

## **BA-Abschlussarbeit**

Echtzeit-Fusion von heterogenen Datenströmen

### **Kurzbeschreibung**

Durch eine steigende Anzahl von Datenquellen, die kontinuierlich zu analysierenden Daten liefern, stößt das Prinzip des Batch-Processings an seine Grenzen. Stream Processing ist eine Methode zur Verarbeitung von Daten, während sie erzeugt werden, und ermöglicht Analysen und Entscheidungen in Echtzeit. Die Stream-Verarbeitung ergänzt häufig Workflows der Batch-Verarbeitung, bei denen eine Menge von Daten auf einmal verarbeitet wird. Während Stream-Prozessoren in der Regel kontinuierlich laufen, werden Batch-Jobs in der Regel nach einem Zeitplan ausgeführt. Apache Flink ist ein Framework und eine verteilte Processing Engine für Stateful Computations über unbeschränkte und beschränkte Datenströme. Flink wurde so konzipiert, dass es in allen gängigen Cluster-Umgebungen läuft und Berechnungen mit In-Memory-Geschwindigkeit und in beliebigem Umfang durchführt. Um einen unbegrenzten Datenstrom zu verarbeiten, werden "Sliding Window" Queries verwendet. Die naive Verarbeitung dieser Queries führt oft zu einer Menge redundanter Berechnungen, die die Leistung unnötig verringern. Im Laufe der Jahre wurden verschiedene Ideen vorgeschlagen, die sich die Eigenschaften von "Sliding Window" Queries zunutze machen, um überflüssige Berechnungen zu reduzieren. In diesem Projekt beginnen wir mit dem Konzept des Sliding Window und untersuchen die Möglichkeit, Informationen aus verschiedenen Quellen mit dieser Technik zu fusionieren. Die Sensorfusion ist ein mehrstufiger Prozess, der sich auf die Assoziation, Korrelation und Kombination von Daten und Informationen aus verschiedenen Quellen konzentriert. Das Ziel ist es, eine höhere Genauigkeit und bessere Erkenntnisse zu erzielen, als dies mit einem einzelnen Sensor allein möglich wäre. Die Herausforderung bei der Datenfusion von Datenströmen liegt in den zeitlichen Einschränkungen der Berechnung, den unterschiedlichen Zeitauflösungen der Sensoren, die die Daten erzeugen, und der Genauigkeit der erwarteten Ergebnisse. Dieses Projekt erforscht die Anwendung traditioneller Algorithmen zur Sensorfusion auf Datenströme, genauer gesagt als Verarbeitung auf Sliding Windows. Es wird ein Benchmark klassischer Algorithmen und deren Bewertung anhand von Beispieldatenströmen durchgeführt.

### **Tasks**

- Einführung in Stream Computation, Sliding Window Queries und Apache Flink
- Einführung in die Datenfusion und Auswahl von Datenfusionsalgorithmen
- Implementierung von Sliding Window Datenfusionsalgorithmen in Apache Flink
- Evaluierung und Tests mit Benchmark-Datenströmen für die Datenfusion

### **Voraussetzungen**

- Gute Programmierkenntnisse (Python, Java)
- Gute Mathematikkennntnisse

### **Betreuer**

Prof. Dr. Ing. Cristian Axenie, M.Sc.