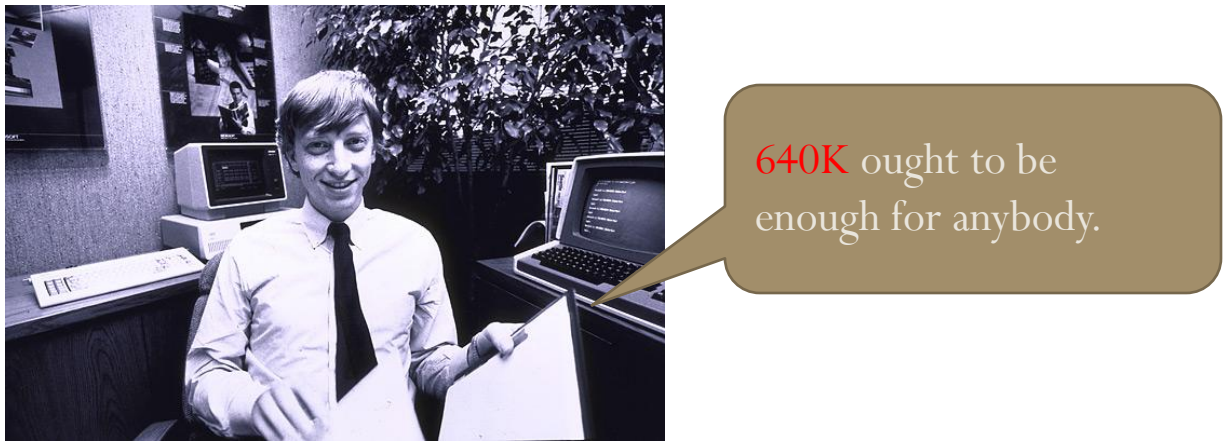
Strategic implementation of Big Data, AI/ML, Blockchain and Cloud in MES

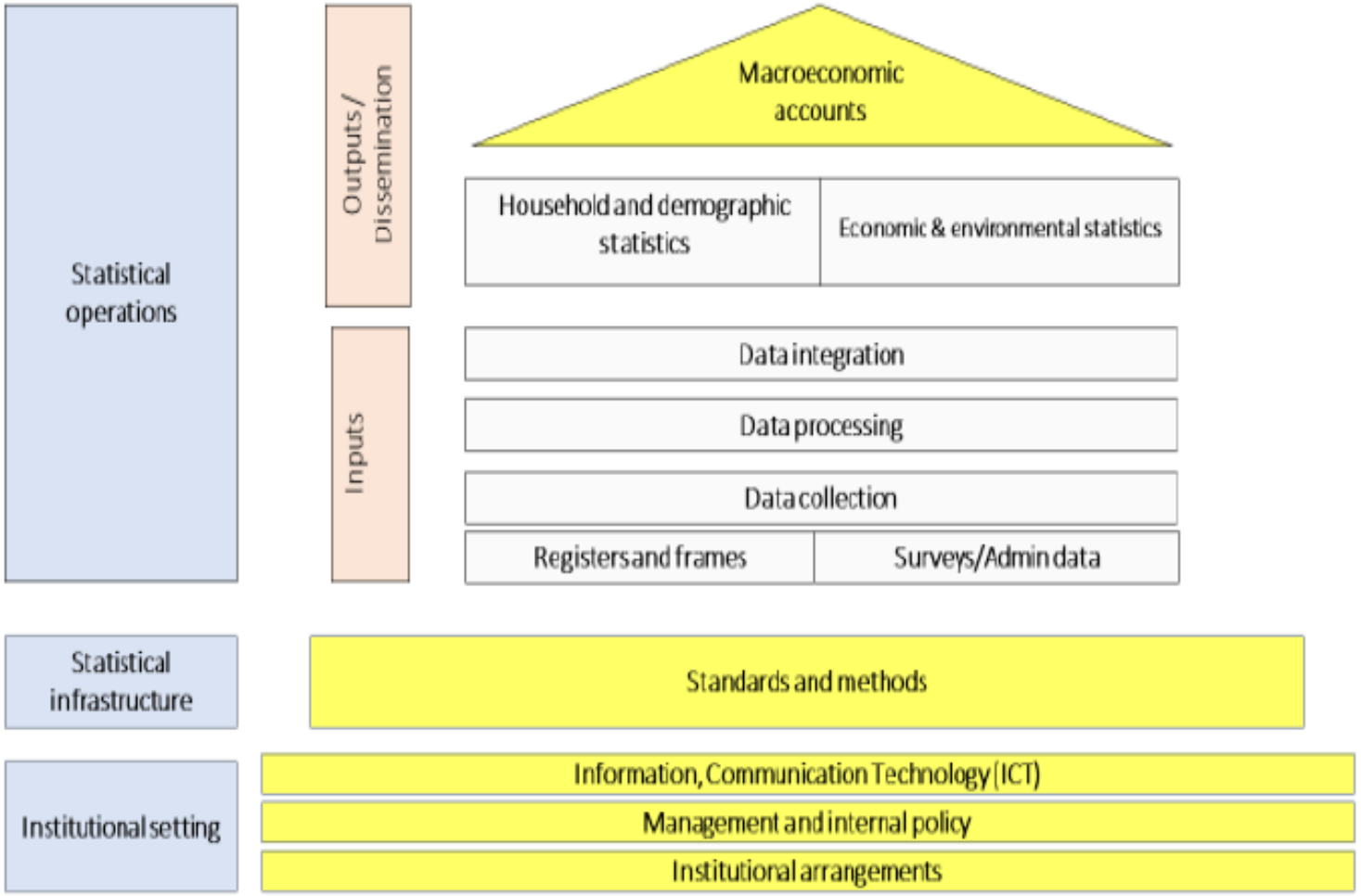
Srinivas Bhoosarapu
Chief Information Security Officer (CISO)
Pension Fund Regulatory Development Authority (PFRDA)

How much data? What is Big Data?

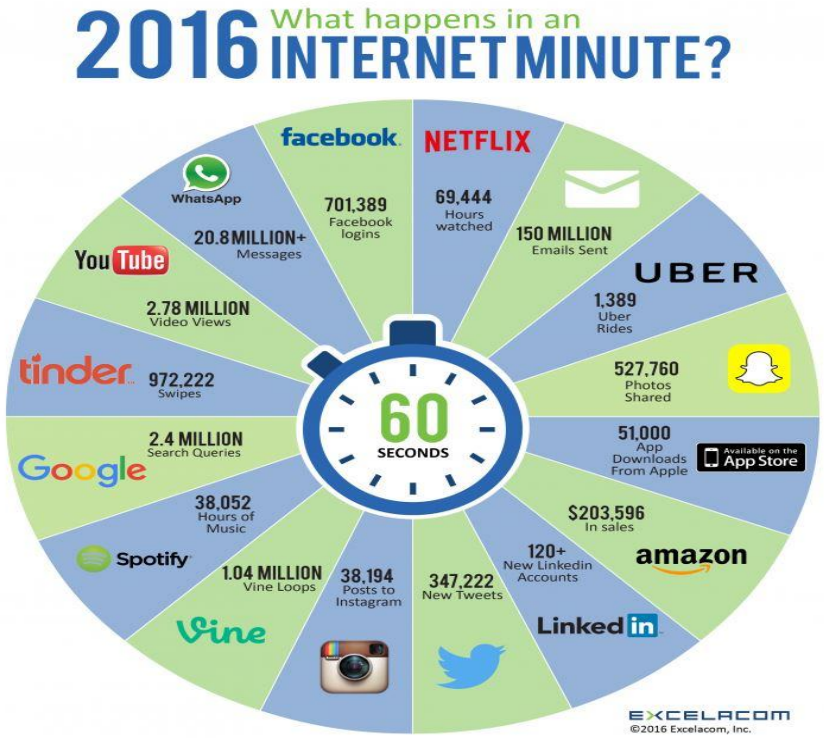
- Google processes 20 PB a day (2008)
- Wayback Machine has 3 PB + 100 TB/month (3/2009)
- Facebook has 2.5 PB of user data + 15 TB/day (4/2009)
- eBay has 6.5 PB of user data + 50 TB/day (5/2009)
- CERN's Large Hydron Collider (LHC) generates 15 PB a year



