# Fujitsu World Tour 2018

May 30, 2018

FUJITSU

shaping tomorrow with you

#FujitsuWorldTour

Human Centric Innovation

Co-creation for Success

# Security and Privacy of Big Data
# A NIST Perspective

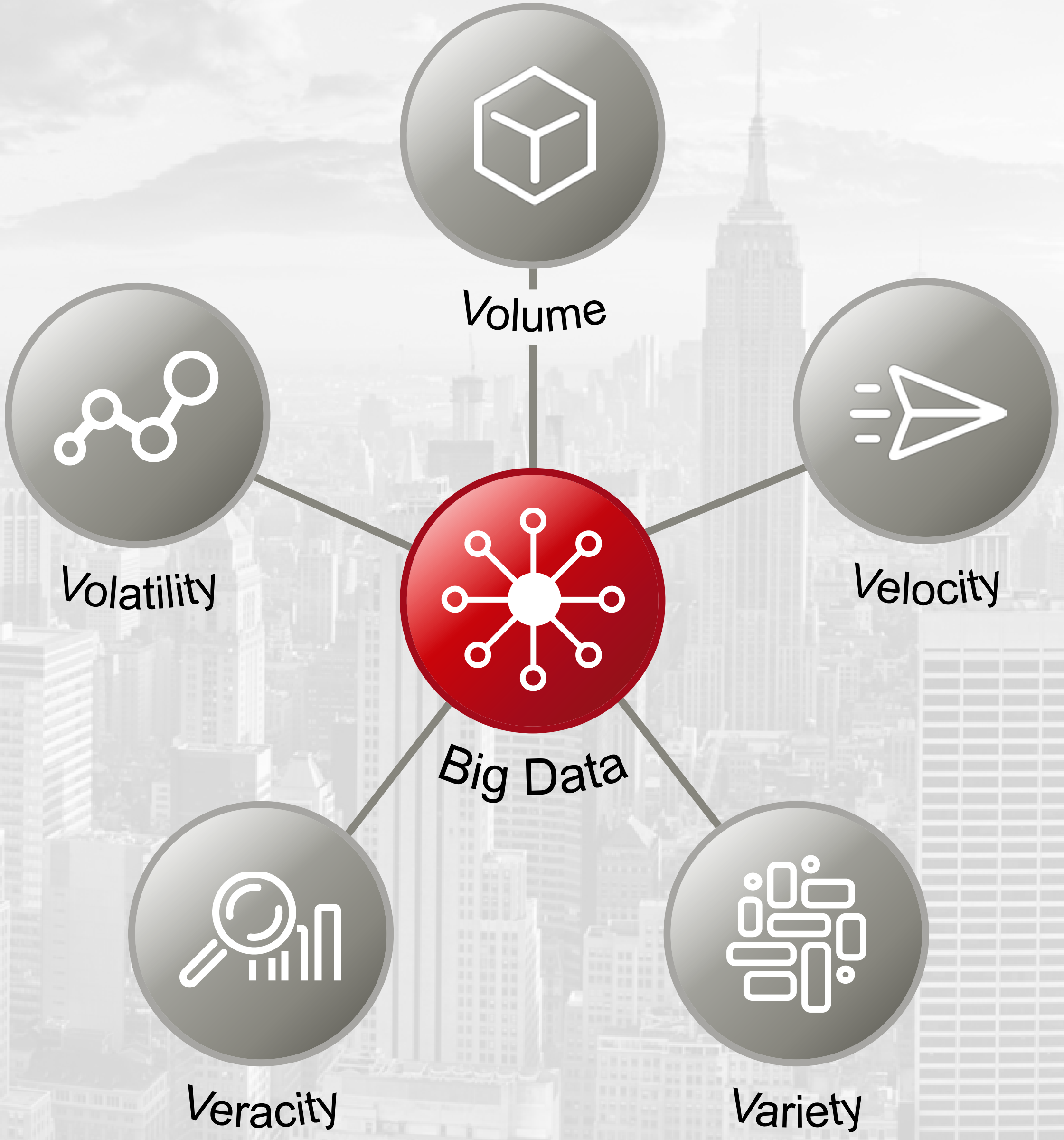
Arnab Roy
Fujitsu Laboratories of America
Co-Chair, NIST Big Data WG: Security and Privacy SG

FUJITSU
shaping tomorrow with you

# What is Big Data?

Big Data consists of extensive datasets - primarily in the characteristics of volume, variety, velocity, and/or variability - that require a scalable architecture for efficient storage, manipulation, and analysis.

[NIST SP1500-1]



# Why are Security and Privacy important for Big Data?

- Volume of data is growing exponentially
  - 90% of the data in the world today was created in the last two years (Source: http://www-01.ibm.com/software/data/bigdata/)
  - Global big data market revenues were forecasted to reach $12.4 Billion in 2014 growing to $23.8 Billion in 2016, according to the firm Visiongain and IDC
- Data breach is costly
  - Average cost of breach for single record is $200
  - With 20% probability 10,000 records get breached (in 2 year time frame) in any organization
  - With 20% probability organization will lose $2M in two years!!!
- Data breach damages company reputation

2014 Cost of Data Breach Study: Global Analysis

Benchmark research sponsored by IBM
Independently conducted by Ponemon Institute LLC
May 2014

Figure 4. Per capita cost by industry classification
Consolidated view (n=314)

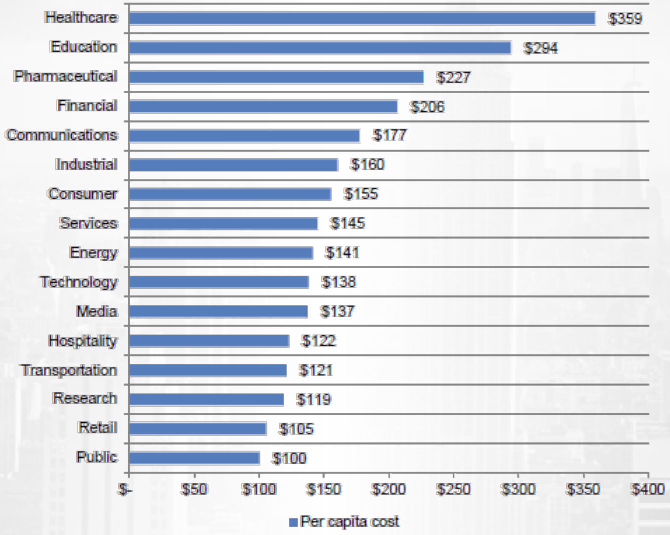
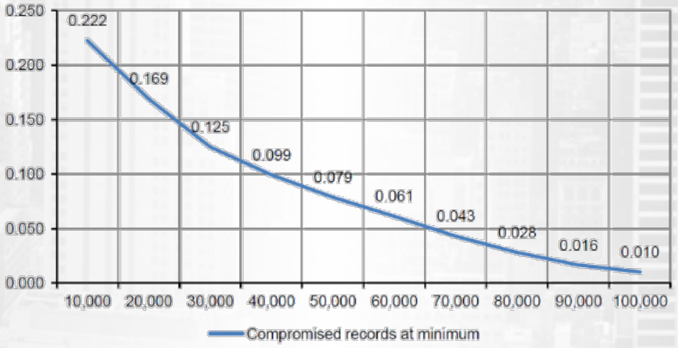



Figure 19. Probability of a data breach involving a minimum of 10,000 to 100,000 records



# NIST Big Data Public Working Group (NBD-PWG)

## Goal

Develop a secured reference architecture that is vendor-neutral, technology- and infrastructure-agnostic to enable any stakeholders (data scientists, researchers, etc.) to perform analytics processing for their given data sources without worrying about the underlying computing environment.

