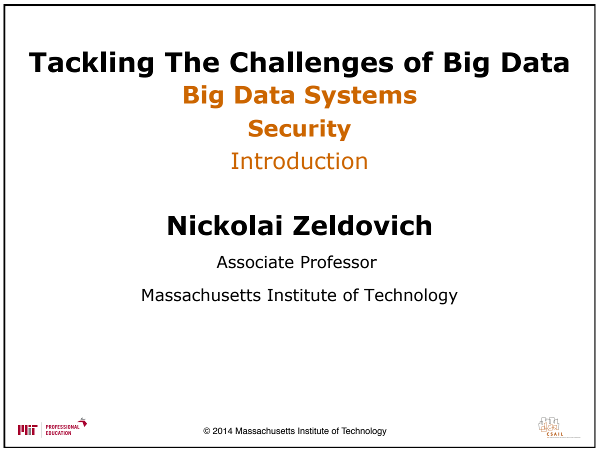
Tackling The Challenges of Big Data

Big Data Systems

Nickolai Zeldovich

Associate Professor

Massachusetts Institute of Technology



Tackling The Challenges of Big Data

Big Data Systems

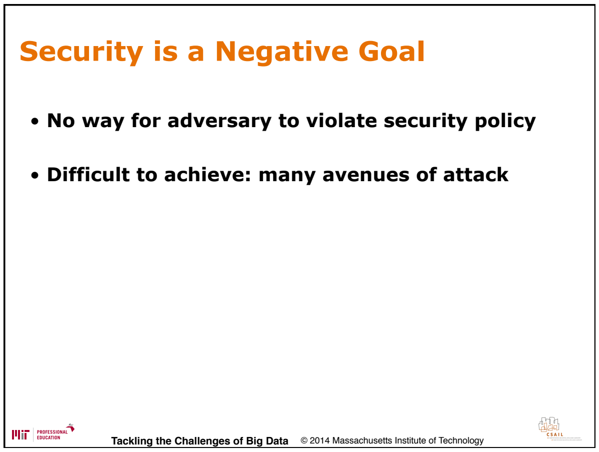
Security

Introduction

Nickolai Zeldovich

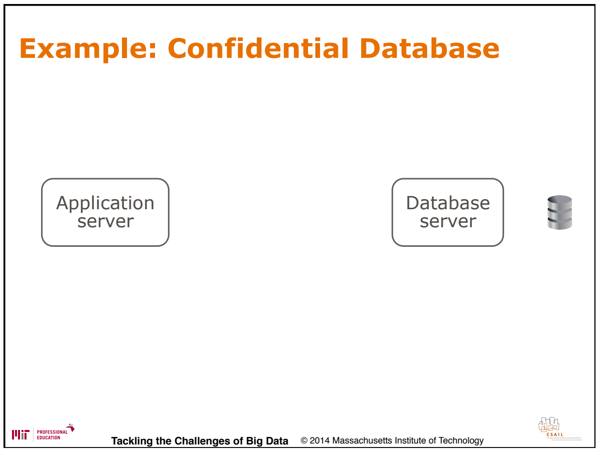
Associate Professor

Massachusetts Institute of Technology



Security is a Negative Goal

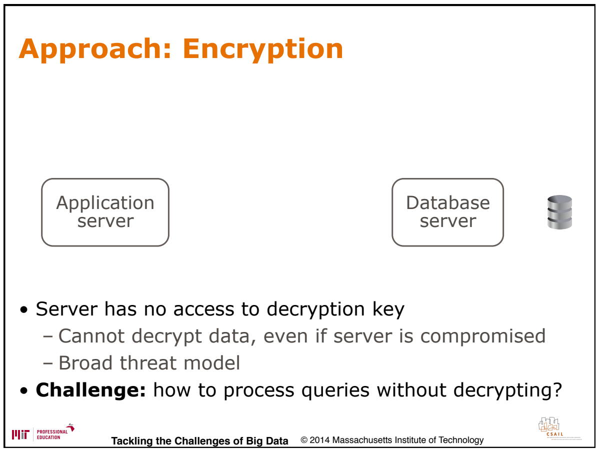
- No way for adversary to violate security policy
- Difficult to achieve: many avenues of attack



Example: Confidential Database

Application server

Database server

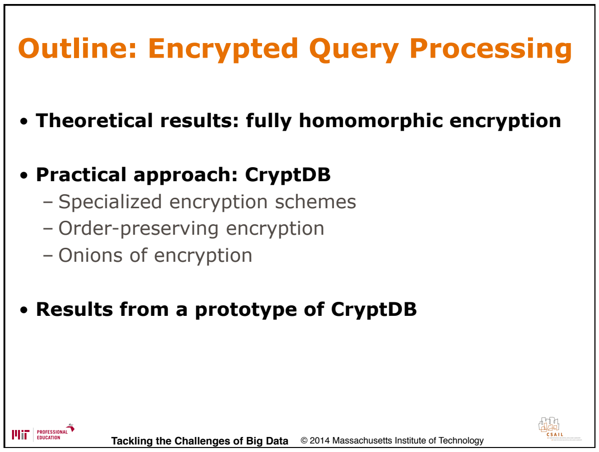


Approach: Encryption

Application server

Database server

- Server has no access to decryption key
  - Cannot decrypt data, even if server is compromised
  - Broad threat model
- **Challenge:** how to process queries without decrypting?



