Secure SQL Processing

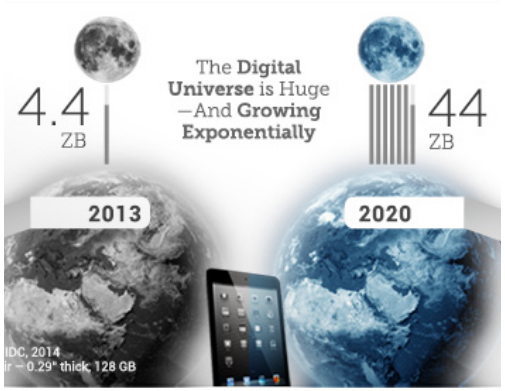
Rui Oliveira
University of Minho & INESC TEC

Horizon 2020 DS-2014-1

73rd IFIP Working Group 10.4 on Dependable Computing and Fault Tolerance - 11-15 January 2018 - Goa, India

Context

- Data is being produced, exchanged, stored and processed in volume never seen before.
- It is driving businesses (BI, AI, Machine learning)
- To cope with this scale companies rely on third-party infrastructures - Cloud Computing



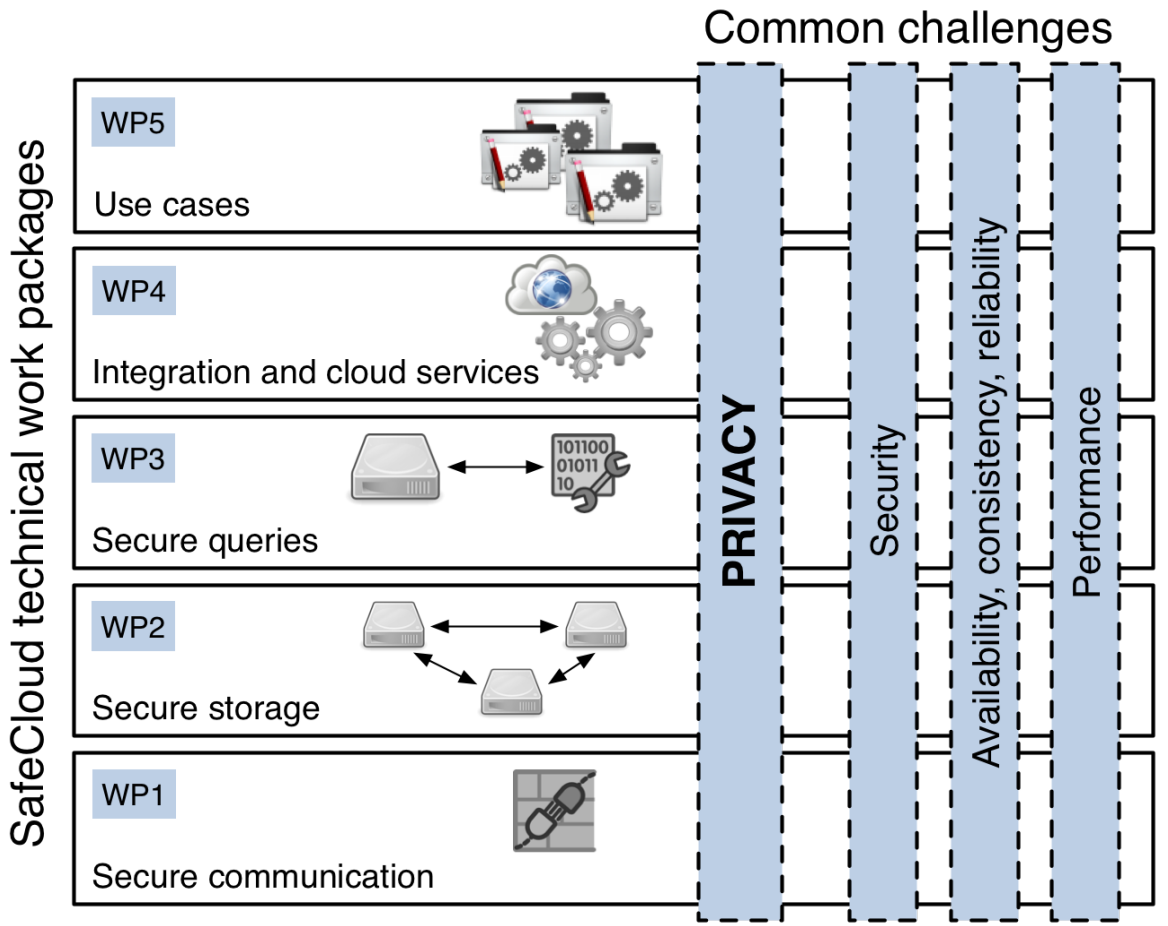
- Serious privacy, integrity and security issues
- Unauthorised government or third-party access to private and sensitive data
- Data is no longer fully controlled by its rightful owner

