

Jan Krol

Data Expert

Ein Bild, das Text, blau, Schild enthält.

Automatisch generierte BeschreibungEin Bild, das Text, Schild, blau enthält.

Automatisch generierte BeschreibungEin Bild, das Text, Schild enthält.

Automatisch generierte BeschreibungEin Bild, das Text, Schild enthält.

Automatisch generierte BeschreibungEin Bild, das Text, Schild enthält.

Automatisch generierte Beschreibung

|  |  |
| --- | --- |
| Persönliche Daten |  |
|  | **Big Data Specialist**    Schwerpunkt: Big Data Cloud-Architektur, Datenmanagement-Plattformen |
|  | **Berufserfahrung:**   * Deutsche Bahn * Daimler Truck * Jungheinrich * Lufthansa Technik * Fielmann * BMW Group |
|  | **Sprachkenntnisse:**  Deutsch – Muttersprache Englisch – verhandlungssicher  Polnisch – verhandlungssicher |

|  |  |
| --- | --- |
| Fähigkeiten | * Spezialist in Big Data Plattformen auf Amazon Web Services & Microsoft Azure * ETL-Prozesse/Pipelines & Data-Engineering * Architektur von Datenmanagement-Plattformen in großen Unternehmen * Aufbau von Data Lakes & Data Lakehouse * Anwendungsmigrationen unter Verwendung von Cloud-Services * Beratung & Implementierung von Automatisierungskonzepten, insbesondere DataOps & DevOps * Integration von Active Directory Sicherheitskonzepten und Compliance-Anforderungen * Python, SQL, TypeScript, Golang |

|  |  |
| --- | --- |
| Technische und methodische Fähigkeiten | * Big Data Cloud Architekturen (AWS & Microsoft Azure) * Data Engineering (Databricks, Synapse Analytics, Fabric, Apache Spark, AWS Glue, Athena, Redshift & EMR) * Infrastructure as Code (Terraform, Pulumi, AWS CDK, ARM) |

