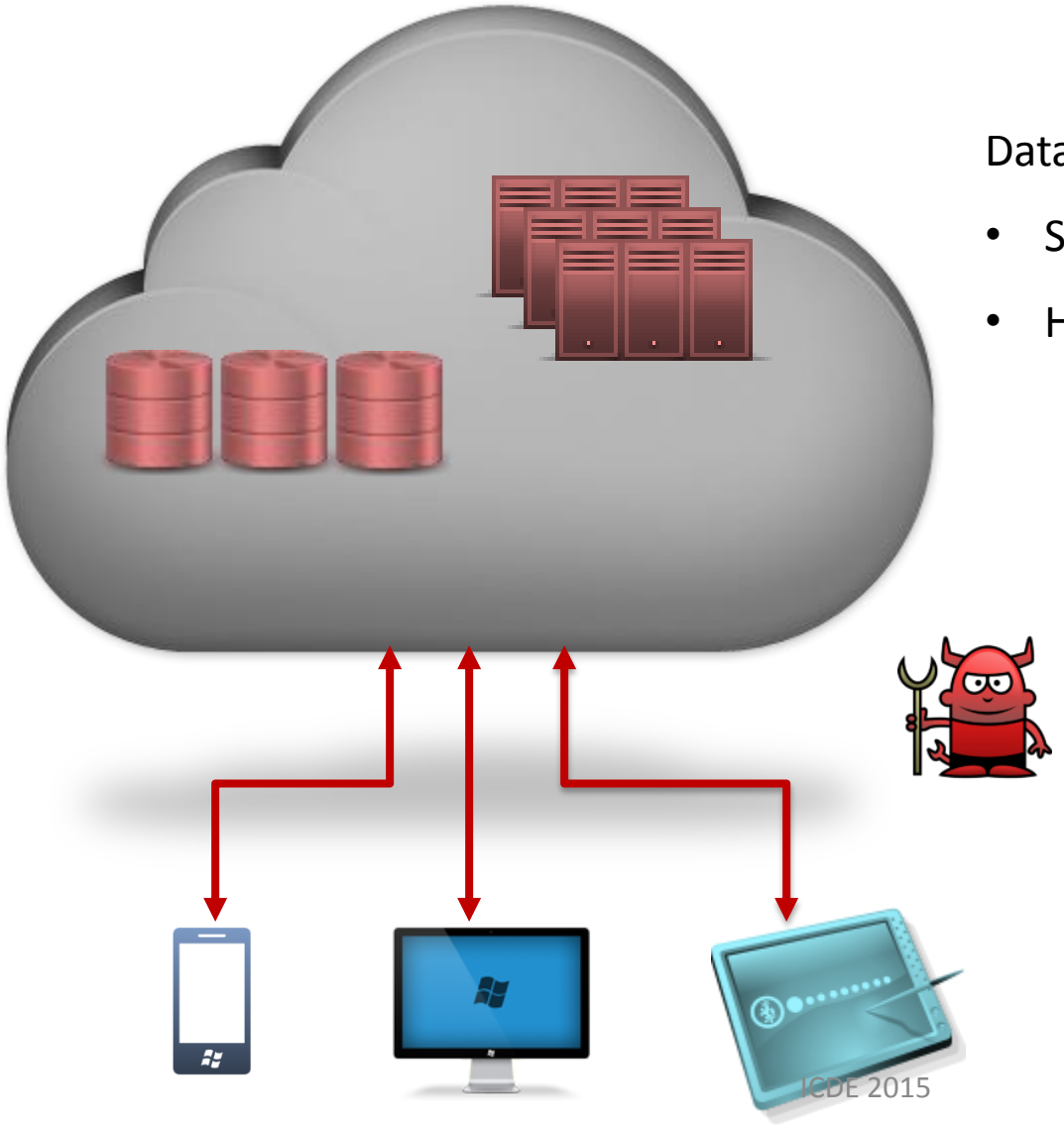
Transaction Processing on Confidential Data using Cipherbase

Arvind Arasu, Ken Eguro, Manas Joglekar*
Raghav Kaushik, Donald Kossmann, Ravi Ramamurthy
Microsoft Research
Stanford University*

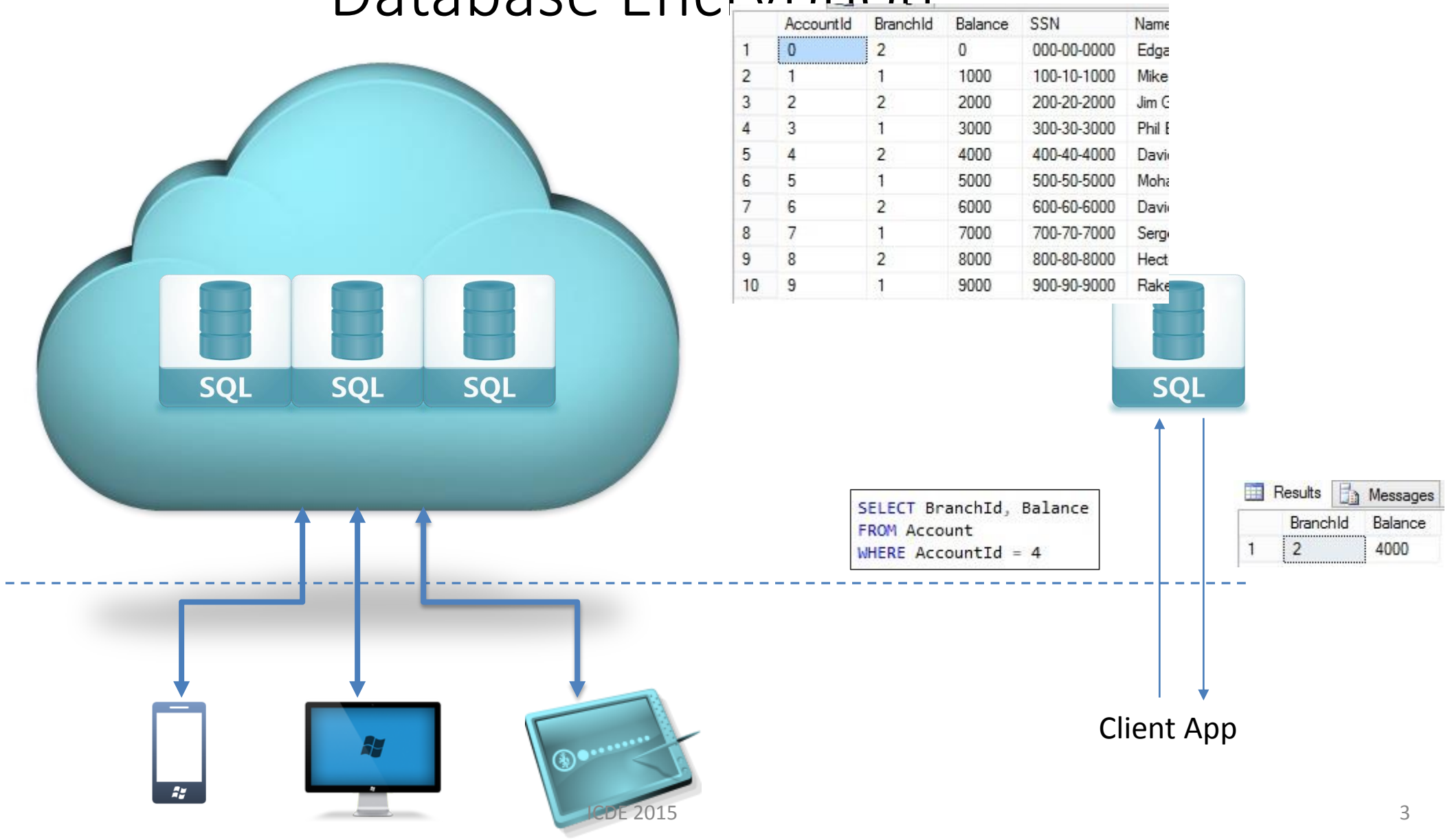
Cloud Data Security Concerns



Data in the cloud vulnerable to:

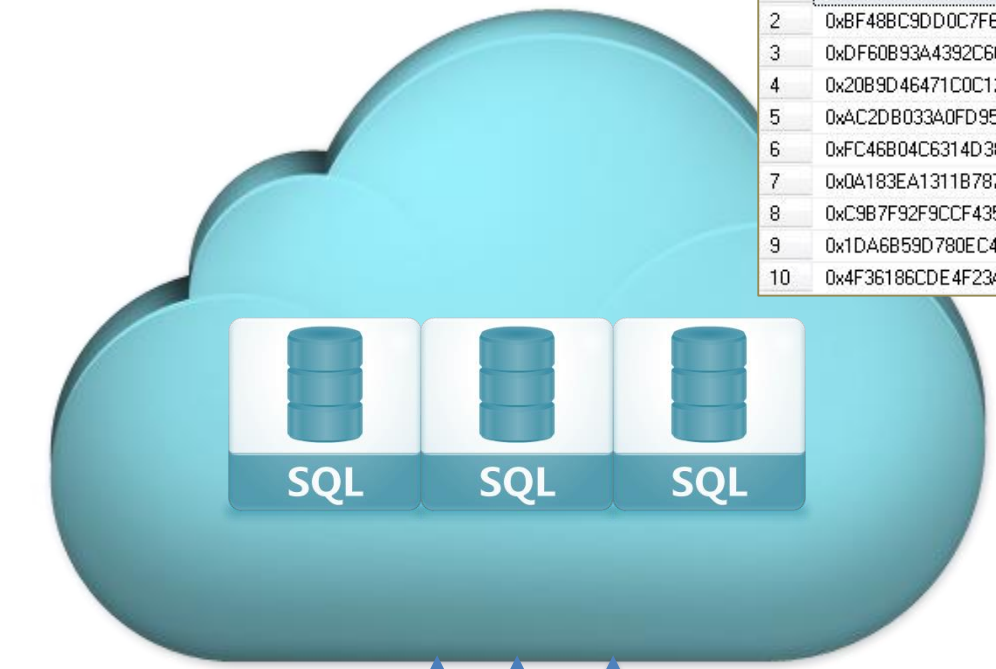
- Snooping administrators
- Hackers with illegal access
  - Compromised servers

