Thinking BIG

Big Data Concepts and Patterns

Dr. Lilia Sfaxi

Data Science Wokshop, ENIT

September, 13th 2017


DATA

DATA

Big Data

Open Data

Data Mining

Data Viz

Data Analytics

Data Science

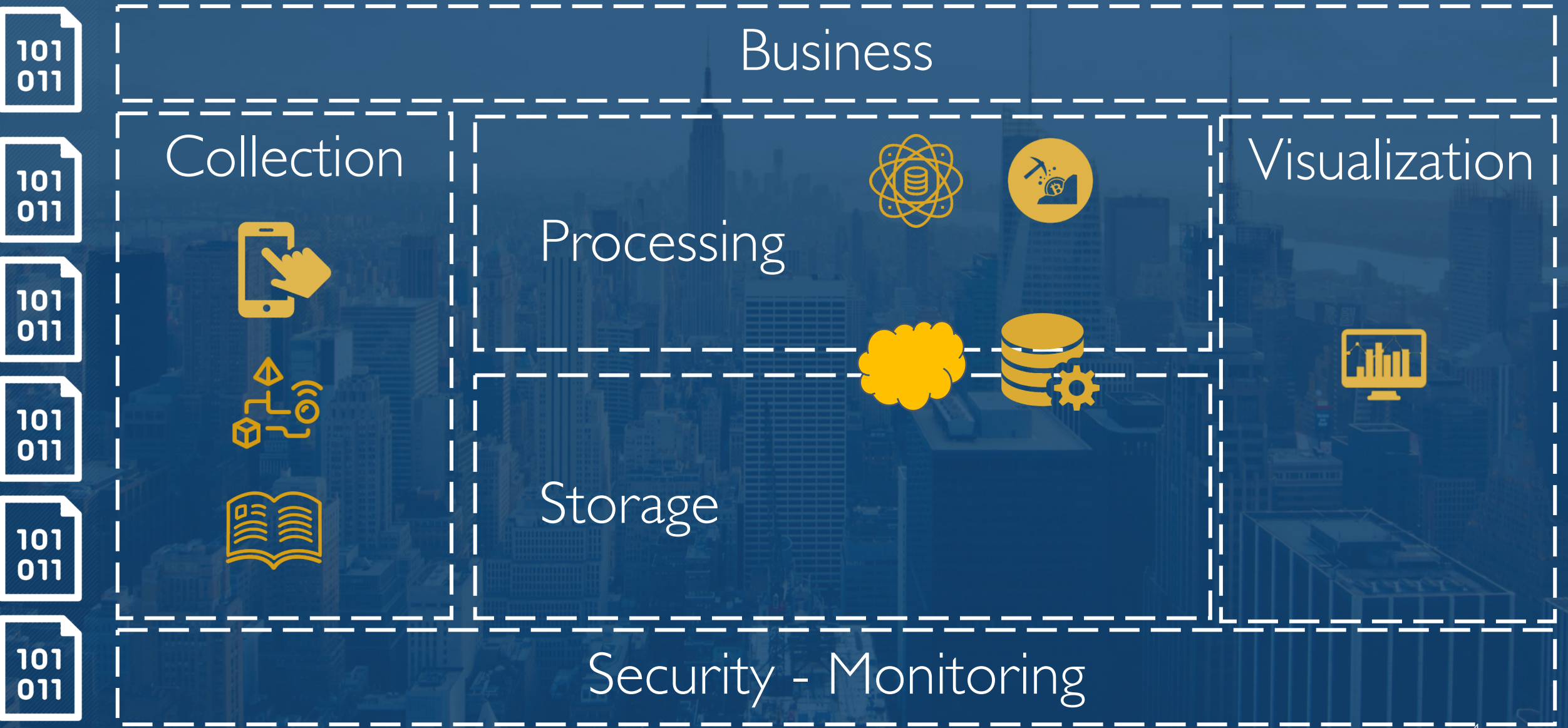
Cloud

Mobile

IoT

BI

Data Issues


101
011
Business
Collection
Processing
Visualization
Storage
Security - Monitoring

Data Issues

Available Data Value Extraction On-Time

Data Issues

Infrastructure

Data Issues

Infrastructure

Scalable

Volume

Available

Velocity

Flexible

Variety

Data Issues

3V

Big Data Principles

Store THEN Process

Data-driven Decisions (NOT Decision-Driven Data)

Redunduncy is GOOD

There is NO NEEDLESS Data

GO Polyglot!