### 5 Subgroups (July 2013 – now)

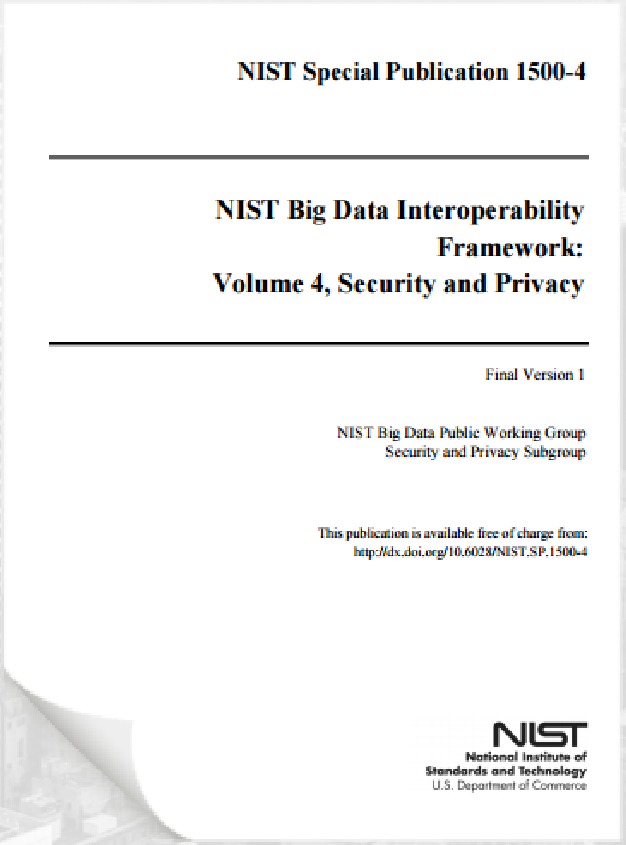
1. Definitions & Taxonomies
2. UC & Requirements
3. Security & Privacy
4. Reference Architecture
5. Standards Roadmap

### Deliverables

1. Big Data Definitions
2. Big Data Taxonomies
3. Big Data Requirements & Use Cases
4. Big Data Security & Privacy
5. Big Data Architectures White Paper Survey
6. Big Data Reference Architecture
7. Big Data Standards Roadmap

# Version 1 Released

## V1 (high-level Reference Architecture components and descriptions) for Big Data Interoperability Framework


NIST Special Publication 1500-4

NIST Big Data Interoperability Framework:
Volume 4, Security and Privacy

Final Version 1

NIST Big Data Public Working Group
Security and Privacy Subgroup

This publication is available free of charge from:
http://dx.doi.org/10.6028/NIST.SP.1500-4

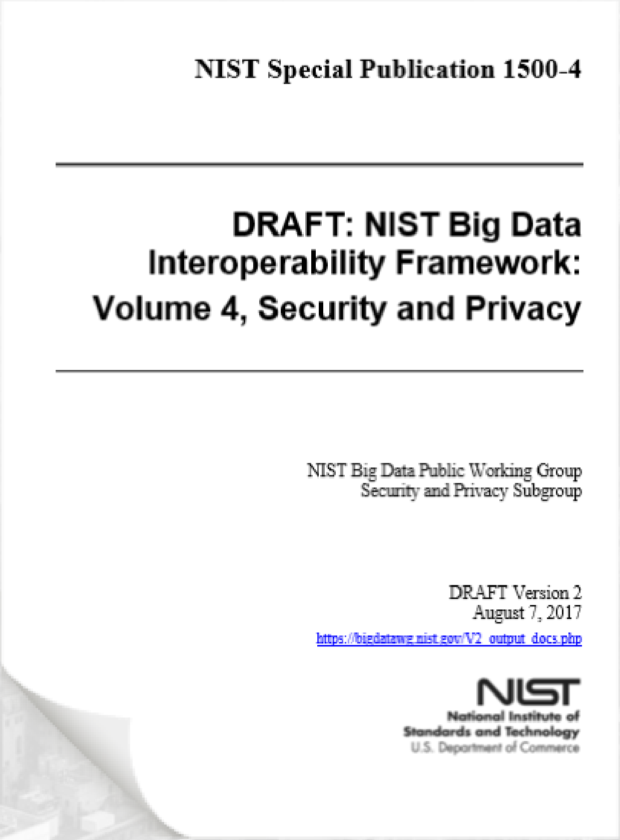
NIST
National Institute of Standards and Technology
U.S. Department of Commerce

Released on September 16, 2015

http://bigdatawg.nist.gov

- NIST SP1500-1 Definitions
- NIST SP1500-2 Taxonomies
- NIST SP1500-3 Use Cases & Requirements
- NIST SP1500-4 Security & Privacy
- NIST SP1500-5 Architecture Survey – White Paper
- NIST SP1500-6 Reference Architecture
- NIST SP1500-7 Standards Roadmap

# Version 2 draft is in NIST review phase


NIST Special Publication 1500-4

**DRAFT: NIST Big Data Interoperability Framework: Volume 4, Security and Privacy**

NIST Big Data Public Working Group
Security and Privacy Subgroup

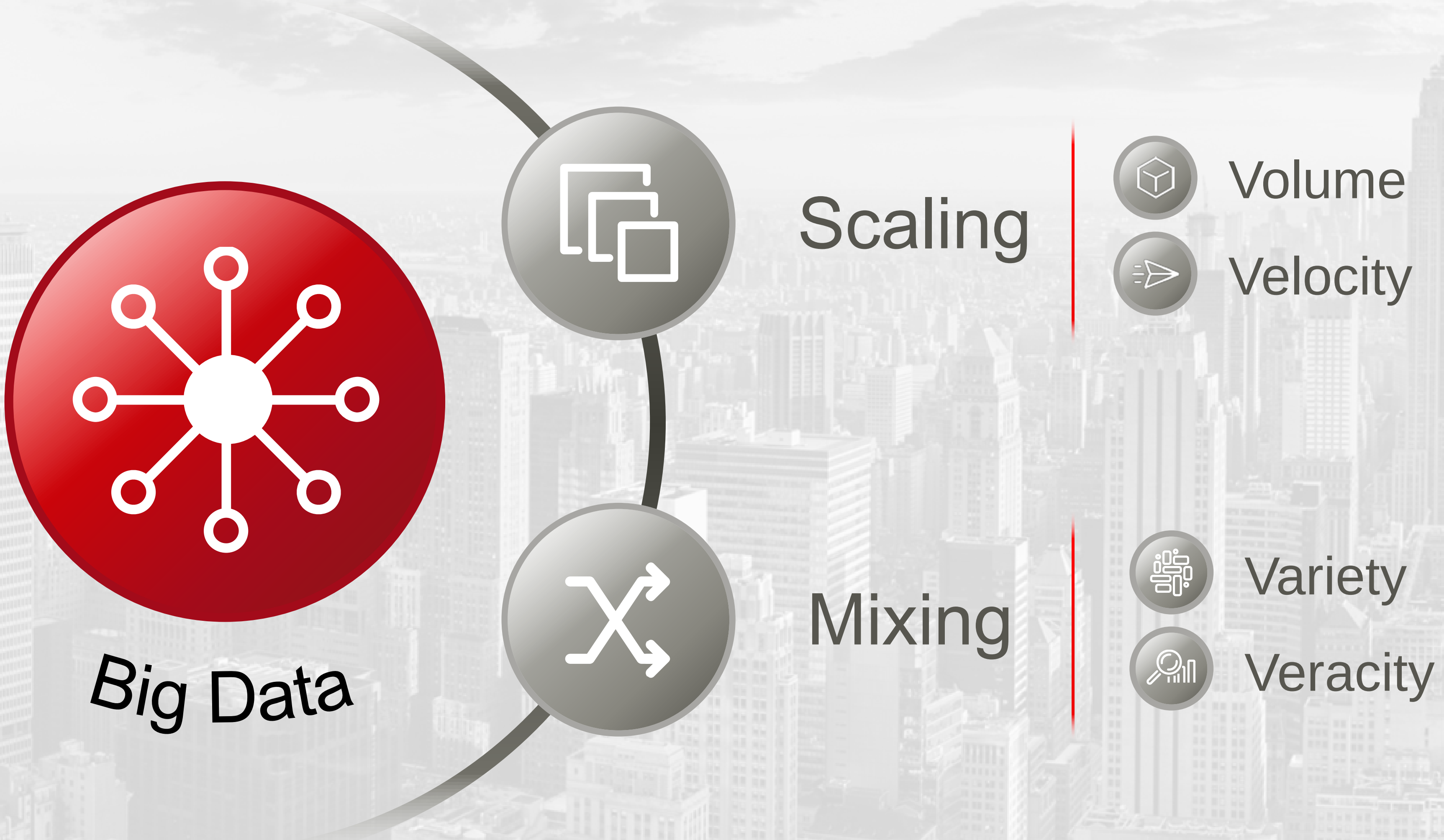
DRAFT Version 2
August 7, 2017
https://bigdatawg.nist.gov/V2_output_docs.php