Database Encryption



Database Encryption

	AccountId	BranchId	Balance
1	0x6C26AB4AC703DCDF1DA14D22BED9DA73B72FDDDEFE6EF6...	0x3CE12E3A28B06208EA28D6A15F73BAEDFB8EB98BFBBC34...	0xC14D28
2	0xBF48BC9DD0C7F63FD0172C1F5680E4AB2AE0A0490C5535446...	0x6D070C5D217050746B87A6895BF2D1167DC97A42BFC60C7...	0xBB401F
3	0xDF60B93A4392C6C0CAF771B37D9D3ECE17771D9816456158A...	0x619E2D7972F82AE764A8A488D1F0204FF697F13441ED6B34...	0xA525AD
4	0x20B9D46471C0C12B6D02FAACCA00CFE0DAA0A4AEBB131262...	0x3B0F8FCFB5443E19F8AAA601B00BDB49CD552A7C5032E34...	0x3AEFAF
5	0xAC2DB033A0FD9547B425A79BA69526DA3885FC1942D2EDA7...	0x87C14686E839F5F875872DAEA5D0D1B4FD22FDB04B3357D...	0x0C4F7E
6	0xFC46B04C6314D38B2DEF5745DA7D725DDBFF01A41AD8EB6B...	0xF79A1559C558BCEB8499916AA0FB1A96AB36582571E582F1...	0x9699020
7	0x0A183EA1311B787CD2E83865E7FA882EF20F3FB8F87FC50491...	0x302C61B667BD221D9BB810FE65919296AA7A1E7ACD93BB3...	0x38AA640
8	0xC9B7F92F9CCF435347CF6E155AFDB11AC3DC55BA9B7026E35...	0x87ACEE700B90F2D5203E1284A4DCB23F04819504A7BD5B0...	0x23716E4
9	0x1DA6B59D780EC407186D9AC7538E6B89DBF5F7161EC157F8F...	0x3F6B565038C35E821C1D1B7DCC3195BFDE2D8A57551D4E...	0x7928165
10	0x4F36186CDE4F23AFDCCAD53850A6529D9FCD3A4784E50DD9...	0xE2B2353C76B9381D52D762E1A9599A567317807655372823...	0xF9EDC9




```
SELECT BranchId, Balance
FROM Account
WHERE AccountId = 0x8de526fdc17ad9ca4ba...
```

Results | Messages

	BranchId	Balance
1	0x34cfd5b...	0x90c2b893...

Client App

Cipherbase Summary

- Data Confidentiality:
  - Strong column-level encryption
  - Decoupled from functionality
  - *Lightweight “trusted module” in secure hardware
- Functionality:
  - Industrial Strength Database system (SQL Server)
  - Concurrency, Recovery, Stored Procedures.
- Performance on TPCC
  - 85% of plaintext for typical encryption
  - 40% of plaintext for “worst case” encryption

No prior work with this

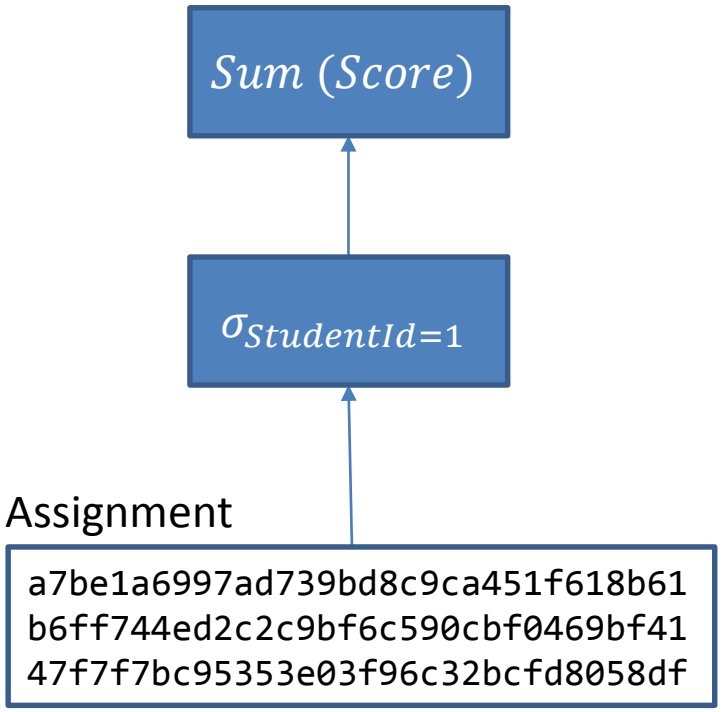
{Confidentiality, Functionality, Performance}

characteristics

Organization

- Introduction
- Solution Landscape & Design Choices
- Cipherbase Design & Engineering
- Evaluation

What Makes Encryption Challenging?



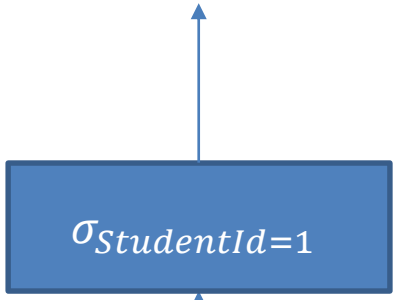

```
Select Sum (Score)
From Assignment
Where StudentId = 1
```

Solution Landscape

- Two fundamental techniques
  - Directly compute over encrypted data
    - Special ***homomorphic*** encryption schemes

Deterministic Encryption

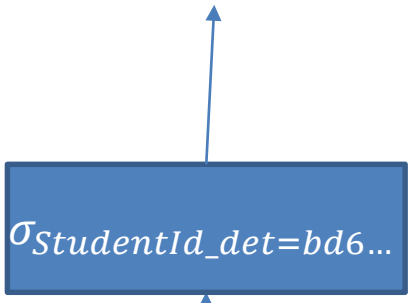
```
select *
from assignment
where studentid = 1
```



StudentId	AssignId	Score
1	1	68
1	2	71
3	4	99
…	…	…

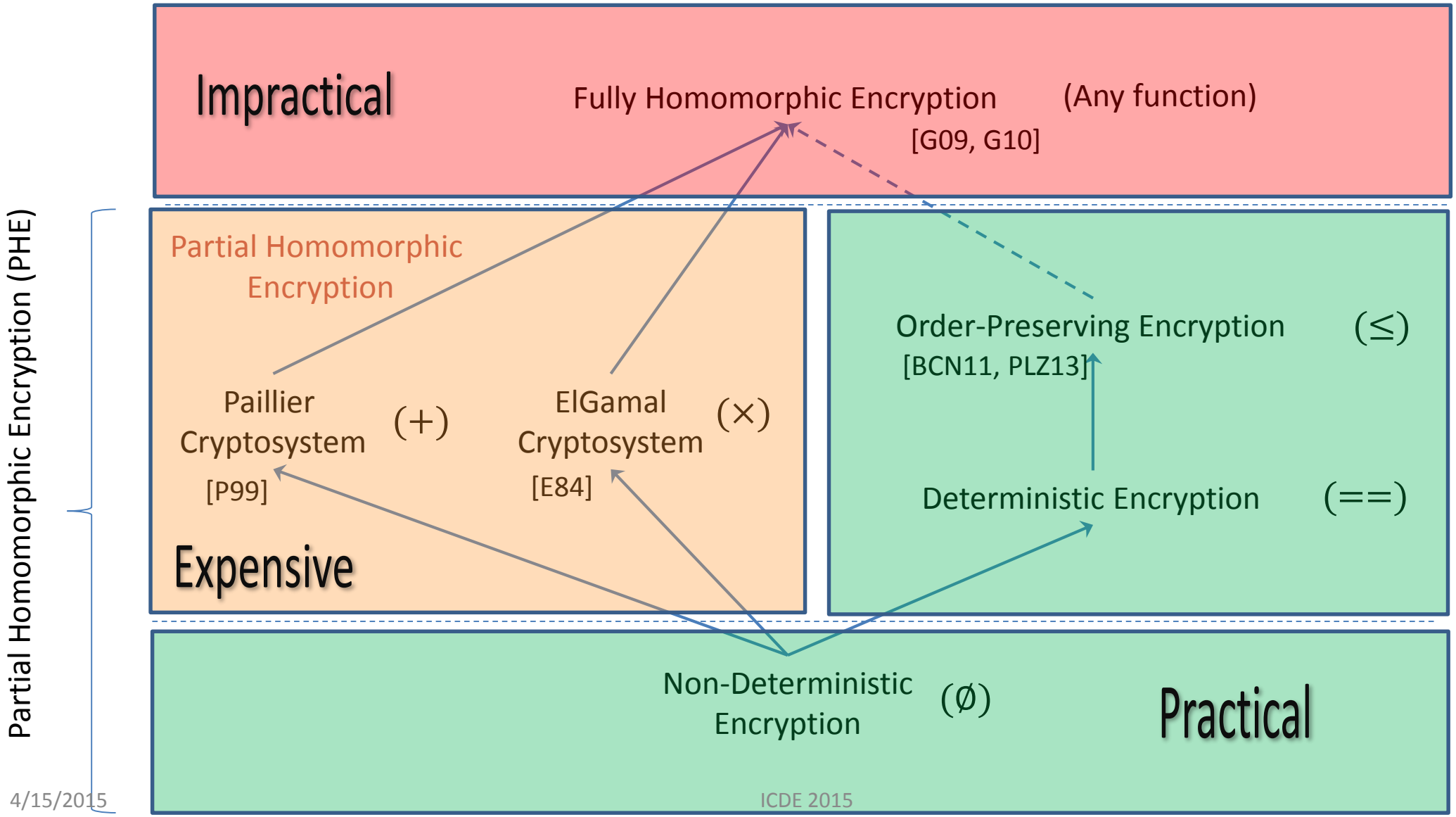
Deterministic Encryption

```
select *
from assignment
where studentid_det = bd6e7c3df2b5779e0b61216e8b10b689
```