Integrated Statistics Approach

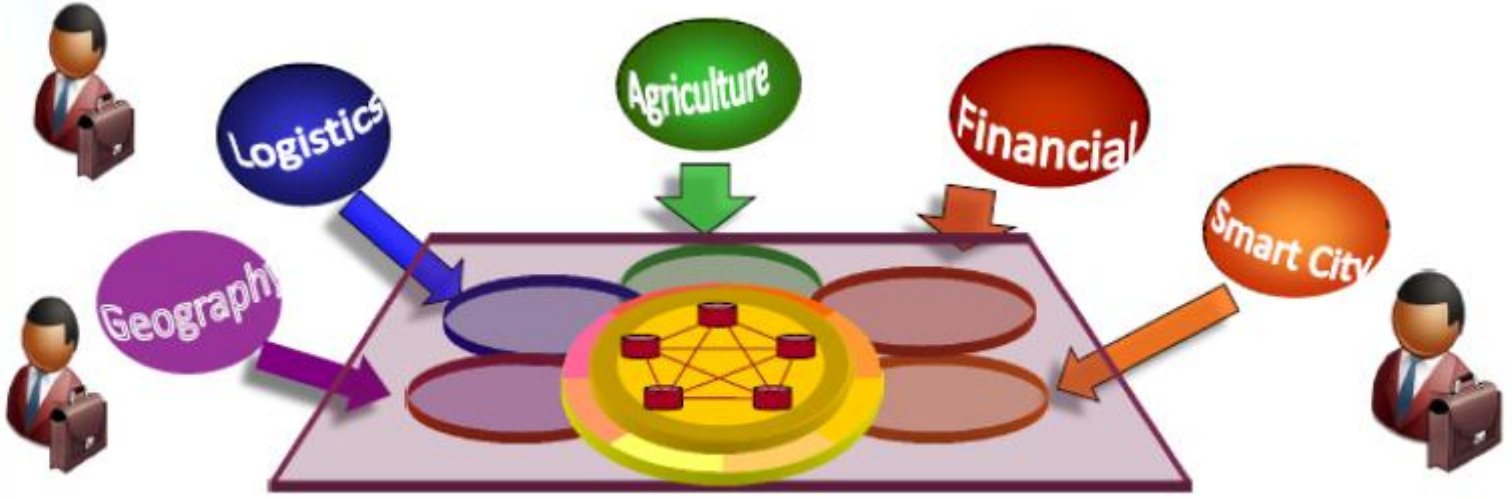


Internet in a Minute

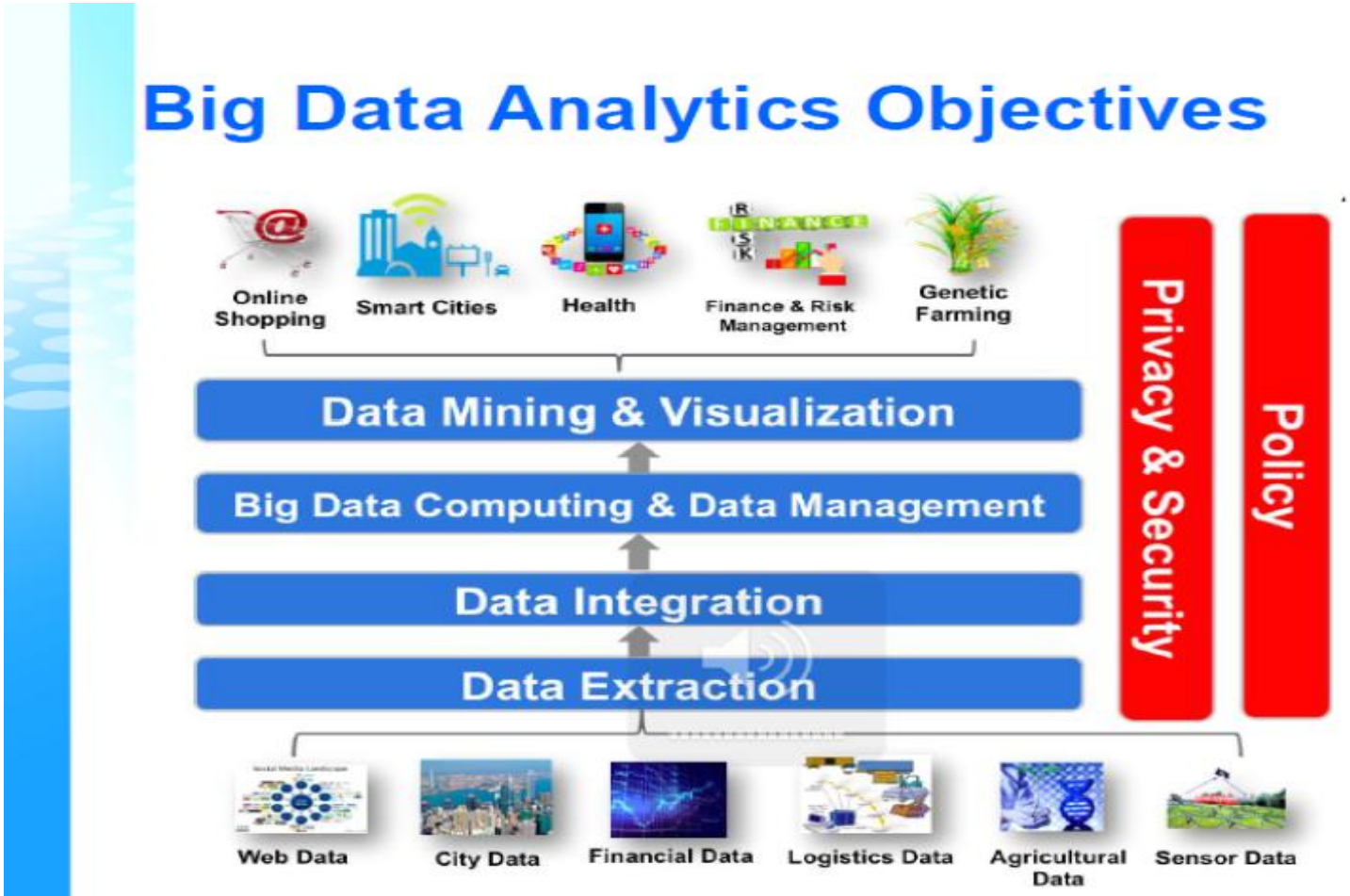


Multi-disciplinary Big-data Analytics

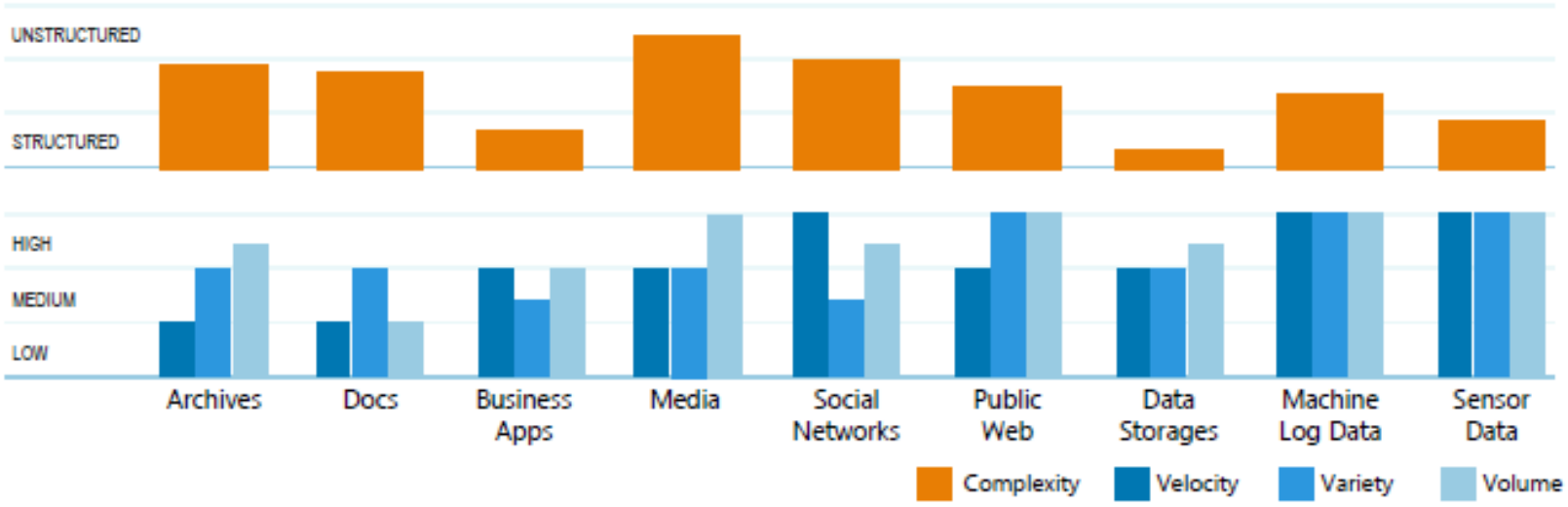
- **Objectives:**
  - Interdisciplinary, multi-university, multi-team research on heterogeneous scientific and technological big data analytics



Big Data Analytics



Big Data Challenges



Archives
Scanned documents, statements, medical records, e-mails etc..

Docs
XLS, PDF, CSV, HTML, JSON etc.

Business Apps
CRM, ERP systems, HR, project management etc.

Media
Images, video, audio etc.

Social Networks
Twitter, Facebook, Google+, LinkedIn etc.

