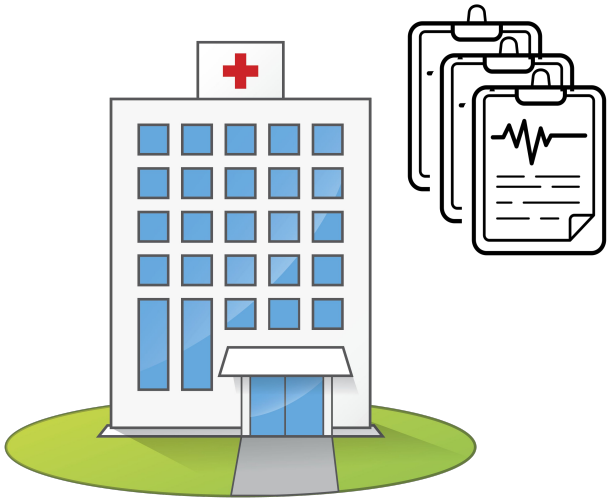
Secure Data Processing at Scale

Kajetan Maliszewski

Technische Universität Berlin

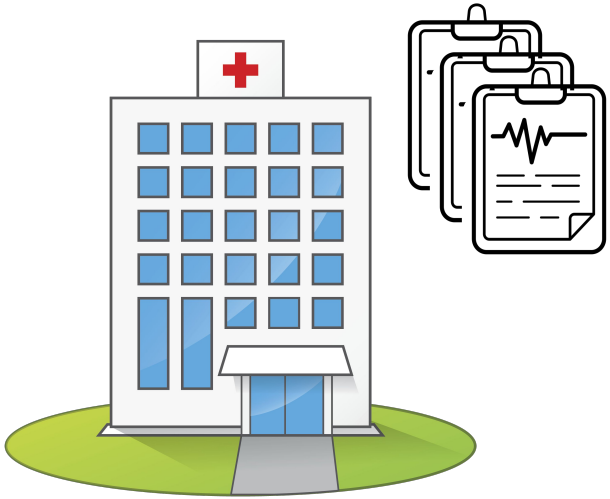
PhD@VLDB 2020

Motivation