StudentId_DET	AssignId	Score
bd6e7c3df2b5779e0b61216e8b10b689	1	68
bd6e7c3df2b5779e0b61216e8b10b689	2	71
7ad5fda789ef4e272bca100b3d9ff59f	4	99
…	…	…

Homomorphic Encryption Schemes



PHE Limitations

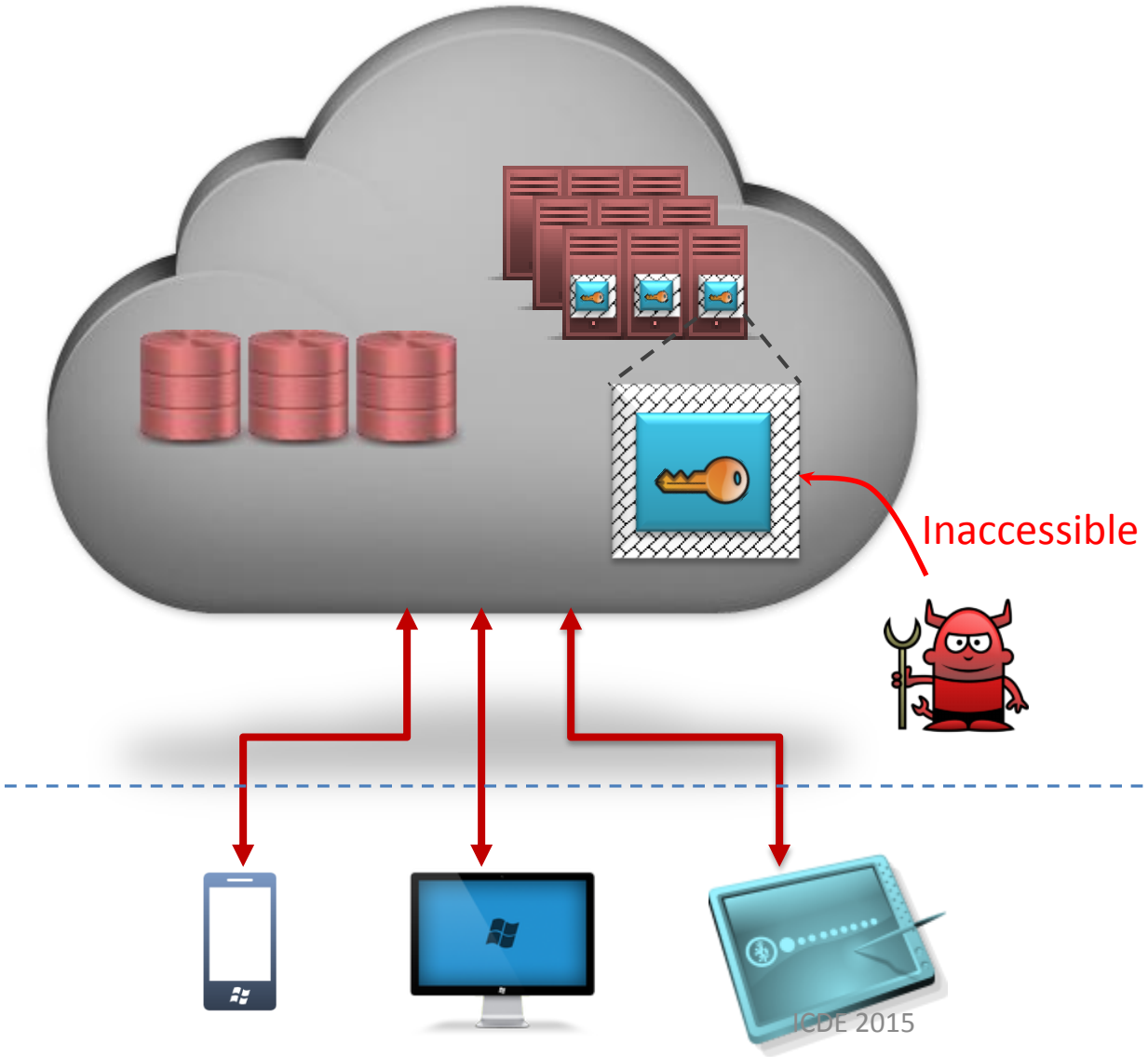
- Limited Server Functionality
  - `SUM(L_EXTENDEDPRICE*(1-L_DISCOUNT)*(1+L_TAX))`
- Data Security tied to functionality
- Lack of Composability
  - `A + B = C`
- Performance
  - ≈ msec for a single addition under Paillier

CryptDB [PRZ+11], Monomi [TFM 13], [HMH08]

Solution Landscape

- Two fundamental techniques
  - Directly compute over encrypted data
    - Special ***homomorphic*** encryption schemes
    - Challenge: limited class of computations
    - Challenge: Not composable
  - Use a “secure” location
    - Hardware provisioned isolation and protection
    - Computations on plaintext

Secure Location

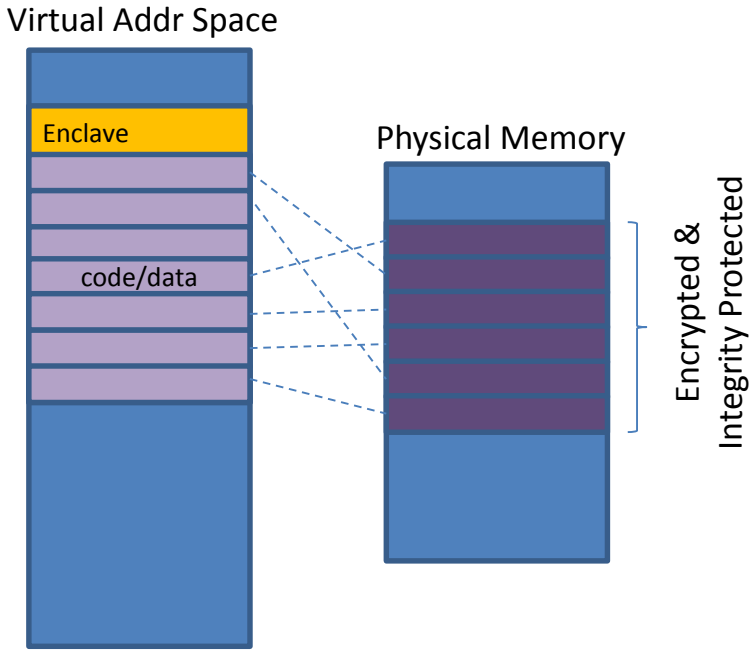


Secure Hardware Landscape

- Long history
  - Banking, Defense Applications
- Becoming mainstream and commoditized
- Players:
  - Crypto co-processors
  - FPGAs
  - Intel SGX
  - TPM, HSM

Intel Software Guard Extensions

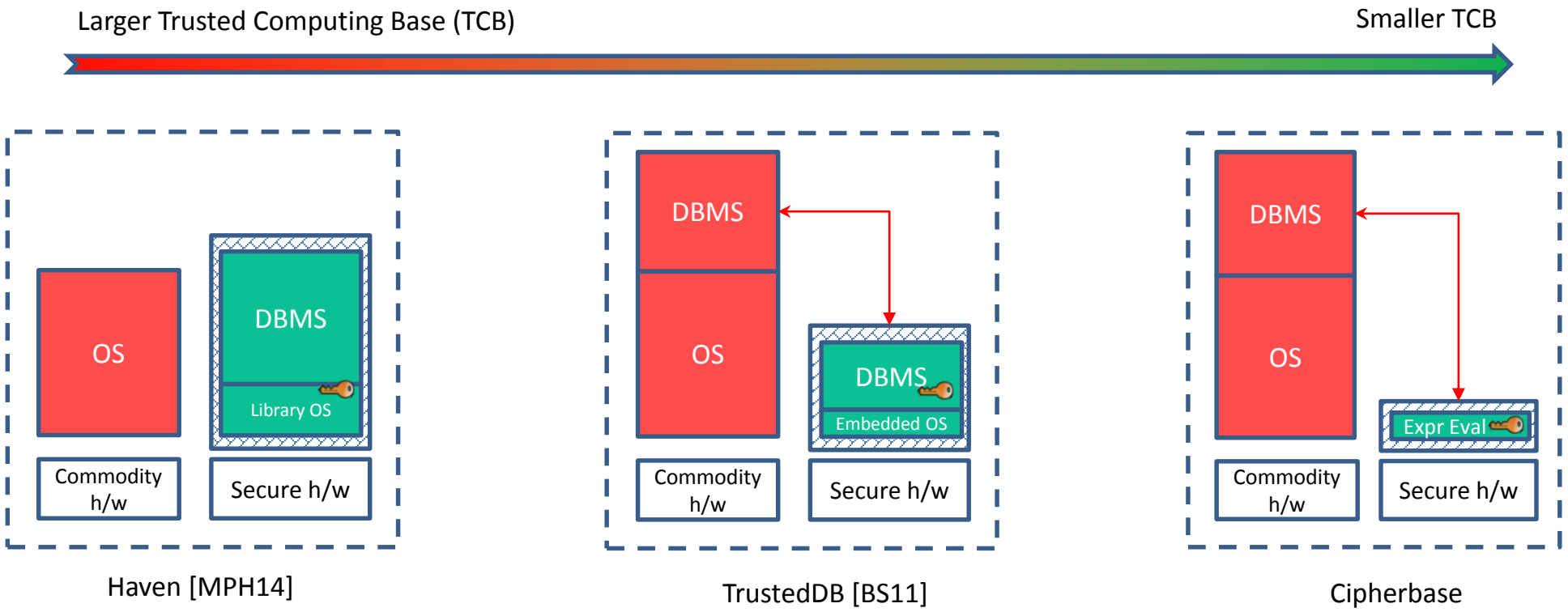
- Extensions to Intel Architecture
- Isolation to code + data within a designated region called *enclave*
  - Confidentiality
  - Integrity



