University of
Zurich^UZH

Department of Informatics

University of Zürich
Department of Informatics
Binzmühlestr. 14
CH-8050 Zürich
Phone. +41 44 635 43 11
Fax +41 44 635 68 09
www.ifi.uzh.ch/dbtg

UZH, Dept. of Informatics, Binzmühlestr. 14, CH-8050 Zürich

Prof. Dr. Michael Böhlen
Professor
Phone +41 44 635 43 33
Fax +41 44 635 68 09
boehlen@ifi.uzh.ch

Zürich, February 19, 2018

Bachelor's Thesis: Implementation and Performance Evaluation of Multi-dimensional Fast Fourier Transform Algorithms on Streaming Data using Apache Flink

Centralized database systems and large distributed systems have served applications for decades, but the rate of data production poses new challenges to data processing. With centralized systems, the database acts as a single source of truth, which makes it slow if there are many data ingestion pipelines and makes it a single point of failure. With distributed systems at scale it gets a challenge to maintain a consistent global state. To overcome these challenges, streaming architectures have been proposed. Such systems allow data records to continuously flow from data sources to applications and between applications. There is no single database that holds the global state for the entire data. Instead continuously moving streams provide local consistency.

There are many domains that have to process massive data streams, one of which is Radio Astronomy. The Australia Square Kilometer Array Pathfinder (ASKAP) roughly produces 2.5 GB/s, approximately 216 TB/day and 100 PB/year. One of the most basic and fundamental operations applied on Radio Astronomy raw data is the Fast Fourier Transform (FFT) to produce images. The goal of this Bachelor thesis is to design, implement and evaluate FFT algorithms on the Apache Flink streaming platform.

What is Discrete Fourier Transform And Fast Fourier Transform?
There are cases where we need to determine the frequency content of a time-domain signal. The discrete Fourier transform (DFT) converts a finite sequence of equally-spaced samples of a function into a same-length sequence of equally-spaced samples of the discrete-time Fourier transform (DTFT), which is a complex-valued function of the frequency.

$$F[n] = \sum_{k=0}^{N-1} f[k] \exp^{-j\frac{2\pi}{N}nk} \quad where\ n = 0\ to\ N-1 \tag{1}$$

F[n] is the Discrete Fourier Transform of sequence f[k]. For 1-D data the complexity of DFT is $O(N^2)$.

An FFT computes the DFT and produces exactly the same result as evaluating the DFT definition directly. The most important difference is that an FFT is much faster with time complexity O(N logN). A two dimensional DFT is given as:

$$F(u,v) = \sum_{m=0}^{N-1}\sum_{n=0}^{N-1} f[m,n]\exp^{-j2\pi\frac{um}{M}+\frac{vn}{N}} \tag{2}$$

2D DFT/FFT is carried out by 1D transforming all rows of the 2D function $f[m,n]$ and then 1D transforming all columns of the resulting matrix. The order of the steps is not important [2].

Tasks

1. Literature study on Apache Flink [1] and Fast Fourier Transform [2][3].
2. Build a streaming pipeline for Calibrated Visibility Dataset as follows.

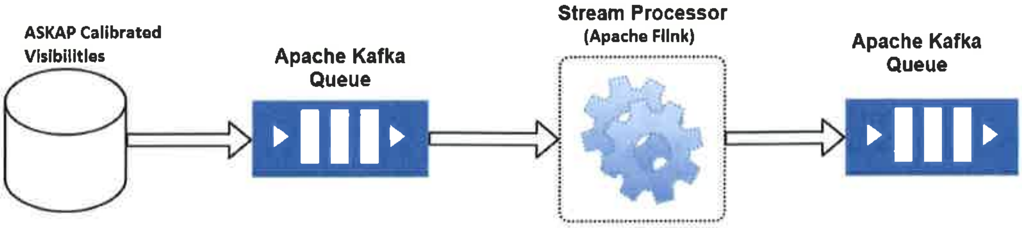
ASKAP Calibrated Visibilities
Apache Kafka Queue
Stream Processor (Apache Flink)
Apache Kafka Queue

figure 1: Streaming Pipeline

3. Implement Radix-2, Radix-4 and Split Radix FFT Algorithms.
4. Evaluate the run-time performance of the above mentioned algorithms.
5. Write a thesis (approximately 50 pages).
6. Present your thesis in a DBTG meeting

Optional Task

1. Implement parallel version of Radix-2 in Apache Flink [4].

References

1. Carbone, P., Katsifodimos, A., Ewen, S., Markl, V., Haridi, S. et al. (2015): Apache flink: Stream and batch processing in a single engine
2. $www.fourier.eng.hmc.edu/e101/lectures/Image_processing/node6.html$
3. $www.engineering.purdue.edu/\ ipollak/ee438/FALL04/notes/Section1.4.pdf$

4. G. Xie and Y. c. Li, "Parallel Computing for the Radix-2 Fast Fourier Transform," 2014 13th International Symposium on Distributed Computing and Applications to Business, Engineering and Science, Xian Ning, 2014, pp. 133-137.

Supervisor: Muhammad Saad (saad@ifi.uzh.ch)
Start Date: 19 February, 2018
End Date: 19 August, 2018
Presentation Date: 21 August,2018 @14.00

University of Zurich
Department of Informatics

Prof. Dr. Michael Böhlen
Professor