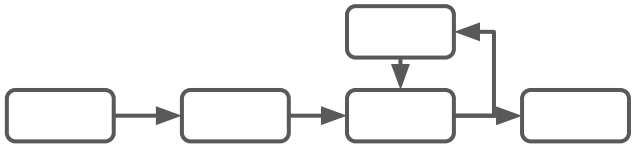


Medical records

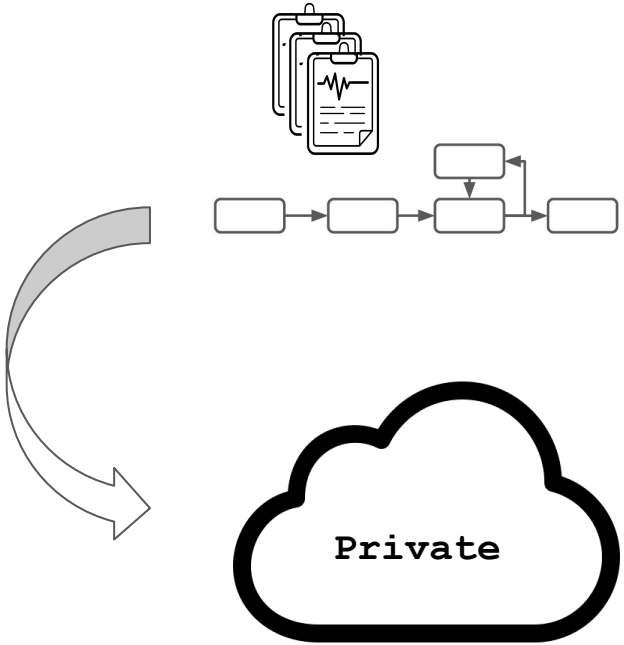
Motivation

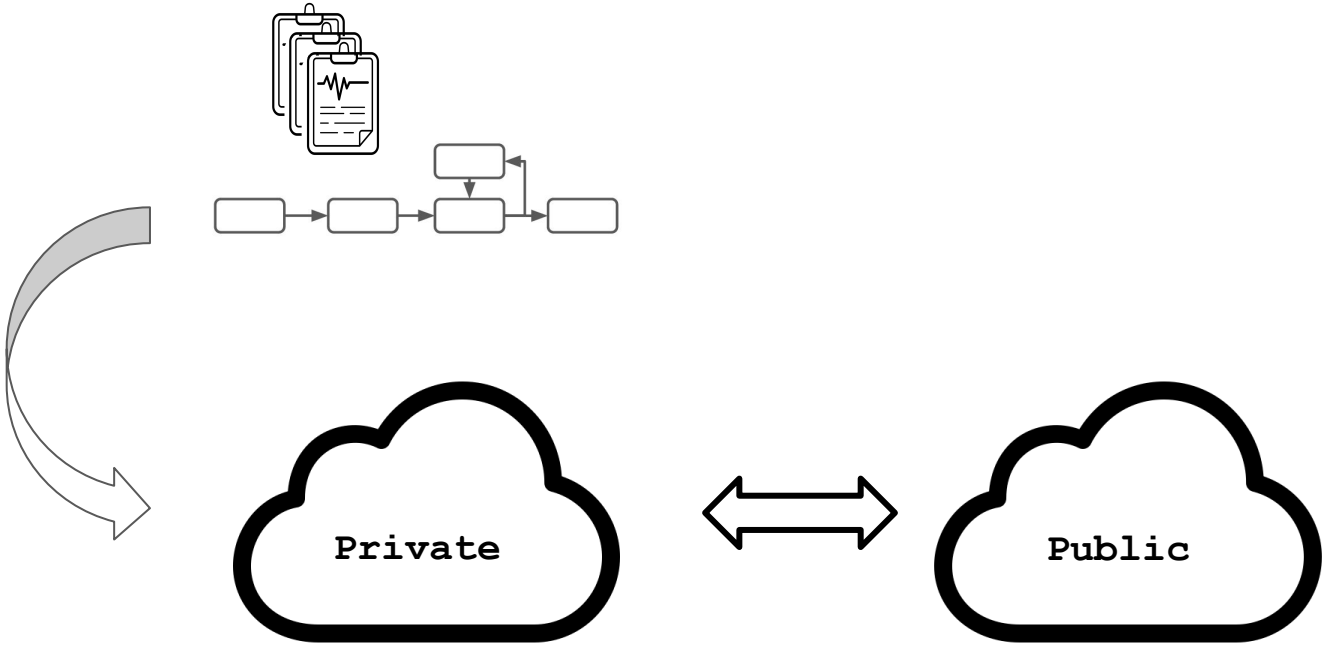


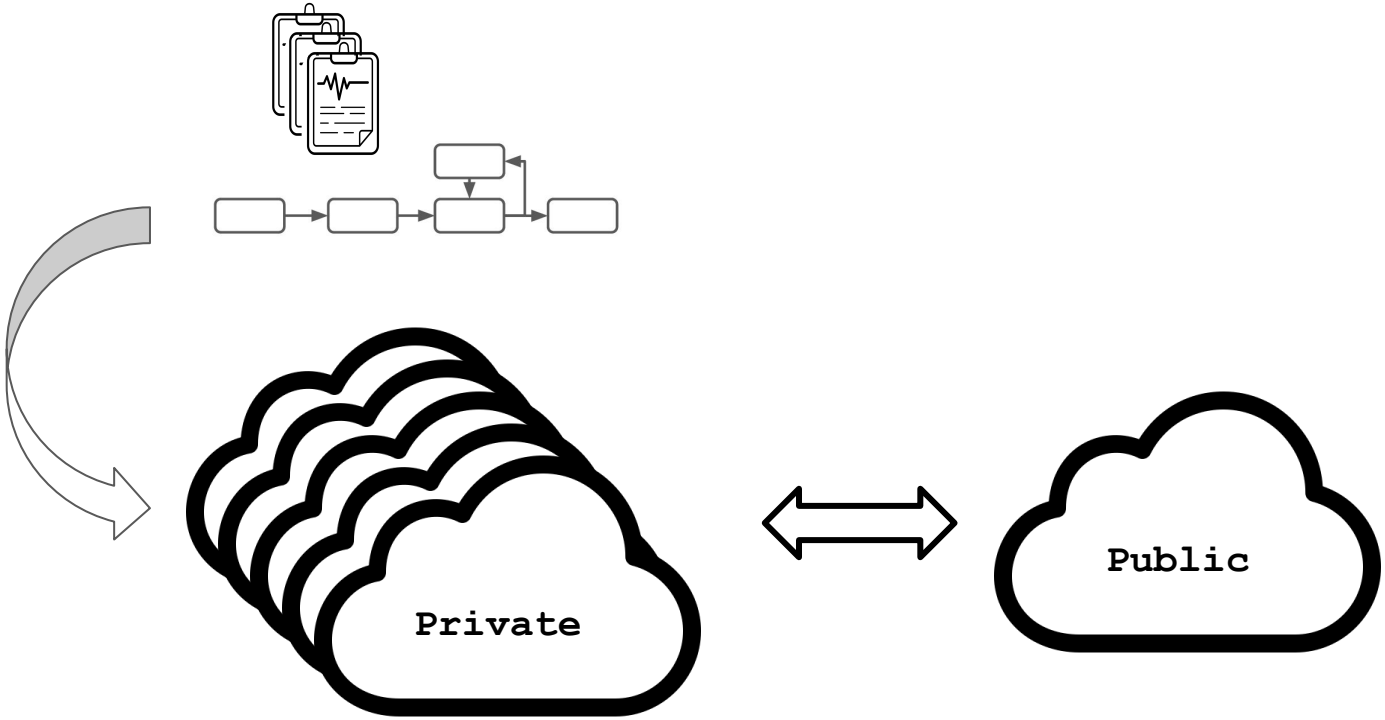