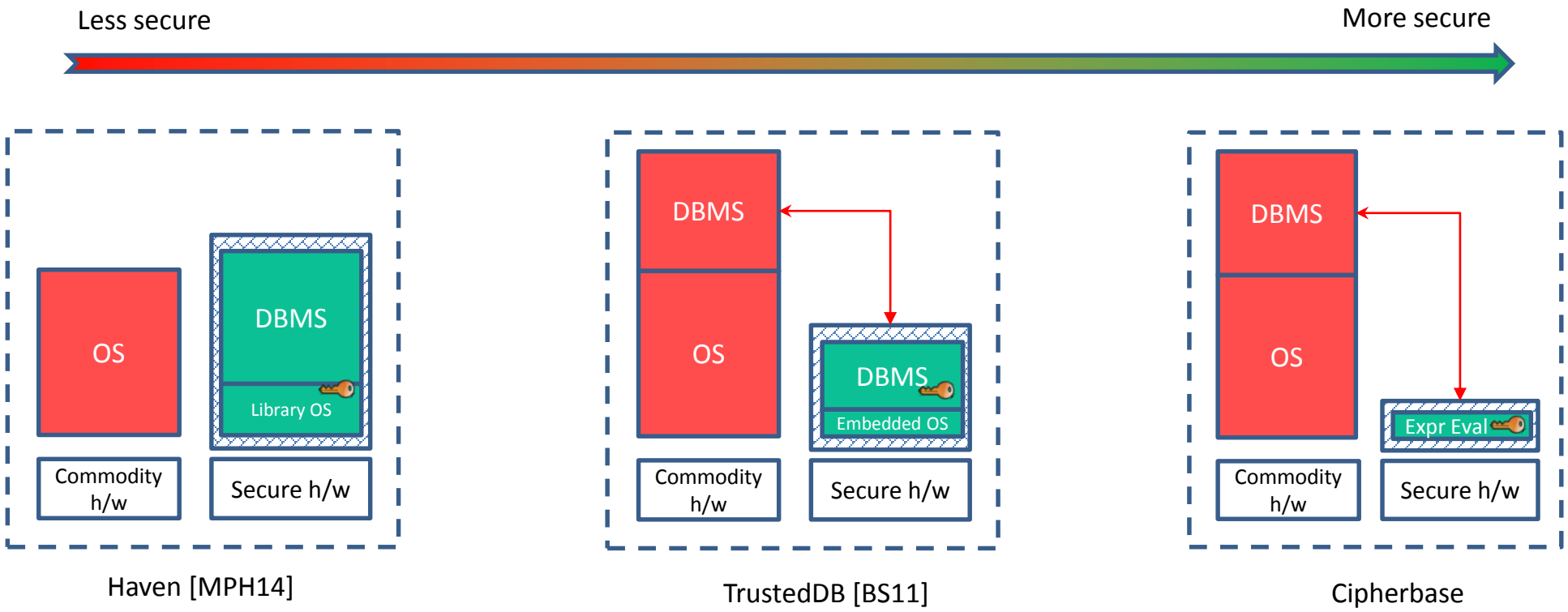
Ack: Andrew Baumann

[MAB+ 13, AGJ+ 13, HLP+ 13]

