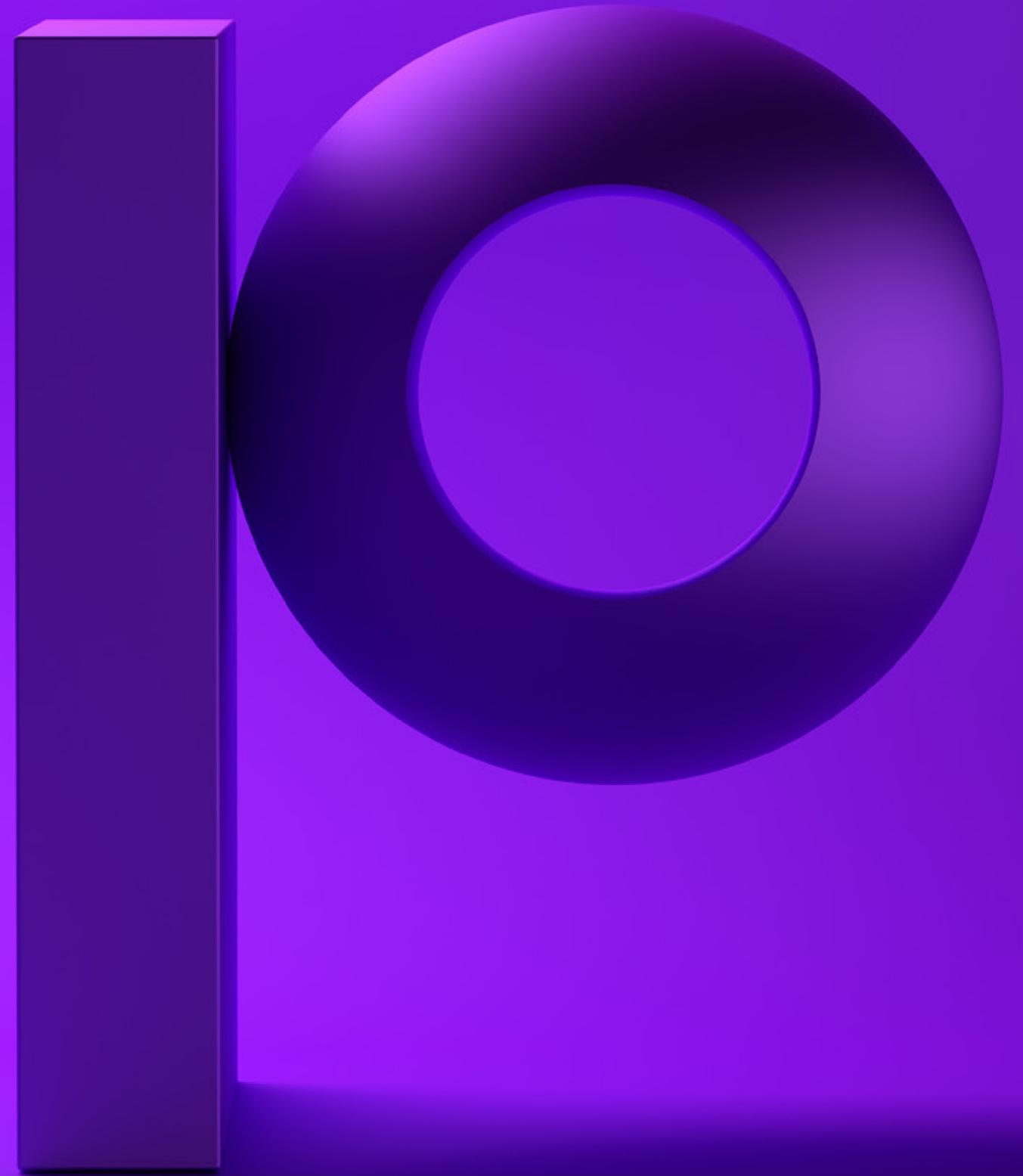


precisely

Daten durch
Datenintegration für
KI vorbereiten



Einleitung

Daten sind der Treibstoff für künstliche Intelligenz (KI). Wenn Daten jedoch qualitativ minderwertig, nicht verfügbar und nicht integriert sind, kann dies zu Verzerrungen und Ungenauigkeiten führen, die Ihrem Unternehmen schaden.