Outline: Encrypted Query Processing

- **Theoretical results: fully homomorphic encryption**
- **Practical approach: CryptDB**
  - Specialized encryption schemes
  - Order-preserving encryption
  - Onions of encryption
- **Results from a prototype of CryptDB**

Tackling The Challenges of Big Data

Big Data Systems

Security

Introduction

THANK YOU

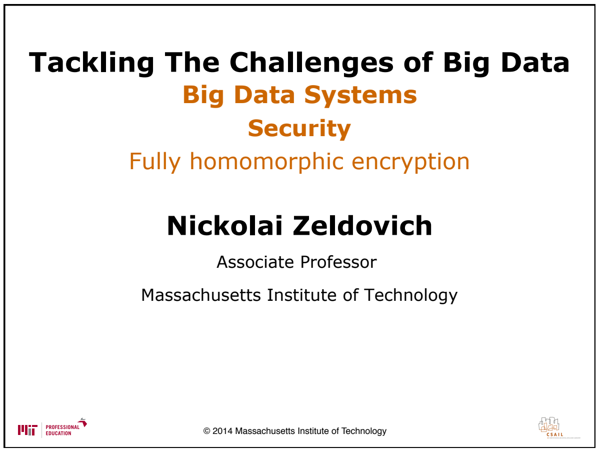
Tackling The Challenges of Big Data

Big Data Systems

Nickolai Zeldovich

Associate Professor

Massachusetts Institute of Technology



Tackling The Challenges of Big Data

Big Data Systems

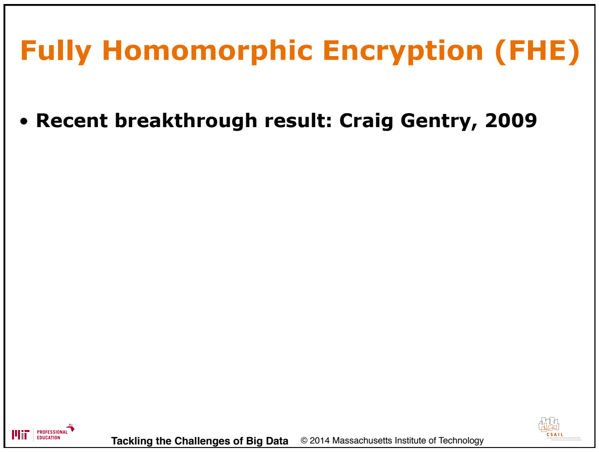
Security

Fully homomorphic encryption

Nickolai Zeldovich

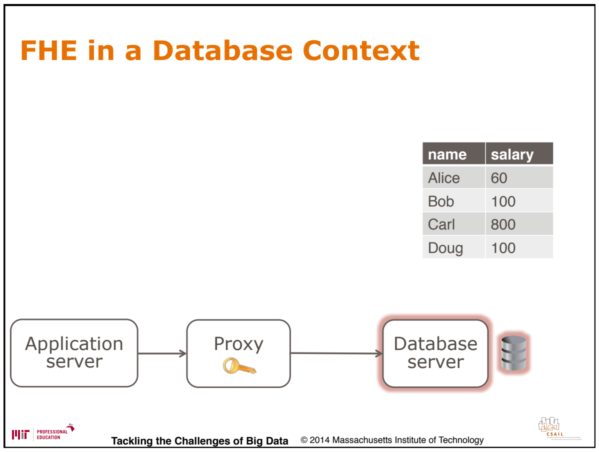
Associate Professor

Massachusetts Institute of Technology



Fully Homomorphic Encryption (FHE)

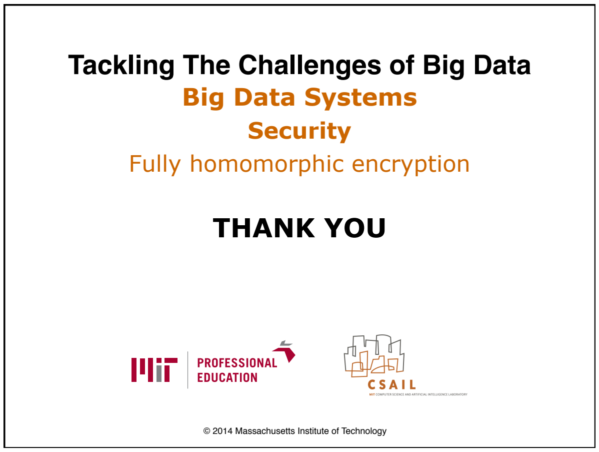
- **Recent breakthrough result: Craig Gentry, 2009**