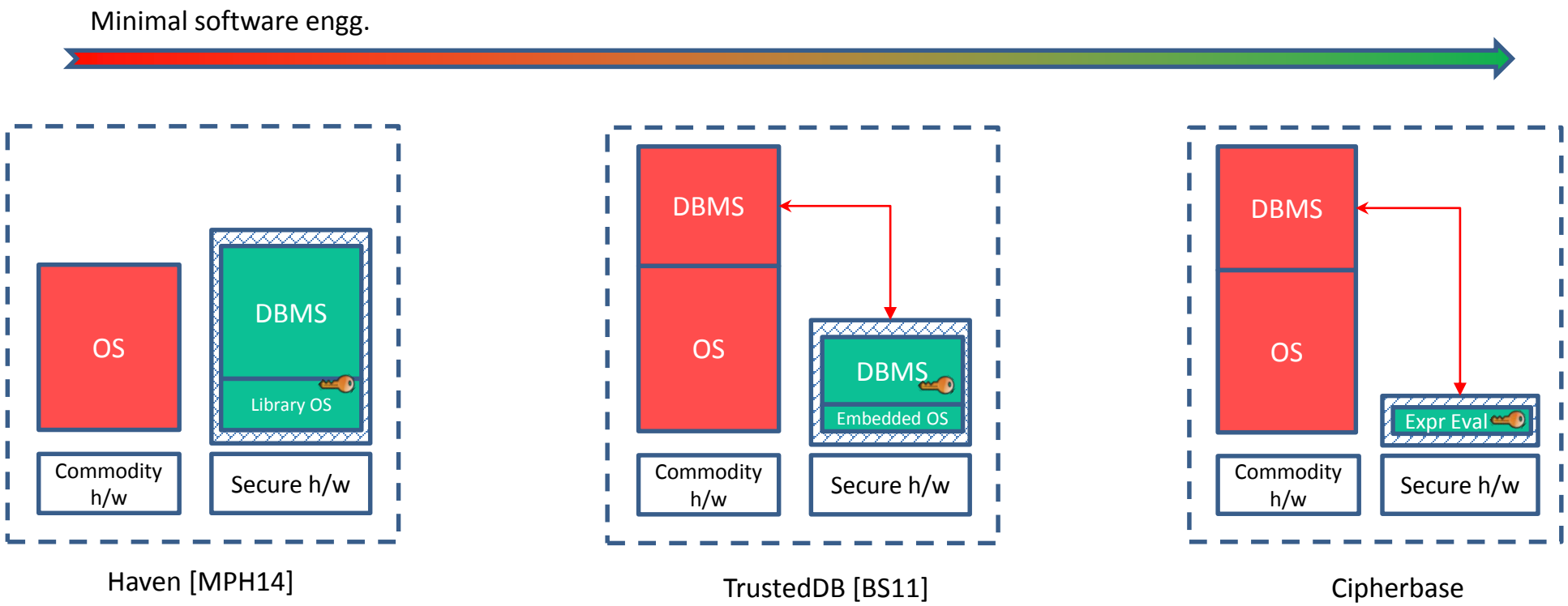
Design Choice: Trusted Functionality



Design Choice: Trusted Functionality



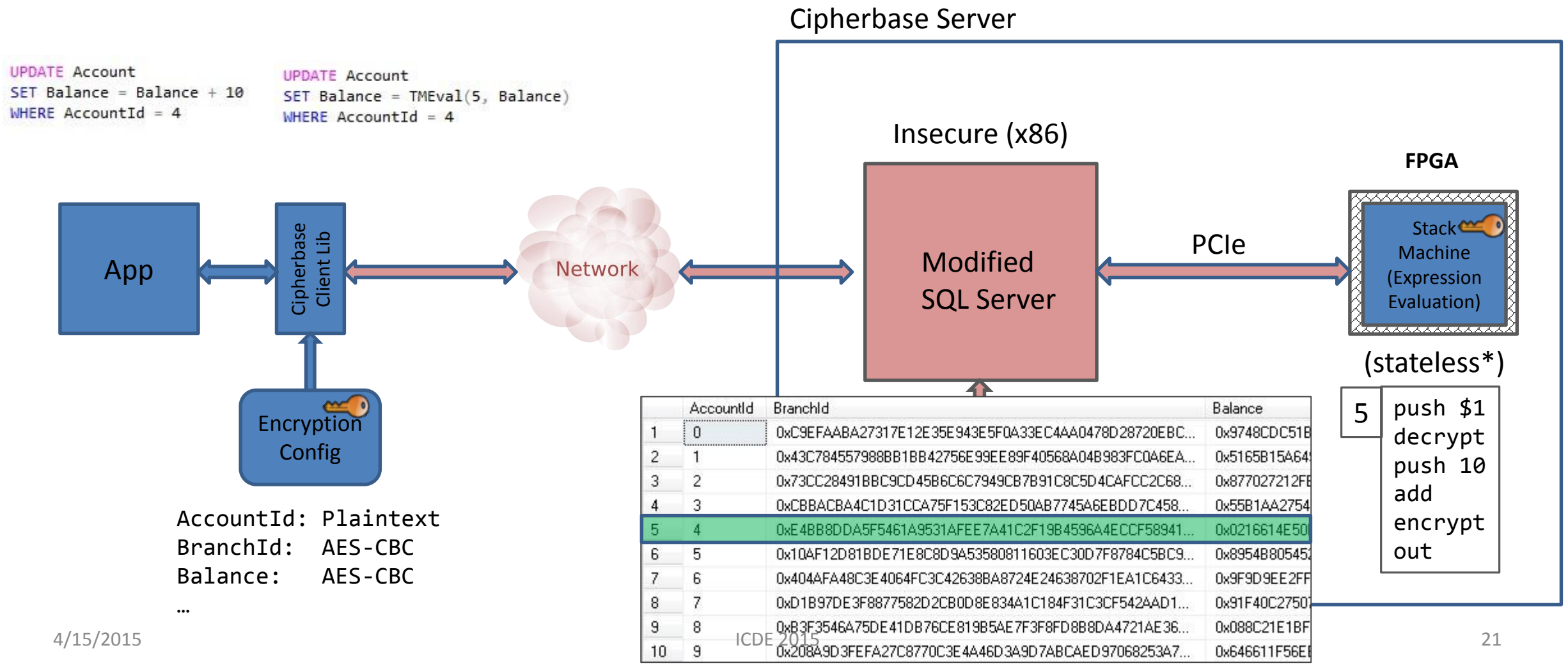
Design Choice: Trusted Functionality



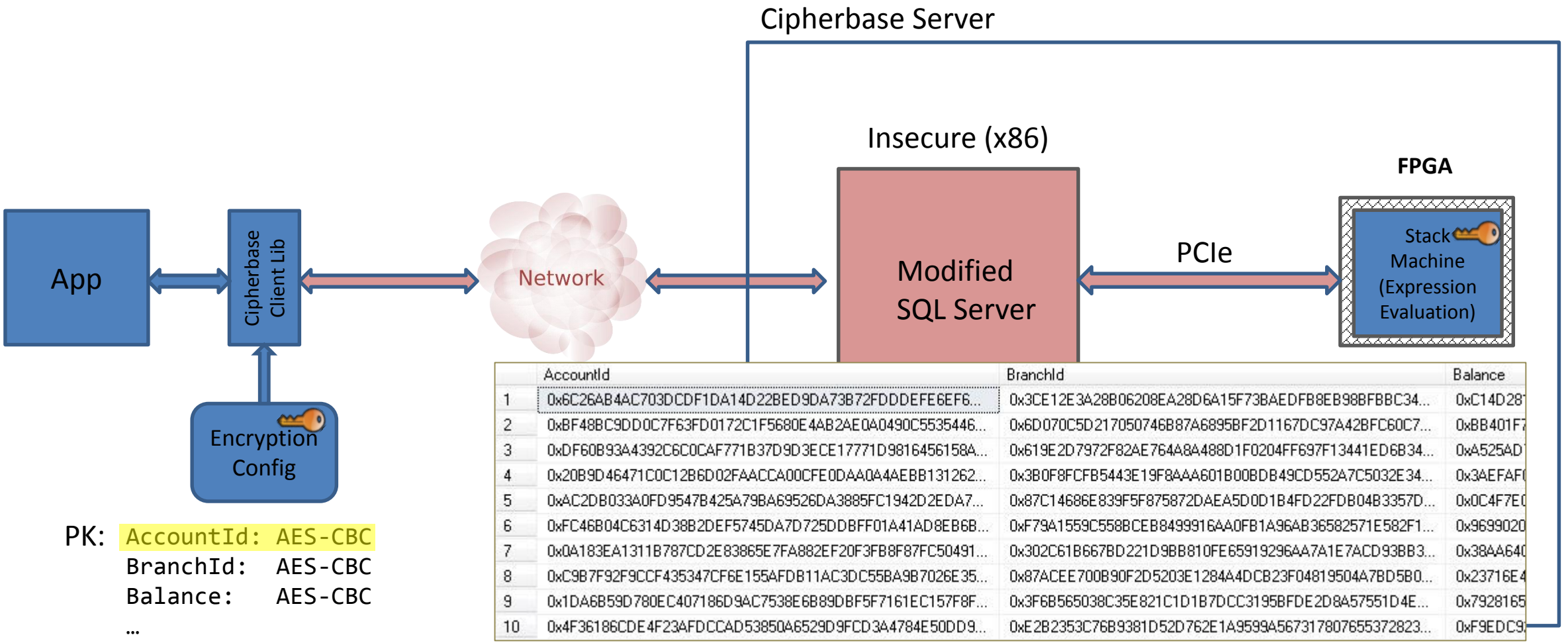
Organization

- Introduction
- Solution Landscape & Design Choices
- Cipherbase Design & Engineering
- Evaluation