Datenintegration ist ein wesentlicher Schritt bei der Vorbereitung Ihrer Daten für die erfolgreiche Nutzung von KI. Durch die Integration von Daten aus verschiedenen Quellen und Formaten in einer einheitlichen Sicht können Unternehmen problemlos auf ihre gesamten Daten zugreifen und sie analysieren, unabhängig von der Herkunft und Organisation der Daten. Im Rahmen der Datenintegration wird außerdem sichergestellt, dass die Daten genau, vollständig und zuverlässig sind – dies sind unverzichtbare Eigenschaften für erfolgreiche KI-Initiativen.

Diese Checkliste bietet einen kurzen Überblick über die wichtigsten Schritte und Erwägungen für die Datenintegration.

Datenintegration für KI

- Ziele und Umfang der Datenintegration definieren
- Aktuelle Datenlandschaft bewerten und Datenquellen ermitteln
- Richtige Strategie, Partner und Tools für die Datenintegration auswählen
- Datenintegrationsarchitektur und -Pipeline entwerfen und implementieren
- Datenqualität und Data Governance während des gesamten Prozesses sicherstellen
- Leistung und Ergebnisse der Datenintegration überwachen und optimieren



Schritt 1: Definieren Sie die Ziele und den Umfang der Datenintegration

Es gibt keine einheitliche Lösung für die Datenintegration; je nach KI-Anwendungsfall können unterschiedliche Ansätze und Verfahren erforderlich sein. Wenn Sie die Ziele und den Umfang der Datenintegration entsprechend Ihren individuellen KI-Zielen definieren, werden auch der Umfang und die Grenzen Ihres Projekts definiert. Sie können dann die erwarteten Ergebnisse, Zeitpläne, Ziele und Erwartungen mit den Stakeholdern und Teammitgliedern teilen.

Stellen Sie sich für Schritt 1 die folgenden Fragen:

- Für welche Probleme oder Chancen Ihres Unternehmens möchten Sie KI nutzen?
- Was sind die erwarteten Ergebnisse oder Vorteile Ihres KI-Projekts?
- Welche Datenquellen und -typen müssen integriert werden?
- Was sind die Herausforderungen und Risiken?
- Was sind die Erfolgskriterien und -kennzahlen?
- Über welche Ressourcen und welches Budget verfügen wir?
- Mit welchen Tools und Partnern werden wir arbeiten?



Schritt 2: Bewerten Sie Ihre aktuelle Datenlandschaft und identifizieren Sie Datenquellen

Wenn Sie Ihre aktuelle Datenlandschaft bewerten und die zu integrierenden Datenquellen identifizieren, können Sie die Verfügbarkeit, die Zugänglichkeit und die Qualität Ihrer Daten ermitteln sowie Lücken und Probleme erkennen, deren Behebung Priorität haben soll.

In dieser Phase sind folgende Aufgaben auszuführen:

- Führen Sie eine Bestandserfassung und Prüfung Ihrer Daten durch, um alle Datenquellen und -typen zu bestimmen und zu dokumentieren, die für Ihr KI-Projekt relevant sind.
- Bewerten Sie die Datenqualität und Vollständigkeit jeder Datenquelle, und ermitteln Sie alle Probleme bei der Datenqualität, z. B. fehlende, ungenaue, inkonsistente oder doppelt vorhandene Daten.
- Bewerten Sie die Datenverfügbarkeit und Sicherheit jeder Datenquelle, und ermitteln Sie Probleme beim Datenzugriff, z. B. Datensilos, Datenfragmentierung oder Datenschutz.
- Bewerten Sie die Konformität und Interoperabilität der Daten jeder Datenquelle, und ermitteln Sie Probleme im Hinblick auf die Datenintegration, z. B. Format, Struktur, Schema oder Semantik der Daten.
- Priorisieren Sie die Datenquellen und -typen, die für Ihr Projekt am wichtigsten und wertvollsten sind, und bestimmen Sie die Reihenfolge und Häufigkeit der Datenintegration.



Schritt 3: Denken Sie darüber nach, welche Datenintegrationsstrategie und -werkzeuge infrage kommen

Die Datenintegration umfasst Methoden und Technologien, die sich je nach den Quellen, Typen, Formaten und Umgebung Ihrer Daten ändern können. In diesem Schritt wählen Sie das beste Verfahren zum Integrieren der generierten und von Ihrem Unternehmen genutzten Daten. Bestimmen Sie die geeigneten Methoden, Tools und Partner entsprechend der Komplexität und dem Umfang Ihres Projekts. Sie vergleichen die Lösungen und wählen diejenigen aus, die eine optimale Funktionalität, Leistung und Kosteneffizienz bieten.

