworks auf dem Markt gibt, gilt es jetzt, die Anwendung von KI in Unternehmen mit Hochdruck voranzutreiben, um aktuelle und zukünftige Geschäftsanforderungen zu erfüllen. Die InterSystems IRIS Data Platform stellt viele der dazu benötigten Funktionalitäten bereit.

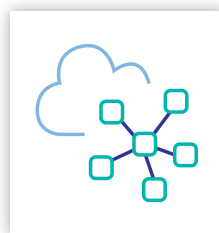
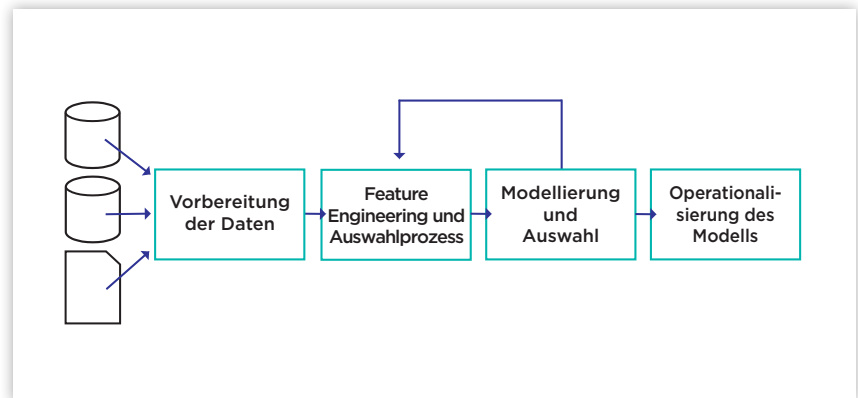
Die Kernfunktionalitäten von InterSystems IRIS im Überblick:

- Horizontal skalierbare Datenbankverwaltungsfunktionen für Hybrid Transactional/Analytical Processing (HTAP) zur Durchführung von Echtzeitanalysen bei sehr großen Datenmengen;
- Integrierte Funktionalität für strukturierte und unstrukturierte Analysen und KI-Modelle;
- Eine flexible Multi-Model-Datenbank, die auf Feature Engineering für KI in großem Umfang ausgelegt ist;
- Umfassende Interoperabilität zur Integration von Daten und Anwendungen, zur Vorverarbeitung von Daten für Analysen sowie zur Erarbeitung und Umsetzung komplexer Geschäftsprozesse;
- Eine zentrale Entwicklungsumgebung, um die Erstellung und Wartung fortschrittlicher analytischer Anwendungen zu erleichtern, inklusive eines optimierten Konnektors zu Best-of-Breed Analytics Werkzeugen.

Die KI-Pipeline

Die Entwicklung und Implementierung KI-basierter Initiativen erfordern folgende Schritte:

- Vorbereitung der Daten
- Feature Engineering und Auswahlprozess
- Modellierung und Konfiguration
- Ausführung und Operationalisierung von KI-Modellen



