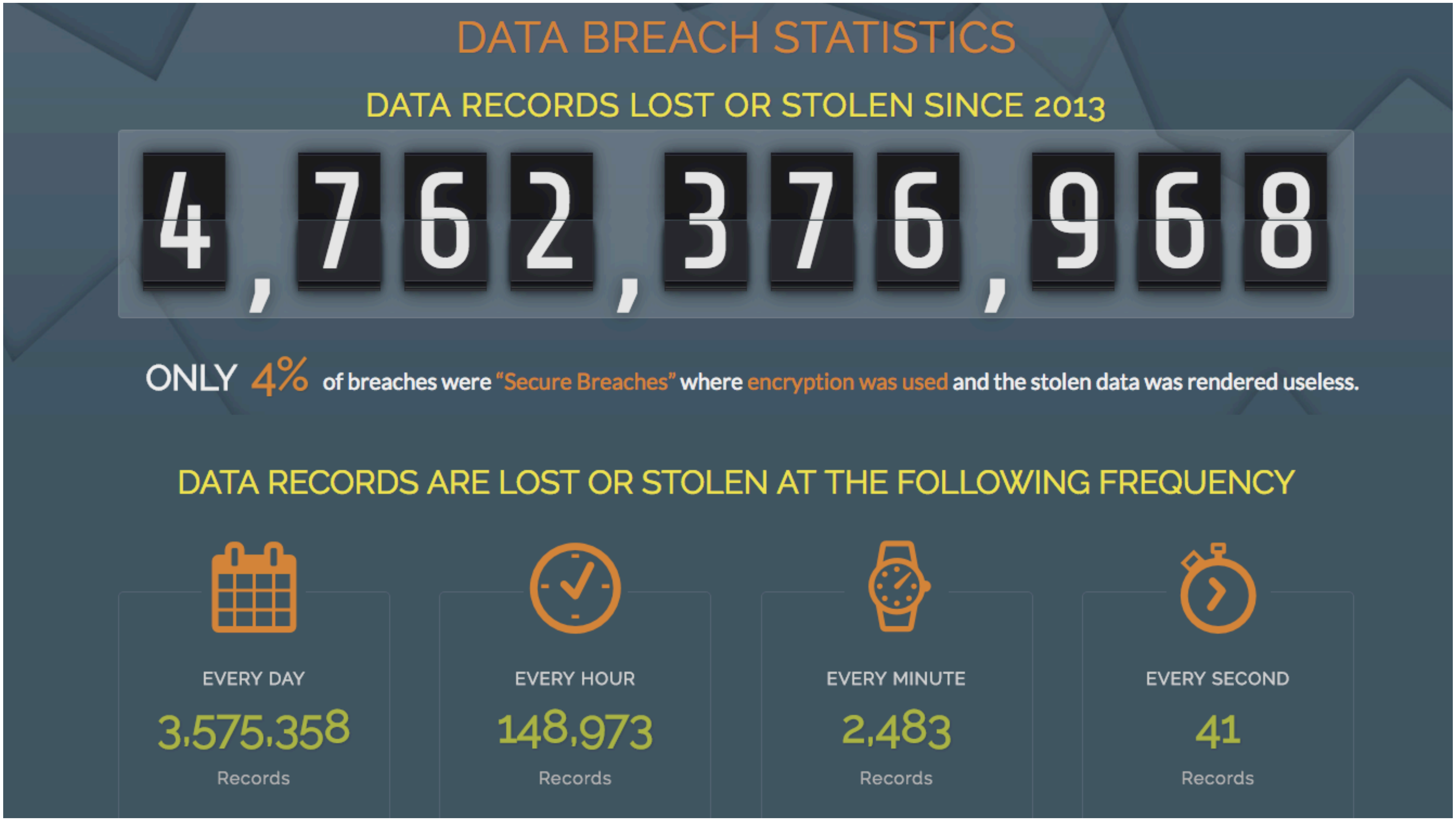
CASSANDRA SUMMIT 2016

Ameesh Divatia

The Promise and Perils of Encrypting Cassandra Data



DATA BREACH STATISTICS

DATA RECORDS LOST OR STOLEN SINCE 2013

4,762,376,968

ONLY 4% of breaches were "Secure Breaches" where encryption was used and the stolen data was rendered useless.

