

Let's Go, Private! Towards a Privacy-Preserving and Distributed Machine Learning System

Cláudia Brito

claudia.v.brito@inesctec.pt

INESC TEC & University of Minho

Supervisors: **João Paulo** (INESC TEC & UMinho) and **Pedro Ferreira** (INESC TEC & Faculty of Sciences, UPorto)



INESC TEC

Cofinanciado por:

COMPETE
2020

PORTUGAL
2020

UNIÃO EUROPEIA
Fundo Europeu
de Desenvolvimento Regional

FCT
Fundação
para a Ciência
e a Tecnologia

Carnegie
Mellon
Portugal

Limitations

- Untrusted third-party infrastructures
- Common cryptographic schemes impose impractical overheads
- TEEs' performance decreases with the increase of computations, I/O operations and, the trusted computing base (TCB)

Challenges

ML datasets and models are stored and processed in plaintext

Reducing the code base running inside enclaves

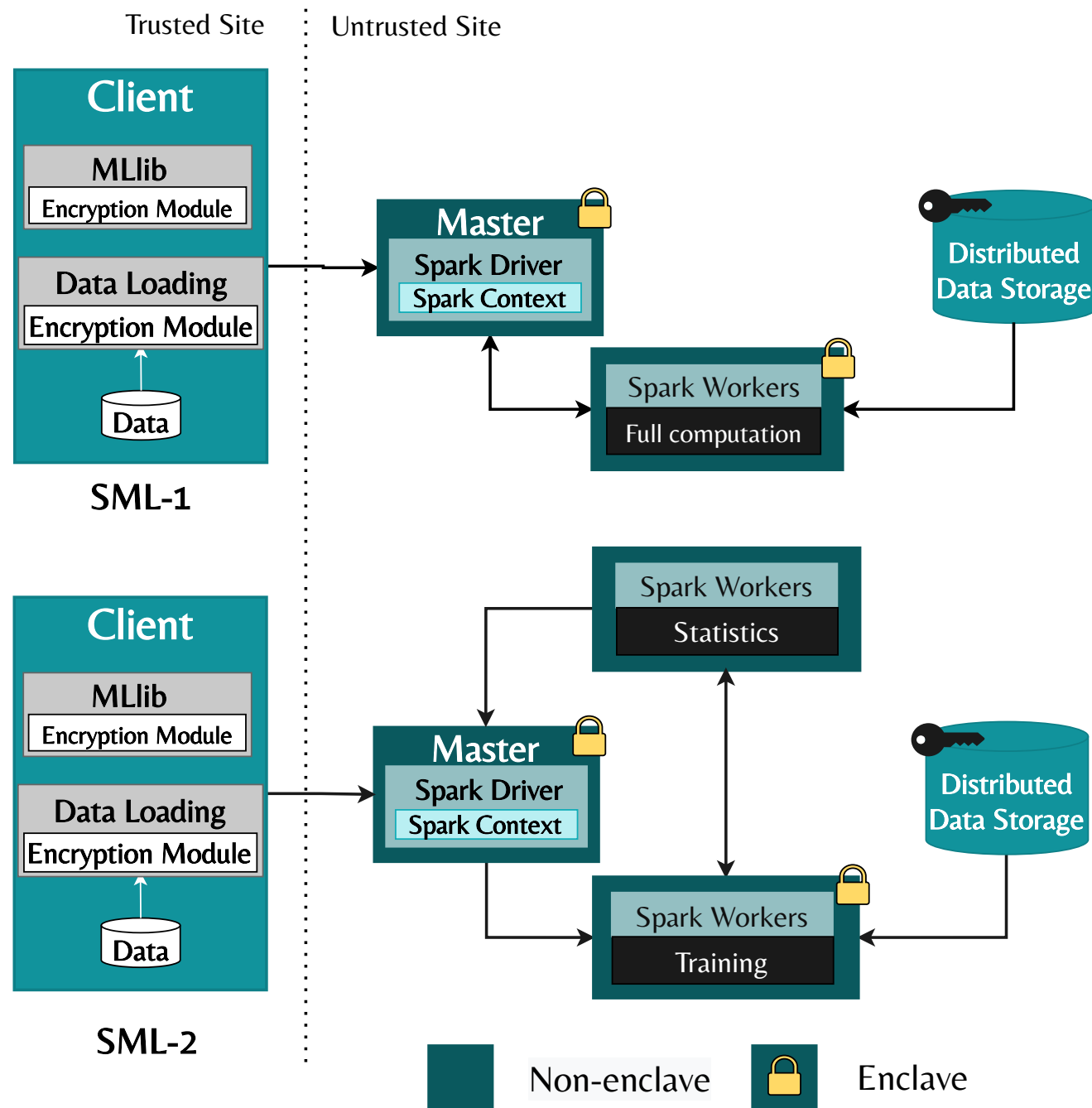
Reducing the number of operations to be performed inside the enclave