FHE in a Database Context

name	salary
Alice	60
Bob	100
Carl	800
Doug	100

Application server → Proxy → Database server



Tackling The Challenges of Big Data

Big Data Systems

Security

Fully homomorphic encryption

THANK YOU

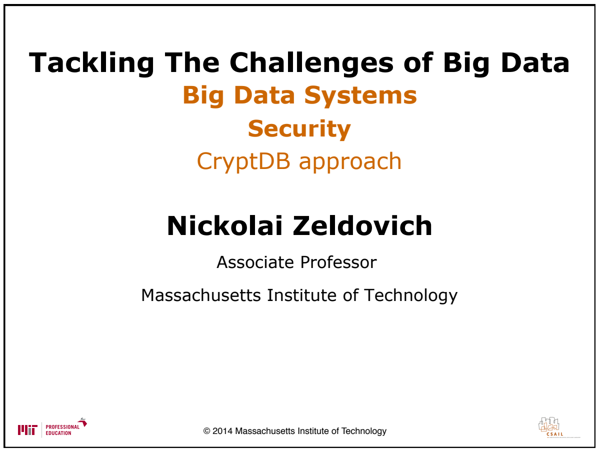
Tackling The Challenges of Big Data

Big Data Systems

Nickolai Zeldovich

Associate Professor

Massachusetts Institute of Technology



Tackling The Challenges of Big Data

Big Data Systems

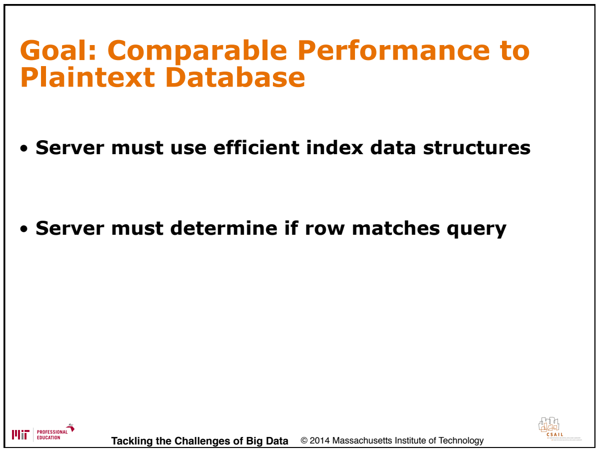
Security

CryptDB approach

Nickolai Zeldovich

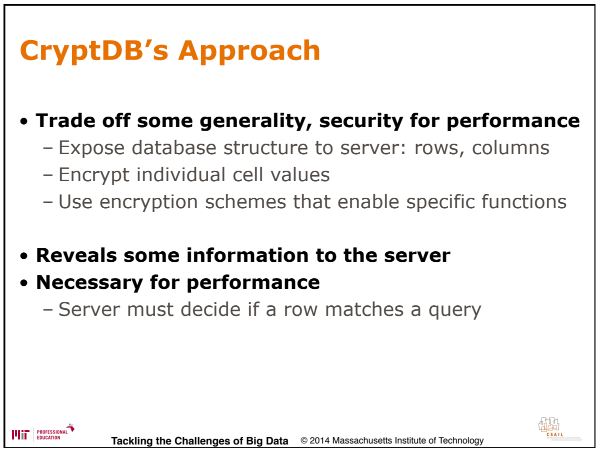
Associate Professor

Massachusetts Institute of Technology



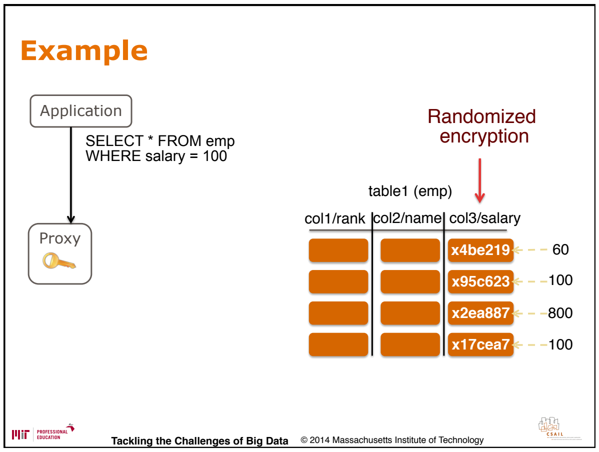
Goal: Comparable Performance to Plaintext Database

- **Server must use efficient index data structures**
- **Server must determine if row matches query**