Vorbereitung der Daten

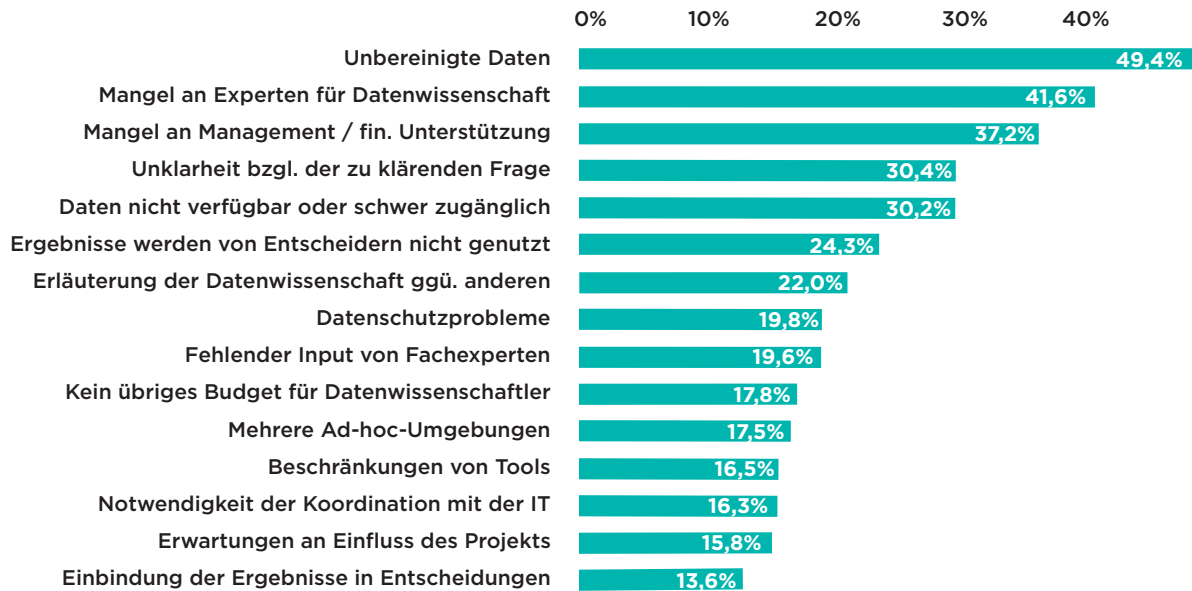
Moderne Unternehmen sammeln riesige Mengen an Daten aus verschiedenen Quellen, in verschiedenen Formaten und mit verschiedenen semantischen Darstellungen. Je mehr Daten in Analysen berücksichtigt werden können, desto größer die Wahrscheinlichkeit eines geschäftlichen Nutzens. Allerdings werden rohe und unbereinigte Daten wahrscheinlich keine präzisen Ergebnisse hervorbringen. Deshalb müssen, um das volle Potenzial von KI ausschöpfen zu können, große Mengen an Daten aus unterschiedlichen Quellen erfasst, integriert und normalisiert werden – und das bedeutet viel Arbeit. Laut IDG verbringen Data Scientists (Datenwissenschaftler) rund 80 Prozent ihrer Zeit mit diesen Aufgaben.⁵

In der im Jahr 2017 von Kaggle, der größten Online Community für Data Scientists und Experten für Machine Learning, durchgeführten Umfrage „State of Data Science & Machine Learning“ zum Status quo der Datenwissenschaft und maschinellem Lernen nannten mehr als 7.000 Personen, die sich mit Data Science befassen, unbereinigte Daten als größtes Problem.⁶

⁵ <https://www.infoworld.com/article/3228245/data-science/the-80-20-data-science-dilemma.html>

⁶ <https://www.kaggle.com/surveys/2017>

Barrieren für die geschäftliche Nutzung der Datenwissenschaft




Quelle: Im Jahr 2017 von Kaggle durchgeführte Umfrage „State of Data Science & Machine Learning“ unter 7.276 Personen

Die meisten Datenwissenschaftler würden ihre Zeit zwar lieber der Analyse von Daten widmen, doch ihre Hauptaufgabe im Zusammenhang mit KI liegt in der Datenaufbereitung, oder anders ausgedrückt, in der Umwandlung von rohen Daten in etwas, was für KI-Zwecke genutzt werden kann.

InterSystems IRIS entlastet Datenwissenschaftler und beschleunigt die Erfassung, Integration und Normalisierung der Daten, damit präzise KI-Modelle erstellt und konfiguriert werden können. Die Plattform bietet umfassende eingebettete Integrationsfunktionen und gewährleistet so die sofortige Anbindung an eine Vielzahl von paketierten Anwendungen, Datenbanken, Industriestandards, Protokollen und Technologien. Flexible Möglichkeiten zur Umwandlung von Daten über eine grafische Benutzeroberfläche vereinfachen die Auflösung von Unterschieden in der Semantik und den Datenmodellen zwischen Datenquellen, Anwendungen und Services. Die Plattform unterstützt außerdem eine Vielzahl von Industriestandards, die in verschiedenen Branchen, wie dem Gesundheitswesen oder der Finanz-, Handels- und Telekommunikationsbranche zum Einsatz kommen, darunter REST-Architekturen und Standards für Webdienste (z. B. JSON, XML, XPATH, XSLT, SOAP und DTDs).