Medical records

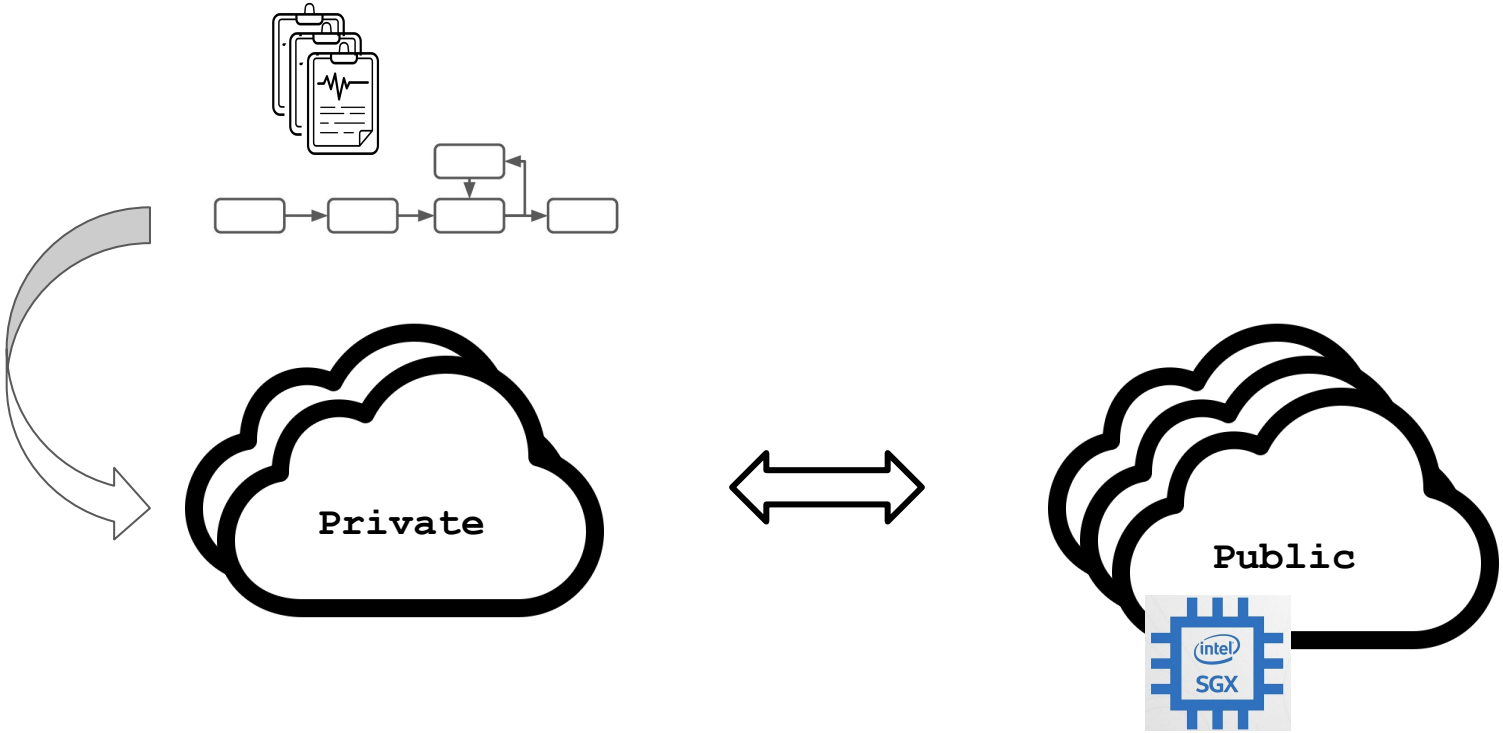


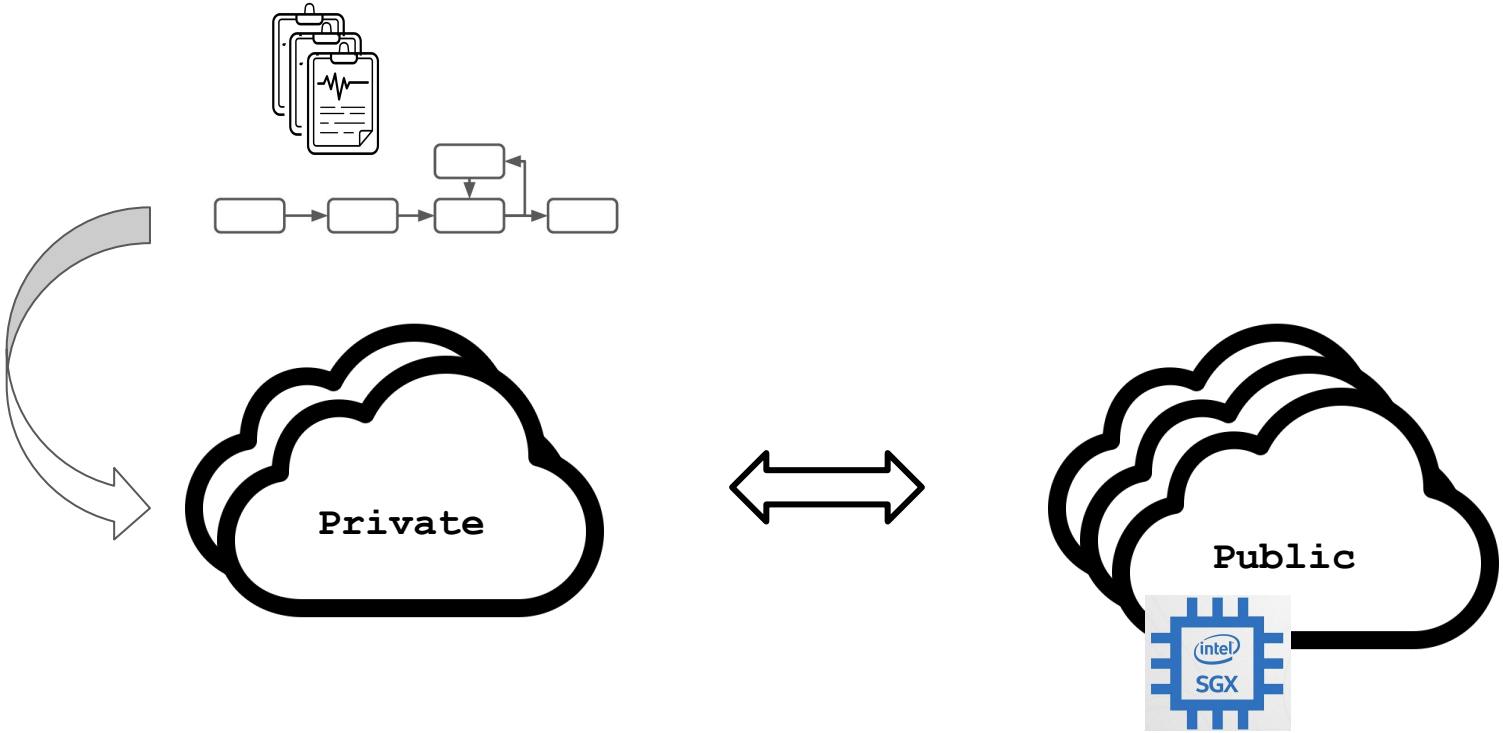
ML pipelines


Private


Private
Public

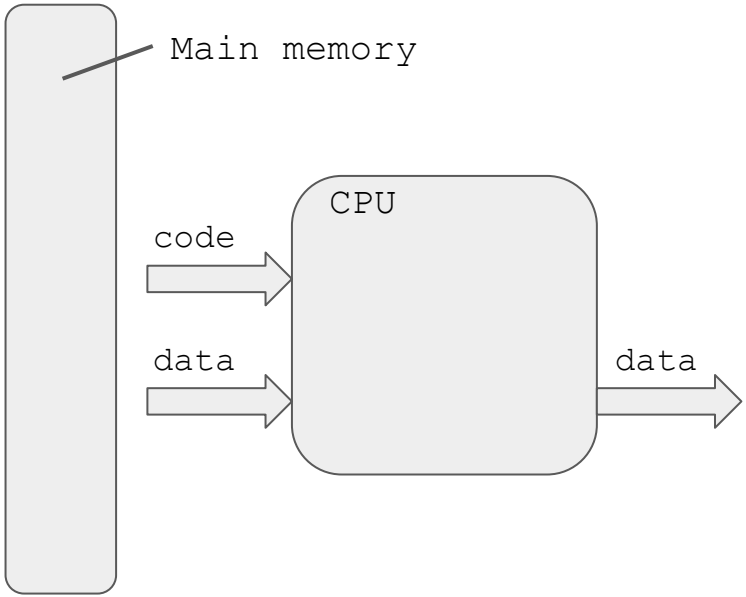

Private
Public