DATA RECORDS ARE LOST OR STOLEN AT THE FOLLOWING FREQUENCY

EVERY DAY	EVERY HOUR	EVERY MINUTE	EVERY SECOND
3,575,358 Records	148,973 Records	2,483 Records	41 Records

Perils of Cloud Hosted Databases

baffle

Cloud hosted database services being adopted by enterprise IT

- Where is the data stored?

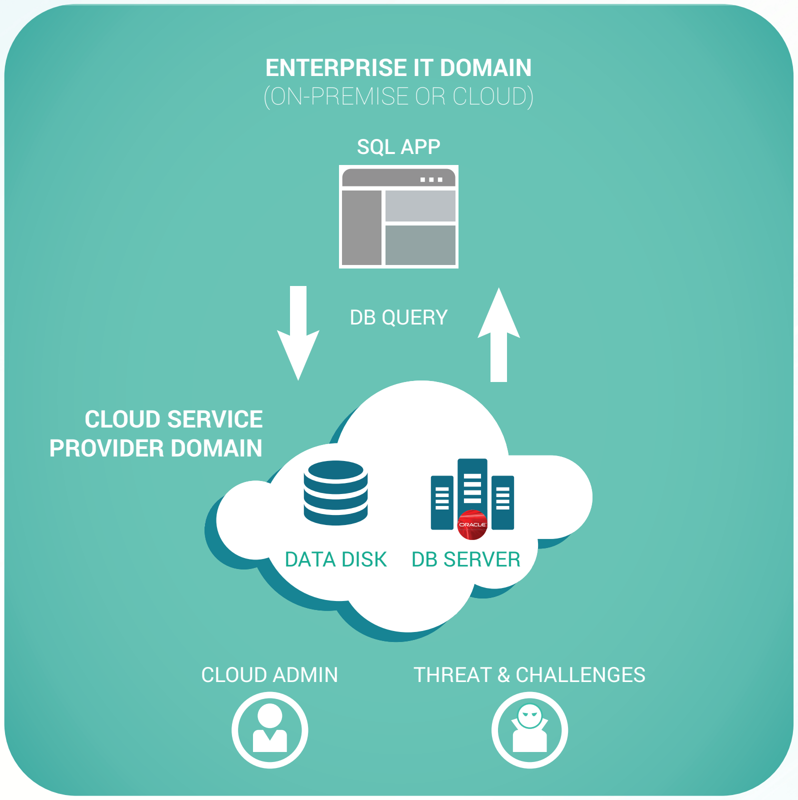
Threats

- Stolen cloud administrator credentials
- Violations of compliance regulations

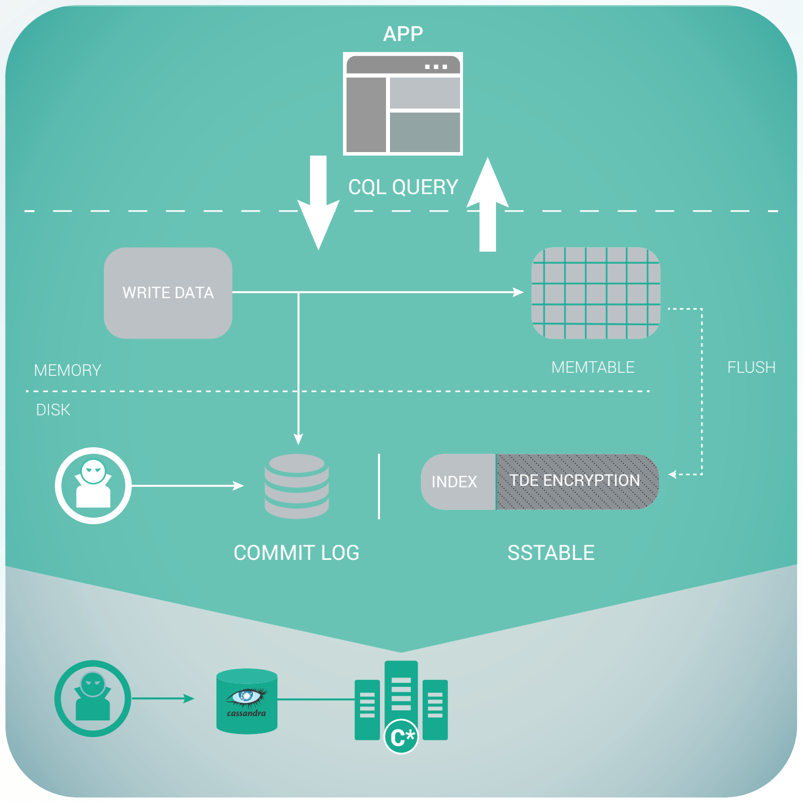
Need

- Complete control of data by enterprise IT at all times
  - Encrypt all data that is uploaded to the cloud

How does enterprise IT deal with this?