Zudem bietet InterSystems IRIS fortschrittliche Sicherheitsfunktionen wie rollenbasierte Sicherheit mit Granularität auf Zeilenebene, um für einen sicheren Zugriff auf sensible Daten zu sorgen. Dank dieser Funktion kann jede einzelne Zeile mit einer Liste autorisierter Benutzer verknüpft werden, sodass ausschließlich Benutzer mit der dazu erforderlichen Berechtigungsstufe auf bestimmte Datenelemente zugreifen können.



„DIE ENTWICKLUNG
VON FEATURES IST
SCHWIERIG, ZEITAUFGEBEND
UND ERFOR-
DERT
EXPERTENWISSEN.
»ANGEWANDTES
MASCHINELLES LERNEN«
IST IM GRUNDE FEATURE
ENGINEERING.“

*Professor Andrew Ng,
Stanford University*

Feature Engineering und Auswahlprozess

Das Feature Engineering kann die Erstellung neuer oder aggregierter Variablen auf Grundlage der Rohdaten und die Umwandlung der Struktur von Variablen umfassen. Dadurch können Informationen bereitgestellt werden, die nützlicher sind als die Rohdaten allein. Beispielsweise könnte ein Feature das Verhältnis zwischen dem Umsatz des vergangenen Monats und dem Umsatz der vergangenen 12 Monate für jeden Kunden nachvollziehen.

Die Entwicklung neuer Features kann die Umwandlung oder Kodierung einiger Variablen erfordern, indem beispielsweise die Struktur von kategorisch in numerisch geändert wird. Zum Beispiel kann die Häufigkeit des Vorkommens eines US-Bundesstaats (z. B. New York oder Kalifornien) relevant für eine Analyse sein. Unter Umständen akzeptiert der Algorithmus es jedoch nicht, wenn die Daten als eine Eigenschaft mit mehreren String-Werten dargestellt werden. Ein Beispiel für eine Umwandlungs- oder Verschlüsselungsaufgabe wäre die Umwandlung der Eigenschaft (US-Bundesstaat) mit 50 unterschiedlichen String-Werten (einer für jeden Bundesstaat) in 50 verschiedene Features (eines für jeden Bundesstaat), jedes mit einem numerischen Wert, der der Anzahl der Vorkommen des Bundesstaats entspricht.

Eine weitere gängige Anforderung im Zusammenhang mit Feature Engineering ist die Extrahierung des Inhalts und der Grundstimmung (Sentiment) aus freiem Text. Zum Beispiel könnte die Häufigkeit des Vorkommens eines Konzepts in einem Dokument Aufschluss über die relative Bedeutung des Konzepts geben, während eine Sentiment-Analyse dem Text eine positive oder negative Grundstimmung zuordnen kann.

InterSystems IRIS beschleunigt und vereinfacht diverse Aufgaben in Verbindung mit Feature Engineering. Sie ermöglicht Benutzern die Erstellung neu berechneter Variablen, bietet integrierte Funktionalität für Business Intelligence (multidimensionales OLAP) und verteilte SQL-Verarbeitung, um komplexe Berechnungen mit sehr umfangreichen Datensätzen und hoher Performanz zu unterstützen. Daten werden einmal in der Datenbank gespeichert, und die Plattform gewährt flexibel sowohl relationalen als auch nicht relationalen Zugriff auf die Daten. Wenn es einfacher ist, mit Daten in Zeilen und Spalten zu arbeiten, können Entwickler über SQL-Befehle mit den Daten interagieren. Alternativ können die Daten als Dokumente, Objekte oder Schlüsselwertdaten abgerufen und bearbeitet werden.