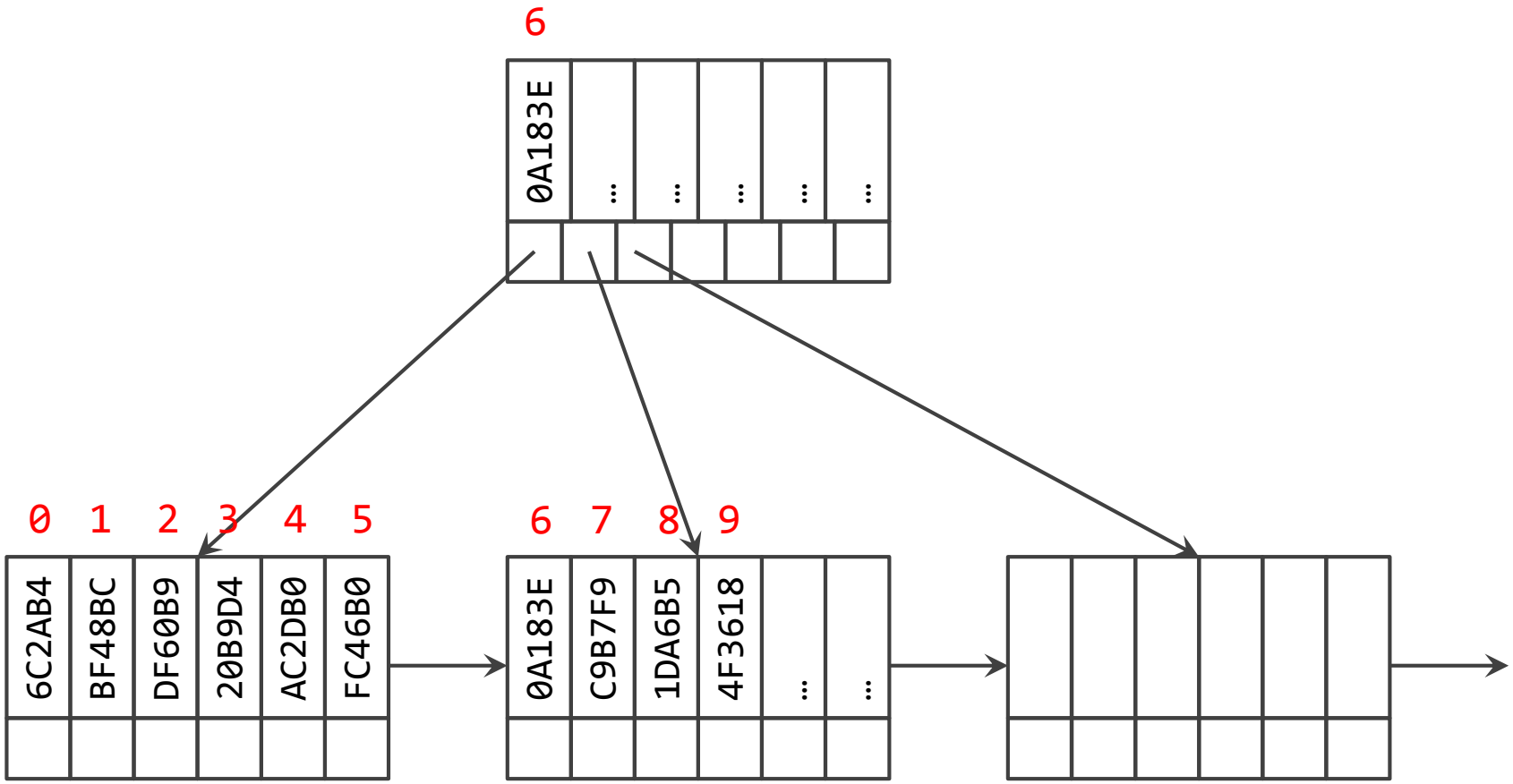
Life of a Query in Cipherbase I



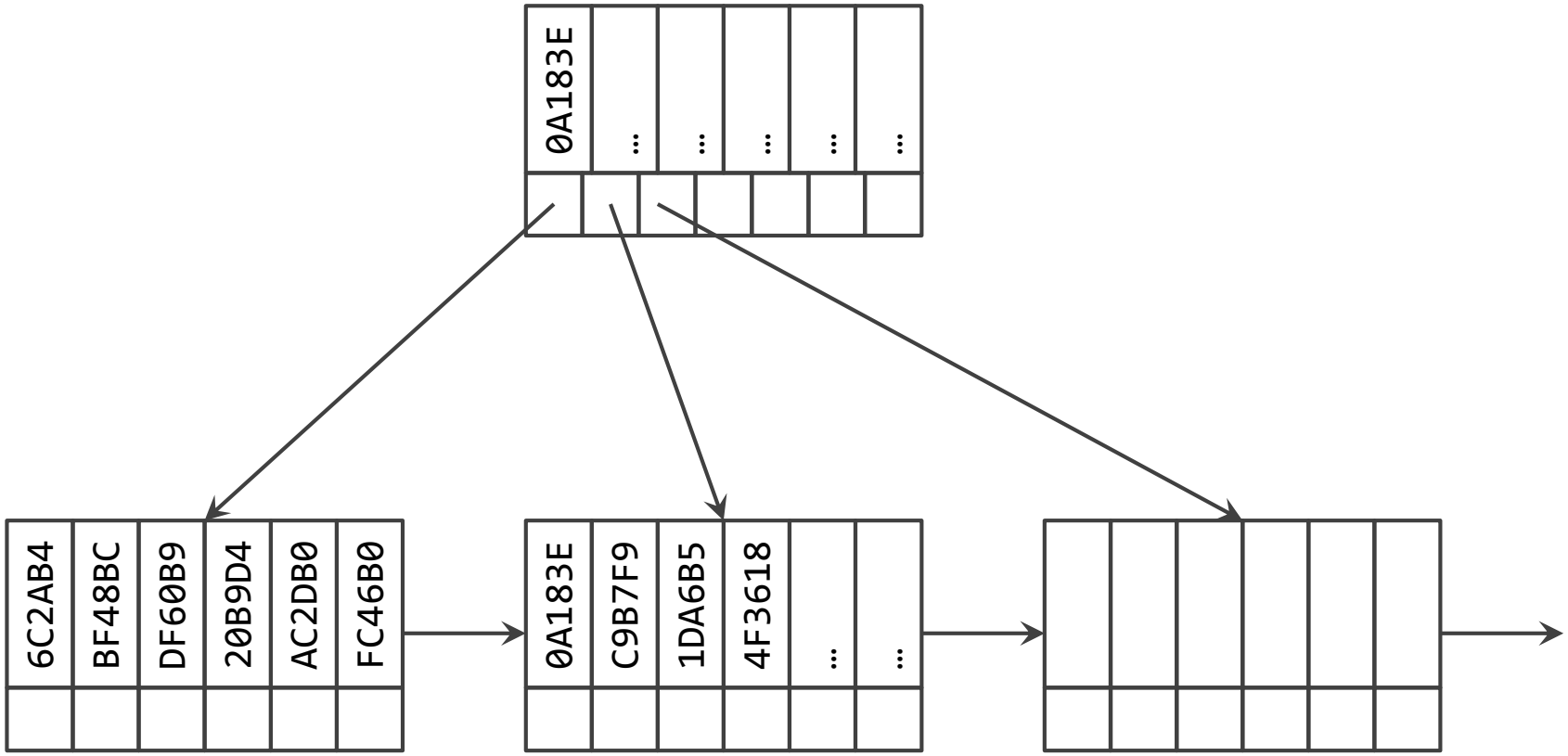
Life of a Query in Cipherbase II



B+-Tree Indexes over Encrypted Data



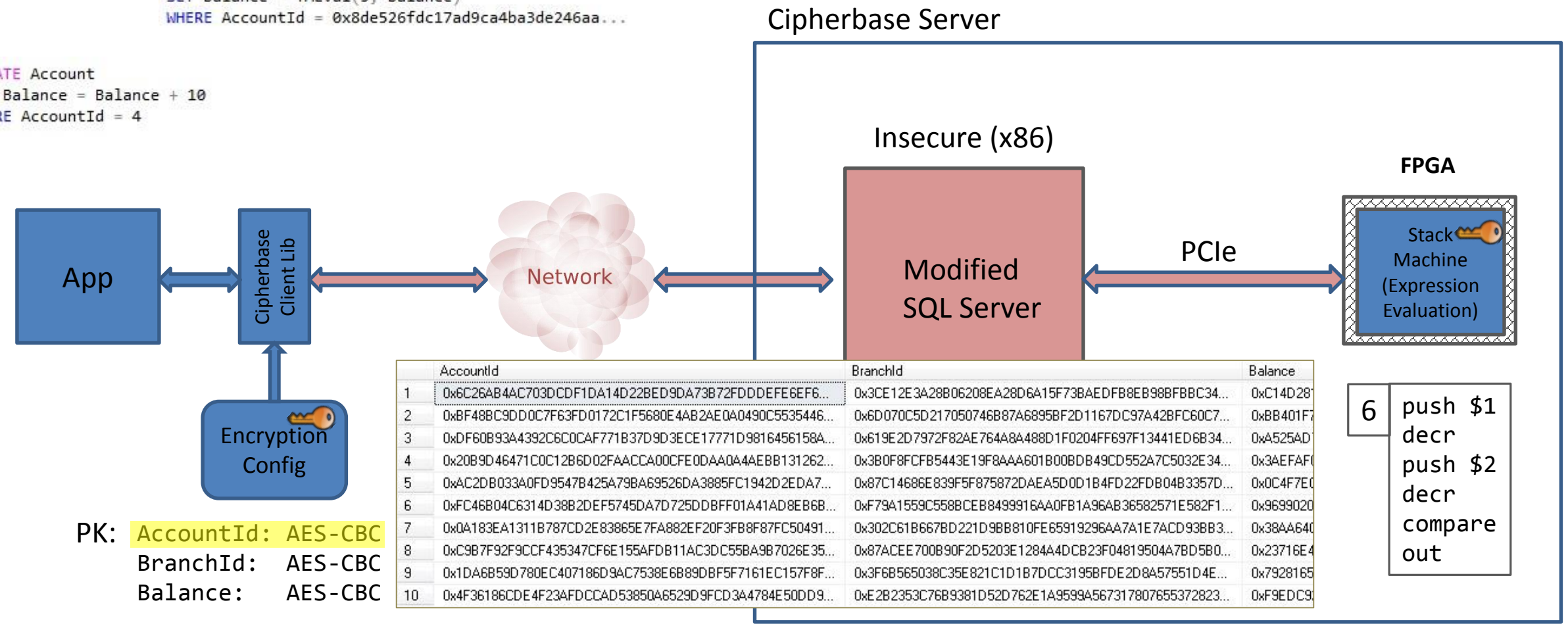
B+-Tree Indexes over Encrypted Data



Life of a Query in Cipherbase II

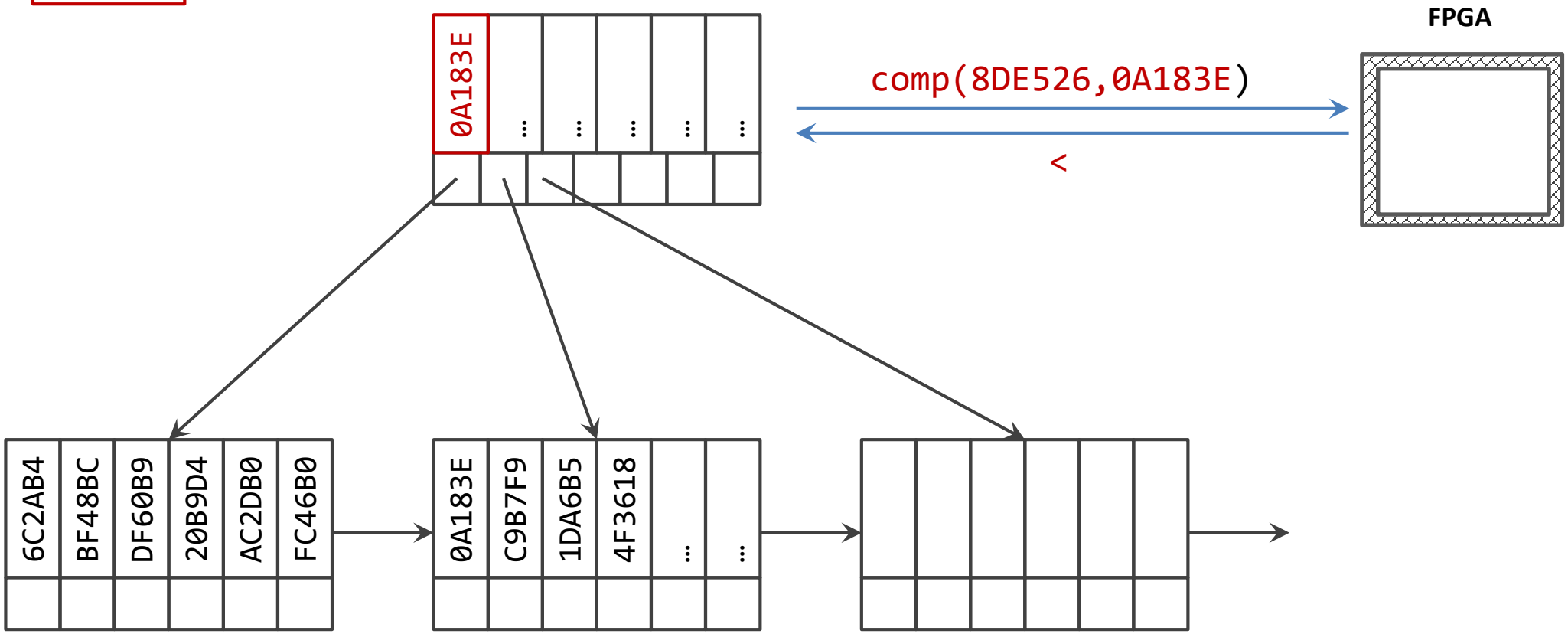
```
UPDATE Account
SET Balance = TMEval(5, Balance)
WHERE AccountId = 0x8de526fdc17ad9ca4ba3de246aa...
```

```
UPDATE Account
SET Balance = Balance + 10
WHERE AccountId = 4
```