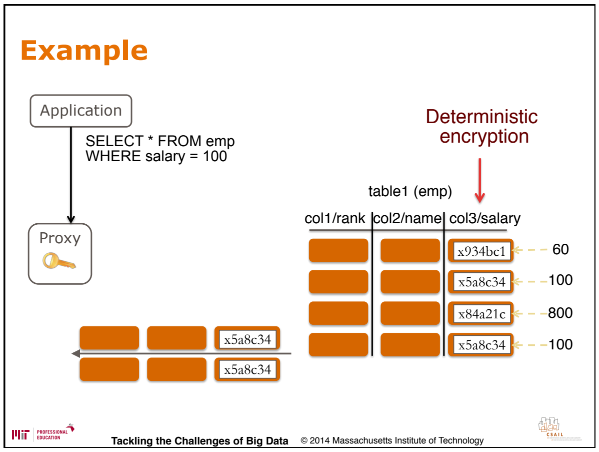


CryptDB's Approach

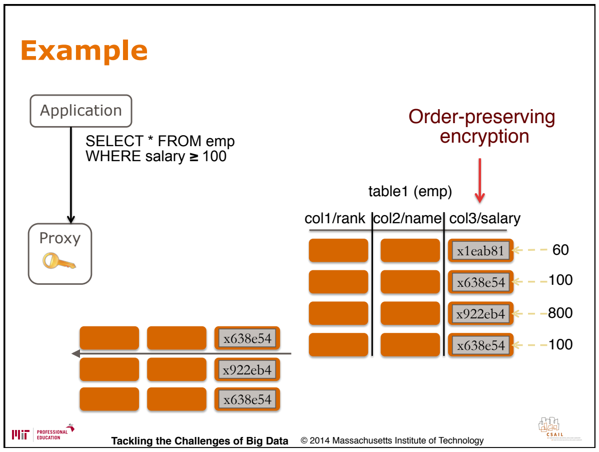
- **Trade off some generality, security for performance**
  - Expose database structure to server: rows, columns
  - Encrypt individual cell values
  - Use encryption schemes that enable specific functions
- **Reveals some information to the server**
- **Necessary for performance**
  - Server must decide if a row matches a query



Example



Example



Example

Application

SELECT * FROM emp
WHERE salary ≥ 100

Proxy

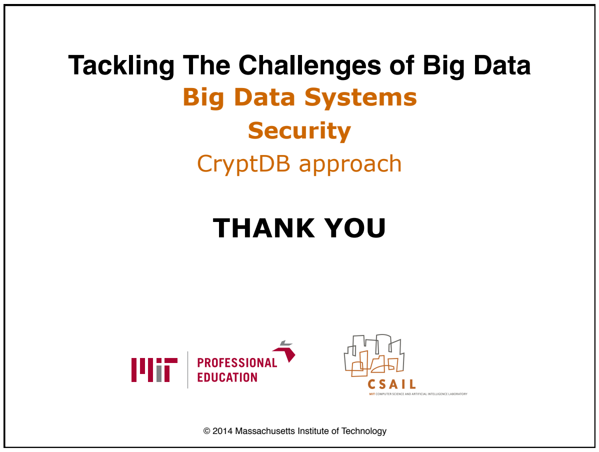
Order-preserving encryption

table1 (emp)

col1/rank	col2/name	col3/salary	
		x1eab81	60
		x638e54	100
		x922eb4	800
		x638e54	100