Important requirements

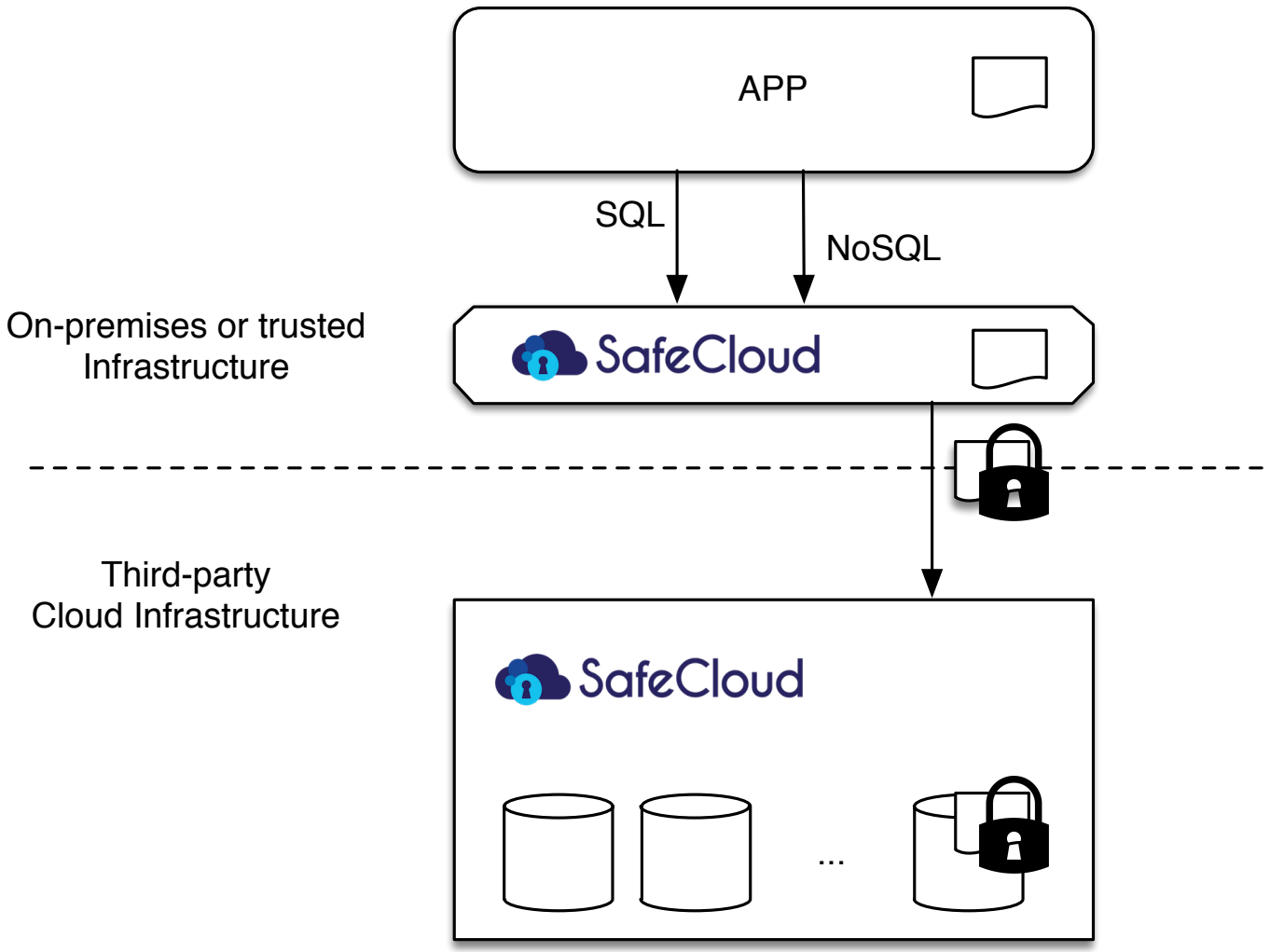
- **Minimally disruptive**. Databases are around for a long time. Any DB technology must provide a backwords-compatible SQL interface.
- **Transparent privacy for the application.** Requiring applications to be rewritten is unrealistic.
- **Flexible**. The system must be compliant with many privacy-preserving mechanisms as there isn’t a “one size fits all” solution for data privacy.

SafeCloud Project

SafeCloud aims at re-architect cloud infrastructures to ensure that data transmission, storage, and processing can be done in a privacy-by-design fashion.