Private
Public
intel
SGX

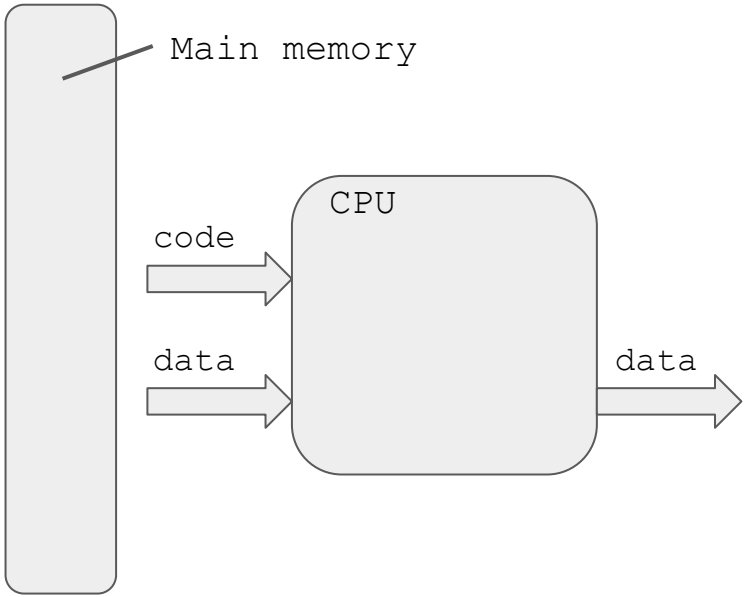


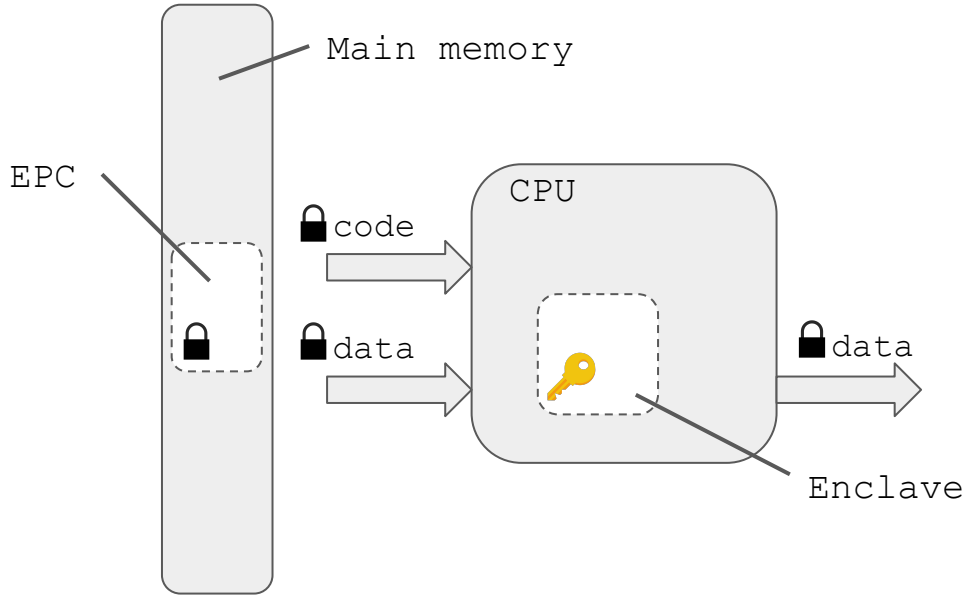
Problem: Enable **efficient** data processing jobs to hybrid compute environments