NIST
National Institute of Standards and Technology
U.S. Department of Commerce

Public comments received on 21 September 2017

https://bigdatawg.nist.gov/home.php

# A 10,000-feet view



# Emergent S&P Considerations

FUJITSU

## (Big) Scaling

Retarget to Big Data infrastructural shift

- Distributed computing platforms like Hadoop
- Non-relational data stores

## (Data) Mixing

Control visibility while enabling utility

- Balancing privacy and utility
- Enabling analytics and governance on encrypted data
- Reconciling authentication and anonymity

# S&P Requirements Emerging due to Big Data Characteristics

## Variety

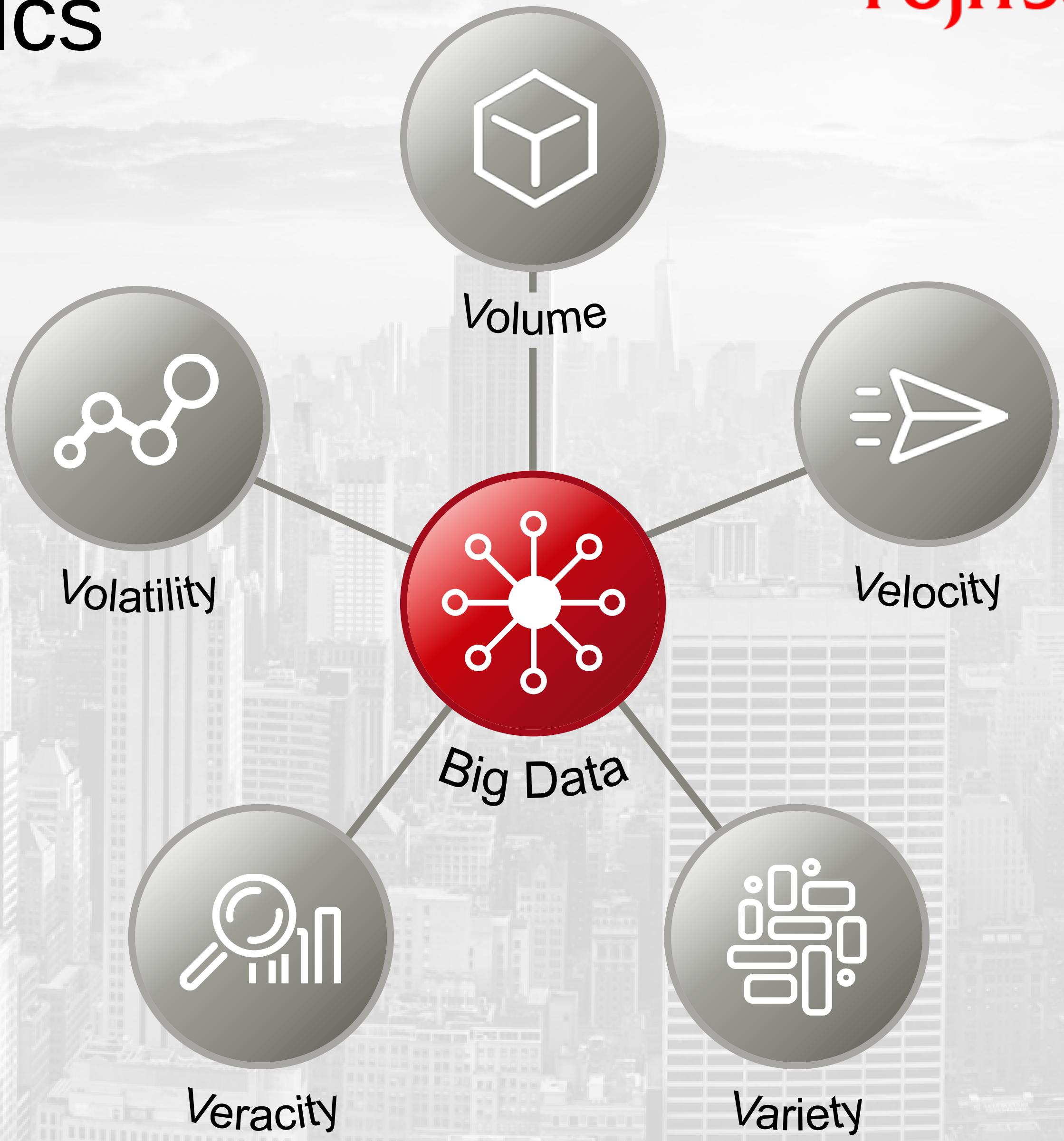
Traditional encryption schemes hinder organization of data based on semantics.

## Volume

Threat models for multi-tiered data storages are complex and evolving.

## Velocity

Distributed computing infrastructures and non-relational data storages require retargeting of traditional security mechanisms.