x638e54
x922eb4
x638e54

Tackling the Challenges of Big Data



Tackling The Challenges of Big Data

Big Data Systems

Security

CryptDB approach

THANK YOU

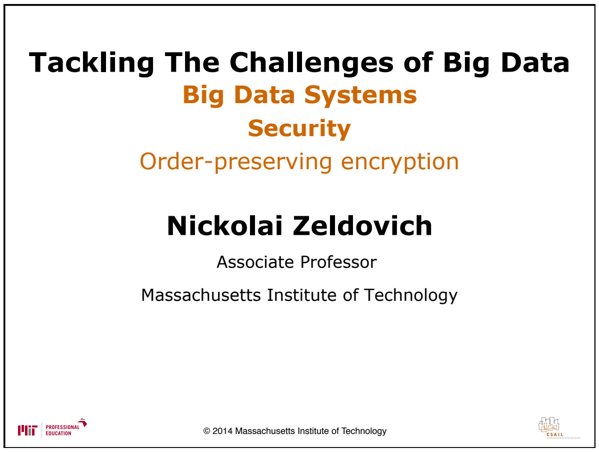
Tackling The Challenges of Big Data

Big Data Systems

Nickolai Zeldovich

Associate Professor

Massachusetts Institute of Technology



Tackling The Challenges of Big Data

Big Data Systems

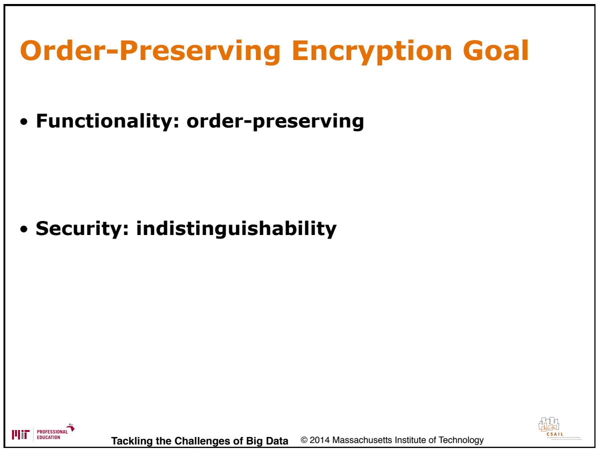
Security

Order-preserving encryption

Nickolai Zeldovich

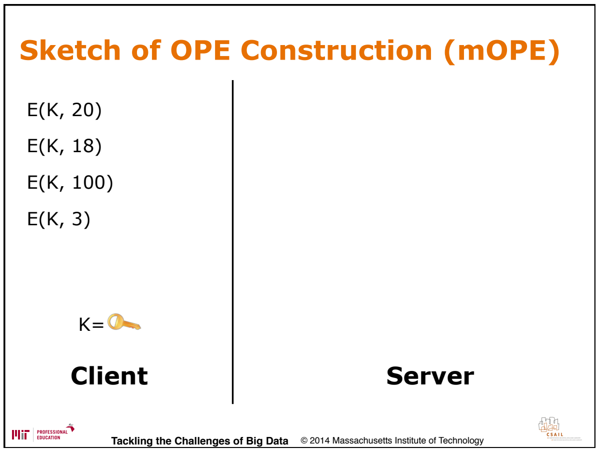
Associate Professor

Massachusetts Institute of Technology



Order-Preserving Encryption Goal

- **Functionality: order-preserving**
- **Security: indistinguishability**



Sketch of OPE Construction (mOPE)

E(K, 20)

E(K, 18)

E(K, 100)

E(K, 3)