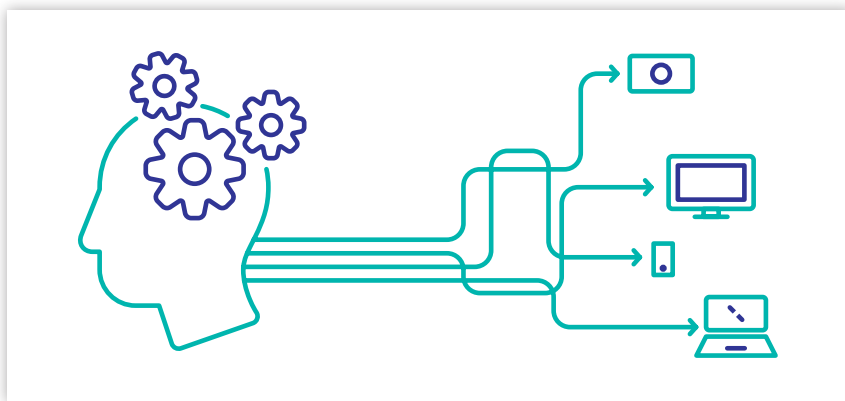
InterSystems IRIS stellt Funktionalität für Natural Language Processing (NLP) bereit, mit der Texte in natürlicher Sprache in ihrem Inhalt und in Bezug auf das Sentiment analysiert werden können. InterSystems IRIS erkennt automatisch Konzepte und Beziehungen in Texten, ohne dass irgendwelche Vorarbeit oder Domänenwissen vonnöten ist. Mit diesen Funktionen sind Datenwissenschaftler in der Lage, Informationen aus Dokumenten, Notizfeldern, Social-Media-Feeds und anderen Quellen zu extrahieren und neue Features für Analysen zu entwickeln.

Feature Engineering kann Zehntausende (oder noch mehr) Matrizen hervorbringen. Ein InterSystems-Kunde aus dem Gesundheitswesen hat 113.000 verschiedene Features erstellt, um die Weiterentwicklung seiner KI-gestützten Patientenrisikomodelle voranzutreiben. Damit solche Aufgaben mit maximaler Effizienz erledigt werden können, speichert InterSystems IRIS Daten als multidimensionale Arrays in der Datenbank und wendet horizontale Skalierungstechniken an. InterSystems IRIS ermöglicht dadurch bei Feature-Engineering-Aufgaben mehr Leistung mit weniger Hardwareressourcen.

Modellierung und Konfiguration

Obwohl InterSystems IRIS hoch performante SQL-Fähigkeiten bietet, werden viele Modellierungen und Konfigurationen mithilfe spezieller Tools vorgenommen. Bei seiner Umfrage aus dem Jahr 2017 listete Kaggle 49 gängige Modellierungstools auf und identifizierte Python als die am häufigsten verwendete Datenanalyse-sprache, gefolgt von R und SQL.⁷

InterSystems IRIS bietet Unterstützung für xDBC-Protokolle und andere Standards und lässt sich über einen Shard-kompatiblen Spark-Konnektor direkt mit Apache Spark integrieren. Somit können InterSystems IRIS-Anwendungen die Spark-Verarbeitung eingliedern und Spark-Anwendungen verteilte Daten aus InterSystems IRIS einbinden. Der Konnektor macht sich die Tatsache zunutze, dass die Datenbank von InterSystems IRIS partitioniert ist, weshalb die Arbeitsknoten von Apache Spark sich automatisch mit den Datenpartitionen verbinden und so parallel an separaten Datenelementen arbeiten können. Diese parallelen Direktverbindungen sorgen für einen hohen Datendurchsatz und unterstützen die sehr schnelle Datenaufnahme in horizontal verteilte Cluster. Dank dieser Architektur können Modelle auch nach und nach in Echtzeit trainiert werden, wenn Daten aus Produktionsanwendungen erfasst und abgerufen werden, sodass Datenwissenschaftler Modelle an neue Geschäftsbedingungen und Umgebungsvariablen anpassen können.