# S&P Requirements Emerging due to Big Data Characteristics

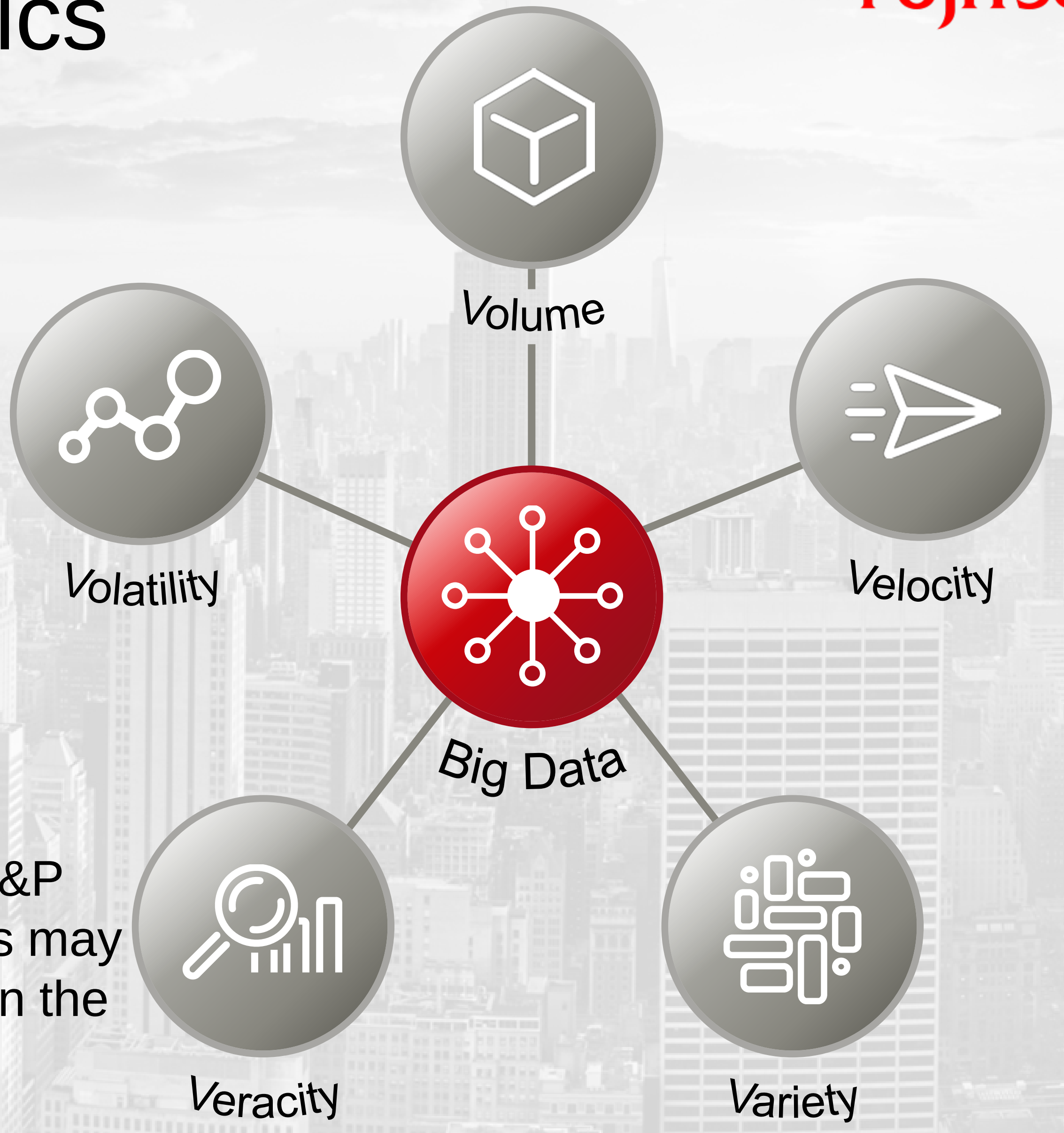
FUJITSU

## Veracity

Keeping track and ensuring integrity of the ownership, source and other metadata of individual data is a complex and sophisticated requirement, given the movement of data between nodes, entities and geographical boundaries.

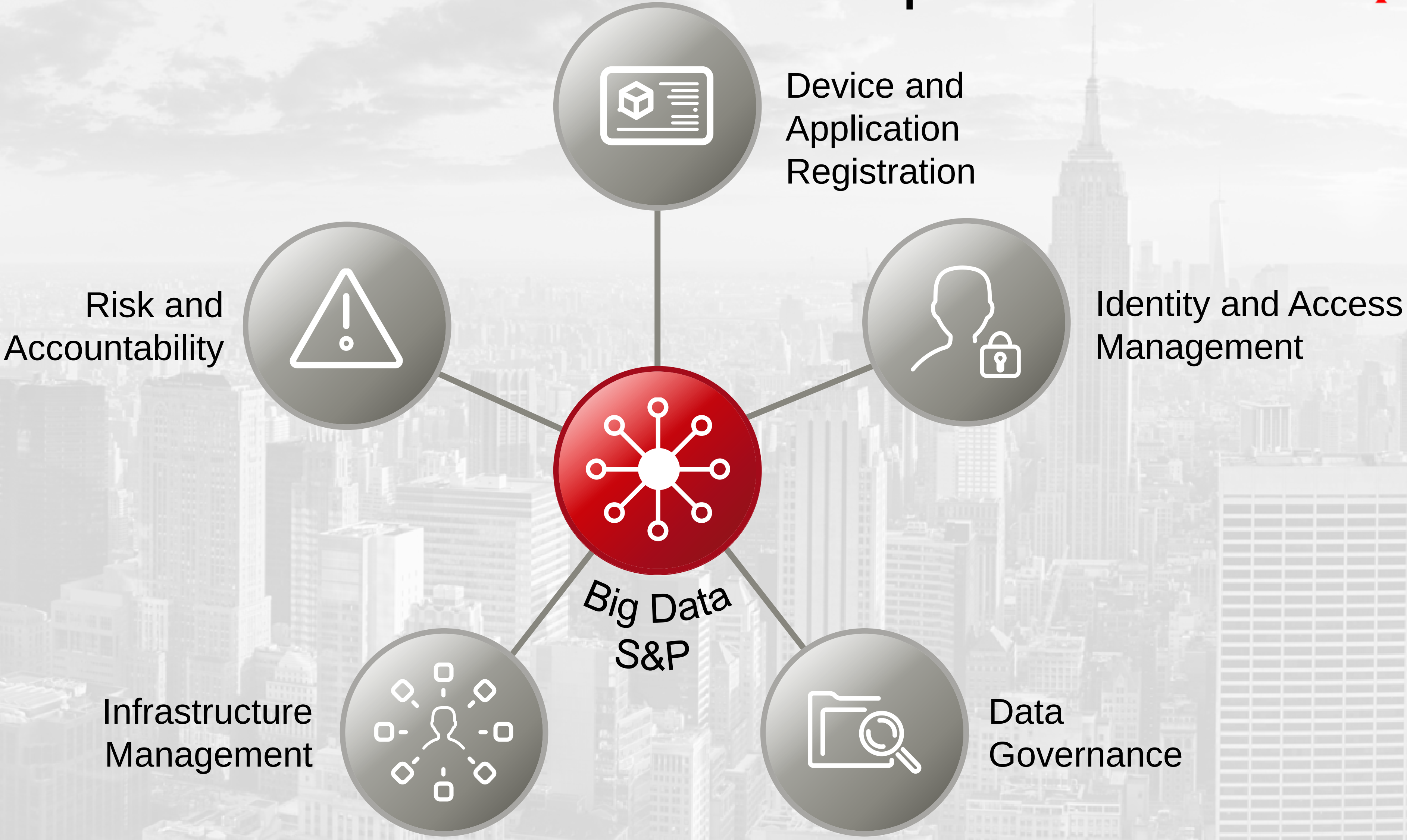
## Volatility

Indefinitely persistent data requires evolving S&P considerations. With the passage of time, roles may evolve and governance may shift depending on the merger and disappearance of responsible organizations.