Trusted Execution Environments (Intel SGX)

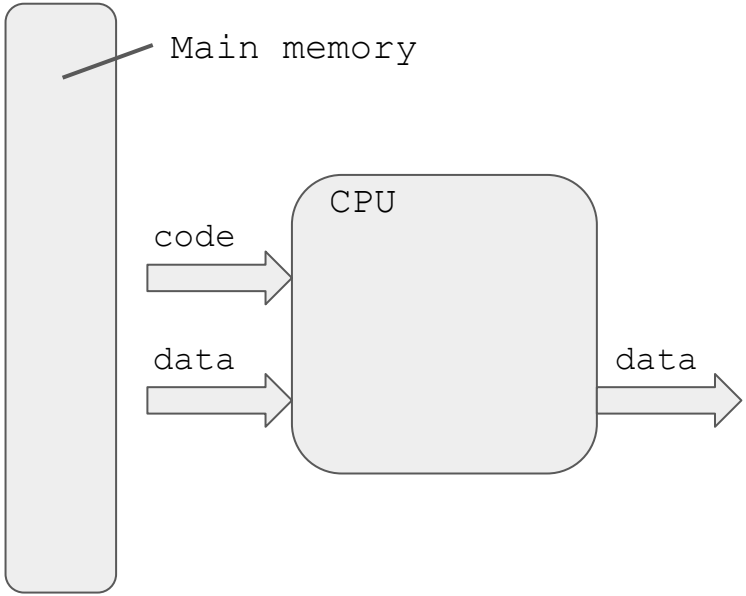


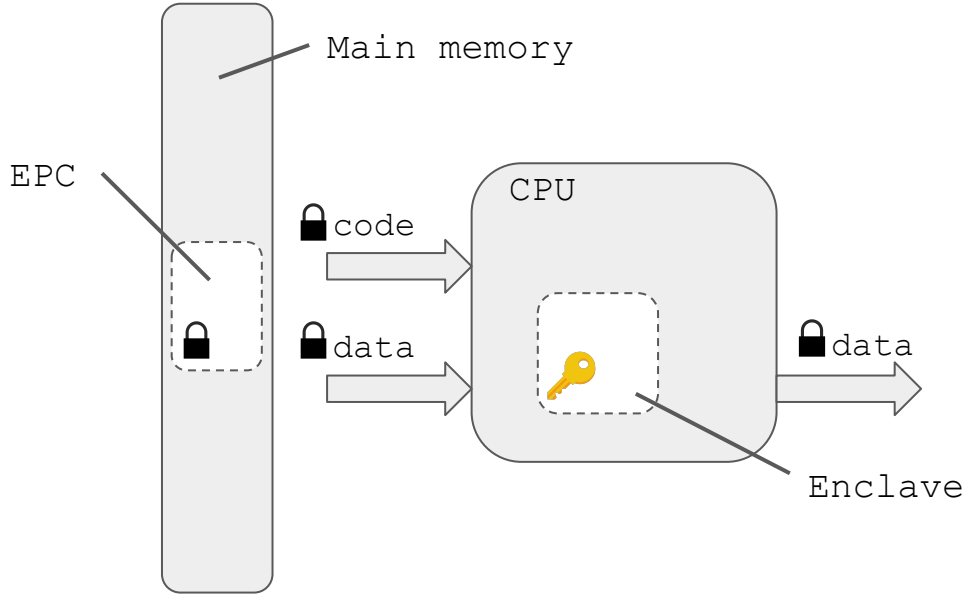
Trusted Execution Environments (Intel SGX)





Trusted Execution Environments (Intel SGX)