⁷ <https://www.kaggle.com/surveys/2017>

Operationalisierung von KI-Modellen

Bei einigen Anwendungen müssen KI-Modelle lediglich unabhängig voneinander im Batch-Modus ausgeführt werden. Viele andere jedoch erfordern die Integration von KI-Modellen in Geschäftsprozesse, die in Echtzeit als Reaktion auf Ereignisse, Transaktionen oder Benutzeraktionen ausgeführt werden. Beispielsweise werden Modelle zur Betrugserkennung oftmals in Systeme zur Echtzeitverarbeitung von Kreditkartentransaktionen eingebunden, um potenziell betrügerische Transaktionen zu erkennen und zu stoppen.

Die Branche bezeichnet die Integration von KI-Modellen in Geschäftsprozesse als „Operationalisierung“. Darüber hinaus hat Gartner in seinem Bericht „Hype Cycle for Data Science and Machine Learning“ aus dem Jahr 2018 einen neuen Begriff geprägt: „kontinuierliche Intelligenz“. Gartner definiert „kontinuierliche Intelligenz“ als „Echtzeitanalysen (...) die in einen Geschäftsprozess eingebunden werden, um aktuelle und historische Daten zu verarbeiten und anschließend Handlungsempfehlungen zu geben. Diese Innovation sorgt für die Automatisierung und Unterstützung der Entscheidungsfindung. Sie bedient sich verschiedener Technologien, darunter Augmented Analytics, Event Stream Processing, Optimierung, Geschäftsregelmanagement und maschinelles Lernen.“ Kontinuierliche Intelligenz zählt zu einer Handvoll von KI-Unterbereichen, die Gartner als „transformativ“ bezeichnet.⁸

InterSystems IRIS stellt eine komplette Suite an Tools für die Entwicklung und Ausführung von daten- und analyseintensiven Geschäftsprozessen in Echtzeit bereit, die mit internen und externen Datenquellen, Anwendungen und Diensten verknüpft werden können. Zudem bietet die Plattform grafische Werkzeuge zur Schematisierung von Prozessen, Regeln und Workflows, um intelligente Geschäftsprozesse zu unterstützen.

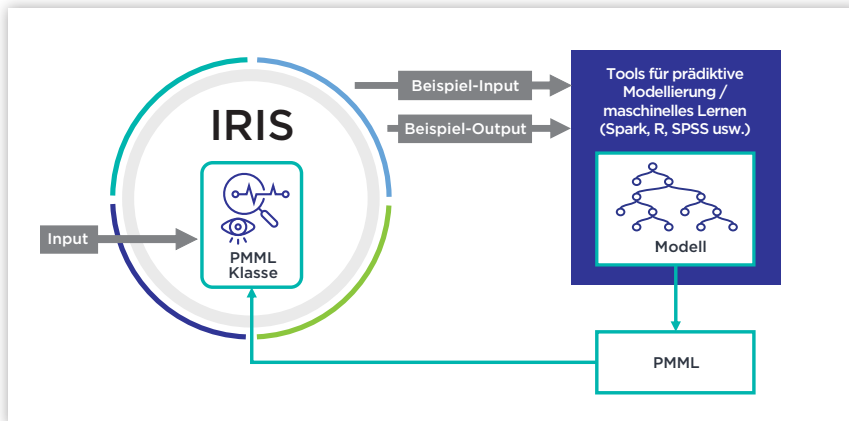
Dank integrierter HTAP-Funktionalitäten können KI-Modelle außerdem durch Ereignisse und Transaktionen in Echtzeit ausgelöst und mit einer Reaktionszeit von wenigen Millisekunden ausgeführt werden. Dies ist unerlässlich für geschäftskritische Anwendungen, wie die Erkennung von Betrugsversuchen während des Zahlungsvorgangs oder des Risikos der Kundenabwanderung, während ein Call-Center-Mitarbeiter mit einem Kunden telefoniert.