Tackling The Challenges of Big Data

Big Data Systems

Security

Order-preserving encryption

THANK YOU

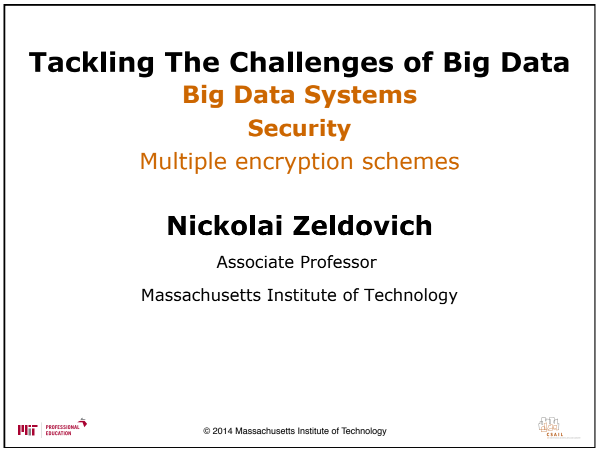
Tackling The Challenges of Big Data

Big Data Systems

Nickolai Zeldovich

Associate Professor

Massachusetts Institute of Technology



Tackling The Challenges of Big Data

Big Data Systems

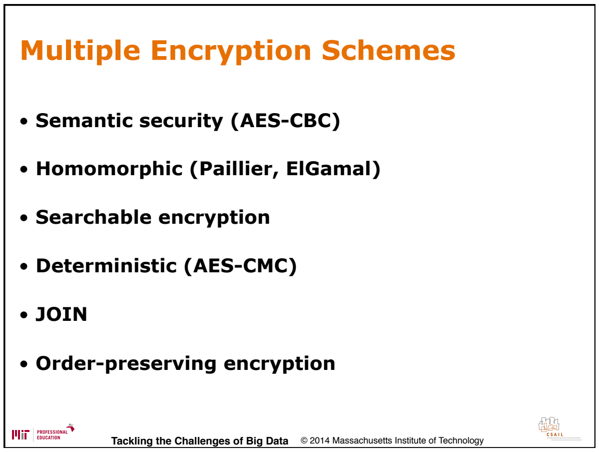
Security

Multiple encryption schemes

Nickolai Zeldovich

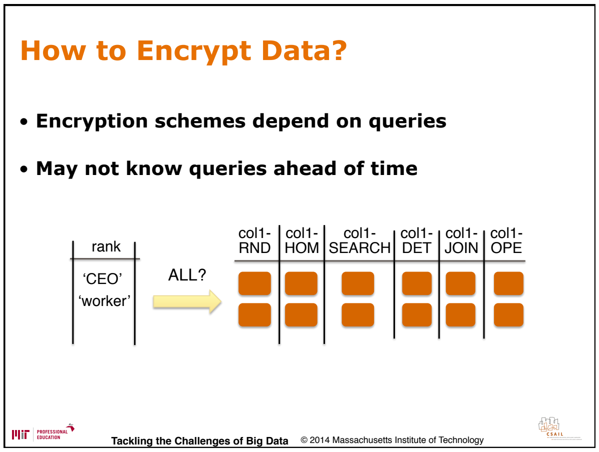
Associate Professor

Massachusetts Institute of Technology



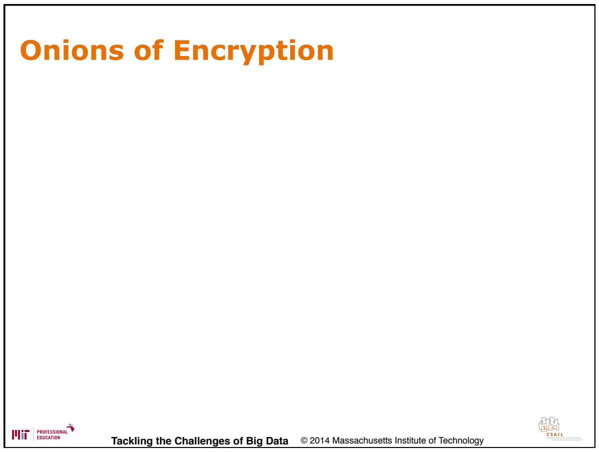
Multiple Encryption Schemes

- Semantic security (AES-CBC)
- Homomorphic (Paillier, ElGamal)
- Searchable encryption
- Deterministic (AES-CMC)
- JOIN
- Order-preserving encryption

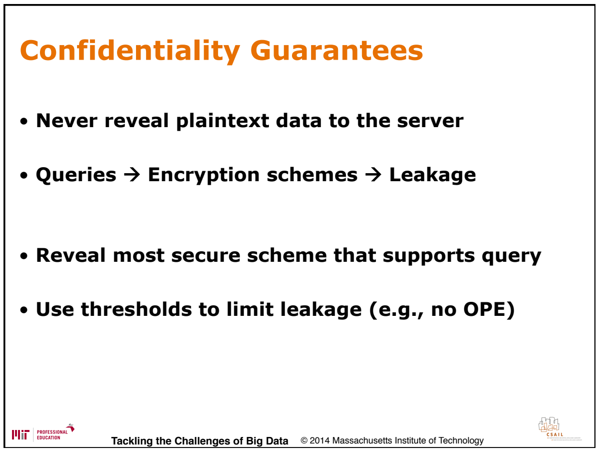


How to Encrypt Data?

- Encryption schemes depend on queries
- May not know queries ahead of time



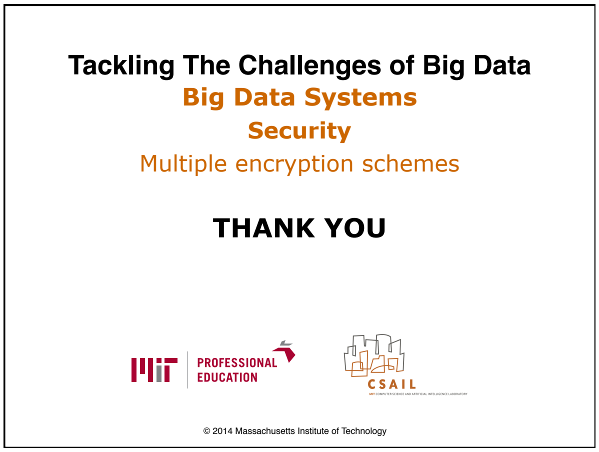
Onions of Encryption



Confidentiality Guarantees

- Never reveal plaintext data to the server
- Queries → Encryption schemes → Leakage
- Reveal most secure scheme that supports query
- Use thresholds to limit leakage (e.g., no OPE)

Tackling the Challenges of Big Data



Tackling The Challenges of Big Data

Big Data Systems

Security

Multiple encryption schemes

THANK YOU

Tackling The Challenges of Big Data

Big Data Systems

Nickolai Zeldovich

Associate Professor

Massachusetts Institute of Technology

Tackling The Challenges of Big Data

Big Data Systems

Security

CryptDB results

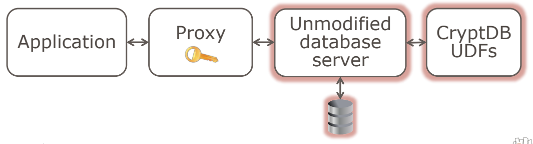
Nickolai Zeldovich

Associate Professor

Massachusetts Institute of Technology

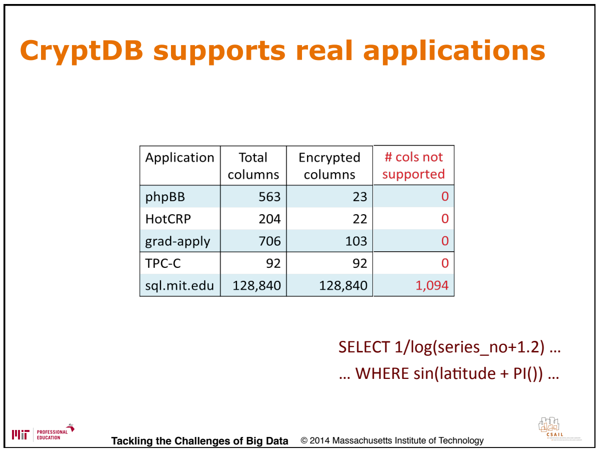
Implementation

- **Proxy rewrites query to run over encrypted data**
  - User-defined functions for crypto (onion adjustment, …)
- **Prototype: 26,000 lines of C++ code**