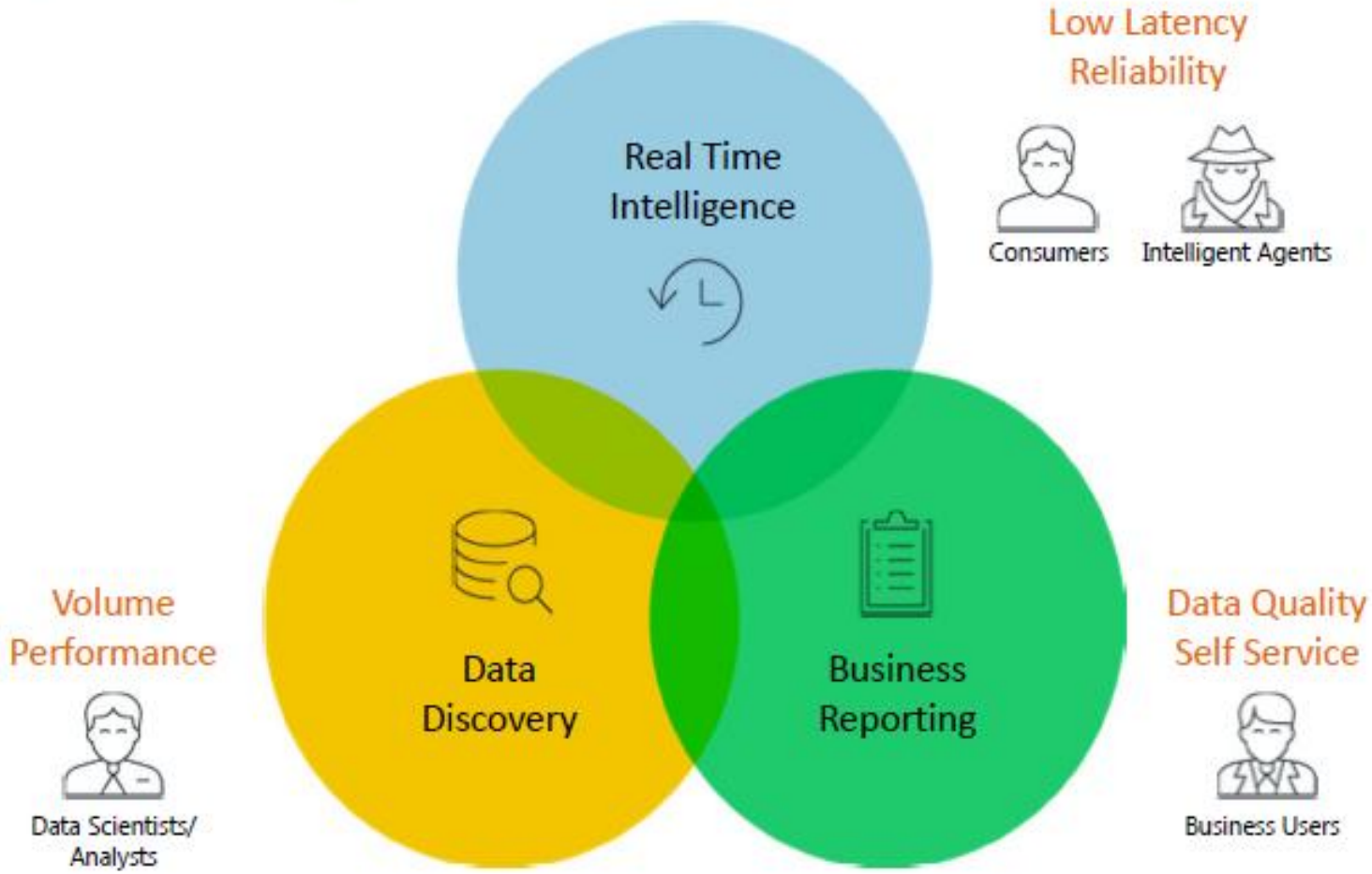
Public Web
Wikipedia, news, weather, public finance etc

Data Storages
RDBMS, NoSQL, Hadoop, file systems etc.

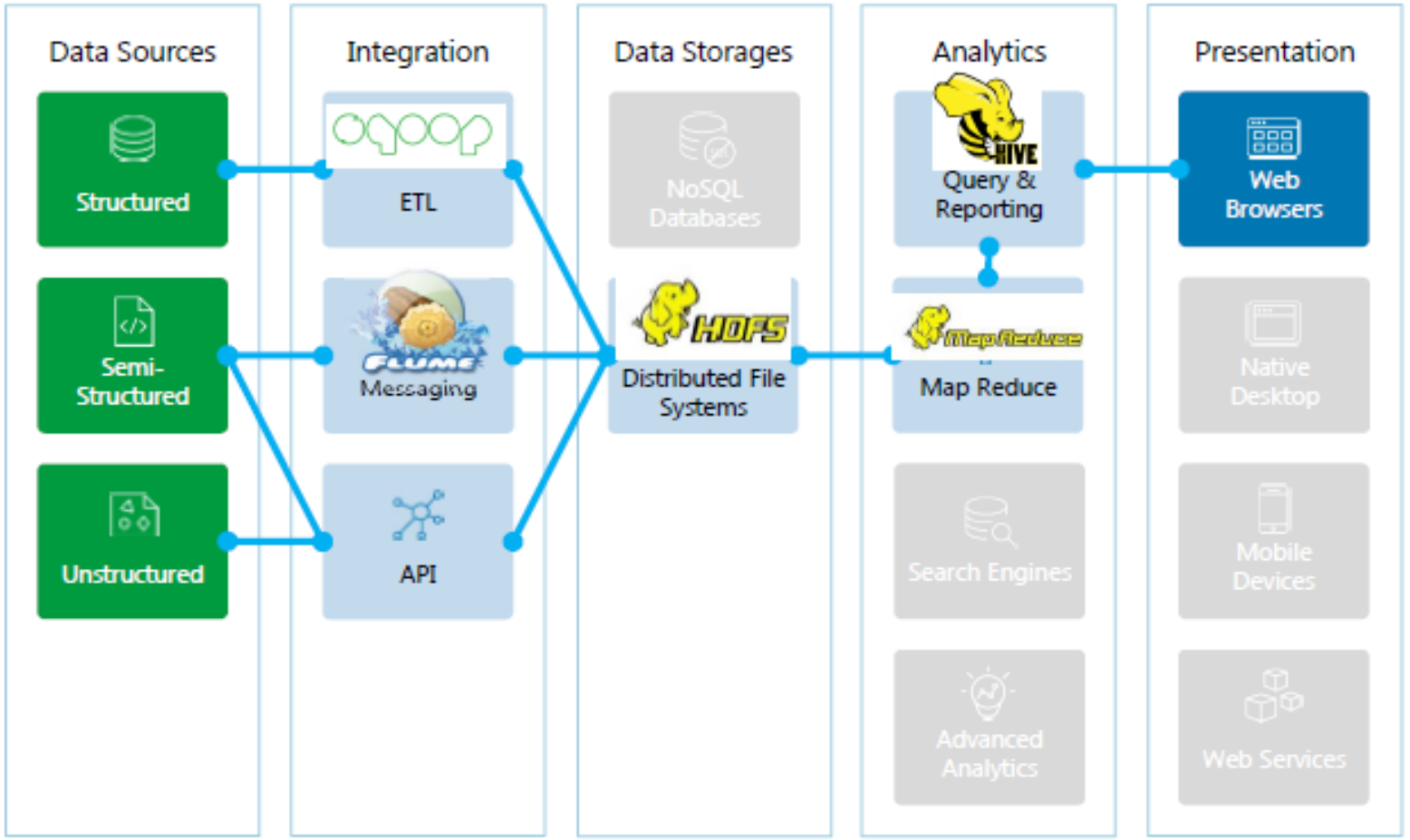
Machine Log Data
Application logs, event logs, server data, CDRs, clickstream data etc.

Sensor Data
Smart electric meters, medical devices, car sensors, road cameras etc.

