

## Apache Hop – Next Generation Data Integration

### Short Facts

Um Euch die Lektüre langer Ausschreibungstexte zu ersparen, hier alle Infos in Kürze:

- **Wir suchen:** Studierende für eine Bachelor-/Masterarbeit aus der Fachrichtung (Wirtschafts-) Informatik - Interesse am Thema und Motivation sind uns wichtig
- **Thema:** Apache Hop - Next Generation Data Integration – Gerne könnt Ihr aber auch mit Vorschlägen an uns herantreten!
- **Standort:** Frankfurt am Main oder Leipzig – Remotearbeit ist nach Absprache möglich
- **Wir bieten:** Eine gute Vergütung, sowie die Teilnahme an Firmenevents (Monatsmeetings, Team-Workshop, Social Events, etc.)
- **Chancen:** Bei erfolgreichem Abschluss der Arbeit besteht eine hohe Chance zur Übernahme in ein Arbeitsverhältnis.

Falls Ihr Lust habt, in einem jungen Beratungsunternehmen zu arbeiten, freuen wir uns auf Eure Bewerbung an [study@integration-factory.de](mailto:study@integration-factory.de) oder in unserer [Stellenbörse](#).

Wenn Ihr weitere Informationen braucht, oder andere Fragen habt, schaut doch einfach mal auf unserer Homepage vorbei ([www.integration-factory.de](http://www.integration-factory.de)) oder schreibt uns eine Mail an die oben genannte Adresse.

### Wer wir sind

integration-factory ist ein mittelständiges Beratungshaus in Frankfurt am Main und Leipzig. Wir bieten Lösungen im Bereich Business Intelligence, Data Warehousing und Big Data und beraten namhafte Kunden aus verschiedensten Branchen (Finanzwirtschaft, Logistik, uvm.) Dabei setzen wir auf innovative Technologien der Datenintegration und begleiten unsere Projekte von der Konzeption bis zur Realisierung.

### Was wir suchen

Mit Hop (Hop Orchestration Platform) hat die Apache Foundation eine auf Metadaten basierende Open-Source-Plattform für Datenintegration, Data Engineering und Datenorchestrierung entwickelt. Im Gegensatz zu vielen am Markt verfügbaren Tools setzt Hop vollständig auf eine visuelle Entwicklung von Workflows und Pipelines per Drag-and-Drop, was das Schreiben von Code nahezu unnötig macht. Laut Apache eignet sich Hop vor allem für das Laden großer Datenmengen in Datenbanken unter Nutzung von massiv-parallelen Verarbeitungsumgebungen in der Cloud oder auf on-Premise-Clustern. Prinzipien wie Slowly-Changing-Dimensions (SCD) und Change Data Capture (CDC) bei der Befüllung von Datawarehouse-Umgebungen werden ebenso unterstützt wie verschiedenste Datenbankarchitekturen (relational, NoSQL ...).

Im Rahmen der Abschlussarbeit soll die Funktionsweise von Apache Hop untersucht und exemplarisch an verschiedenen Use Cases aus den Bereichen DWH und Big Data verprobt werden. Dazu sollen die notwendigen Datenintegrationsprozesse/-pipelines mittels Hop entwickelt werden. Des Weiteren soll der Funktionsumfang von Hop mit anderen Tools zur Datenintegration bspw. Informatica verglichen und Vor- und Nachteile ausgearbeitet werden.