Secure data processing in SafeCloud

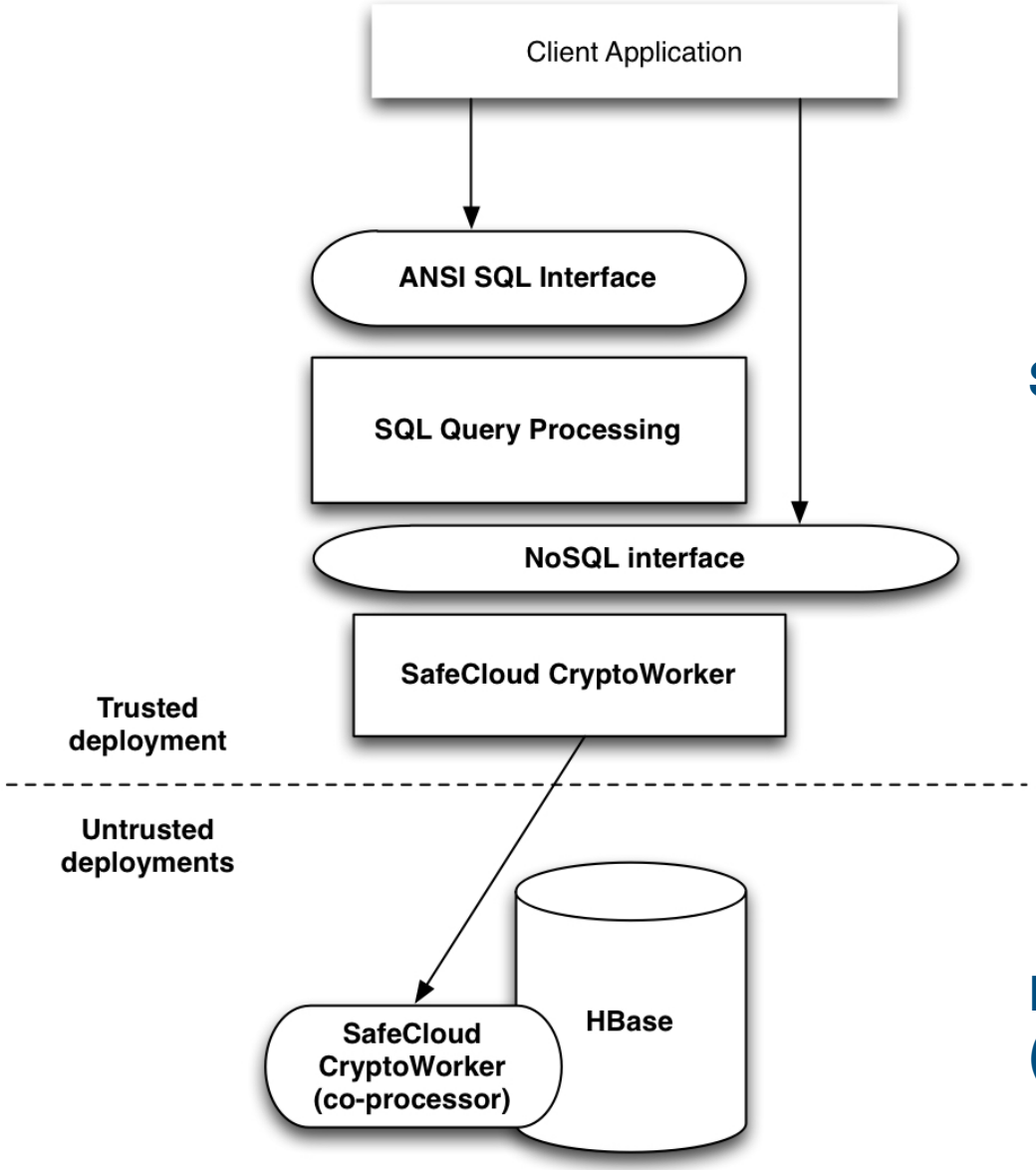


Existing privacy-preserving techniques

Scheme	Construction	Properties	Operations
Standard Encryption	AES-128 CBC w/o IV	None	Insertions
Deterministic Encryption	AES-128 CBC w/ IV	Equality	Reads, equality filters
Order-Preserving Encryption	Boldyreva et al. ‘09	Equality, Order	Searches, equality and order filters

Enough for NoSQL data processing !

Zooming in



SQL to NoSQL translation

Data protection layer (encryption)

Data processing layer (over plain and encrypted data)

SafeCloud DB is

- Compliance: SQL and NoSQL compatible. Applications do not need rewriting.
- Elastic and fast: Designed for large scale data. SafeCloud Database can scale and grow on demand.
- Secure: offers privacy-by-design and data stays private while in transit, at rest and during processing.

Ongoing research

- How fast and scalable can the database be? We know it can be fast when OPE is minimally used. Can fast OPE mechanisms be devised?

- Hybrid deployment ensures privacy for the backend data. Data handled by the business logic of the application is still plain text. How to migrate entire applications to the Cloud?

- Intel SGX. The SafeCloud architecture is compliant with a SGX-based version of the privacy-preserving components. Will this represent a performance improvement?

Ongoing fund raising…

- Processing over encrypted data allows precise computation but can have high overheads and, in particular cases, data isn’t allowed to leave even if encrypted
- Aproximate, privacy preserving, processing is a complementary approach we are willing to explore and integrate into the DBMS:
  - Differential privacy techniques to handle complete but partially anonymized/blurred/obfuscated data
  - Processing "systematically incomplete data" and "selective learning" techniques to handle incomplete data