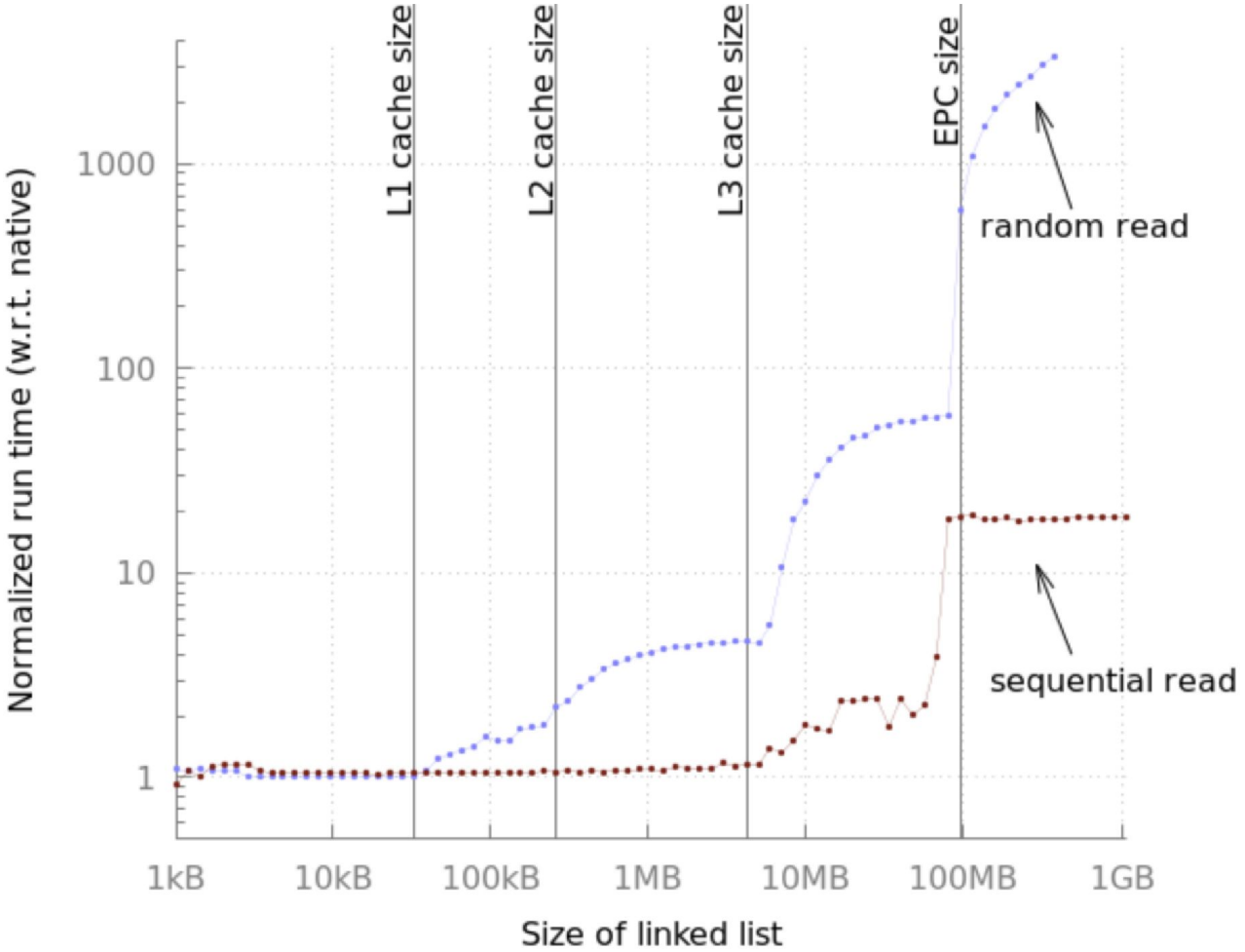


\+ fast
\+ a lot of memory
\- not secure

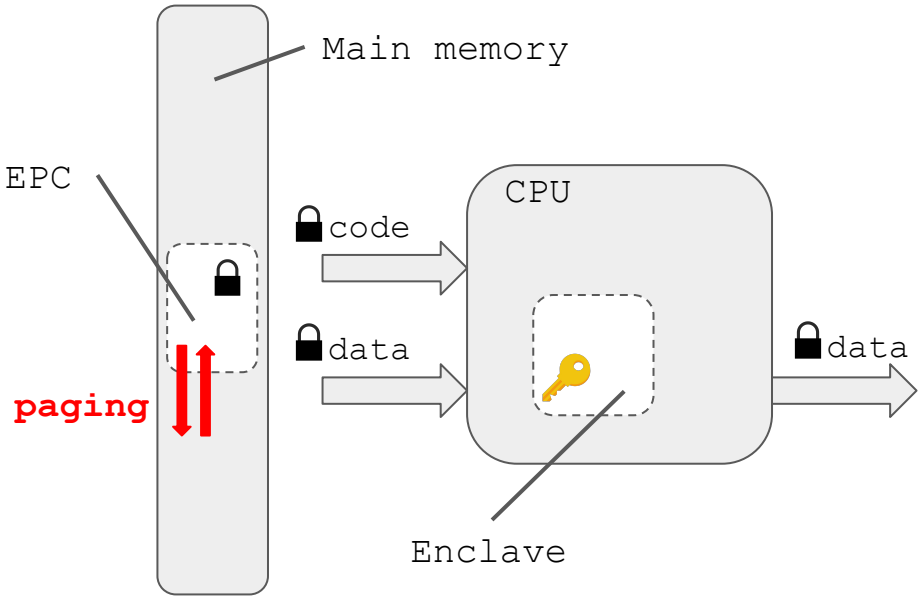
\- slow
\- little memory
\+ secure

Challenges

- Very low main memory (~90 MB)
- Expensive CPU instructions set
- No system calls

