

Während das Verlangen nach Daten bei Post Immobilien (IMS) wächst, kämpfen die Datenspezialisten mit wachsenden Datenbeständen, grösserer Komplexität, steigenden Kosten und eingeschränkten Datenzugriffen wegen Sicherheitsbedenken. Mit der Einführung einer agilen Analytics-Methode sollen Menschen, Prozesse und Technologien so eingesetzt werden, dass der Datenaufbereitungsprozess nicht zum Flaschenhals wird und die Barrieren für den Datenzugang minimiert werden können.

Aktuelle Herausforderungen in der Datenanalyse bei IMS

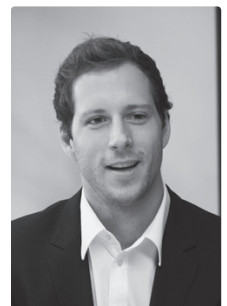
In dem heutigen schnelllebigen Markt ändern sich die Anforderungen an das Datenanalyse-Team und an die Anwendungsfälle ständig. Nebst den volatilen Anforderungen steigt die Menge der zu verarbeitenden Daten exponentiell an. Die Daten werden oftmals in separaten Datensilos gesammelt, welche normalerweise nicht miteinander kommunizieren und zudem unterschiedliche Technologien, Plattformen und Schnittstellen nutzen. Zudem unterscheiden sie sich in Format, Granularität und sind in der Regel nicht so strukturiert, dass sie für die effiziente Erstellung von Analysen geeignet sind. Datenfehler und schlechte Datenqualität führen zu Problemen in der Datenpipeline oder zu falschen Erkenntnissen. Der Datenpipelineunterhalt endet durch aktualisierte Datenquellen, Schema-Erweiterungen oder Verbesserung der Analysen nie und das Überwachen der Pipeline ist aufwändig. Die Datenintegration, Bereinigung, Transformation, Qualitätssicherung, das Deployment neuer Analysen und das Bereitstellen von Entwicklungsumgebungen sind schwerfällige und zeitaufwändige manuelle Prozesse.

Durch die Leistungsbezugspflicht von ICT-Dienstleistungen, müssen sämtliche Datenanalyse-Projekte bei Post Informatik (I) bezogen werden. Die Koordination der unterschiedlichen Teams auf Seite I ist schwierig, das Branchenwissen fehlt und die Personal-Ressourcen sind beschränkt. Die Projekte werden nach der Wasserfallmethodik umgesetzt, was zu mangelnden Projektergebnissen, langen Implementierungszeiten und zu hohen Kosten führt.

Die Herausforderungen überwinden mit DataOps

In dieser Masterarbeit wird eine agile Datenanalyse-Methode auf Basis von DataOps und der agilen Projektmethodik von IMS erarbeitet. Während Agile4IMS den bei IMS bekannten agilen Rahmen liefert, bringt DataOps die Vorzüge von DevOps, statistischer Prozesskontrolle aus Lean Manufacturing

und agiler Entwicklung mit ein. Das Ergebnis ist ein Framework für den End-to-End-Datenlebenszyklus, bei dem neue Prozesse, Methoden und Tools, auf die Datenaufbereitung und Datenauswertung angewendet werden, um die Kommunikation und Integration von Datensilos, Teams und Systemen zu fördern. Dabei wird der Datenanalyse-Prozess als Pipeline-Prozess betrachtet. Es gibt zwei Pipelines; eine Wertschöpfungs-pipeline und eine Entwicklungspipeline. Bei der Wertschöpfungs-pipeline fließen Daten kontinuierlich auf der einen Seite der Pipeline hinein, durchlaufen eine Reihe von orchestrierten Schritten und produzieren auf der anderen Seite einen Wert, zum Beispiel in Form von Berichten, Modellen oder Visualisierungen. Das Verwalten, Orchestrieren und Überwachen dieser Pipeline ist die operative Seite (Ops) der Datenanalyse. Die Entwicklungspipeline regelt das Weiterentwickeln und Aktualisieren der Wertschöpfungs-pipeline. Hier veröffentlicht das DataOps-Team neue oder aktualisierte Datenprodukte in kurzen Abständen in sogenannten Sprints. Mit kontinuierlicher Integration und kontinuierlicher Bereitstellung werden zwei grundlegende Konzepte von DevOps angewendet. Anders als in der Wertschöpfungs-pipeline, wo der Code gleich bleibt und sich die Daten ständig verändern ist es in der Entwicklungspipeline umgekehrt. Die einzelnen Datenverarbeitungsschritte werden als parametrisierbare und containerisierte Micro Services entwickelt. Diese kommunizieren über einen Publish/Subscribe-Message-Broker miteinander. Dabei kommt das Smart Endpoints and Dumb Pipes-Pattern zur Anwendung. Die Micro Services werden so zerstückelt, dass die Wiederverwendbarkeit möglichst gross ist. Bei der Entwicklung wird darauf geachtet, dass Stories so geschrieben werden, dass nach jeder Iteration ein Wert entsteht. Dies gelingt, indem die Stories anhand der Datenwertpyramide geschnitten werden. Das interdisziplinäre DataOps-Team ist für die gesamte Data Factory verantwortlich und stellt die Weiterentwicklung und Qualität beider Pipelines sicher.



Kevin Gyga