Secure RDD for SparkFHE

Sai Sujith Kammari
Dr. Peizhao Hu (Advisor)
TABLE OF CONTENTS

01 Project Refresh
   What is Secure RDD

02 Milestone 1 Refresh
   How I started my project

03 Milestone 2
   Status of my project

04 Next Steps
   To take the project to closure
01
Project Refresh

What is Secure RDD for SparkFHE
Secure RDD for SparkFHE

- Apache Spark is the most popular open-source, distributed processing system used for big data workloads.

- Workloads in Apache Spark
  a. Core
  b. MLlib
  c. Streaming
  d. SQL
  e. GraphX

- Apache Spark provides high-level APIs in Java, Scala, Python and R, and supports code reuse across multiple workloads.

- Implementing homomorphic encryption in the RDDs which are used in MLlib.
02

Milestone 1 refresh

How I started my project
Milestone1 Refresh

- Apache Spark
- Resilient Distributed Datasets
- Data types in MLlib
- Homomorphic Encryption
- Implementation of existing SparkFHE
- Basic Linear Algebra Subprograms
Milestone 2

Status of my project
Milestone 2

- Basic Linear Algebra Subprograms
  - fhe_axpy
  - fhe_scal
  - fhe_syr
  - fhe_spr

- Spark MLlib
  - CtxtVectors
  - CtxtBLAS

- SparkFHE-Examples
  - Examples with the test cases
Next steps

To take the project to closure
Summary

K-Means
Implement a working usecase for the homomorphic encryption friendly Vector and BLAS operations.

Sparse vector
Until now, no implementation is started for sparse vector operations.

Complete Report
Complete report and poster presentation
Any Questions
THANKS

CREDITS: This presentation template was created by Slidesgo, including icons by Flaticon, and infographics & images by Freepik. Please keep this slide for attribution.