Unter anderem sind folgende Faktoren zu berücksichtigen:

- Architektur und Design, z. B. zentral, dezentral oder hybrid
- Verfahren und Methode, z. B. ETL (Extrahieren, Transformieren und Laden), ELT (Extrahieren, Laden und Transformieren), CDC (Change Data Capture) oder Datenvirtualisierung
- Modus und Häufigkeit, z. B. Batch, Echtzeit oder Streaming
- Plattform und Tool-Umgebung, z. B. cloudbasiert (öffentlich, privat oder Multi-Cloud), lokal oder hybrid
- Funktionalität und Fähigkeiten, z. B. Datenerfassung, Datentransformation, Datenbereinigung, Datenanreicherung, Datenzuordnung, Datenvalidierung, Datenlieferung oder Datenüberwachung
- Skalierbarkeit und Leistung, z. B. Menge, Geschwindigkeit, Latenz, Durchsatz oder Zuverlässigkeit der Daten
- Anforderungen, die für Governance, Datenzugriff durch Stakeholder und Compliance erfüllt werden müssen
- Partnerschaften, die für die ordnungsgemäße Implementierung der Datenintegrationsmuster und Anwendung der Daten erforderlich sind



Schritt 4: Entwerfen und implementieren Sie eine Architektur und Muster für die Datenintegration

Der nächste Schritt ist das Entwerfen und Implementieren einer Datenintegrationsarchitektur und von Datenintegrationsmustern, mit denen Ihre Datenintegrationsziele erreicht werden können. Eine Datenintegrationsarchitektur definiert, wie die Datenquellen verbunden, transformiert und für Ihre KI-Anwendungen zur Verfügung gestellt werden. Ein Datenintegrationsmuster ist der Workflow, der die Datenintegrationsarchitektur implementiert und die Aufgaben und Prozesse der Datenintegration ausführt. Mit diesen Schritten können Sie auch sicherstellen, dass Ihr Datenintegrationsmuster zweckmäßig, zuverlässig und sicher ist und die für erfolgreiche KI-Ergebnisse erforderliche Datenqualität aufweist.

Schritte in dieser Phase:

- Definieren Sie die Eingaben und Ausgaben der Datenintegration, z. B. Datenquellen, Datentypen, Formate und Ziele.
- Definieren Sie die Transformationen und Regeln für die Datenintegration, z. B. für die Bereinigung, Anreicherung, Zuordnung und Validierung von Daten.
- Definieren Sie die Flows und Sequenzen der Datenintegration, z. B. die Erfassung, Lieferung und Synchronisierung der Daten.
- Definieren Sie die Kontrollen und Standards der Datenintegration, z. B. Datenqualität, Governance und Sicherheit.
- Implementieren Sie das Datenintegrationsmuster mit den ausgewählten Datenintegrations-Tools- und -Plattformen.
- Testen und überprüfen Sie das Datenintegrationsmuster mit Beispieldaten und -szenarien.



Schritt 5: Stellen Sie Datenqualität und Data Governance während des gesamten Integrationsprozesses sicher

Datenqualität und Data Governance sind für eine erfolgreiche Datenintegration unverzichtbar, insbesondere für KI-Anwendungen, die genaue, vollständige und vertrauenswürdige Daten erfordern. Datenqualität bezeichnet den Grad, in dem Ihre Daten die Erwartungen und Anforderungen Ihres KI-Projekts erfüllen. Data Governance bezeichnet die Richtlinien und Verfahren, mit denen die korrekte Verwaltung und Nutzung Ihrer Daten sichergestellt werden. Datenqualität und Data Governance müssen unbedingt während des gesamten Datenintegrationsprozesses von der Datenquelle bis zum Datenziel sichergestellt werden, um die Zuverlässigkeit, die Benutzerfreundlichkeit und den Wert Ihrer Daten zu verbessern. Dadurch wird auch die Einhaltung der Datenbestimmungen und -standards für Ihre Daten und Branche verbessert.