Evaluation Questions

- **Can CryptDB support real queries and applications?**
- **What is the resulting level of confidentiality?**
- **What is the performance overhead of CryptDB?**

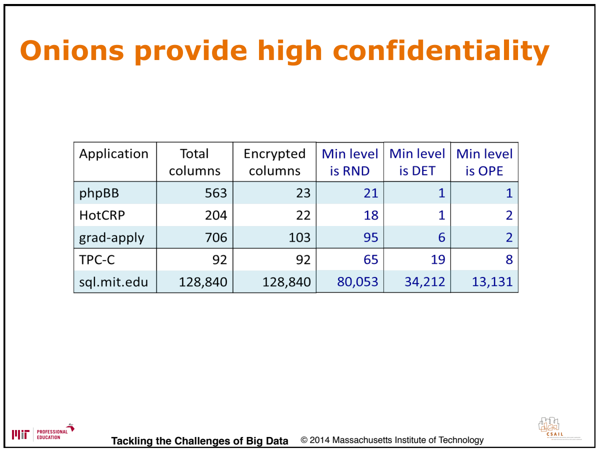


CryptDB supports real applications

Application	Total columns	Encrypted columns	# cols not supported
phpBB	563	23	0
HotCRP	204	22	0
grad-apply	706	103	0
TPC-C	92	92	0
sql.mit.edu	128,840	128,840	1,094

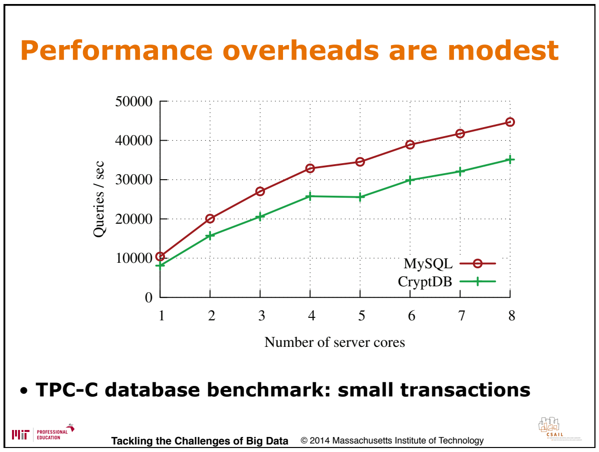
SELECT 1/log(series_no+1.2) …

… WHERE sin(latitude + PI()) …



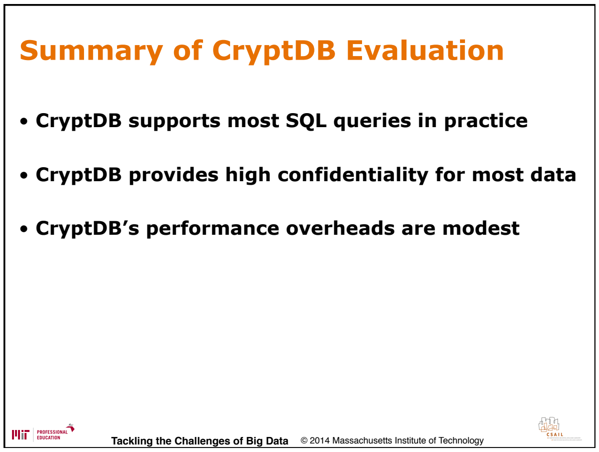
Onions provide high confidentiality

Application	Total columns	Encrypted columns	Min level is RND	Min level is DET	Min level is OPE
phpBB	563	23	21	1	1
HotCRP	204	22	18	1	2
grad-apply	706	103	95	6	2
TPC-C	92	92	65	19	8
sql.mit.edu	128,840	128,840	80,053	34,212	13,131



Performance overheads are modest

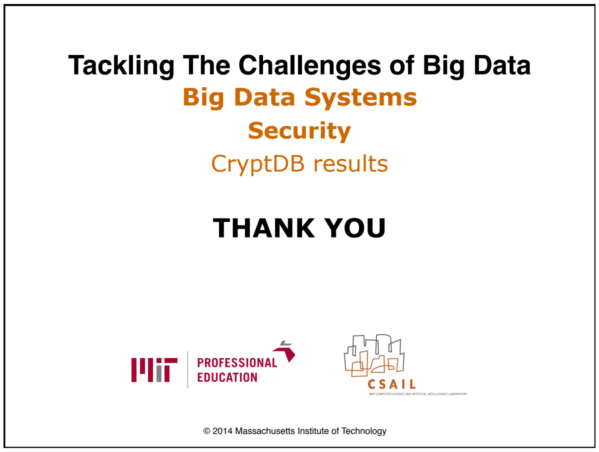
- **TPC-C database benchmark: small transactions**



