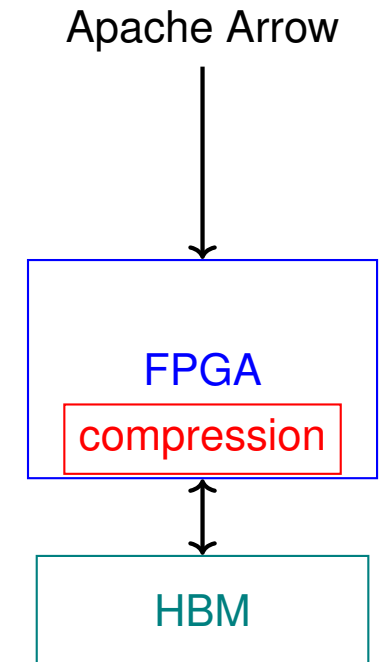


Column Compression on HBM

Background Apache Arrow defines a language-independent, columnar in-memory format for tabular data. Each column is represented by one or more contiguous memory buffers, enabling efficient analytical processing and data exchange.

While Arrow's memory layout offers high performance, it incurs significant memory overhead when processing large-scale datasets, thereby consuming substantial memory resources.

Arrow buffers are uncompressed by default, but the format supports compact representations such as dictionary encoding and run-end encoding. In this work we want to see what FPGA-friendly compression techniques can be applied efficiently.



Suggested Literature

- Integrating Lightweight Compression Capabilities into Apache Arrow
- RLE, FSST: fast random-access string compression, ...

Prerequisites: Basic knowledge in C++ and VHDL or Verilog

Type of Work: Theory (30%), Conception (30%), Implementation (40%) (BA, MA)

Supervisor: Chen Lei (lei.chen@tu-ilmenau.de)