Scalability

Scale UP vs Scale OUT

Classical : Scale UP

Scalability

Scale UP vs Scale OUT

Big Data : Scale OUT

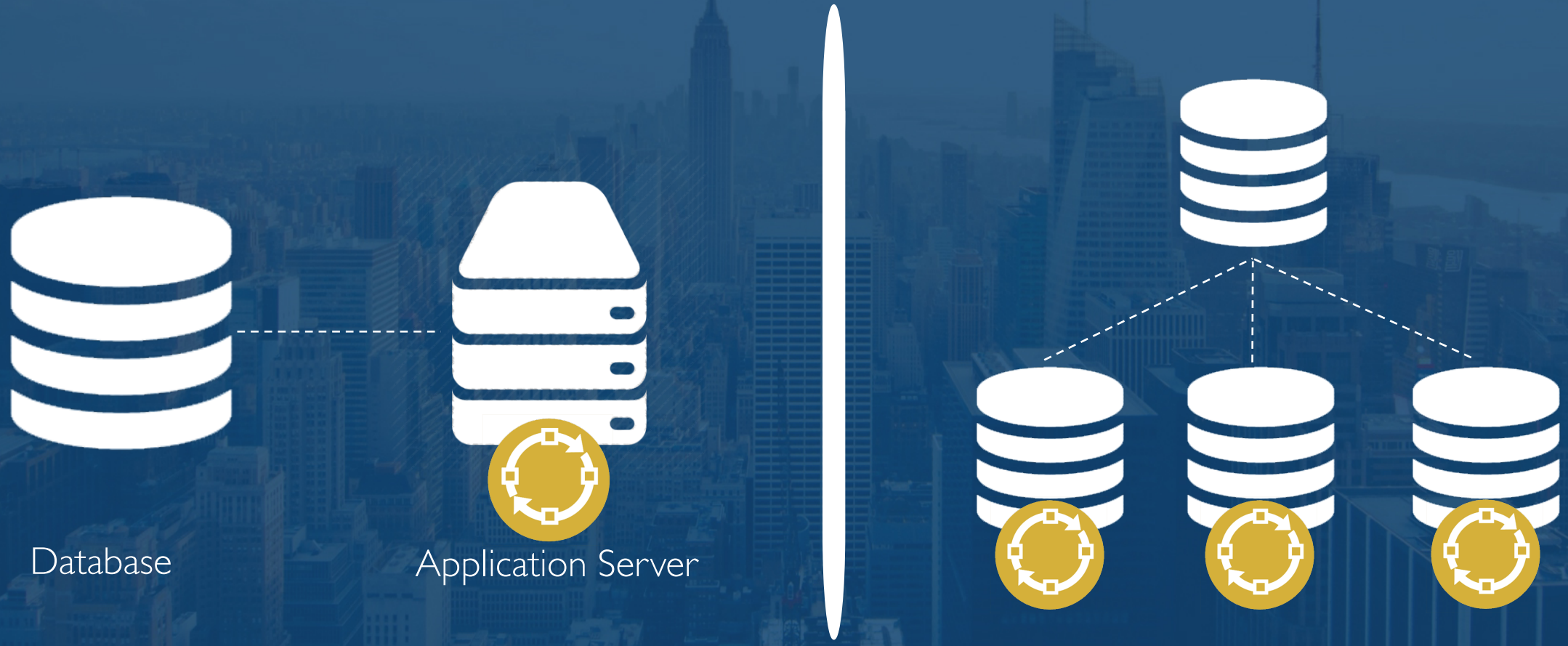
Scalability
Architectures

Ring

Master - Slave

Scalability

Co-Localization of Processing & Storage



Classical Architectures

Big Data Architecture

Scalability

Fault Tolerance

Data Replication

Scalability

Fault Tolerance

Data Replication

Cluster Replication

Scalability

Fault Tolerance

Data Replication

Cluster Replication

Rack Awareness

Availability

CAP Theorem

Consistency

Availability

Partition Tolerance

Availability

CAP Theorem

Atomicity

Consistency

Isolation

Durability

ACID

BASE

Basically Available

Soft-State

Eventual Consistency

Availability

Time

Stream Processing Support

Dynamic and Interactive Charts and Reports

In-Memory Processing

In-Memory Storage

Flexibility

ONE application can support…

Diverse Data Sources

Schema-less Data

Multiple Processing Paradigms

Multiple Storage Systems

Research Domains

In the Big Data Domain

EVERYTHING

Is Yet to Be Done

Research Domains

Optimization

- Processing Time Optimization
- Storage Size and Compression
- Data Access Optimization
- Tradeoff bw Consistency and Availability

Research Domains

Data Science

- Distributed Algorithms for Machine Learning
- Semantic & Sentiment Analysis
- Visualization Algorithms
- Data Mining, Data Prediction, Data Analytics

Research Domains

Big Data Design

- Design Methodologies for Big Data Systems
- Standardization of Big Data Architectures
- Design and Architectural Patterns
- Modeling Language(s) for Schema-less Data

Research Domains

Big Data Security

- Non-Relational Databases Security
- Logs Gathering and Analysis
- Source Data Validation and Filtering
- Access Control and Cryptography

Research Domains

Big Data & Other Trends

- Big Data & Business Intelligence
- Big Data & Cloud Computing
- Big Data & Internet of Things
- Big Data & Mobile

Research Domains

Big Data & Business Fields

- Big Data in Education
- Big Data in Health
- Big Data in Art
- Big Data in Finance

In Other Words…

Big Data is Here to STAY!

Thinking BIG

Big Data Concepts and Patterns

Dr. Lilia Sfaxi

April 2017