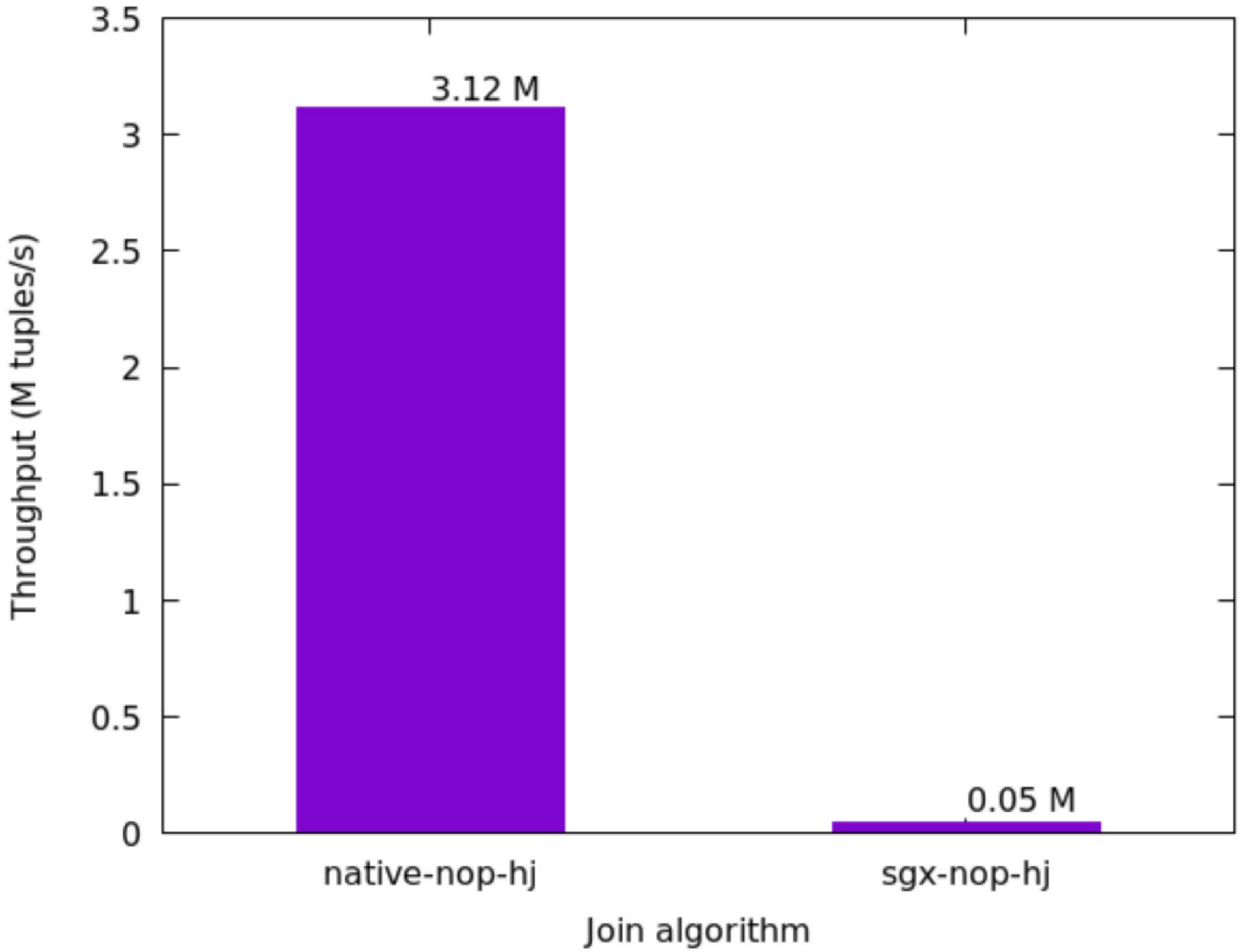


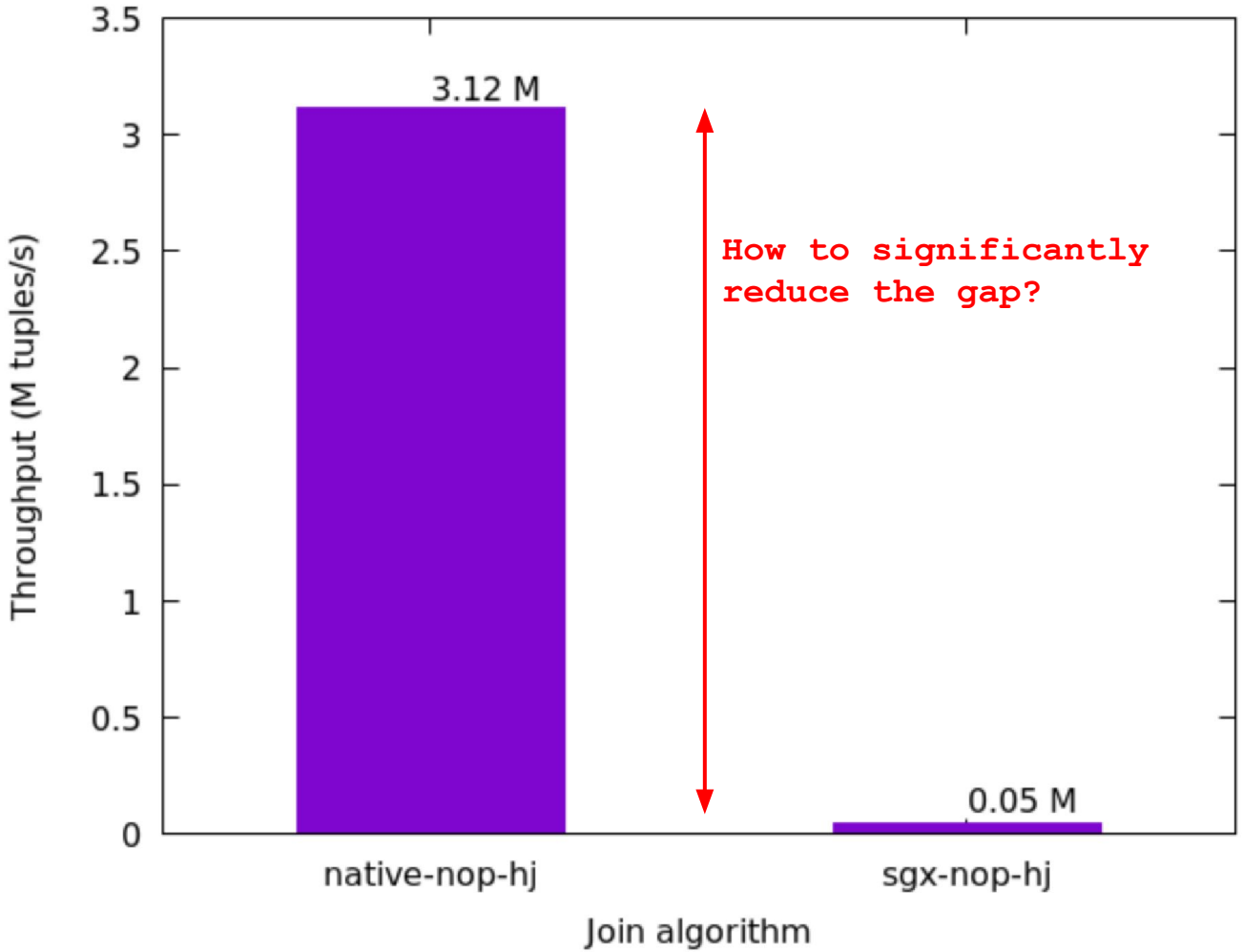
Similar results: Arnautov et al. “SCONE: Secure Linux Containers with Intel SGX”. OSDI 2016