Cassandra Security



Use Case – Apache Cassandra

- Raw customer data stored in multi-tenant clusters as clear text
- Encryption available as an option only at rest

Vulnerabilities

- Data in memory is in the clear
- Encryption key is in the clear in memory risking the entire data store

Anatomy of a Hack

baffle

Threat Scenario

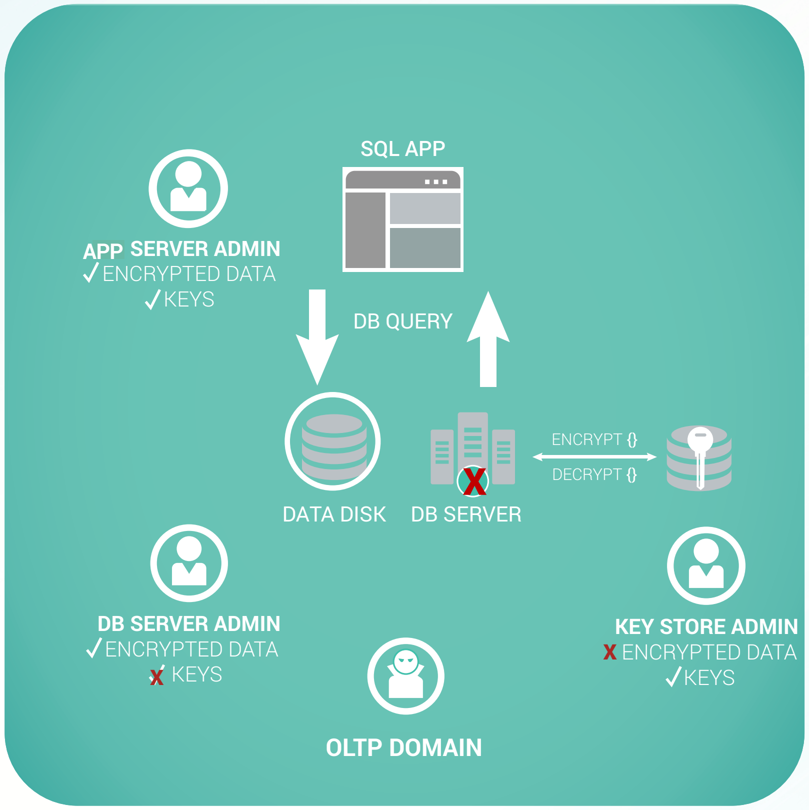
- Database admin has access to data and keys used for 'at rest' security

Vulnerabilities

- Administrator's credentials are stolen by a hacker
- Hacker has access to the data as well as the key

Need

- Separation of keys and data into separate domains
- Never decrypting data in a single compute instance
- **Operate on encrypted data**



Encryption Adoption Challenges

Complexity: Perception that once the data is encrypted, the customer loses control

Key management: What if the keys are lost, stolen or stagnant?