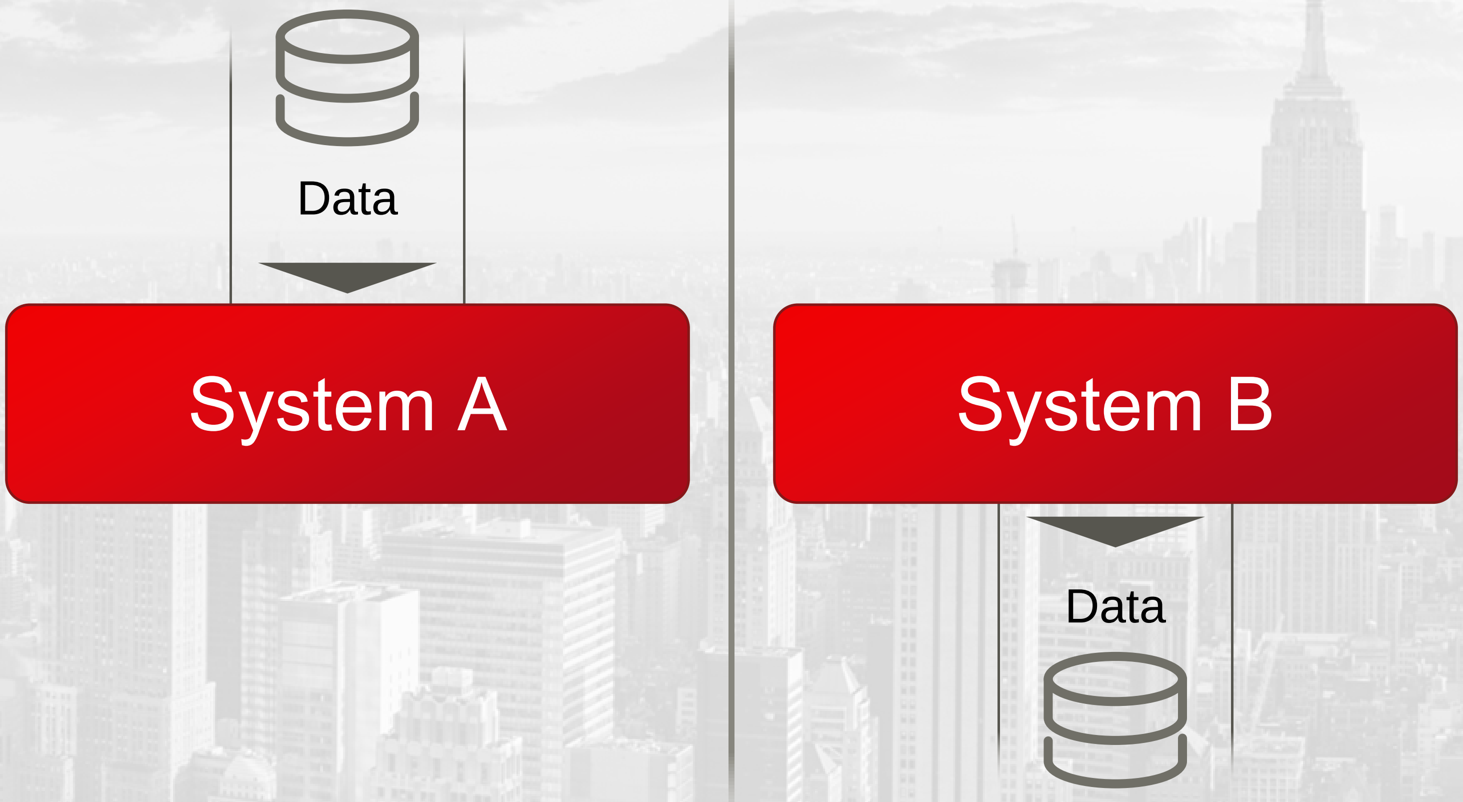


# Conceptual Classification of S&P Topics

# Operational Classification of S&P Topics

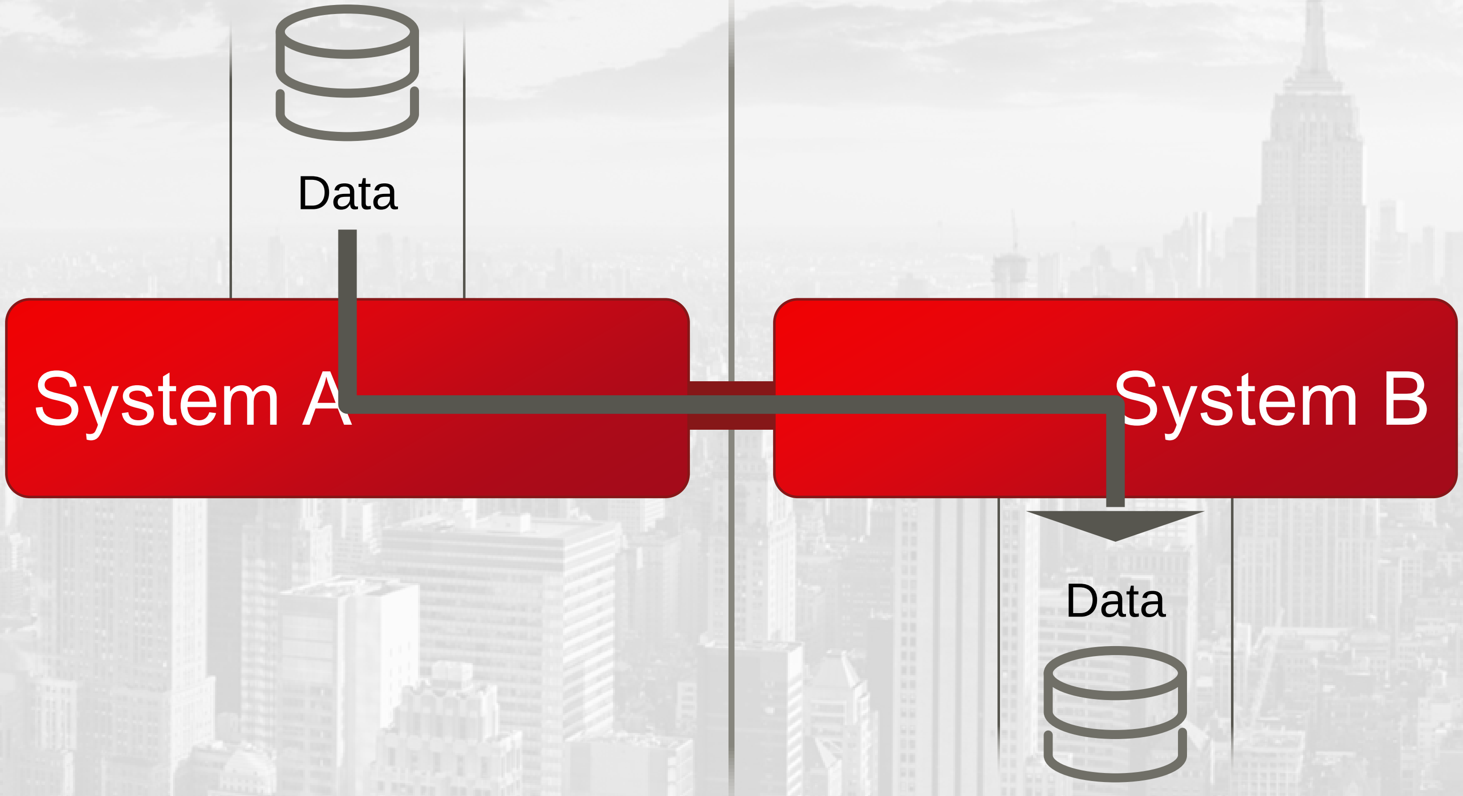


# S&P doesn't compose!



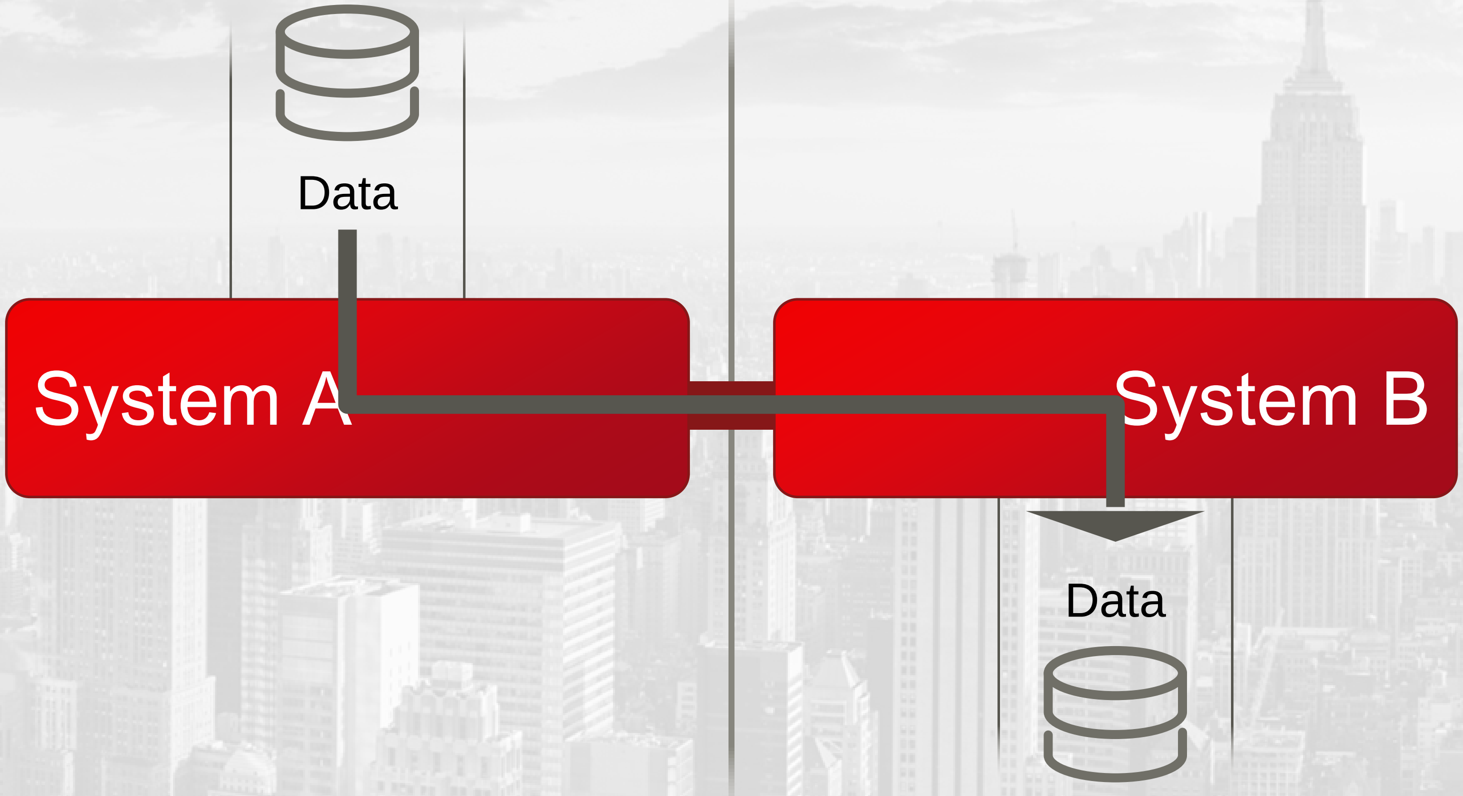