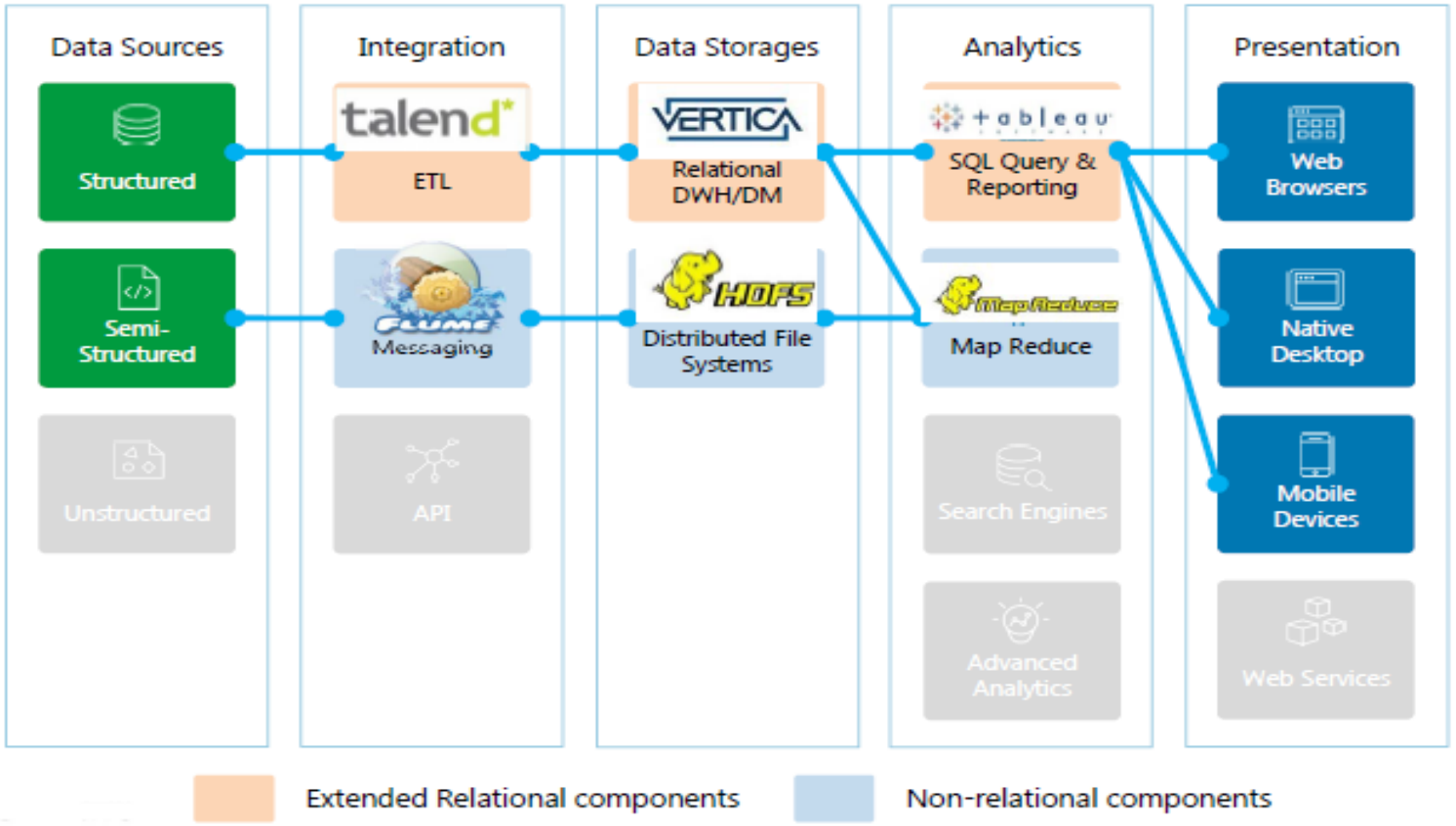
Big Data Analytics Use Cases



Data Discovery: Non-Relational Architecture



Business Reporting: Hybrid Architecture



Architectural Decisions

Architecture Drivers:

- Volume (45 TB)
- Sources (Semi-structured - JSON)
- Throughput (> 20K/sec)
- Latency (1 hour)
- Extensibility (Custom tags)
- Data Quality (Not critical)
- Reliability (24/7)
- Security (Multitenancy)
- Self-Service (Canned reports, Data science)
- Cost (The less the better ☺)
- Constraints (Public Cloud)

Trade-off:

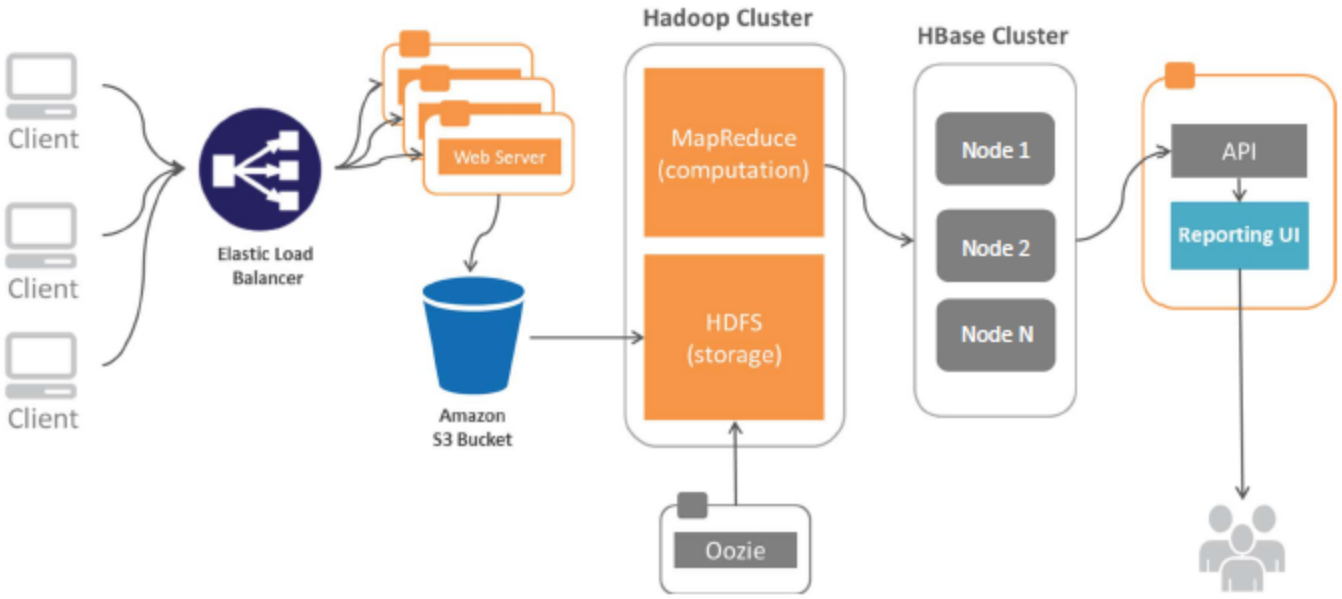
	Extended Relational	Non-Relational
Volume/Scalability	+/-	+
Throughput	+	+
Self-Service	+	+/-
Extensibility	-	+

- ✓ **Non-Relational Architecture**
- ✓ Reporting via **Materialized View** pattern

Solution Architecture

Technologies:

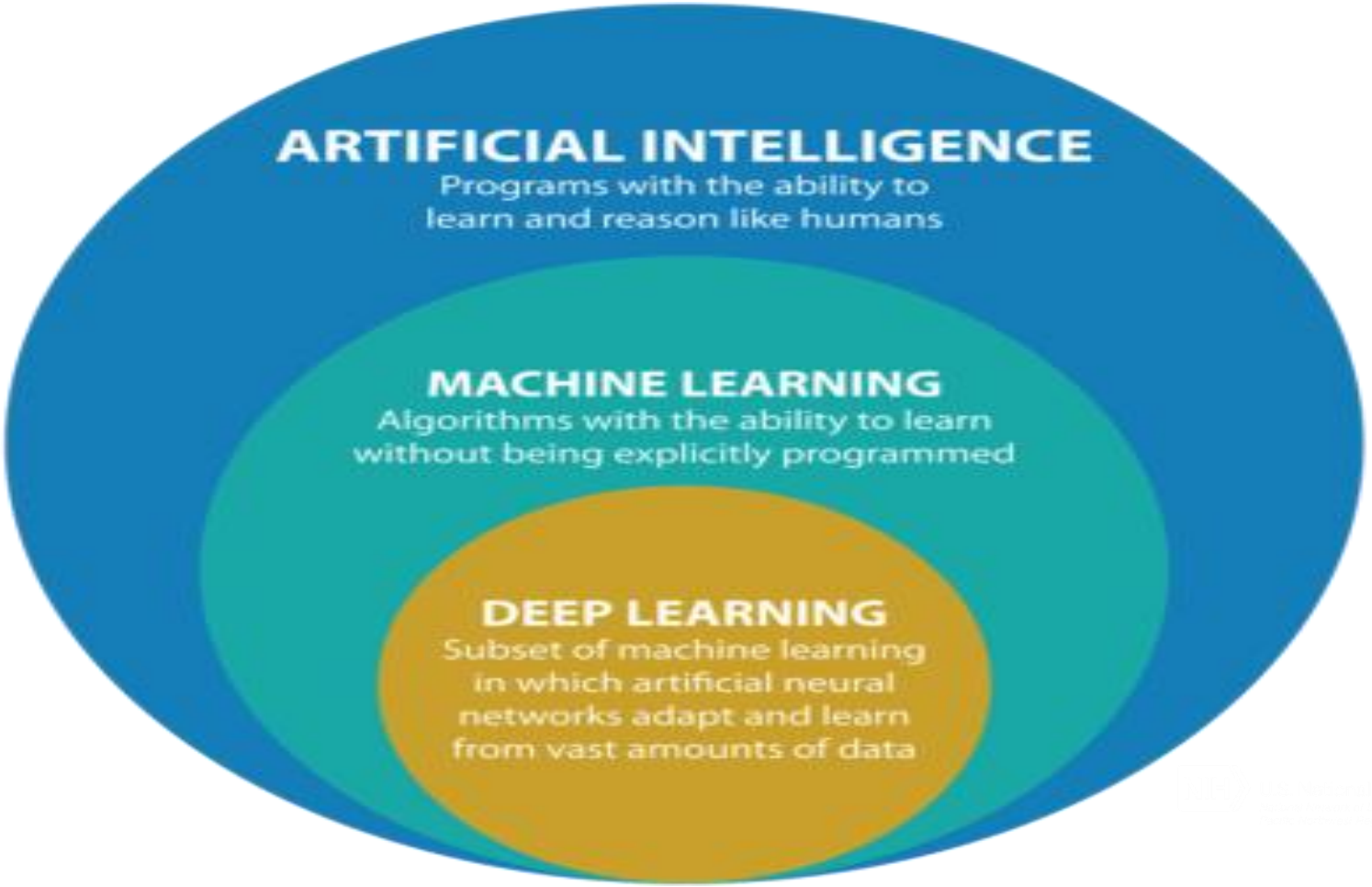
- Amazon S3
- Flume
- Hadoop/HDFS, MapReduce
- HBase
- Oozie
- Hive