Goal

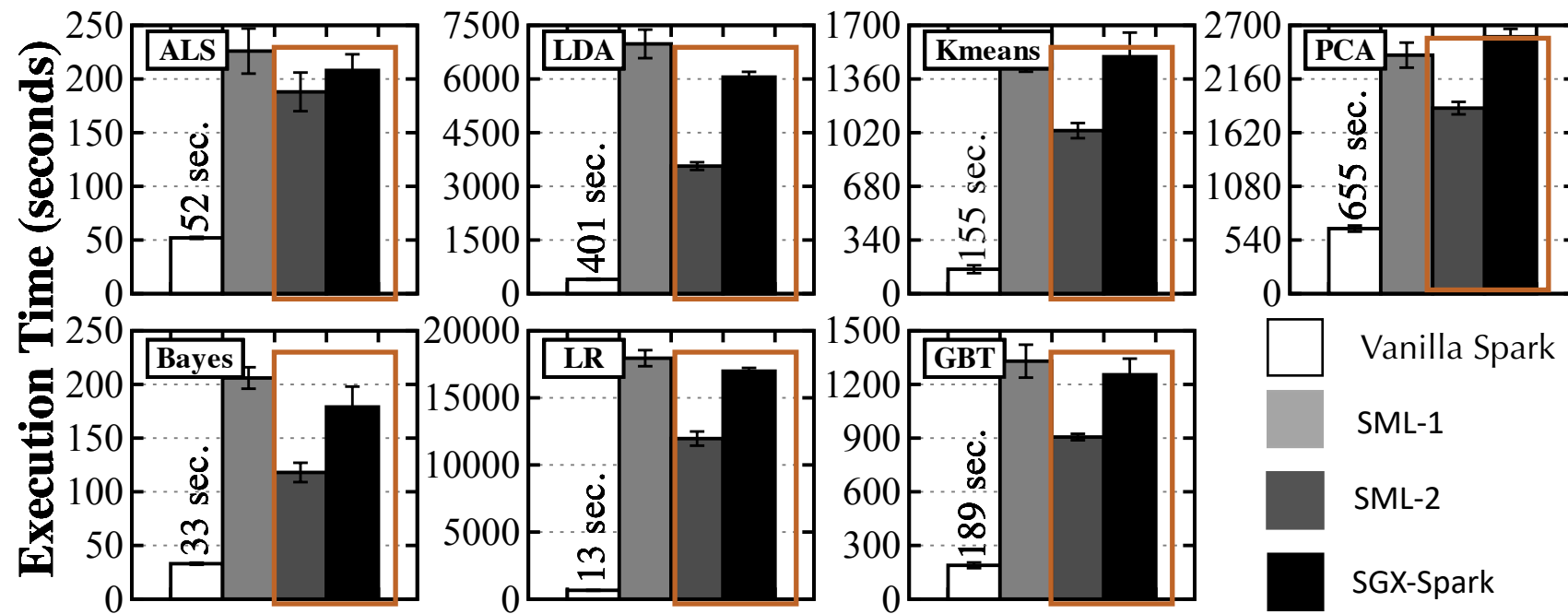
Design an **end-to-end privacy-preserving and distributed machine learning framework**

- Private large scale machine learning and data analysis
- Clients should trust third-party infrastructures while knowing that the computation performed over their data will **not reveal** any sensitive information