System A and System B have known data flow restrictions

# S&P doesn't compose!



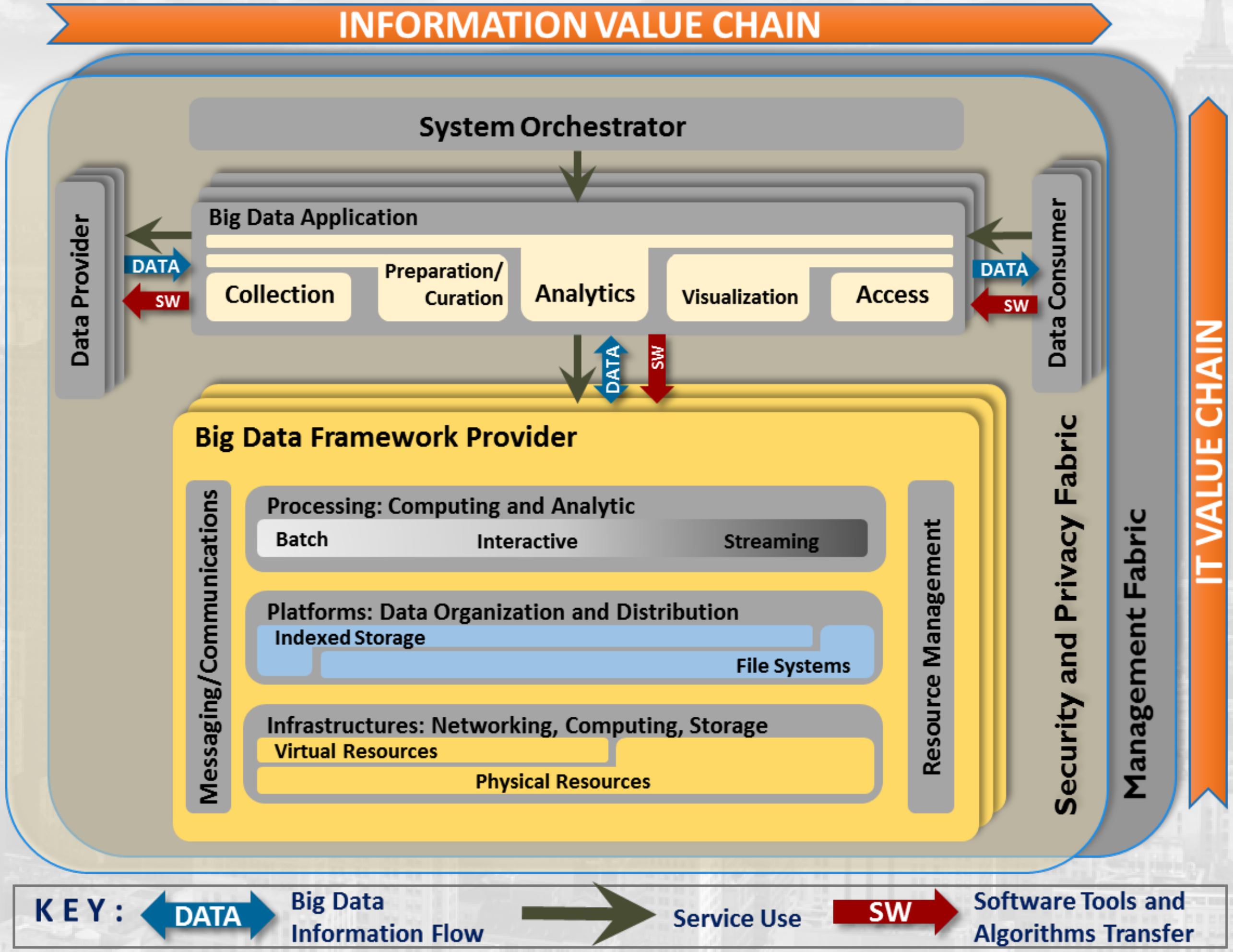

System A and System B have known data flow restrictions.
Combined system can have unexpected data flows!