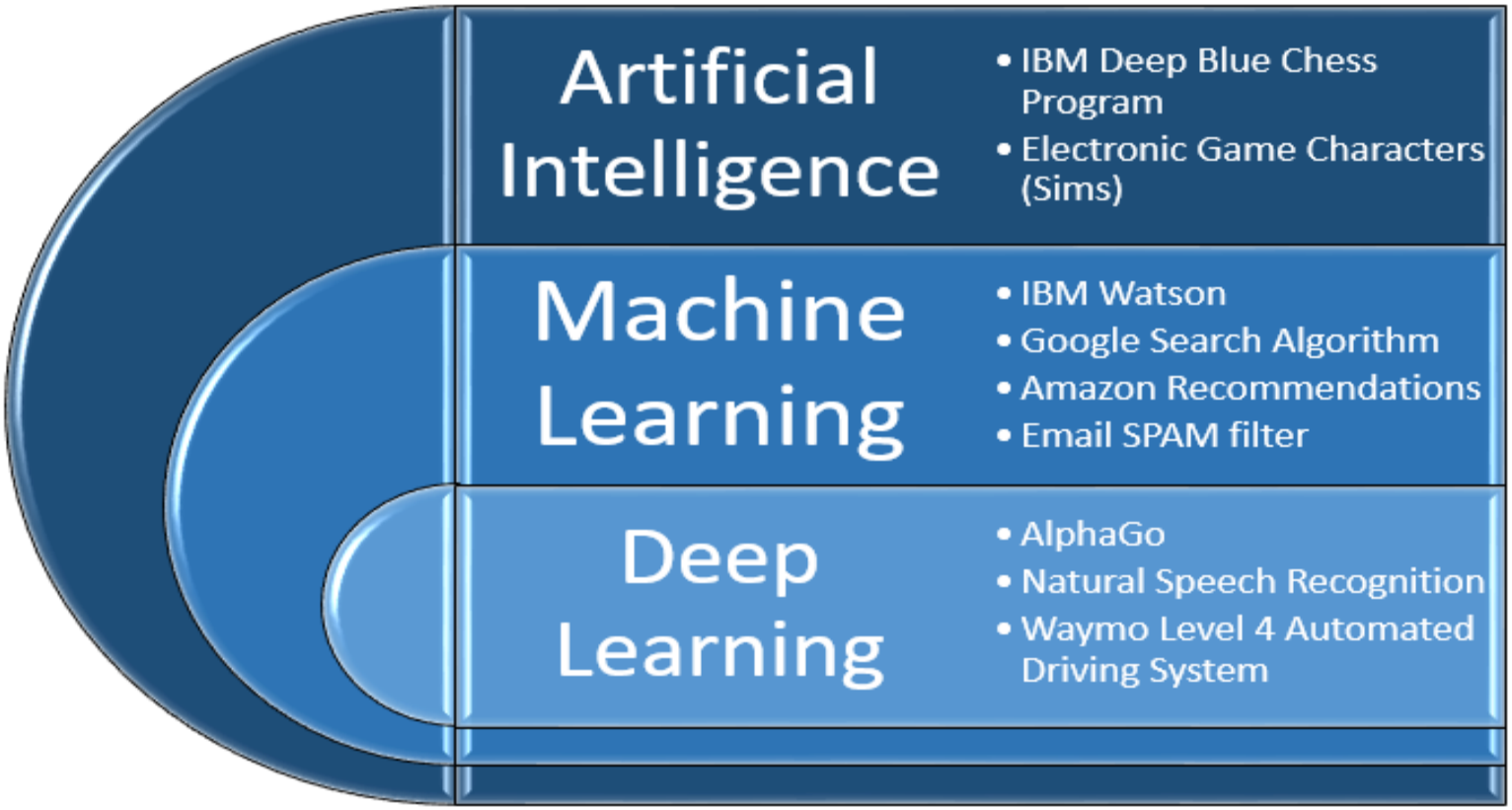
Tips for Designing Big Data Solutions

- Understand data users and sources
- Discover architecture drivers
- Select proper reference architecture
- Do trade-off analysis, address cons
- Map reference architecture to technology stack
- Prototype, re-evaluate architecture
- Estimate implementation efforts
- Set up devops practices from the very beginning
- Advance in solution development through "small wins"
- Be ready for changes, big data technologies are evolving rapidly

What is AI/ML/DL?



AI/ML/DL Usage



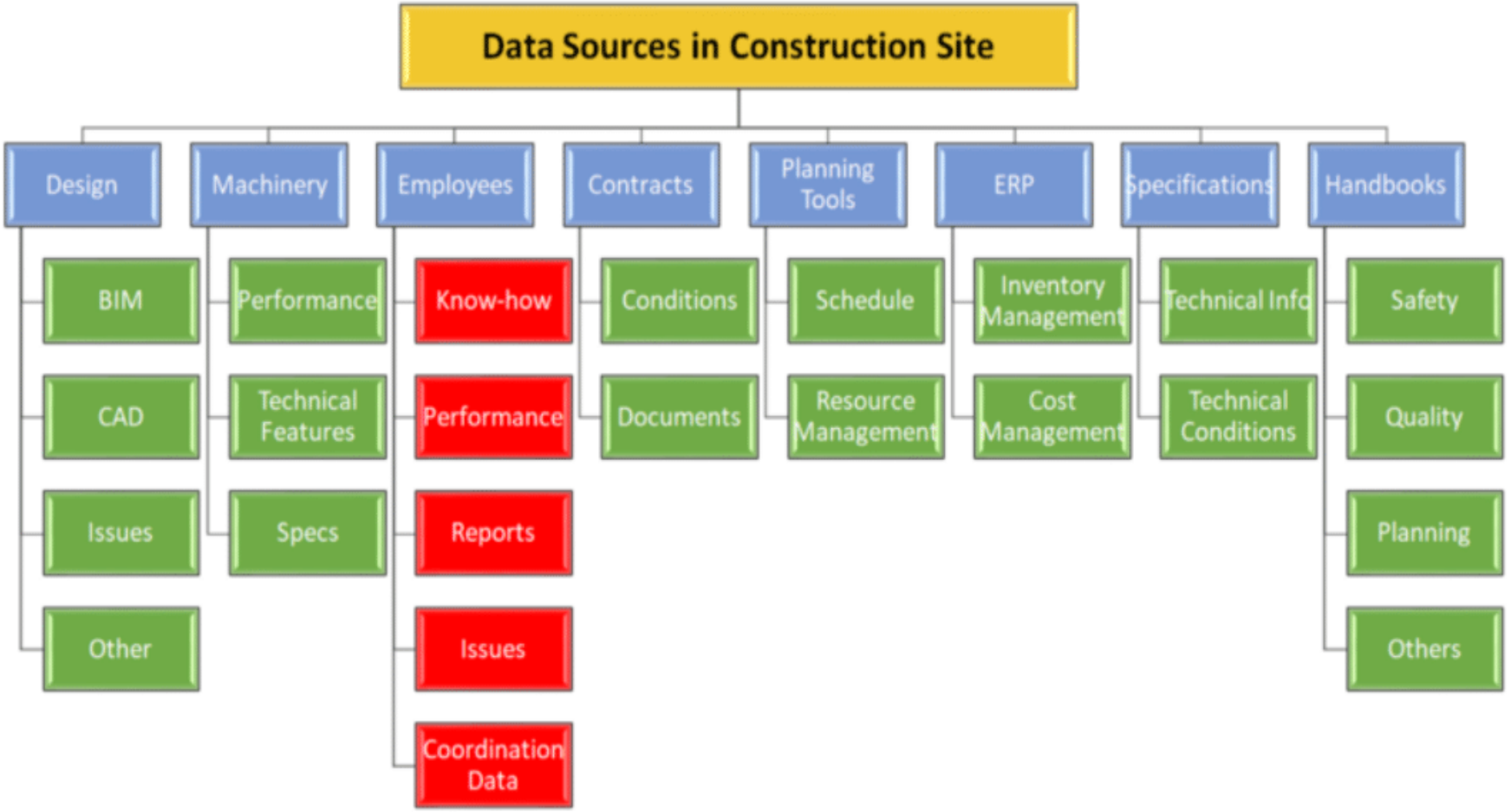
AI in Project Management

- Predict and mitigate risks before they impact project margins
- Identify high risk issues and automatically classify them into actionable categories
- Identify high risk subcontractors based on real-time data as well as past performance and other factors
- Identify and prioritize potential safety concerns across the project lifecycle
- Tag existing safety hazards based on visual data coming out of the job site

