Project SEEED

Processing of Encrypted Data in SAP HANA
Outsourcing Data to the Cloud

- What do you think are the problems?
Outsourcing Data to the Cloud

- What do you think are the problems?
  - Legally
  - Organisationally
  - Technically
    - Provisioning
    - Lifecycle Management
    - ......
  - Security
    - Attacker Models?
Agenda

What is Project SEEED?

Basic Architecture & Attacker Model

DEMO

eColumns / Proxy re-encryption

Detailed Architecture in SAP HANA

Research Outlook
Processing SQL directly over encrypted data in the cloud is feasible.

- SEEED is not only encryption at rest but directly processing encrypted data.
- Different cryptographic algorithms support different SQL operators (select, =, <, >, sum(), count(), groupby(), …).
- Primary keys NEVER leave the client!
- Even if the cloud operator makes a memory dump he will not learn anything.
Searching Encrypted Data in HANA - DEMO

User View

HANA View

aggregating encrypted columns
joining encrypted columns
Try it yourself in our Sandbox

Sample query 1: query with COUNT, EXISTS subquery, JOIN, GROUP BY and ORDER BY

```
SELECT
  O_ORDERPRIORITY,
  COUNT(*) AS ORDER_COUNT
FROM ORDERS
WHERE O_ORDERDATE >= '1993-07-01'
AND O_ORDERDATE < '1993-10-01'
AND EXISTS (SELECT *
  FROM LINEITEM
  WHERE L_ORDERKEY = O_ORDERKEY
  AND L_COMMITDATE < L_RECEIPTDATE)
GROUP BY O_ORDERPRIORITY
ORDER BY O_ORDERPRIORITY
```

<table>
<thead>
<tr>
<th>O_ORDERPRIORITY</th>
<th>COUNT(*)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-URGENT</td>
<td>6733</td>
</tr>
<tr>
<td>2-HIGH</td>
<td>6741</td>
</tr>
<tr>
<td>3-MEDIUM</td>
<td>6899</td>
</tr>
<tr>
<td>4-NOT SPECIFIED</td>
<td>6765</td>
</tr>
<tr>
<td>5-LOW</td>
<td>6840</td>
</tr>
</tbody>
</table>

Query execution time on 3 millions rows (in ms):
on encrypted data: 315
on plain data: 205
Processing Encrypted Data in the Cloud
Process Summary

No decryption needed!

Keys stay with customer!

Encrypted Data

SQL Query (e.g. Total Sales)

Encrypted Result

Decryption

Cleartext Result

Total Sales: 42

On-Premise

HANA

HANA encrypted data

SQL query

Encrypted result

Decryption

Cleartext result

Total Sales: 42

On-Premise

No decryption needed!

Keys stay with customer!
Attacker Model: Technical Consequences
Encryption can take place in different layers

Traditionally Encrypted Database

Client

Query Optimization and Execution

Relational Operators

File and Access Methods

Buffer Management

Disk Space Management

Disk

SEEED

Client

Query Optimization and Execution

Relational Operators

File and Access Methods

Memory Management

Persistency
Accepting the Trade-offs

Performance

Security  Functionality
## Performance Figures
(Median of Server Runtimes in Milliseconds for Table size 1,000,000 rows)

<table>
<thead>
<tr>
<th>Test Case</th>
<th>SEEED</th>
<th>Plain</th>
<th>Impact</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Server-Side Only</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Exact Search</td>
<td>2.0</td>
<td>1.7</td>
<td>1.2 x</td>
</tr>
<tr>
<td>Equi-Join</td>
<td>49.7</td>
<td>33.3</td>
<td>1.5 x</td>
</tr>
<tr>
<td>Grouping with Aggregation (Sum)</td>
<td>674.1</td>
<td>57.8</td>
<td>11.7 x</td>
</tr>
<tr>
<td><strong>Incl. Client-Side</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Order by Aggregate (Sum)</td>
<td>870.1</td>
<td>56.3</td>
<td>15.4 x</td>
</tr>
<tr>
<td><strong>TPCH</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q4</td>
<td>2,402</td>
<td>235</td>
<td>10.2 x</td>
</tr>
<tr>
<td>Q5</td>
<td>1,373</td>
<td>207</td>
<td>6.6 x</td>
</tr>
</tbody>
</table>
## Analysis

<table>
<thead>
<tr>
<th></th>
<th>TPCH</th>
<th>TPCC</th>
<th>Customer X</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Queries</td>
<td>22</td>
<td>20</td>
<td>406</td>
</tr>
<tr>
<td>Total Tables</td>
<td>8</td>
<td>9</td>
<td>2</td>
</tr>
<tr>
<td>Total Columns</td>
<td>61</td>
<td>71</td>
<td>248</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>TPCH</td>
<td>TPCC</td>
<td>Customer X</td>
</tr>
<tr>
<td>RND</td>
<td>17</td>
<td>49</td>
<td>157</td>
</tr>
<tr>
<td>DET</td>
<td>24</td>
<td>17</td>
<td>74</td>
</tr>
<tr>
<td>OPE</td>
<td>20</td>
<td>5</td>
<td>17</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>RND</td>
<td></td>
<td>63,3%</td>
</tr>
<tr>
<td></td>
<td>DET</td>
<td></td>
<td>29,8%</td>
</tr>
<tr>
<td></td>
<td>OPE</td>
<td></td>
<td>7,9%</td>
</tr>
</tbody>
</table>
SEEED Architecture – Components

Web Application Server

SEEED JDBC Driver

Query Rewriting Logic

Crypto Library

HANA JDBC

Key Store

Temporary Database

Web Application

OLAP Engine

Encrypted Sum Operator

Crypto Library

Database Server

Decrypt on UDF

Database Tables

SQL Engine
SEEED – Customers retain full Data Ownership when moving from OP to HANA as a Cloud DB

- Obtain guarantees about deletion of cloud data (NB: No feature of SEEED yet, unsure about R&D effort)

- Automated analysis, labeling data & specifying policies (eg on financial or health care data)

- HANA-based massively parallelized Encryption

- Process SQL directly over encrypted data
- Enforce access control through cryptography
- Allow for standard cloud maintenance operations

Secure Deletion of Cloud Data
On-Premise to Cloud Provisioning
Processing Encrypted Data in the Cloud
SEEED Encryption Cluster

Encryption of complete clear text tables
• Distributed data encryption across several clusters
• Currently implemented using Apache’s Hadoop*

Exemplary use case
• Largest table has 16 columns and 6 Mio rows
• Total: 9 Tables 14.7 Mio rows

Time for encryption
• Single threaded: ~ 6 months
• Distributed Encryption: ~ 48 hours

*Apache Hadoop: http://hadoop.apache.org/
Summary

- SEEED is not only encryption at rest but directly processing encrypted data.
- Different cryptographic algorithms support different SQL operators.
  (select, =, <, >, sum(), count(), groupby(), …).
- Primary Keys NEVER leave the client!
- Even if the cloud operator makes a memory dump he will not learn anything.
- We support the entire SQL syntax and 17 out of 22 TPCH benchmark queries.
Contact

Dr. Andreas Schaad
Research Manager
andreas.schaad@sap.com

Dr. Florian Kerschbaum
Chief Expert
florian.kerschbaum@sap.com

The Team: Axel, Isabelle, Mathias, Martin H., Walter, Tobias, Patrick, Florian H.