Best Practices für diese Phase:

- Legen Sie Rollen und Verantwortlichkeiten für die Datenqualität und Data Governance fest, z. B. Datenbesitzer, Data Stewards, Datenanalysten oder Datennutzer.
- Bestimmen Sie Kennzahlen und Indikatoren für Datenqualität und Data Governance, z. B. Genauigkeit, Vollständigkeit, Konsistenz, Aktualität oder Relevanz der Daten.
- Legen Sie Regeln und Standards für Datenqualität und Data Governance fest, z. B. Datendefinitionen, Datenformate, Datenschemas, Datenwerte oder Datenherkunft.
- Implementieren Sie Tools und Verfahren für Datenqualität und Data Governance, z. B. Datenprofilerstellung, Datenbereinigung, Datenanreicherung, Datenvalidierung, Datenprüfung oder Datenüberwachung.
- Überwachen und messen Sie die Leistung und Ergebnisse der Datenqualität und Data Governance, z. B. mit Berichten, Dashboards, Warnungen oder Feedback zur Datenqualität.
- Verbessern und optimieren Sie die Prozesse und Verfahren für Datenqualität und Data Governance, z. B. mit Plänen zur Optimierung der Datenqualität, Maßnahmen zur Verbesserung der Datenqualität, Best Practices für die Datenqualität oder Erkenntnissen zur Datenqualität.

Schritt 6: Überwachen und optimieren Sie die Leistung und die Ergebnisse der Datenintegration

Die Datenintegration ist ein fortlaufender Prozess, der eine kontinuierliche Überwachung und Optimierung erfordert. Da sich die Datenquellen, Datentypen und Datenmengen im Laufe der Zeit ändern, ändern sich auch die Anforderungen und Herausforderungen der Datenintegration. Durch die Überwachung und Optimierung der Leistung und der Ergebnisse der Datenintegration stellen Sie sicher, dass die Datenintegrations-Pipeline den erwarteten Mehrwert und Nutzen bringt. Dadurch können Sie auch die Chancen und Innovationen der Datenintegration erkennen und nutzen, mit denen sich Ihre KI-Funktionen und Ihre Wettbewerbsfähigkeit verbessern lassen.

Schritte in dieser Phase:

- Überwachen und messen Sie die Leistung und Ergebnisse der Datenintegration, z. B. ihre Geschwindigkeit, Effizienz, Zuverlässigkeit oder Qualität.
- Identifizieren und analysieren Sie Probleme und Engpässe bei der Datenintegration, z. B. Fehler, Verzögerungen oder Anomalien.
- Implementieren Sie die Verbesserungen und Optimierungen der Datenintegration, z. B. Erweiterungen, Upgrades oder Fehlerkorrekturen.
- Überprüfen Sie die Ergebnisse der Datenintegration und das entsprechende Feedback, z. B. Berichte, Dashboards, Warnungen oder Umfragen zur Datenintegration, und werten Sie diese aus.
- Aktualisieren und verfeinern Sie die Ziele und den Umfang der Datenintegration, z. B. Vorgaben, Ergebnisse oder Zeitleisten.
- Dokumentieren und kommunizieren Sie die Erkenntnisse und Best Practices zur Datenintegration, z. B. Dokumente, Schulung oder Wissensvermittlung für die Datenintegration.

Fazit

Datenintegration ist für den Erfolg mit KI von wesentlicher Bedeutung. Durch das Integrieren der Daten aus unterschiedlichen Quellen und in unterschiedlichen Formaten in ein einheitliches und verfügbares Datenökosystem können Sie Ihre Daten für die KI-Bereitstellung vorbereiten. Die Datenintegration stellt außerdem sicher, dass Ihre Daten genau, vollständig und vertrauenswürdig sind. Diese Eigenschaften sind für KI-Anwendungen, die sich auf datengestützte Erkenntnisse stützen, von grundlegender Bedeutung.

Mithilfe dieser Checkliste können Sie Ihre Datenmanagement-Verfahren optimieren, Risiken der Datenintegration ermitteln und mindern, Ihre Dateninfrastruktur für die KI-Bereitstellung vorbereiten und Erkenntnisse zur Nutzung der Datenintegration für erfolgreiche KI und Innovation gewinnen.



Über Precisely

Als weltweit führendes Unternehmen im Bereich Datenintegrität stellt Precisely sicher, dass Ihre Daten genau, konsistent und kontextbezogen sind. Unser Portfolio, einschließlich der Precisely Data Integrity Suite, hilft Ihnen, Daten zu integrieren, die Datenqualität zu verbessern, die Datennutzung zu regeln, Standortdaten zu geokodieren und zu analysieren sowie mit ergänzenden Datensätzen anzureichern, um sichere Geschäftsentscheidungen zu treffen. Über 12.000 Unternehmen in mehr als 100 Ländern, darunter 95 der Fortune 100, vertrauen auf Software, Daten und Strategiedienstleistungen von Precisely, um KI-, Automatisierungs- und Analyseinitiativen voranzutreiben. Erfahren Sie mehr unter www.precisely.com/de.

www.precisely.com/de