Encryption Adoption Challenges

baffle

Performance degradation: Perception that application performance will slow down

Workflow impact: Changes the current paradigm of app to database communication

A Practical Approach

Driver level insertion for data protection

- Transparent application operation

Seamless key management with Amazon KMS

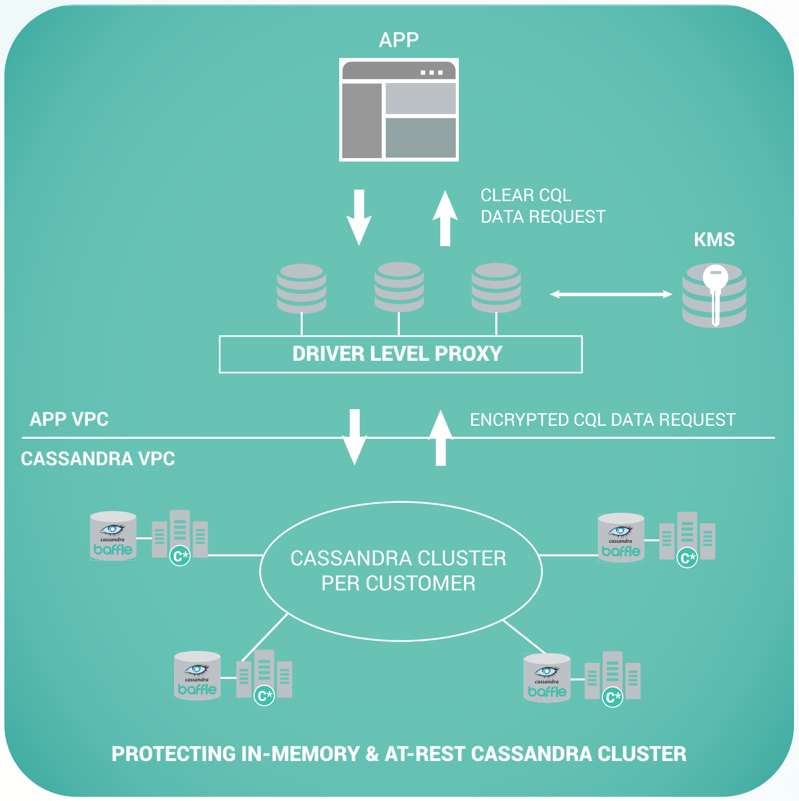
- Generate, use, store, rotate and retire keys

Performance at scale using massive parallelization

- Elastic load balancing in response to demand

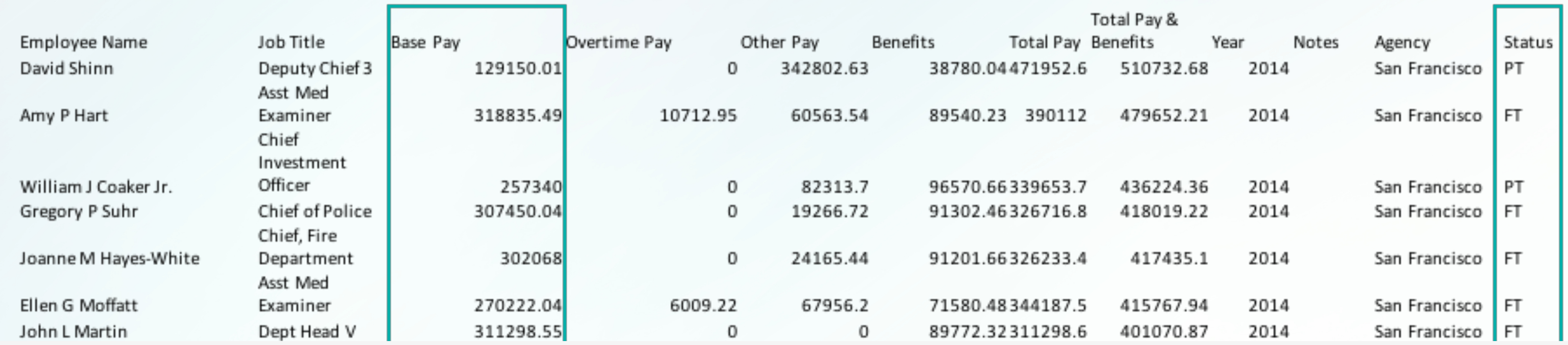
Frictionless integration into enterprise application workflows

- Active monitoring of service components to ensure high availability and failover