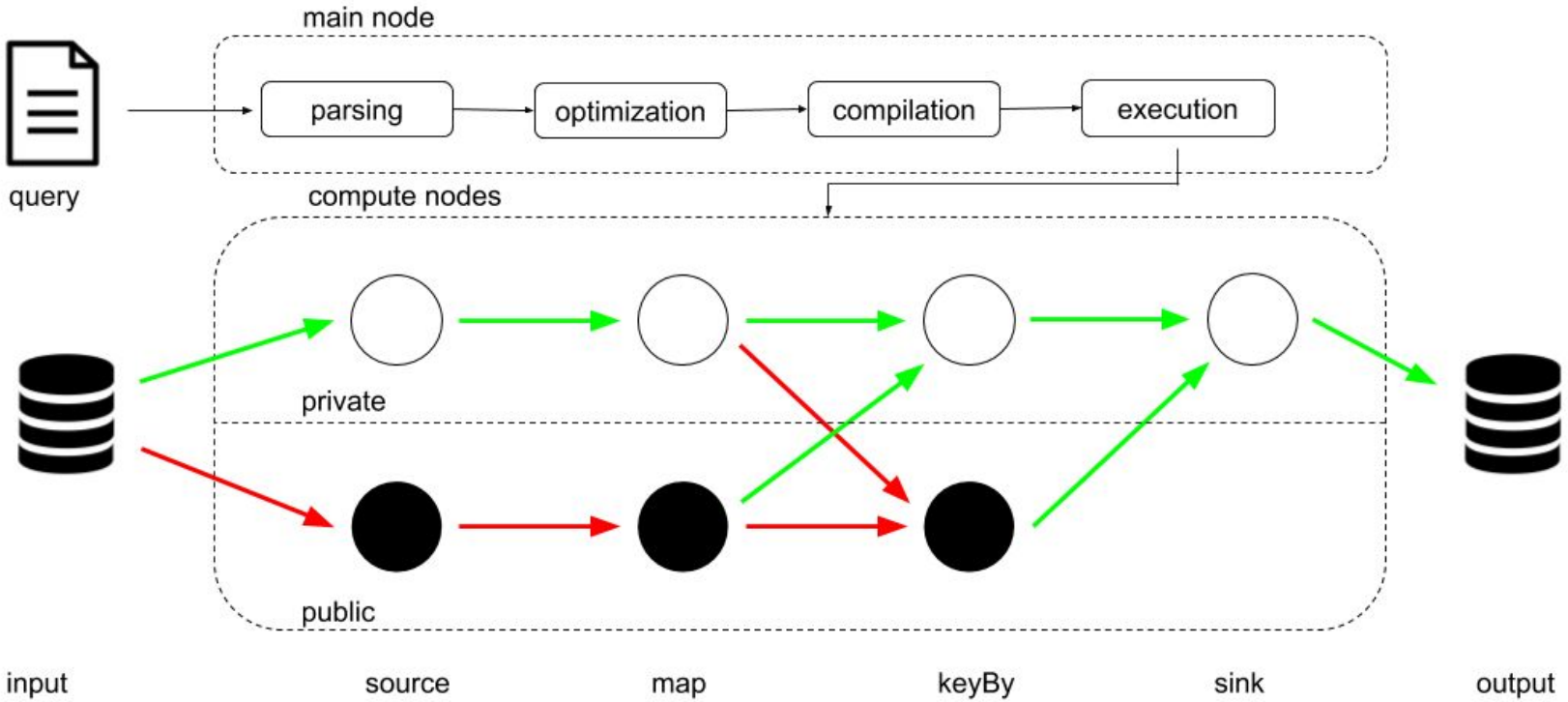


For data larger than EPC:

- EPC paging becomes the bottleneck
- EPC hit vs. miss latencies - 200 cycles vs. 40 K cycles [1]


3.5
3
2.5
2
1.5
1
0.5
0
Throughput (M tuples/s)
3.12 M
0.05 M
native-nop-hj
sgx-nop-hj
Join algorithm


3.5
3
2.5
2
1.5
1
0.5
0
Throughput (M tuples/s)
3.12 M
0.05 M
How to significantly reduce the gap?
native-nop-hj
sgx-nop-hj
Join algorithm


main node
parsing
optimization
compilation
execution
query
compute nodes
private
public
input
source
map
keyBy
sink
output

Next steps

- Understand data processing on SGX
- Experimental evaluation of relational operators on SGX
- New implementations for secure enclaves

References

1. Taassori, Meysam, Ali Shafiee, and Rajeev Balasubramonian. "VAULT: Reducing paging overheads in SGX with efficient integrity verification structures." In ASPLOS 2018.
2. hospital: https://cliparts.co/cliparts/ziX/eRy/ziXeRy78T.png
3. medical record by popcornarts from the Noun Project
4. ai brain: kishore kumar / Getty Images
5. Cloud by Aya Sofya from the Noun Project
6. infrastructure by Yamini Ahluwalia from the Noun Project
7. intel sgx: https://software.intel.com

Next steps

- Understand data processing on SGX
- Experimental evaluation of relational operators on SGX
- New implementations for secure enclaves