# S&P doesn't compose!

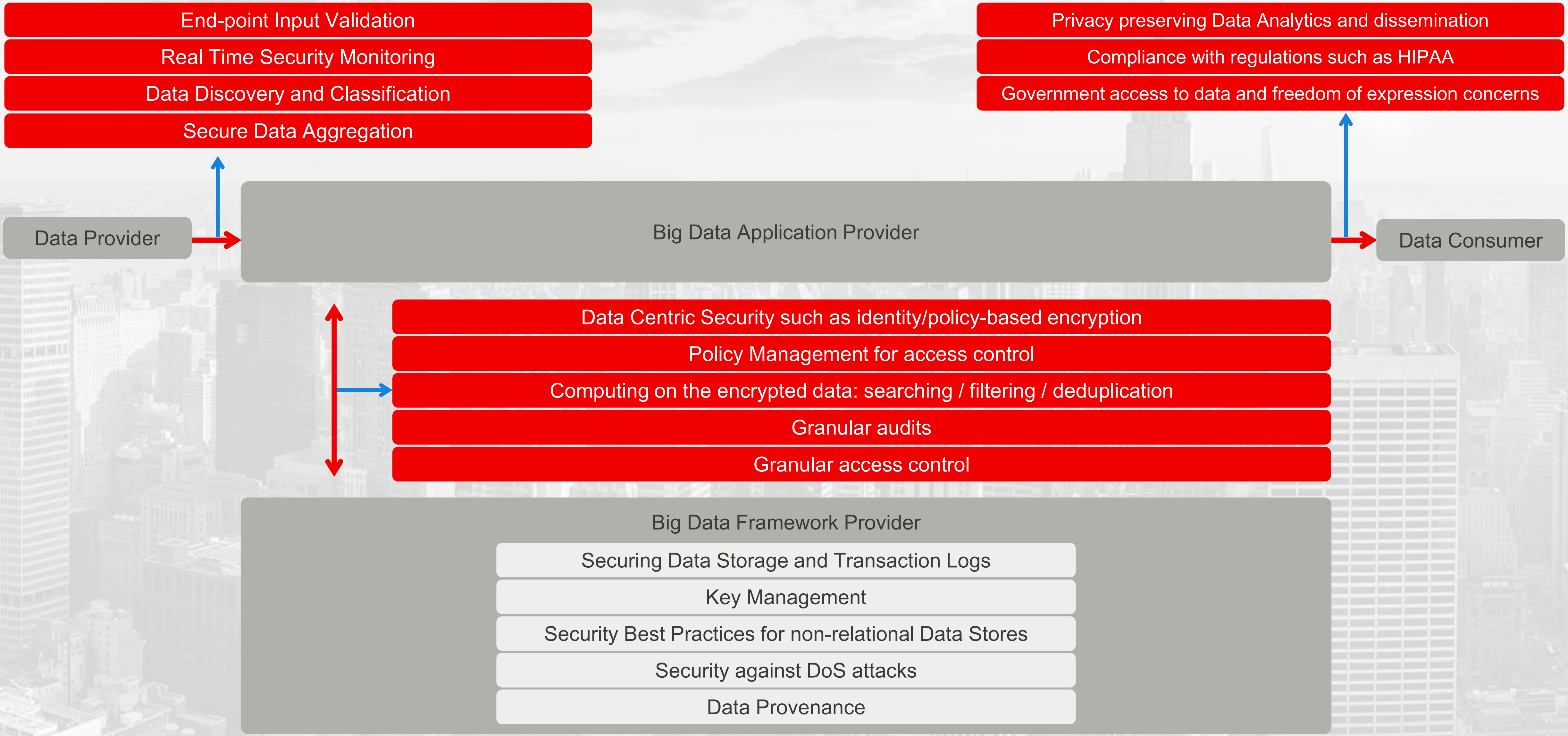



There is a need for Architectural Thinking

# NIST Big Data Reference Architecture



# Big Data Security Reference Architecture



# Use Cases

FUJITSU

## Retail/Marketing
- Consumer Digital Media Usage
- Nielsen Homescan: Family level Retail Transactions
- Web Traffic Analysis

## Healthcare
- Health Information Exchange
- Genetic Privacy
- Pharma Clinical Trial Data Sharing

## Cyber-security
- Network Protection

## Government
- Military
- Education

## Industrial
- Aviation: Sensor Data Storage and Analytics
- Transportation: Cargo Shipping

# Emerging Cryptographic Technologies

FUJITSU

| Technology | Data Provider | Application Provider | Feature | Visibility |
|---|---|---|---|---|
| Homomorphic Encryption | Encrypts data | Stores encrypted data | Capability to perform computations | Only at Data Provider |
| Functional Encryption | Encrypts data | Stores encrypted data | Capability to perform computations | Result of allowed computations visible at Application Provider |
| Access Control Policy-Based Encryption | Encrypts data | Stores encrypted data | No capability to perform computations | Only for entities which have a secret key satisfying the access control policy |
| Secure Multi-Party Computation | Plaintext data | Stores plaintext data | Collaborative computation among multiple Application Providers | Application Providers do not learn others' inputs. They only learn the jointly computed function. |
| Blockchain | Plaintext or encrypted data | Decentralized | Immutable decentralized database | Transaction logging in a decentralized, untrusted environment |
| Hardware primitives for secure computations | Encrypts data | Stores encrypted data | Capability to perform computations. Verified execution. | Controllable visibility at Application Provider. |

# Secure Outsourcing of Computation

Suppose you want to send all your sensitive data to the cloud: photos, medical records, financial records, etc.

You could send everything encrypted

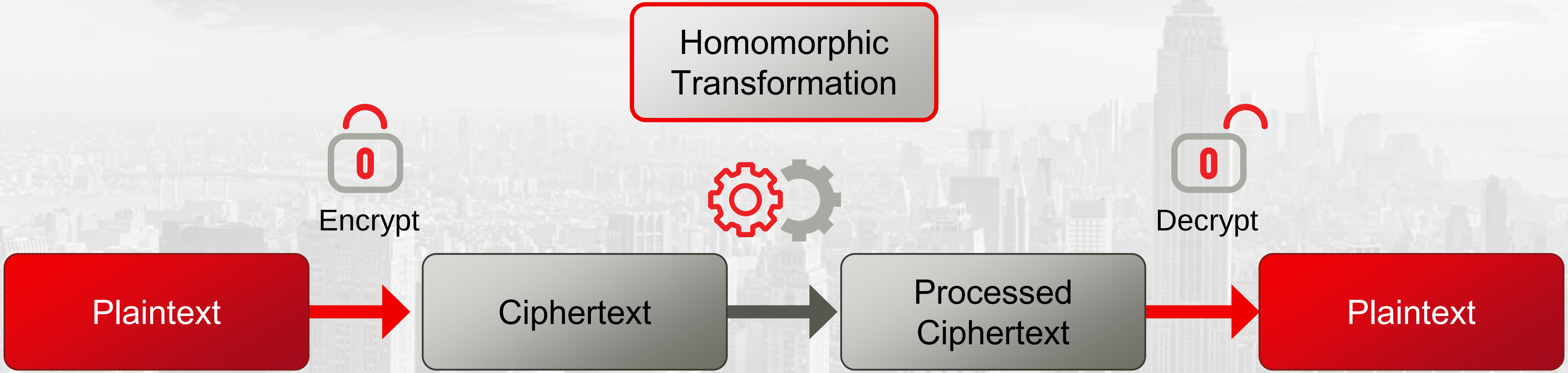
- But wouldn't be much use if you wanted the cloud to perform some computations on them
- What if you wanted to see how much you spent on movies last month?

Solution: Fully Homomorphic Encryption

- Cloud can perform any computation on the underlying plaintext, all the while the results are encrypted!
- Cloud has no clue about the plaintext or the results

# Fully Homomorphic Encryption (FHE)

FUJITSU




- With FHE, computation on plaintext can be transformed into computation on Ciphertext
- As a use case, a cloud can keep and process customer's data without ever knowing the contents
  - Only customer can decrypt the processed data
  - End to end security of customer data

# Access Control Policy-based Encryption

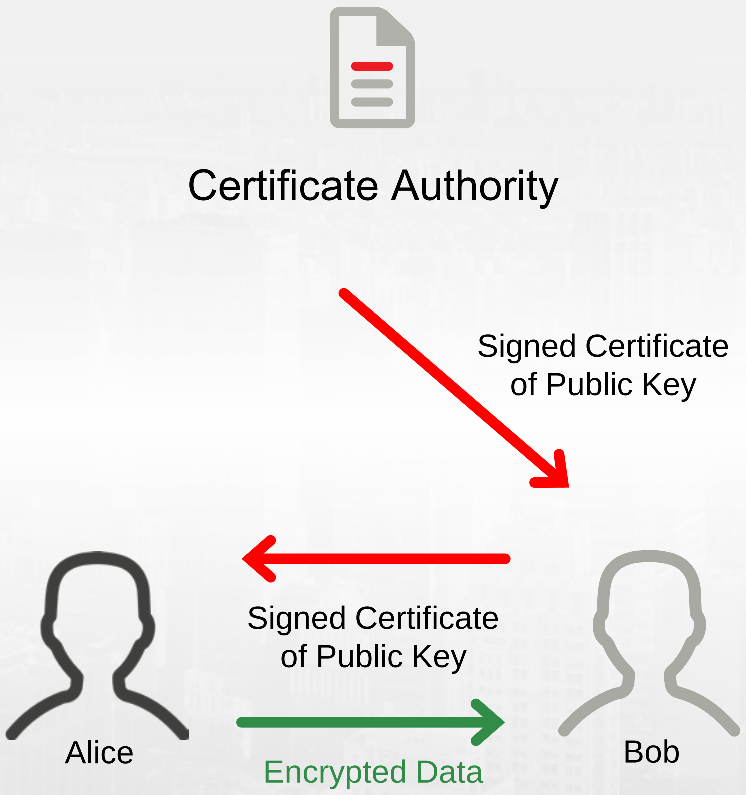
FUJITSU

- Traditionally access control has been enforced by systems – Operating Systems, Virtual Machines
  - Restrict access to data, based on access policy
  - Data is still in plaintext
  - Systems can be hacked!
  - Security of the same data in transit is ad-hoc