KI-Modelle, die mithilfe externer Tools und Anwendungen erstellt wurden, können über Predictive Model Markup Language (PMML) in InterSystems IRIS-Anwendungen integriert werden. PMML ist ein XML-Standard, der alle Parameter eines Modells definiert, das mittels einer externen KI-Anwendung bzw. eines externen KI-Frameworks entwickelt wurde. Die meisten KI-Tools und -Frameworks unterstützen den Export von Modellen über den PMML-Standard. InterSystems IRIS umfasst eine integrierte PMML-Laufzeitumgebung zur Ausführung von Modellen, die mithilfe externer Tools entwickelt wurden, ohne dass zusätzliche Tools eingesetzt oder Daten zwischen Systemen ausgetauscht werden müssen.

Dieser Architekturansatz ermöglicht die nahtlose Ausführung von KI-Modellen im Rahmen komplexer Echtzeit-Geschäftsprozesse.

⁸ Gartner, 2018: „Magic Quadrant for Data Science and Machine Learning Platforms“

Operationalisierung von KI-Modellen zur Ausführung innerhalb von InterSystems IRIS-Anwendungen



Auf Skalierbarkeit ausgelegt

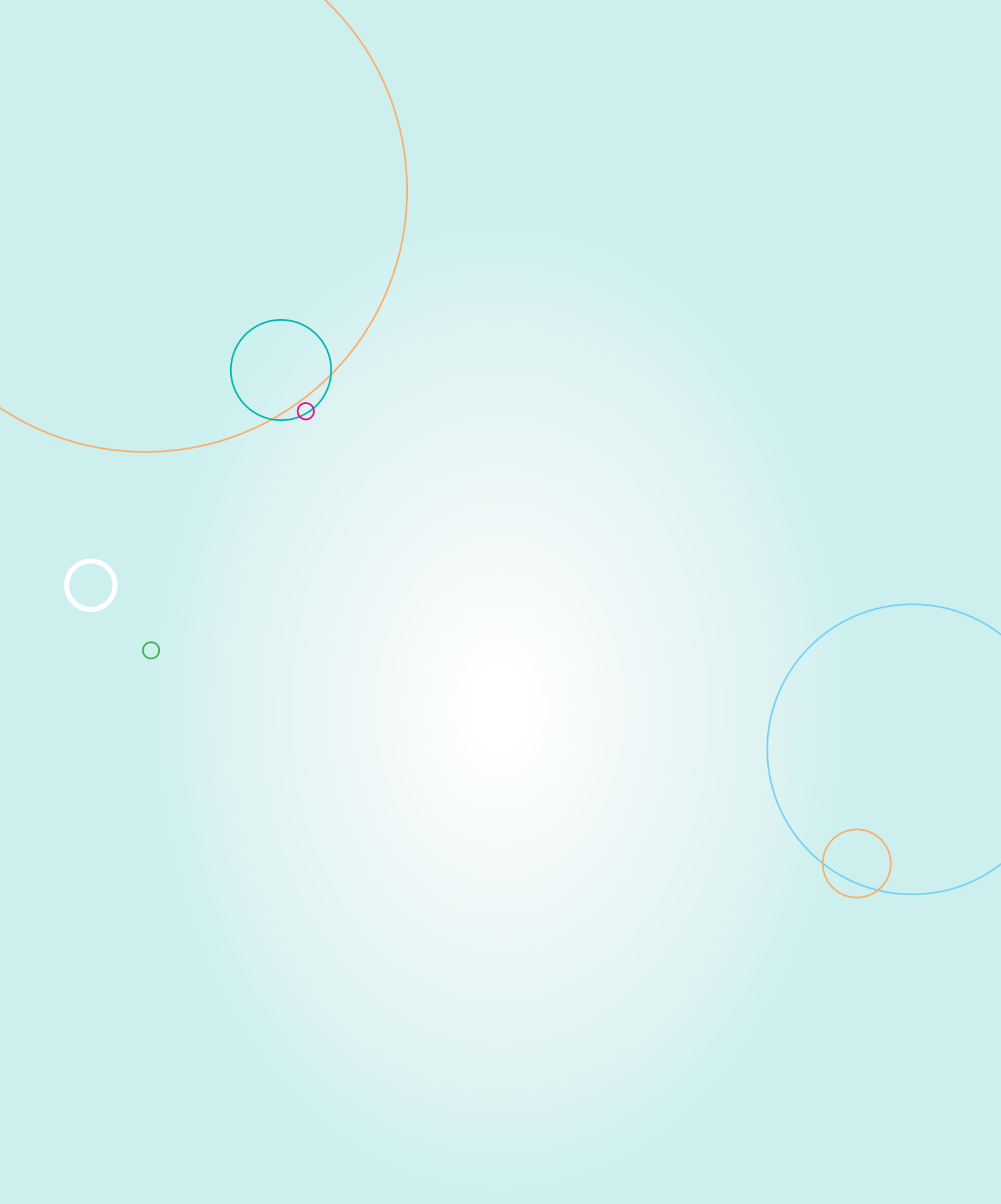
Die Datenbankverwaltungsfunktionen von InterSystems IRIS bieten nachweislich eine hohe Performanz bei geschäftskritischen Unternehmensanwendungen, sowohl in Bezug auf transaktionale (OLTP) als auch analytische Workloads. InterSystems IRIS unterstützt eine besonders schnelle Datenaufnahme und kann gleichzeitig Analysen für Echtzeitdaten und sehr große Mengen an Nicht-Echtzeit-Daten (z. B. historische und Referenzdaten) unter Einsatz von Standardhardware durchführen. Damit ist die Lösung ideal für Anwendungen, die Entscheidungen in Echtzeit treffen müssen, indem sie Muster und Anomalien in beliebig großen Datenbeständen erkennen. Diese Architektur ist immer absolut zuverlässig, auch dann, wenn die Workloads plötzlich stark ansteigen.