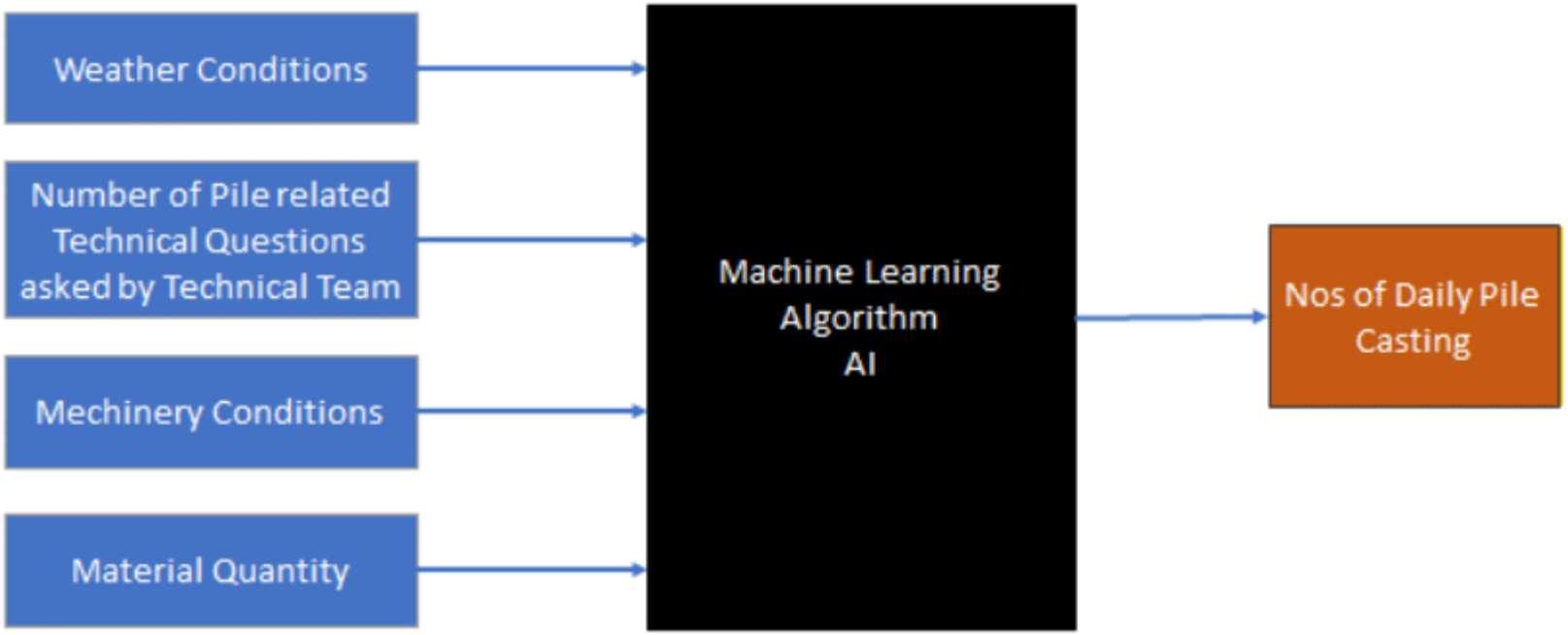
AI for safety in Construction

AI in Construction

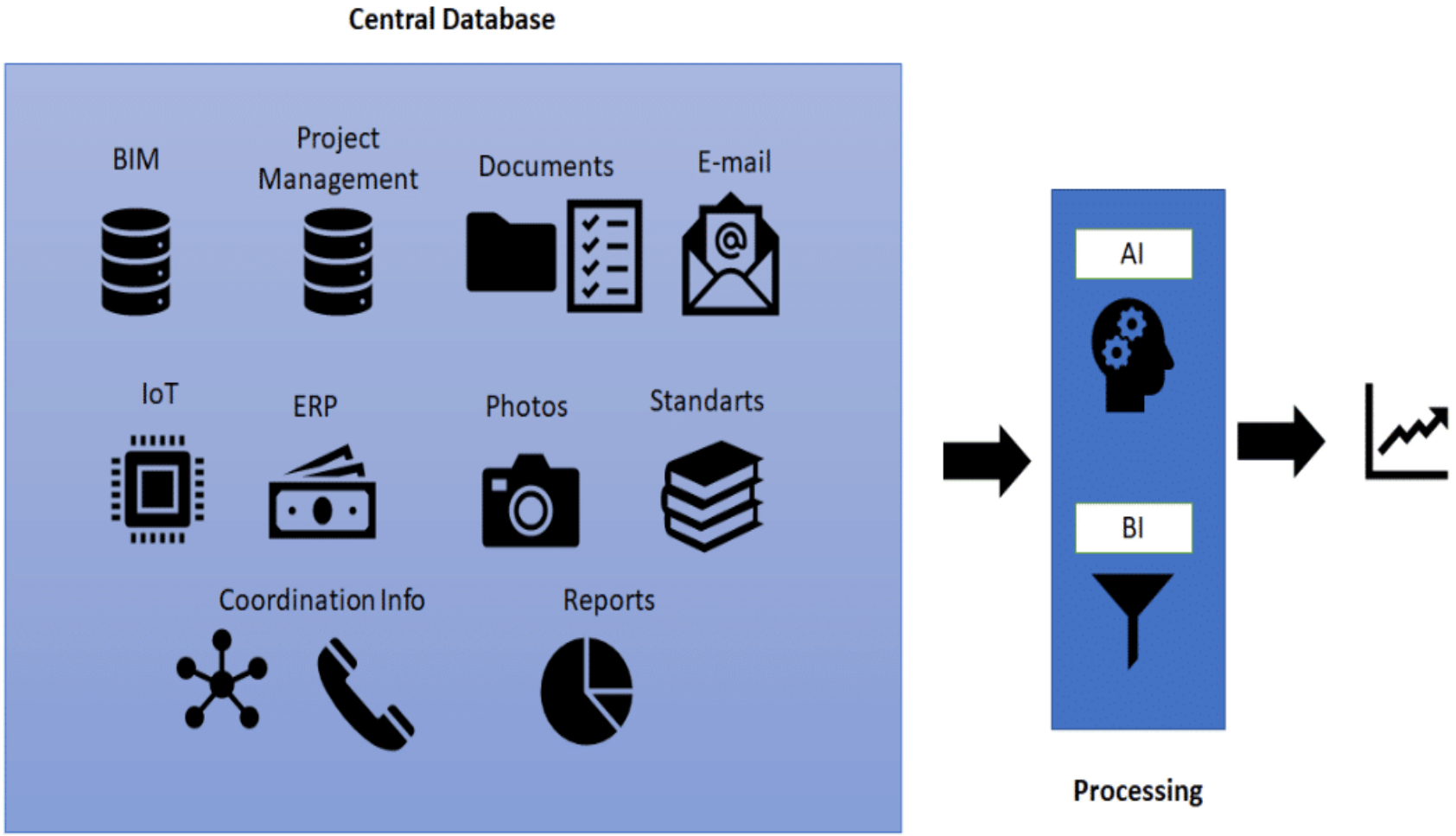
AI in Construction

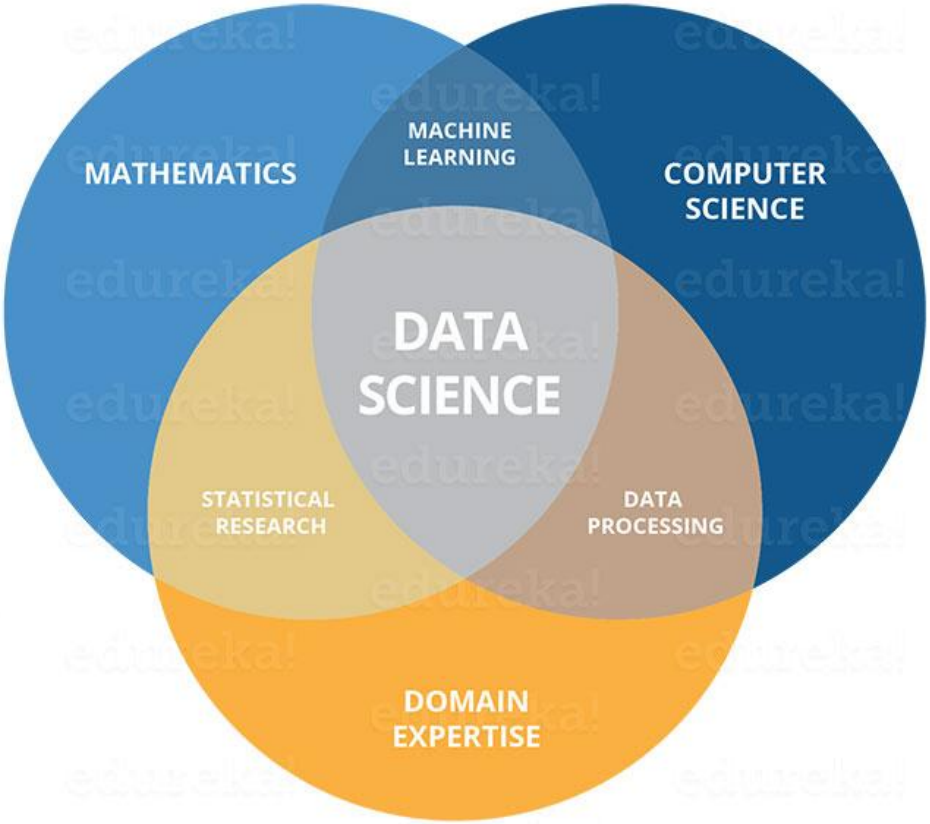


AI in Construction