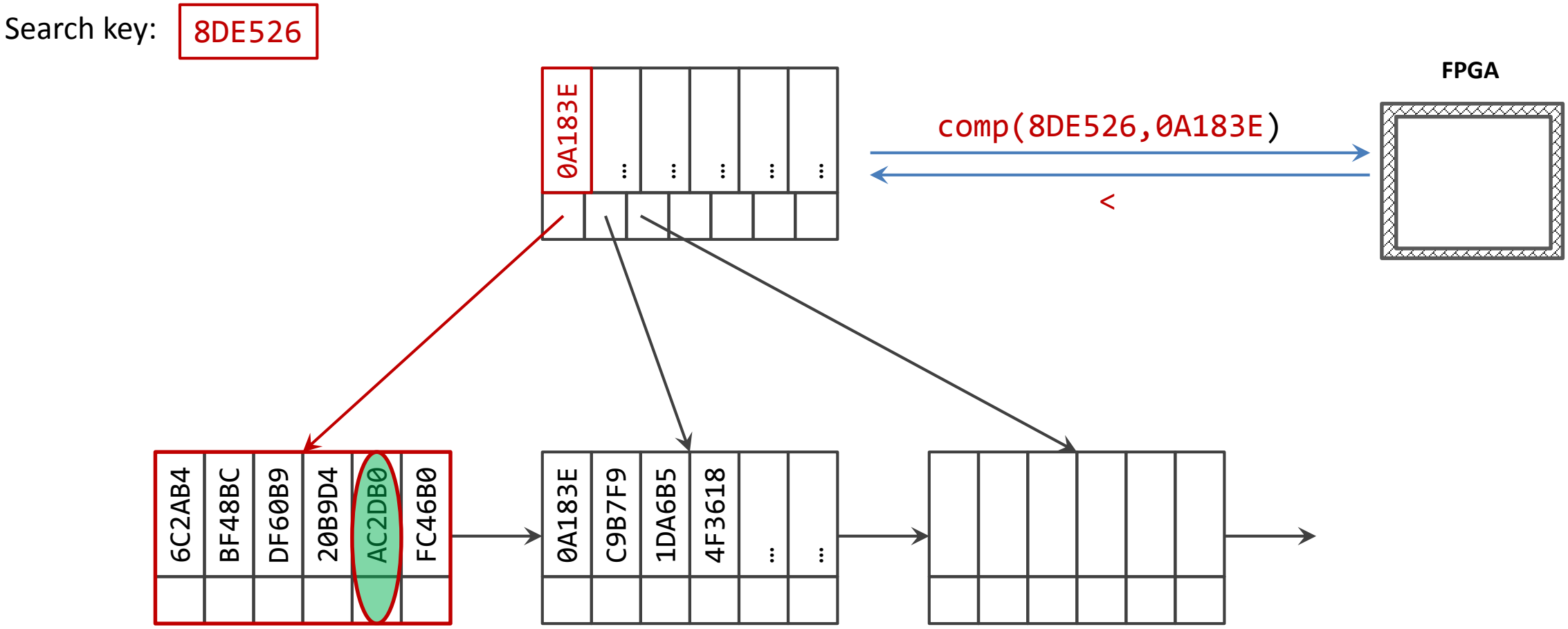


B+-Tree Indexes over Encrypted Data

Search key: 8DE526



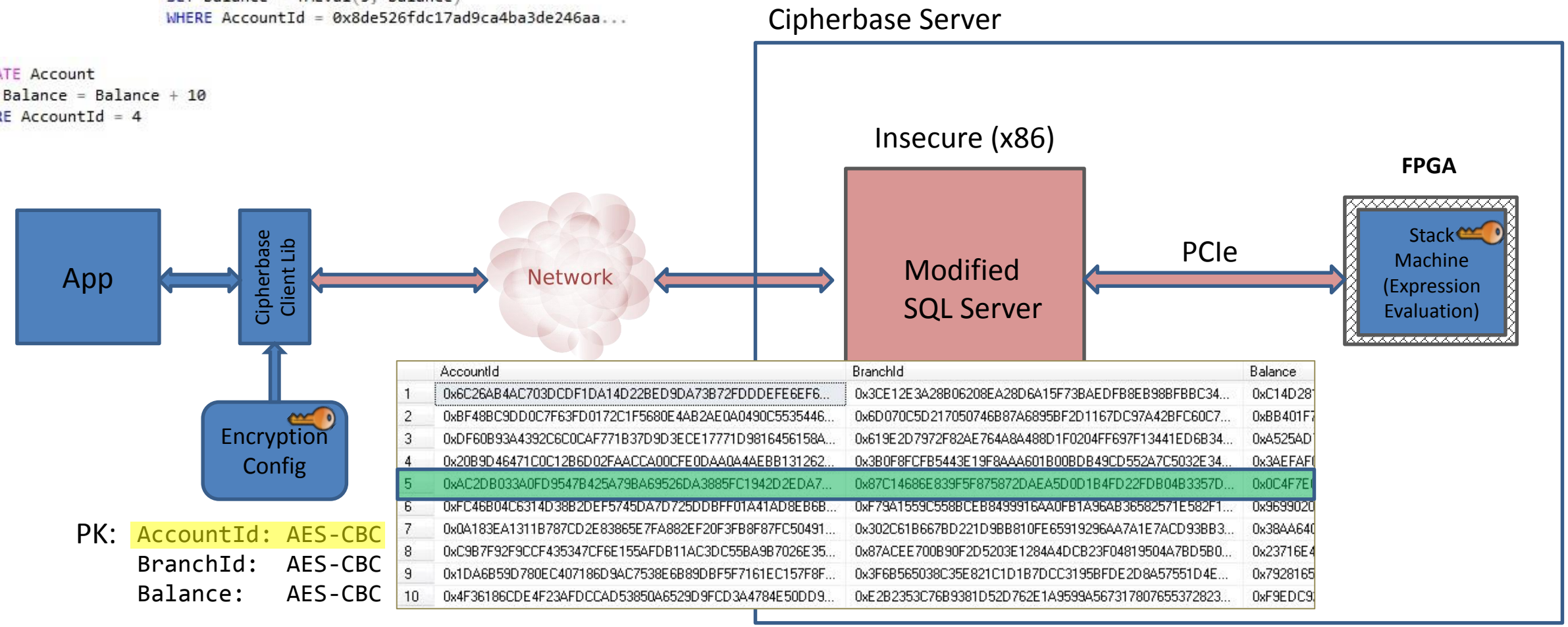
B+-Tree Indexes over Encrypted Data



Life of a Query in Cipherbase II

```
UPDATE Account
SET Balance = TMEval(5, Balance)
WHERE AccountId = 0x8de526fdc17ad9ca4ba3de246aa...
```

```
UPDATE Account
SET Balance = Balance + 10
WHERE AccountId = 4
```



	AccountId	BranchId	Balance
1	0x6C26AB4AC703DCDF1DA14D22BED9DA73B72FDDDEFE6EF6...	0x3CE12E3A28B06208EA28D6A15F73BAEDFB8EB98BFBBC34...	0xC14D28
2	0xBF48BC9DD0C7F63FD0172C1F5680E4AB2AE0A0490C5535446...	0x6D070C5D217050746B87A6895BF2D1167DC97A42BFC60C7...	0xBB401F
3	0xDF60B93A4392C6C0CAF771B37D9D3ECE17771D9816456158A...	0x619E2D7972F82AE764A8A488D1F0204FF697F13441ED6B34...	0xA525AD
4	0x20B9D46471C0C12B6D02FAACCA00CFE0DAA0A4AEBB131262...	0x3B0F8FCFB5443E19F8AAA601B00BDB49CD552A7C5032E34...	0x3AEFAF
5	0xAC2DB033A0FD9547B425A79BA69526DA3885FC1942D2EDA7...	0x87C14686E839F5F875872DAEA5D0D1B4FD22FDB04B3357D...	0x0C4F7E
6	0xFC46B04C6314D38B2DEF5745DA7D725DDBFF01A41AD8EB6B...	0xF79A1559C558BCEB8499916AA0FB1A96AB36582571E582F1...	0x9699020
7	0x0A183EA1311B787CD2E83865E7FA882EF20F3FB8F87FC50491...	0x302C61B667BD221D9BB810FE65919296AA7A1E7ACD93BB3...	0x38AA640
8	0xC9B7F92F9CCF435347CF6E155AFDB11AC3DC55BA9B7026E35...	0x87ACEE700B90F2D5203E1284A4DCB23F04819504A7BD5B0...	0x23716E4
9	0x1DA6B59D780EC407186D9AC7538E6B89DBF5F7161EC157F8F...	0x3F6B565038C35E821C1D1B7DCC3195BFDE2D8A57551D4E...	0x7928165
10	0x4F36186CDE4F23AFDCCAD53850A6529D9FCD3A4784E50DD9...	0xE2B2353C76B9381D52D762E1A9599A567317807655372823...	0xF9EDC9

PK: `AccountId: AES-CBC`
`BranchId: AES-CBC`
`Balance: AES-CBC`
…

Operational Security

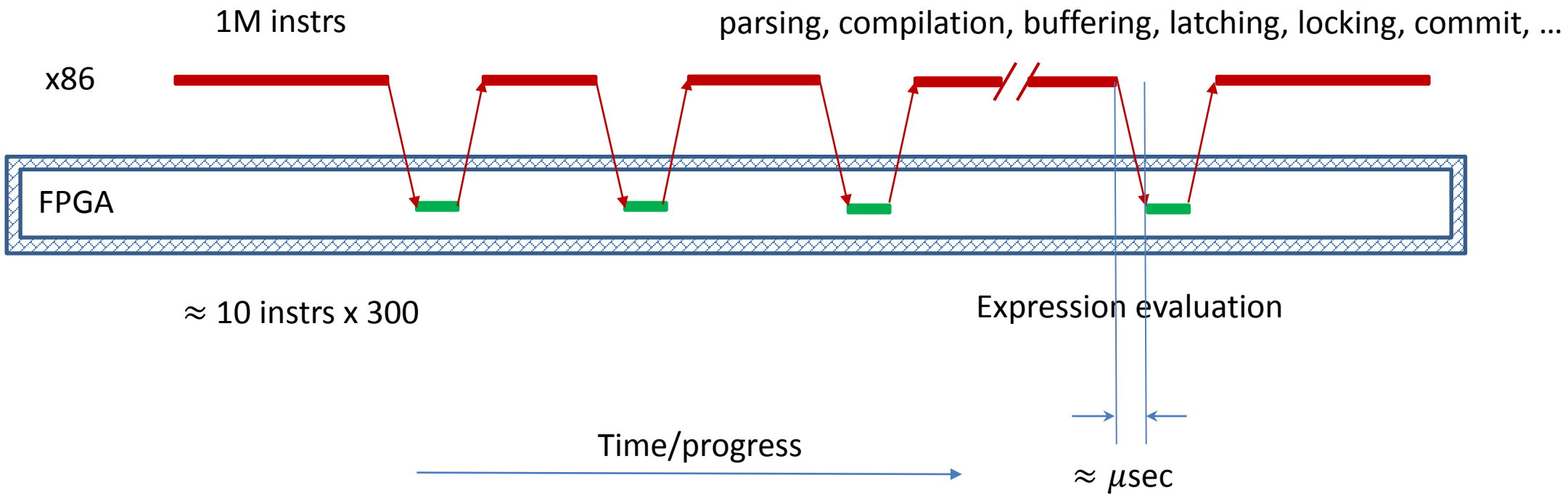
Operation	Adversary Learns
$\sigma_{A=5}$(R)	Unknown predicate p(A) over R tuples
$R \bowtie_A S$ (hash-based)	The join graph and the equivalence relation over R(A) and S(A) for joining A values
$\pi_{A+B}(R)$	Nothing
$Groupby_A^{SUM(B)}(R)$	The equivalence relation over R(A)