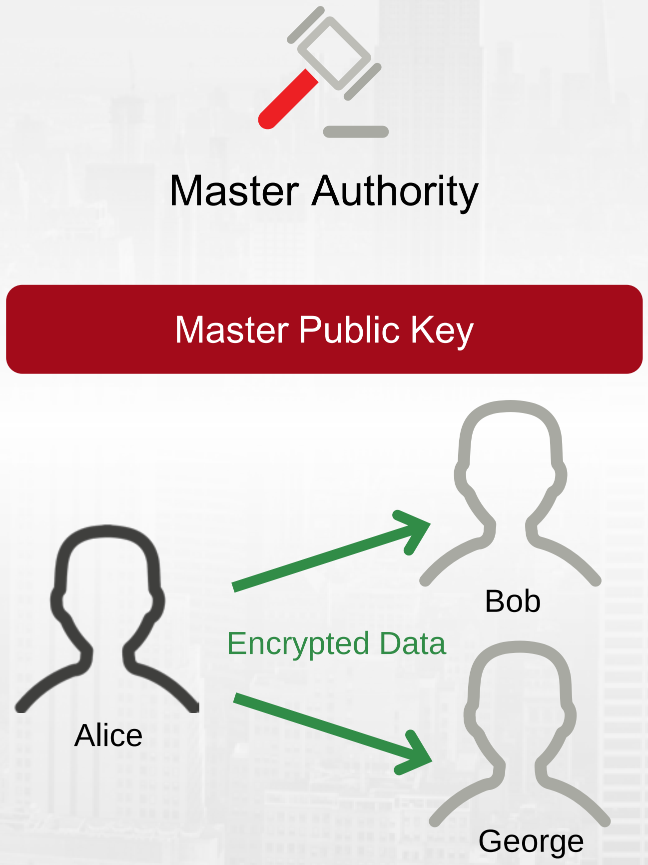
- What if we protect the data itself in a cryptographic shell depending on the access policy?
  - Decryption only possible by entities allowed by the policy
  - Keys can be hacked! – but much smaller attack surface
  - Encrypted data can be moved around, as well as kept at rest – uniform handling

# Identity-based Encryption

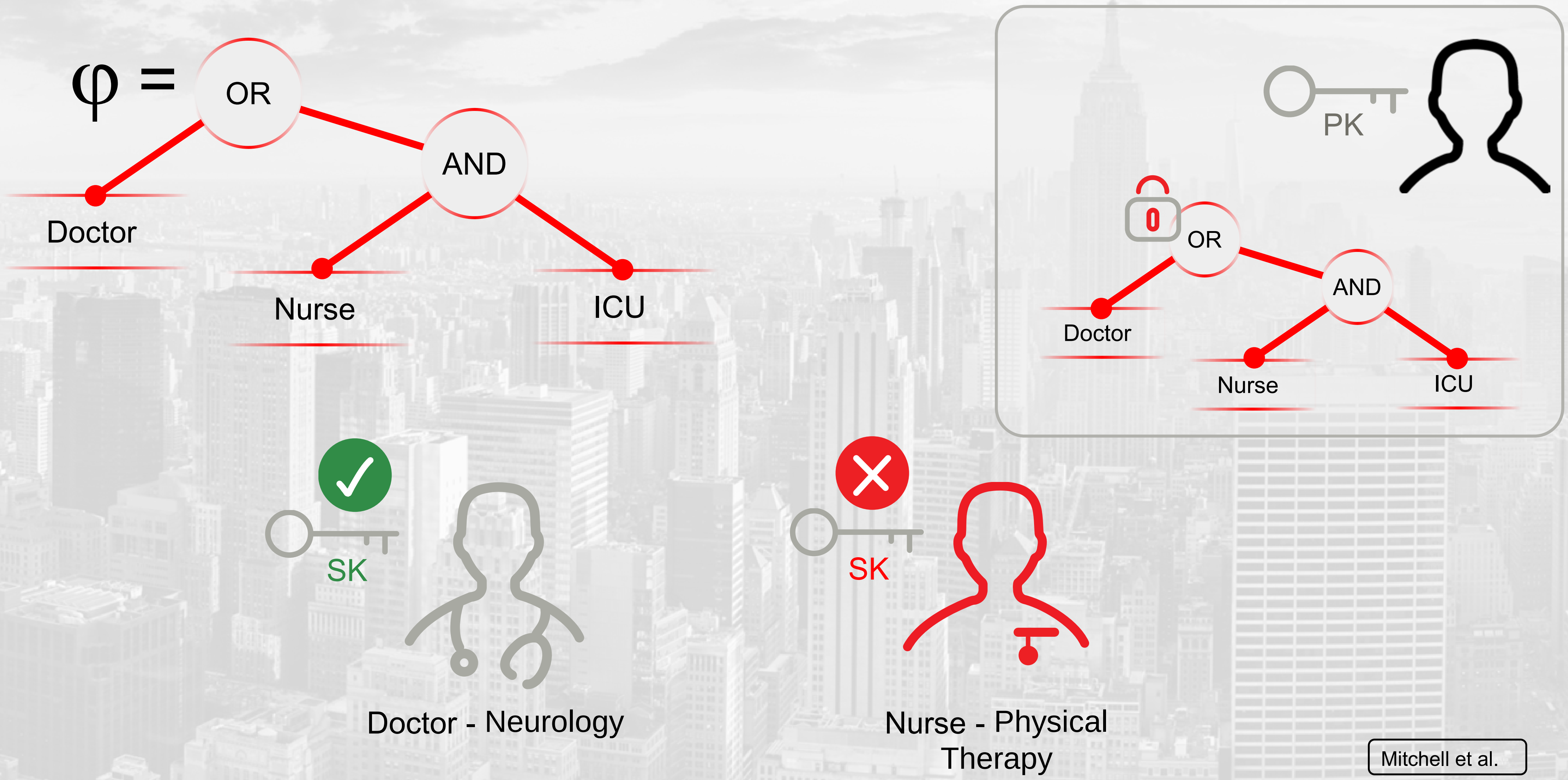
## Public-Key Encryption




## ID-based Encryption



# Policy-Based Encryption




Mitchell et al.

# Blockchain Technology in Practice Today

FUJITSU

## Cryptocurrency

bitcoin

litecoin

ripple

Asset and ownership management

Transaction logging for audit and transparency

Bidding for auctions and contract enforcement

## Smart Contracts

ETHEREUM

Codius

# Recommendations

- Many technologies address S&P requirements of Big Data projects
- Which technology to use involves a lot of risk/benefit analysis
- Consider sensitivity of the data, cost of breach and cost of securing systems when doing this analysis
- For example, for the task of running software on encrypted data at rest, there are at least three possibilities:

| Decrypt the data in the cloud and run software | Run software on the data decrypted inside an HSM | Run software on encrypted data using Fully Homomorphic Encryption |
|---|---|---|
| Pro: fast execution | Pro: less fast, but still practical | Pro: cryptographically secure, no side channel attacks |
| Con: if decryption key is leaked, all the data is exposed | Con: side channel attacks | Con: very slow at this point, except limited operations |

# Take Away Points

- Think S&P at the time of architecting the overall system
  - Not as an afterthought
- In S&P, 1+1 = 0, NOT 2, definitely NOT 4!
  - Does not ~~compute~~ compose
  - Re-analyze S&P when adding new features or joining systems
- Cryptography magic is on the way
  - Stay tuned and patient
- Read NIST Big Data Interoperability Framework SP1500 documents
  - Especially Volume 4: Security and Privacy

# Thank you!  Any Questions?

FUJITSU

Arnab Roy
Fujitsu Laboratories of America

aroy@us.fujitsu.com

shaping tomorrow with you