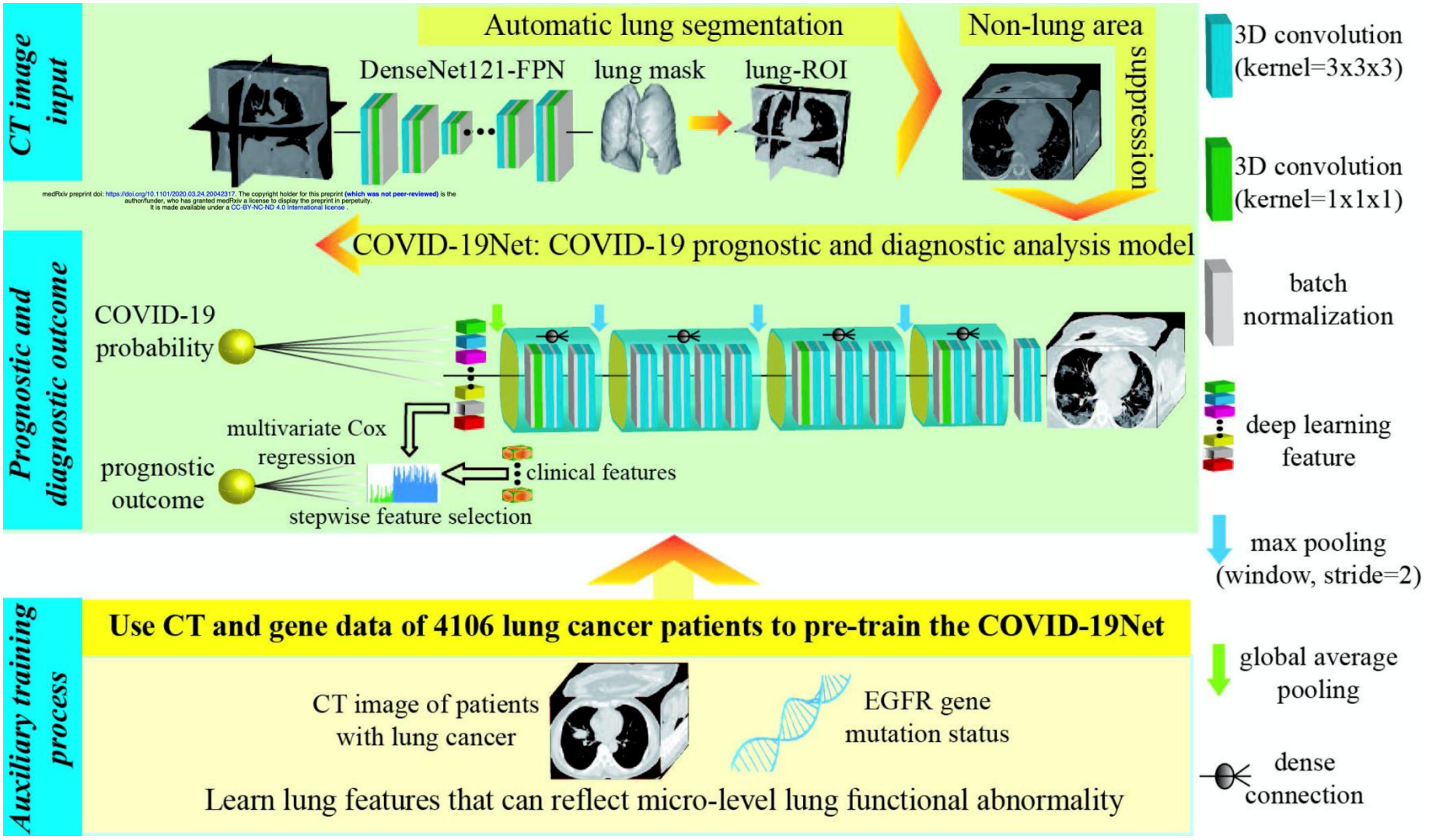


AI in Construction



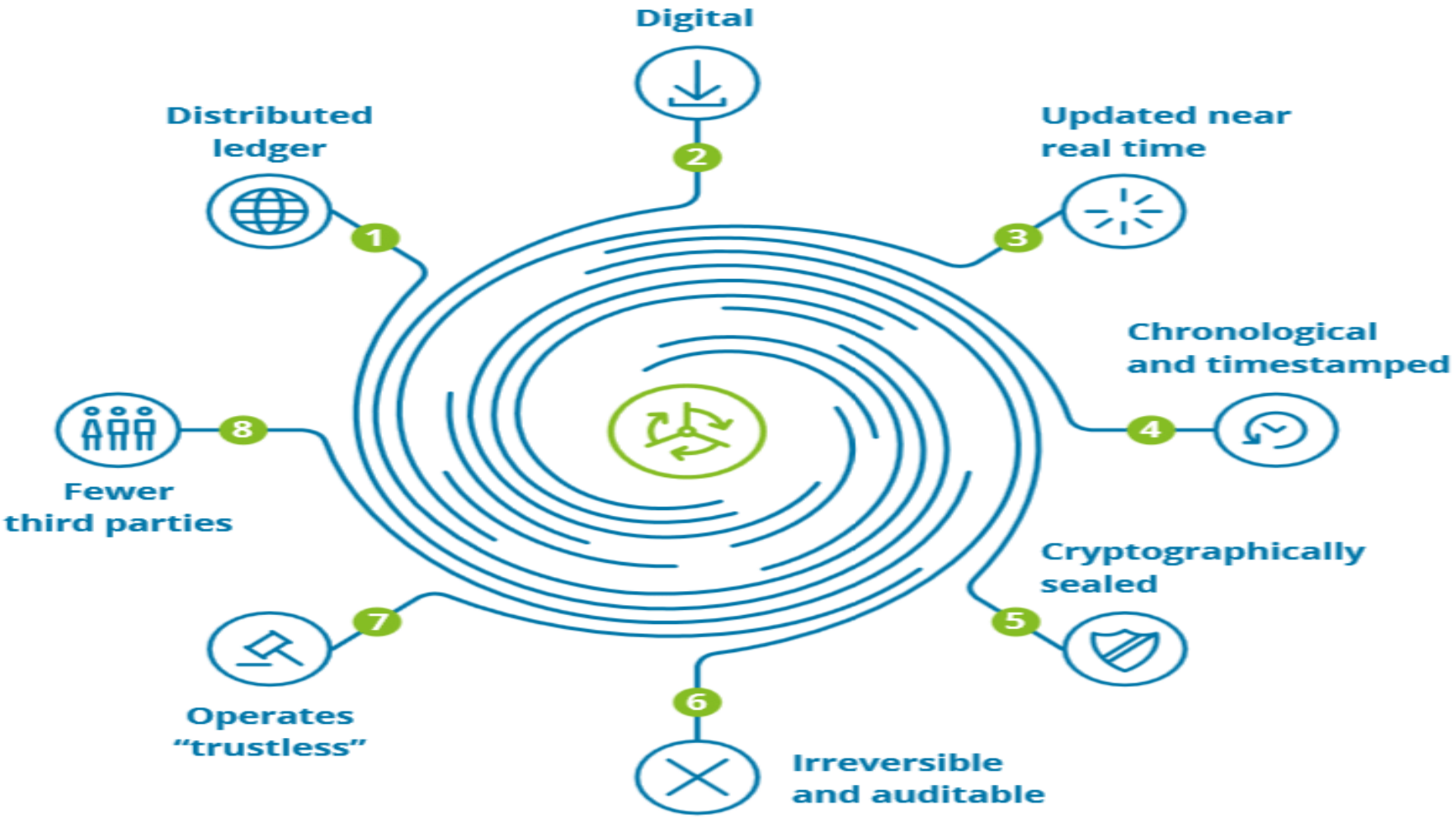

MATHEMATICS
MACHINE LEARNING
COMPUTER SCIENCE
DATA SCIENCE
STATISTICAL RESEARCH
DATA PROCESSING
DOMAIN EXPERTISE

Covid 19 Deep Learning