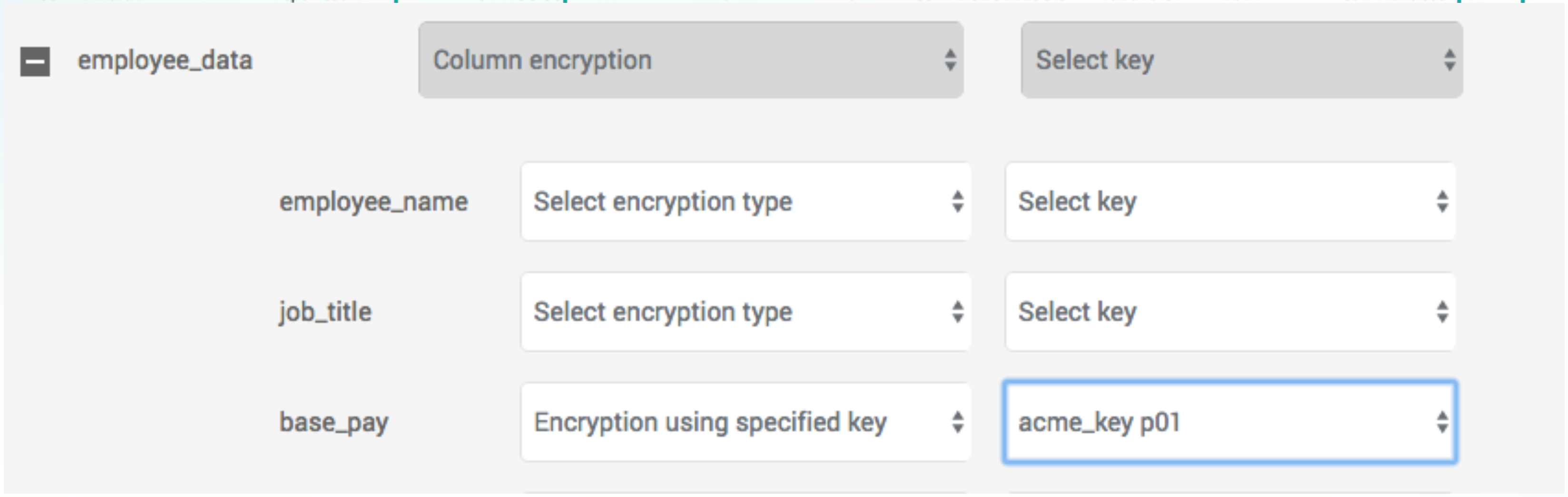

APP
CLEAR CQL DATA REQUEST
KMS
DRIVER LEVEL PROXY
APP VPC
ENCRYPTED CQL DATA REQUEST
CASSANDRA VPC
CASSANDRA CLUSTER PER CUSTOMER
PROTECTING IN-MEMORY & AT-REST CASSANDRA CLUSTER