Summary of CryptDB Evaluation

- CryptDB supports most SQL queries in practice
- CryptDB provides high confidentiality for most data
- CryptDB's performance overheads are modest

Tackling the Challenges of Big Data



Tackling The Challenges of Big Data

Big Data Systems

Security

CryptDB results

THANK YOU

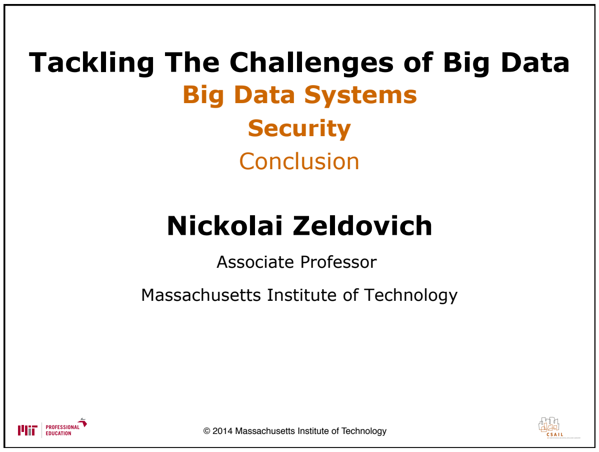
Tackling The Challenges of Big Data

Big Data Systems

Nickolai Zeldovich

Associate Professor

Massachusetts Institute of Technology



Tackling The Challenges of Big Data

Big Data Systems

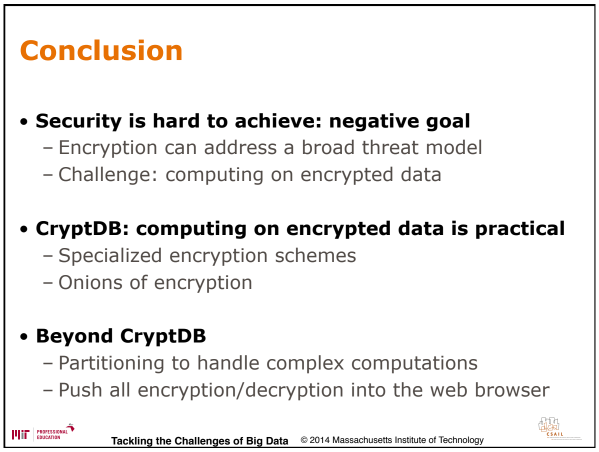
Security

Conclusion

Nickolai Zeldovich

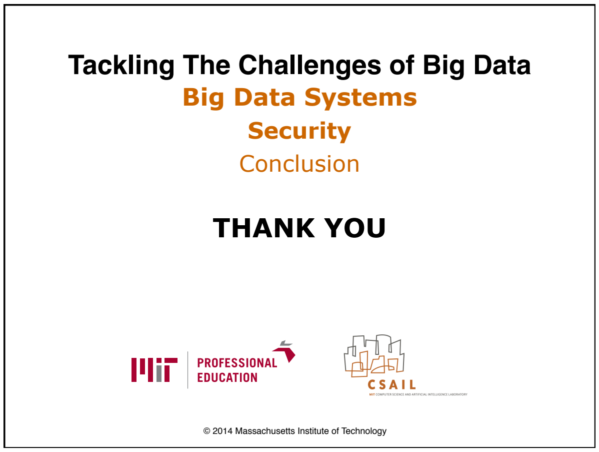
Associate Professor

Massachusetts Institute of Technology



Conclusion

- **Security is hard to achieve: negative goal**
  - Encryption can address a broad threat model
  - Challenge: computing on encrypted data
- **CryptDB: computing on encrypted data is practical**
  - Specialized encryption schemes
  - Onions of encryption
- **Beyond CryptDB**
  - Partitioning to handle complex computations
  - Push all encryption/decryption into the web browser



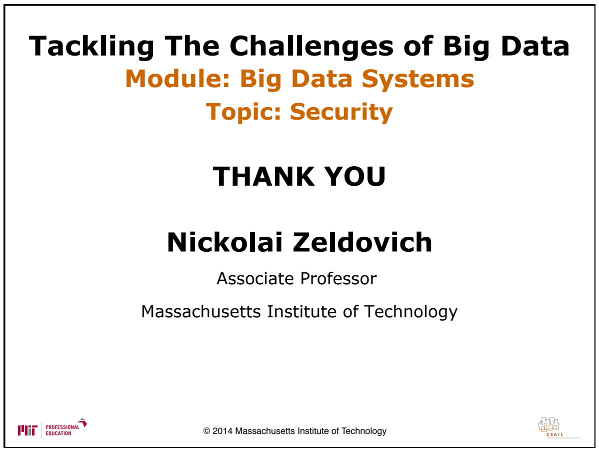
Tackling The Challenges of Big Data

Big Data Systems

Security

Conclusion

THANK YOU



Tackling The Challenges of Big Data

Module: Big Data Systems

Topic: Security

THANK YOU

Nickolai Zeldovich

Associate Professor

Massachusetts Institute of Technology