Fazit

InterSystems IRIS ist eine umfassende Datenplattform, die die Entwicklung, Bereitstellung und Wartung echtzeitbasierter, datenintensiver Analyseanwendungen erleichtert und beschleunigt. Sie stellt wichtige Funktionen für die Vorverarbeitung von Daten bereit, um die KI-Modellierung und -Konfiguration zu unterstützen und KI-Modelle in Echtzeit-basierte Geschäftsprozesse einzubinden. Auf diese Weise hilft InterSystems IRIS Organisationen dabei, mithilfe von KI und Big Data innovative Dienstleistungen bereitzustellen, neue Umsatzströme zu erschließen, Kundenerlebnisse zu verbessern und Geschäftsprozesse zu optimieren. So können diese sich einen Wettbewerbsvorteil verschaffen – und ihre Wertschöpfung beschleunigen.

Beschleunigen Sie die Umsetzung Ihrer KI-Initiativen mit der InterSystems IRIS Data Platform.

InterSystems ist die treibende Kraft hinter den wichtigsten IT-Anwendungen der Welt. Im Gesundheitswesen, im Finanzsektor, in der öffentlichen Verwaltung und in vielen anderen Bereichen, wo viel auf dem Spiel steht, ist InterSystems „the power behind what matters“. Das 1978 gegründete, privat gehaltene Unternehmen mit Niederlassungen rund um den Globus hat seinen Sitz in Cambridge, Massachusetts (USA). Die Softwareprodukte von InterSystems werden tagtäglich von Millionen Menschen in über 80 Ländern genutzt.



The power behind what matters.