|  |  |
| --- | --- |
| Zertifizierungen | * Azure Solutions Architect Expert:   + AZ-300: Microsoft Azure Architect Technologies   + AZ-301: Microsoft Azure Architect Design * Databricks Certified Associate Developer   + for Apache Spark 3.0 * HashiCorp Certified: Terraform Associate * AWS Certified Machine Learning – Specialty * AWS Certified Solutions Architect – Associate * AWS Certified Cloud Practitioner * AWS Technical Professional * AWS Business Professional |
| Projekte  **06/2024**  **| @ Industrie**  **laufend**  **03/2023**  **| @ Intralogistik**  **05/2024**  **01/2022**  **| @ Logistik**  **02/2023**  **09/2021**  **| @ E-Commerce**  **01/2022**  **04/2020**  **| @ E-Commerce**  **09/2021**  **02/2019**  **| @ Transport & Logistik**  **04/2020**  **09/2018**  **| @ Transport & Logistik**  **02/2019**  **vor 2018** | * **Modernisierung der Datenarchitektur mit Azure Databricks Live Tables und Databricks Asset Bundle** * ***Projektzusammenfassung****:* * *In diesem Projekt fungierte ich als technischer Leiter eines Teams von 3 Offshore-Entwicklern bei der Implementierung skalierbarer und robuster Datenlösungen in Azure Databricks. Ein Hauptaugenmerk lag auf der Einführung von Databricks Live Tables für Schema- und Tabellenverwaltung sowie der Implementierung von Databricks Asset Bundle nach dem Infrastructure as Code (IaC)-Prinzip. Ich entwarf und optimierte die Medallion-Datenarchitektur zur Verbesserung der Datenverarbeitungsabläufe und arbeitete eng mit verschiedenen Geschäftsbereichen zusammen, um sicherzustellen, dass die Datenlösungen ihren spezifischen Anforderungen entsprachen.*   + Technische Führung und Teammanagement     - Führung und Mentoring eines Teams von 3 Offshore-Entwicklern     - Etablierung von Codierungsstandards und Best Practices für das Entwicklungsteam     - Durchführung von Code-Reviews und Bereitstellung technischer Anleitungen     - Durchführung von Wissenstransfer und technischen Weiterbildungssitzungen   + Architektur und Implementierung     - Entwurf und Verfeinerung der Medallion-Datenarchitektur (Bronze, Silber, Gold)     - Einführung von **Databricks Live Tables** für effiziente Schemaevolution und Tabellenverwaltung     - Einführung von **Databricks Asset Bundle** mit IaC-Prinzipien für automatisierte, versionskontrollierte Bereitstellungen     - Entwicklung skalierbarer ETL-Pipelines in Azure Databricks     - Erstellung optimierter Datenspeicherlösungen mit Blick auf zukünftige Skalierbarkeit   + Infrastructure as Code und DevOps-Integration     - Einrichtung eines vollständigen IaC-Workflows für Datenplattformkomponenten     - Versionskontrolle und CI/CD-Integration für Databricks Asset Bundles     - Automatisierte Bereitstellung von Tabellenschemata, Jobs und Notebooks durch IaC-Pipelines     - Implementierung von Umgebungsförderungsstrategien (Dev/Test/Prod)     - Konfigurationsmanagement für umgebungsübergreifende Konsistenz   + Services     - **Azure Databricks**     - **Databricks Live Tables**     - **Databricks Asset Bundle**     - **Azure Data Factory**     - **Delta Lake**     - **Spark SQL**     - **Azure Key Vault**     - **Azure Storage**     - **Power BI (Integration von Datenkonsumenten)** * **Entwicklung einer umfassenden Datenstrategie und eines Governance-Frameworks für eine Datenmanagementplattform auf Databricks** * ***Projektzusammenfassung****:*   *In diesem „Leuchtturm“-Projekt leitete ich die Entwicklung einer robusten Datenstrategie und eines Governance-Frameworks mit dem Ziel, die Datenverarbeitungskapazitäten der Organisation zu optimieren und zu verbessern. Kern des Projekts war der Aufbau einer hochperformanten Datenmanagementplattform auf Databricks, ergänzt durch das Design und die Implementierung einer effizienten Data-Hub-Ingest-Plattform.*   * + Leitung des Designs und der Etablierung einer unternehmensweiten Datenstrategie, ausgerichtet an Geschäftszielen und technologischen Fortschritten   + Entwicklung eines umfassenden Daten-Governance-Frameworks zur Gewährleistung von Datenqualität, Datenschutz und Einhaltung von Branchenstandards   + Überwachung der Bereitstellung und Anpassung der Datenmanagementplattform auf Databricks, Verbesserung der Datenverarbeitung, Analyse und Reportingfähigkeiten mit Power BI   + Entwicklung eines robusten Data-Hubs mit hochperformanter Ingest-Pipelines basierend auf AWS EventBridge,   + Optimierung des Datenflusses von verschiedenen Quellen zu zentralisierten Speichersystemen (Data Lake House auf Azure)   + Zusammenarbeit mit funktionsübergreifenden Teams zur Integration der Datenmanagementplattform in die bestehende IT-Infrastruktur und Geschäftsprozesse   + Durchführung von Schulungen und Workshops für neue Teams, Förderung einer datengetriebenen Kultur und Verbesserung der Datenkompetenz in der gesamten Organisation   + Services:     - **Azure Databricks**     - **Databricks Data Catalog**     - **AWS EventBridge**     - **Kinesis**     - **Event Hub**     - **Structured Streaming (Apache Spark)** * **Innovative Integration und Analyse von Logistikdatenströmen mit PySpark Structured Streaming und Data Mesh-Implementierung** * ***Projektzusammenfassung:***   *Dieses Projekt konzentrierte sich auf die anspruchsvolle Integration von Logistikdatenströmen mit Event Hub und Kafka unter Verwendung von PySpark Structured Streaming. Unser Ansatz revolutionierte die Art und Weise, wie Logistikdaten in Echtzeit erfasst, verarbeitet und durch den Graphen-Ansatz verknüpft werden konnten. Durch den Einsatz von Technologien wie GraphFrame, Azure Synapse Analytics, Apache Spark und PowerBI etablierten wir ein robustes System, das nicht nur eine hohe Datenqualität und reibungslose Übertragung gewährleistet, sondern auch den IT-Governance-Prinzipien entspricht.*   * + Integration eines Logistikdatenstroms mit Kafka durch PySpark Structured Streaming   + Definition der notwendigen Datenstrukturen für den Datenstrom   + Robuste und effiziente Integration des Logistikdatenstroms mit Event Hubs   + Echtzeitnutzung von Logistikdaten für Analyse und Weiterverarbeitung   + Entwurf und Implementierung von Pipelines zur Erfassung, Verarbeitung und Weiterleitung des Datenstroms   + Effiziente Datenverarbeitung mit PySpark Structured Streaming   + Konfiguration und Initialisierung des PySpark Streaming-Jobs   + Durchführung umfassender Test- und Überwachungsmechanismen   + Gewährleistung einer reibungslosen Datenübertragung und hoher Datenqualität   + Services:     - **Azure Synapse Analytics**     - **Purview Data Catalog**     - **Event Hub**     - **GraphFrame**     - **PowerBI** * **Verbesserte Datenverarbeitungs- und Integrationssysteme für E-Commerce mit serverlosen und verteilten Data-Mesh-Architekturen** * ***Projektzusammenfassung:***   *In diesem Projekt war meine Hauptaufgabe, verschiedene interne E-Commerce-Produktteams bei der Entwicklung, Implementierung und Wartung leistungsfähiger Datenverarbeitungs- und Integrationssysteme zu leiten und zu unterstützen. Der Schwerpunkt lag auf der Migration bestehender Datendienste und -pipelines zu einer neuen, verbesserten Architektur, wobei die Entwicklung eines eventbasierten Systems unter Verwendung serverloser Technologien und Big-Data-Frameworks betont wurde.*   * Unterstützung und Anleitung bei der Migration der bestehenden Datendienste, Pipelines und Assets auf eine neue und weiterentwickelte Architektur * Entwicklung eines eventbasierten Systems * Verwendung von Lambda-Funktionen und PySpark * Integration mit Kafka * Design- und Architekturplanung * Implementierung von Lambda-Funktionen und PySpark-Jobs * Konfiguration und Verbindung mit Kafka * Serverlose Architektur für Skalierbarkeit und Verfügbarkeit * Verarbeitung und Analyse von Ereignisdaten in Echtzeit * PySpark-Transformationen, Filterungen und Aggregationen * Effiziente und zuverlässige Verbindung mit Kafka * Konfiguration, Sicherheitseinstellungen und Integration mit anderen Komponenten * Umfangreiche Tests und Überwachungsmechanismen * Hochperformantes und skalierbares Event-System * Gewinnung wertvoller Erkenntnisse aus Ereignisdaten * Datengetriebene Entscheidungsfindung * Services:   + **AWS Glue**   + **Apache Spark**   + **Data Catalog**   + **Athena**   + **Redshift**   + **Lambda**   + **ECS**   + **Step Functions** * Umsetzung verteilter Data Mesh Architekturen, damit verschiedene Produktteams effizient mit Daten arbeiten können * Datenverarbeitung mit Big-Data-Frameworks und Datenbanktechnologien * Erarbeitung serverlose/elastische Cloud-Architektur (AWS) * Bereitstellung der Architektur nach DevOps-Best-Practices und Infrastructure-as-Code (AWS CDK & Terraform) * **Migration und Verbesserung der E-Commerce-Datenplattform zur AWS Data Lakehouse-Architektur** * ***Projektzusammenfassung:***   *Dieses Projekt beinhaltete die strategische Entwicklung und Migration bestehender Analytics-Datenpipelines in eine Data Lakehouse-Architektur unter Nutzung von AWS-Diensten. Ein wesentlicher Aspekt war die Verbesserung der Big-Data-Lake-Umgebung und die Gewährleistung strenger Datenqualitäts- und Compliance-Standards, insbesondere im Hinblick auf die DSGVO.*   * Weiterentwicklung der Big Data Lake Umgebung in **AWS** * Umsetzung und Konzeption eines **Data Lakehouse** * explorative Analyse sowie Algorithmenentwicklung durch Datenbereitstellung und -aufbereitung (**AWS Glue, Spark, Lambda**) * Data Ingestion * Entwicklung von Data Pipelines, ETL-Jobs zur Bereitstellung von konsumfertigen Datenquellen **(AWS Glue, AWS Redshift, Spark, PySpark)** * Regressionstest und Qualitätsprüfung in den Datenstrecken und Data Lake * Orchestrierung und Anbindung von Datenquellen * Umsetzung von automatisieren Deployments mittels DevOps-Best-Practices (**AWS Codebuild + Codepipeline**, **GitHub Actions**) * Aufbau der Infrastruktur mittels IaC (AWS **CDK**) * Systembetreuung (inkl. Kostenmonitoring) * **Entwicklung einer Architektur und Implementierung einer Big-Data-Umgebung für konzernweite, standardisierte Plattformdienste auf Basis von Azure** * ***Projektzusammenfassung:***   *Dieses Projekt umfasste die Entwicklung und Implementierung einer standardisierten Big-Data-Architektur für konzernweite Plattformdienste im Transport- und Logistiksektor unter Verwendung verschiedener Azure-Dienste. Meine Rolle war entscheidend, um die Integration von Datentransparenz, Datenqualität, DataOps, die Einhaltung von Datenvorschriften und die Implementierung agiler Methodologien sicherzustellen.