Selective Privacy Example

baffle

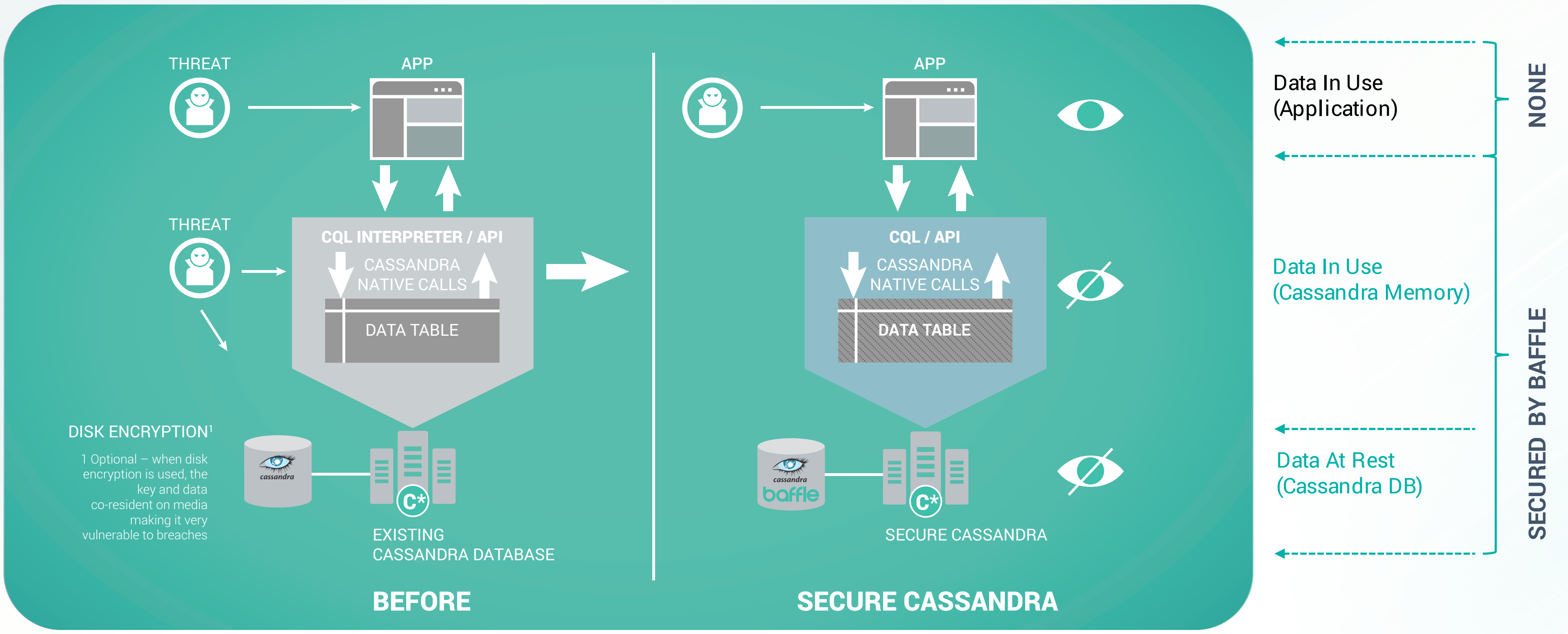


Employee Name	Job Title	Base Pay	Overtime Pay	Other Pay	Benefits	Total Pay	Total Pay & Benefits	Year	Notes	Agency	Status
David Shinn	Deputy Chief 3	129150.01	0	342802.63	38780.04	471952.6	510732.68	2014		San Francisco	PT
Amy P Hart	Asst Med Examiner	318835.49	10712.95	60563.54	89540.23	390112	479652.21	2014		San Francisco	FT
William J Coaker Jr.	Chief Investment Officer	257340	0	82313.7	96570.66	339653.7	436224.36	2014		San Francisco	PT
Gregory P Suhr	Chief of Police	307450.04	0	19266.72	91302.46	326716.8	418019.22	2014		San Francisco	FT
Joanne M Hayes-White	Chief, Fire Department	302068	0	24165.44	91201.66	326233.4	417435.1	2014		San Francisco	FT
Ellen G Moffatt	Asst Med Examiner	270222.04	6009.22	67956.2	71580.48	344187.5	415767.94	2014		San Francisco	FT
John L Martin	Dept Head V	311298.55	0	0	89772.32	311298.6	401070.87	2014		San Francisco	FT



employee_data	Column encryption	Select key
employee_name	Select encryption type	Select key
job_title	Select encryption type	Select key
base_pay	Encryption using specified key	acme_key p01

Secure Cassandra – Before and After


THREAT
APP
THREAT
CQL INTERPRETER / API
CASSANDRA NATIVE CALLS
DATA TABLE
DISK ENCRYPTION[1]
1 Optional – when disk encryption is used, the key and data co-resident on media making it very vulnerable to breaches
cassandra
EXISTING CASSANDRA DATABASE
BEFORE
APP
CQL / API
CASSANDRA NATIVE CALLS
DATA TABLE
cassandra
baffle
SECURE CASSANDRA
SECURE CASSANDRA
Data In Use (Application)
Data In Use (Cassandra Memory)
Data At Rest (Cassandra DB)
NONE
SECURED BY BAFFLE

Driver Modifications

Added couple of pipeline stages for encryption / decryption in the java driver

```
Connection.java

pipeline.addLast("messageDecryptor", new MessageDecryptor(transformDB));

pipeline.addLast("messageEncryptor", new MessageEncryptor.Encryptor(transformDB));
```

Tiers of Security

baffle

PARAMETER	STOCK CASSANDRA	SECURE CASSANDRA
PROTECTION		
Data on disk	**Encrypted**	**Encrypted**
Data in Cassandra server memory	Clear	**Encrypted**
Data in App memory	Clear	Clear
Vulnerability	No data Protection (App, DB Server, Disk)	Data Protection below CQL Interface (DB Server & Disk)
IMPACT		
Driver	Unchanged	Added pipeline stages
DB Binaries	Unchanged	Add support for custom types
OPERATIONS ON ENCRYPTED DATA	None	Search, Sort, Aggregate

Cassandra Modifications

Added custom types for encrypted values with serializers/ deserializers

Added custom functions for operations such as compares and aggregates on the encrypted values

```
Custom Integer Type:
Db/marshal/EncIntType.java
public class EncIntType extends AbstractType<BlindInt>
{
    public static final EncIntType instance = new EncIntType();
    private EncIntType()
    {
        super(ComparisonType.CUSTOM);
    }
…
}
public synchronized int compareCustom(ByteBuffer o1, ByteBuffer o2)
{
```

Conclusion

Cloud hosted databases have *perils* for stored data

Practical approach to protecting Cassandra data

- Driver level modifications to enable encryption
- Support custom types and functions for encrypted operations

The *promise* of keeping data encrypted…always! is attainable

- No threats to availability
- No impact on stability
- Minimal impact on performance

CASSANDRA SUMMIT 2016

Thank You

baffle

Make data breaches irrelevant

www.baffle.io