Data Security depends on the operations performed

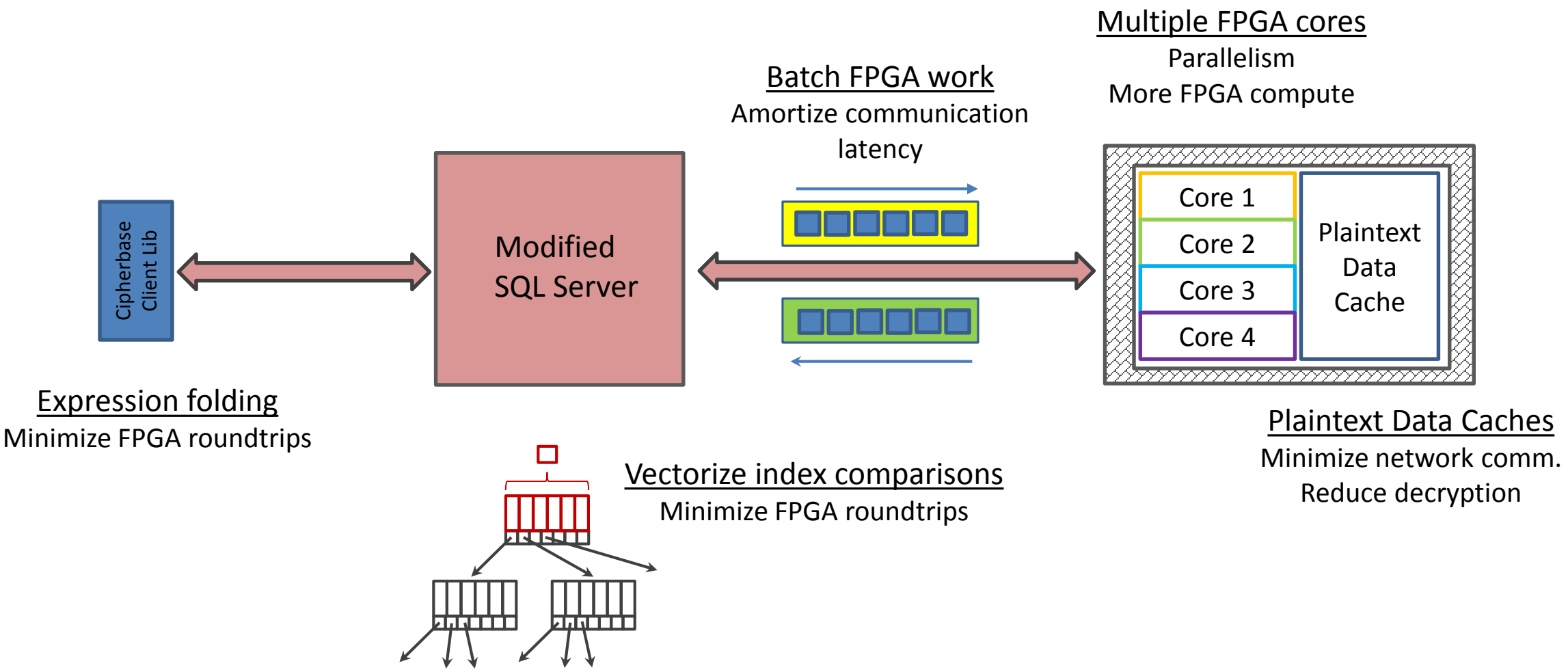
Transaction Processing Performance Challenges

Life of a transaction

TPCC New Order:



Summary of Performance Optimizations



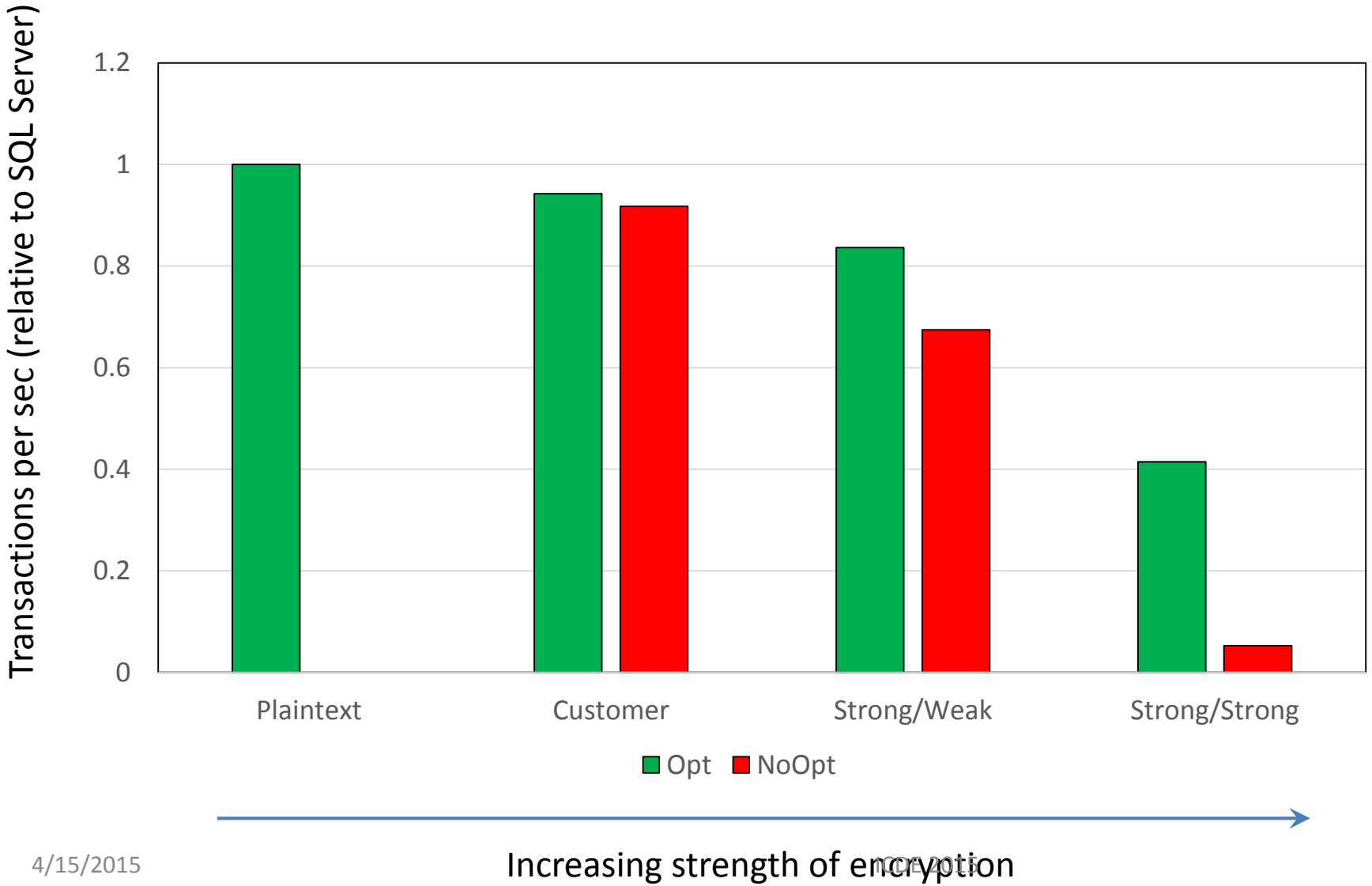
Organization

- Introduction
- Solution Landscape & Design Choices
- Cipherbase Design & Engineering
- Evaluation

Cipherbase Prototype

- SQL Server code
  - Basic functionality
    - ≈ 1000 LoC
    - Localized to expression evaluation module
  - Optimizations
    - ≈ 5000-10000 LoC
    - Localized to FPGA driver, indexing
  - Unchanged: everything else

Performance on TPCC



Encryption schemes:

Customer: Customer PII data strongly encrypted

Strong/Weak: Index columns deterministic, all others strongly encrypted

Strong/Strong: All columns strongly encrypted

Cipherbase Summary

- Security:
  - Strong encryption
  - Decoupled from functionality
- Functionality:
  - Industrial Strength Database system (SQL Server)
  - Transaction Processing
- Performance on TPCC
  - 85% of plaintext for typical encryption
  - 40% of plaintext for “worst case” encryption
- Lightweight “trusted module” in secure hardware

http://research.microsoft.com/en-us/projects/cipherbase/