*   * + Erarbeitung von Lösungen in Projekten zu Azure und Automatisierung und Vorstellung/ Diskussion derselben   + Azure-Services: Azure Data Catalogue, Azure Synapse Analytics, Azure Data Factory, Azure Databricks   + Automatisierter Aufbau der Infrastruktur mit Infrastructure as Code (Terraform) und Ansible   + Scrum, JIRA, Gitlab, Docker   + Umsetzung Echtzeit-Datenübertragung mit Apache Kafka   + Beratung zu Azure Plattformstrategie bzgl. Referenzarchitekturen   + Entwicklung von Mechanismen und Automatismen zur proaktiven Beseitigung von Schwachstellen der Azure- und Kubernetes-Komponenten auf Basis standardisierter Cluster (Security by default)   + Konzeptionelle Weiterentwicklung der architektonischen und technologischen Plattform im Bereich Container-Orchestrierung auf Basis von Kubernetes, Continuous Integration & Continuous Deployment   + Erstellung Benutzer- und Berechtigungskonzepte unter Berücksichtigung der Konzernvorgaben   + Betriebsführung der angebotenen Services   + Agil arbeitendes Team   + Services:     - **Azure Data Catalogue (Purview),**     - **Azure Synapse Workspace Analytics,**     - **Azure Data Factory,**     - **Azure Databricks**     - **Terraform**     - **Gitlab Runner**     - **Azure DevOps** * **AWS-Infrastrukturberatung und -implementierung für globale Prozessoperationen im Transport- und Logistikbereich** * ***Projektzusammenfassung:***   *Dieses Projekt umfasste die Beratung und die praktische Implementierung einer AWS-Infrastruktur zur Unterstützung eines Prozessbetriebsteams, das für mehrere internationale Anwendungen im Transport- und Logistiksektor verantwortlich ist. Meine Rolle war entscheidend bei der Identifizierung und Implementierung von Optimierungen, der Entwicklung und Wartung der kritischen Systeminfrastruktur sowie der umfassenden Unterstützung und Schulung interner Teams.*   * + Bereitstellung und Betrieb von Servern, Betriebssystemumgebungen und Datenbanksystemen in AWS   + Erkennen von Optimierungspotenzialen in kaufmännischer als auch technischer Sicht   + Erarbeitung und Präsentation von optimierten Prozessen   + Implementierung von Optimierungen (AWS Lambda boto3)   + Eigenverantwortliches Handeln und vertreten der erarbeiteten Ergebnisse im Team und gegenüber den Projektleitern/Auftraggebern   + Administration und Wartung der bereitgestellten Systeme   + Erarbeitung von Wartungs- und Monitoringkonzepten für diese Systeme   + Unterstützung und Beratung von Entwicklungsprojekten zu Nutzung, Konfiguration und Optimierung der bereitgestellten Systeme   + Beratung zu Architekturen und Betriebskonzepten unter Nutzung der AWS Cloud Infrastrukturen   + Beim Einsatz neuer AWS Services: Schulung der internen Mitarbeiter für die veränderten Arbeitsweisen   **Anwendungsmigration für einen Geschäftsbereich (Transport & Logistik**) inkl. Aufbau AD   * + Bereitstellung von AWS Infrastruktur: Datenbanken (SQL) & EC2-Maschinen, sowie Lambda Services   + Deployment mittels Terraform   + Planung und Ausführung der Anwendungsmigration   + Ausrollen von Berechtigungen   + Infrastrukturbereitstellung mitAWS Cloudformation   **Realisierung und technische Projektunterstützung** **einer Web-Applikation zur Verwaltung des Zertifizierungsprozesses unter Berücksichtigung von Compliance-Vorgaben für einen namhaften Automobilhersteller**   * + Infrastrukturbereitstellung auf AWS (MySQL Server)   + Front End Entwicklung in React.js   + Back End Realisierung mit Java EE | glassfish   + Refactoring und Code Optimization   + Agile Arbeitsweise durch Scrum mit Jira   + CI/CD mit Jenkins   + Dokumentation in Confluence   **Aufbau und Consulting einer Marktplatzplattform basierend auf Microsoft Azure Services**   * + Integration Payment-Provider   + Planung und Architektur von Microsoft Azure Services   + Realisierung/Beratung von Back End und Front End   + Erstellung User Management mittels Active Directory   + Implementierung Upload-Tool zum Hochladen von sehr großen Dateien direkt aus dem Webbrowser   + Security Engineering   + **Realisierung Prozessautomatisierung: Implementierung eines automatischen Ticketgenerators basierend auf Schwachstellen-Scan (Vulnerability Assessment)**   + Design und Planung der kosteneffizientesten Infrastrukturkomponenten   + Implementierung von Python-Logik in AWS Lambda   + Bereitstellung der Infrastruktur durch AWS Cloudformation via YAML-Templates   + Code Optimization   + Senden von Benachrichtigungsmails via AWS SNS |
|  |  |
|  |  |
|  |  |
|  |  |