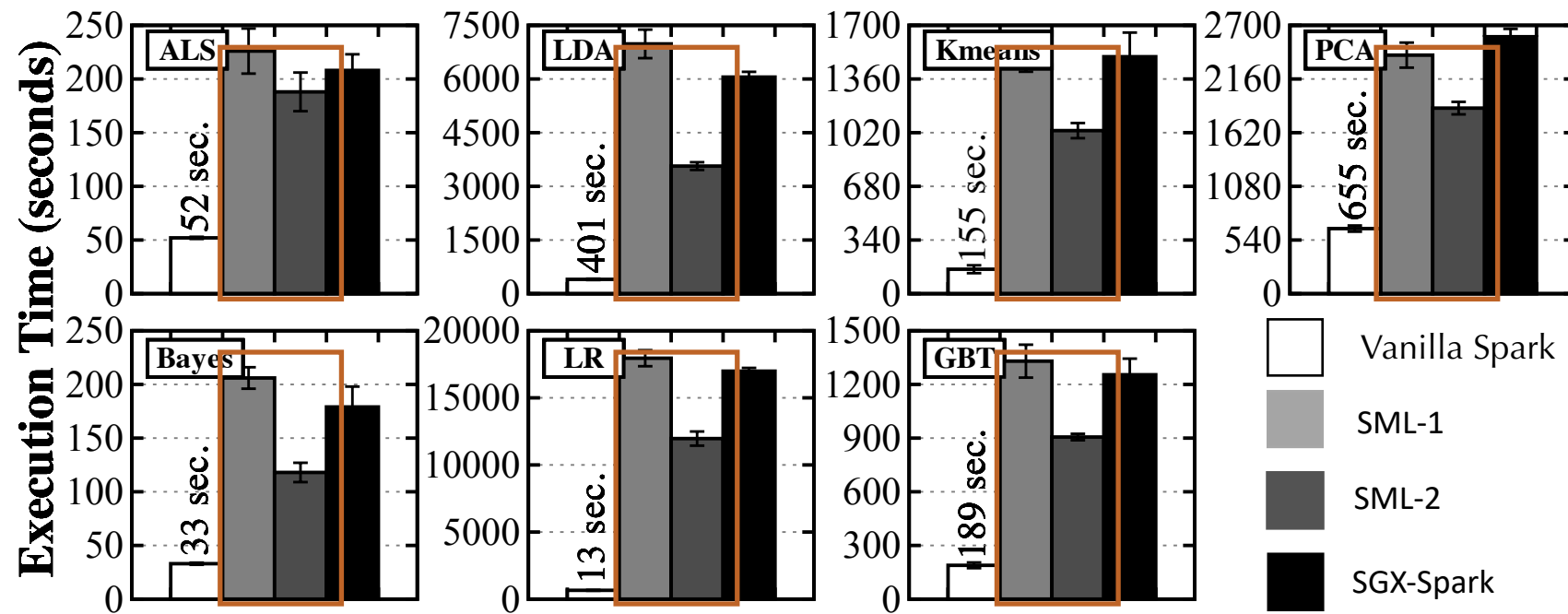
Solution



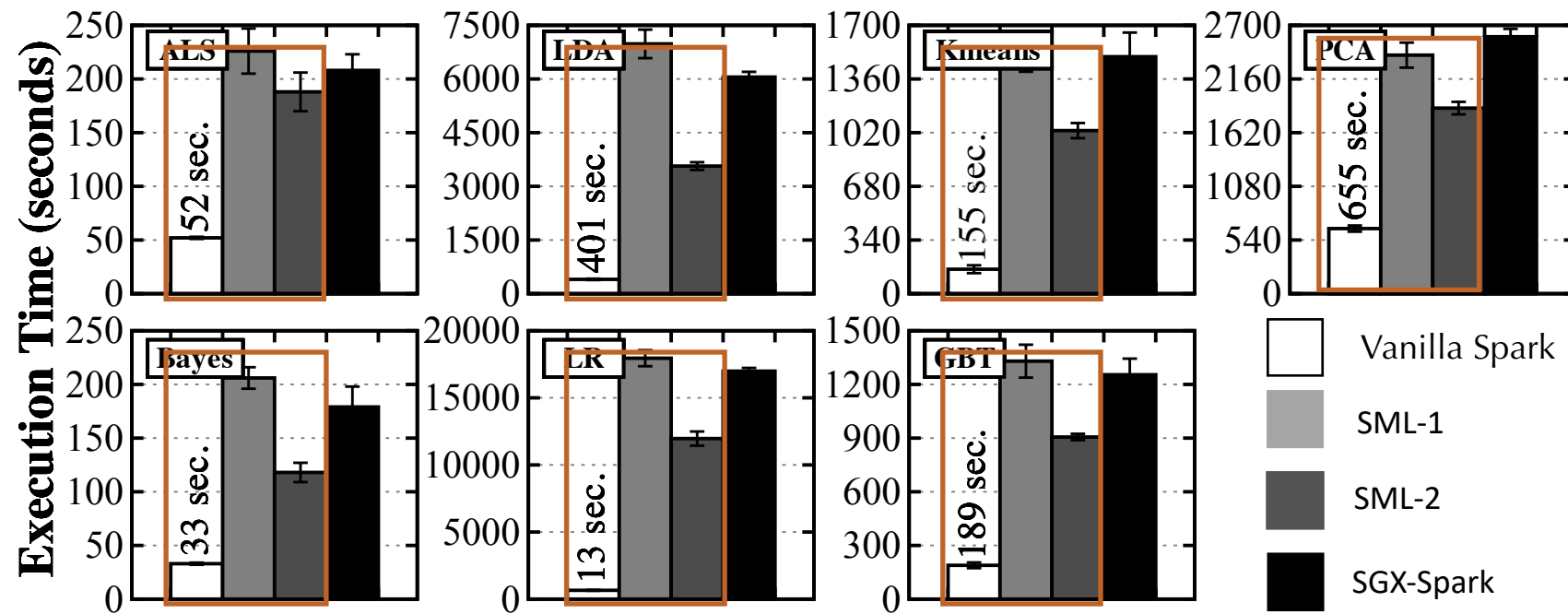
Results



Results



Results



Next Steps

- Focus on white-box access attacks
 - Increase security measures with focus on ORAM and Differential Privacy
- Real data use cases with a focus on genomic data