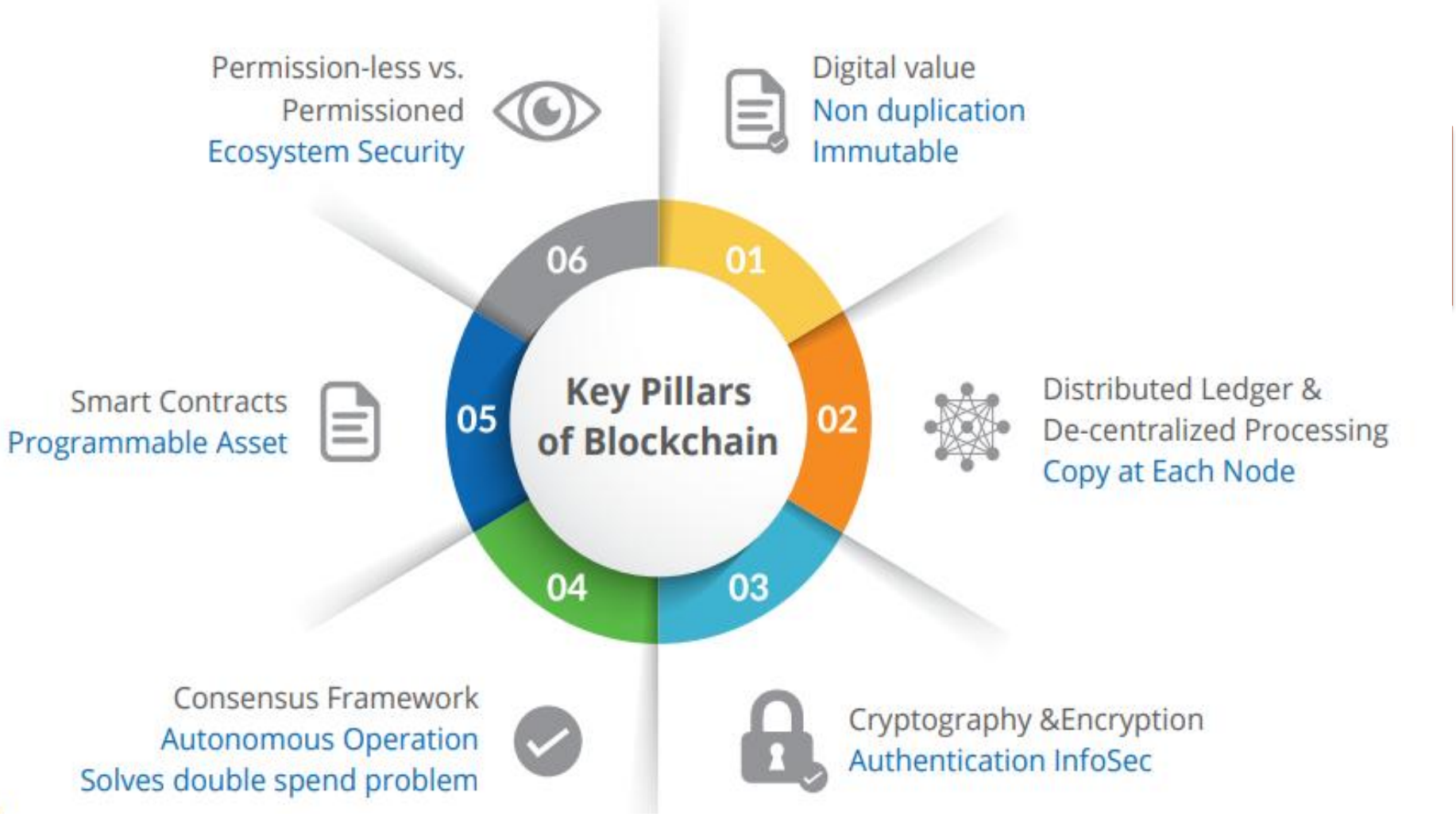


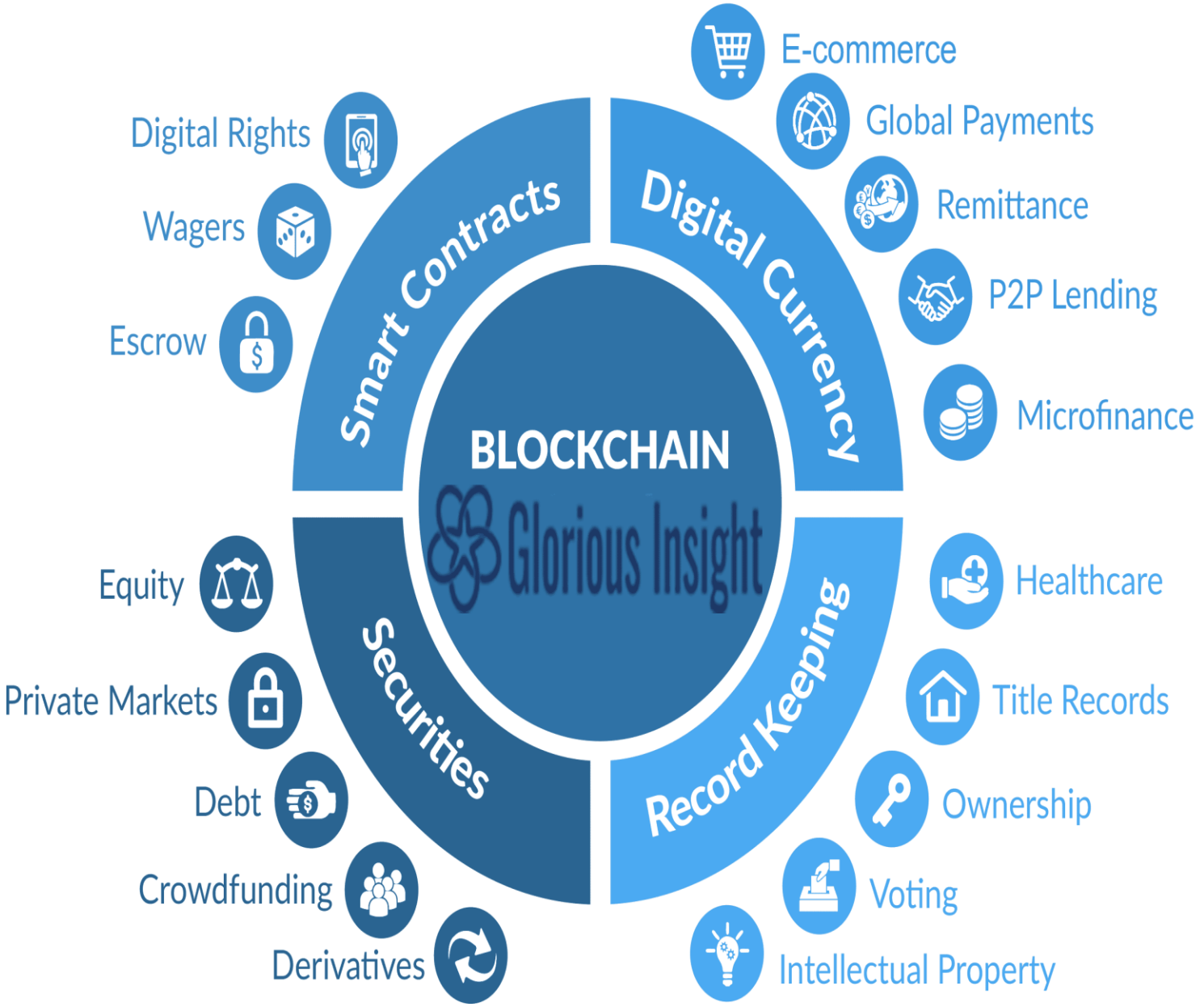
medRxiv preprint doi: https://doi.org/10.1101/2020.03.24.20042317. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

What is Blockchain?



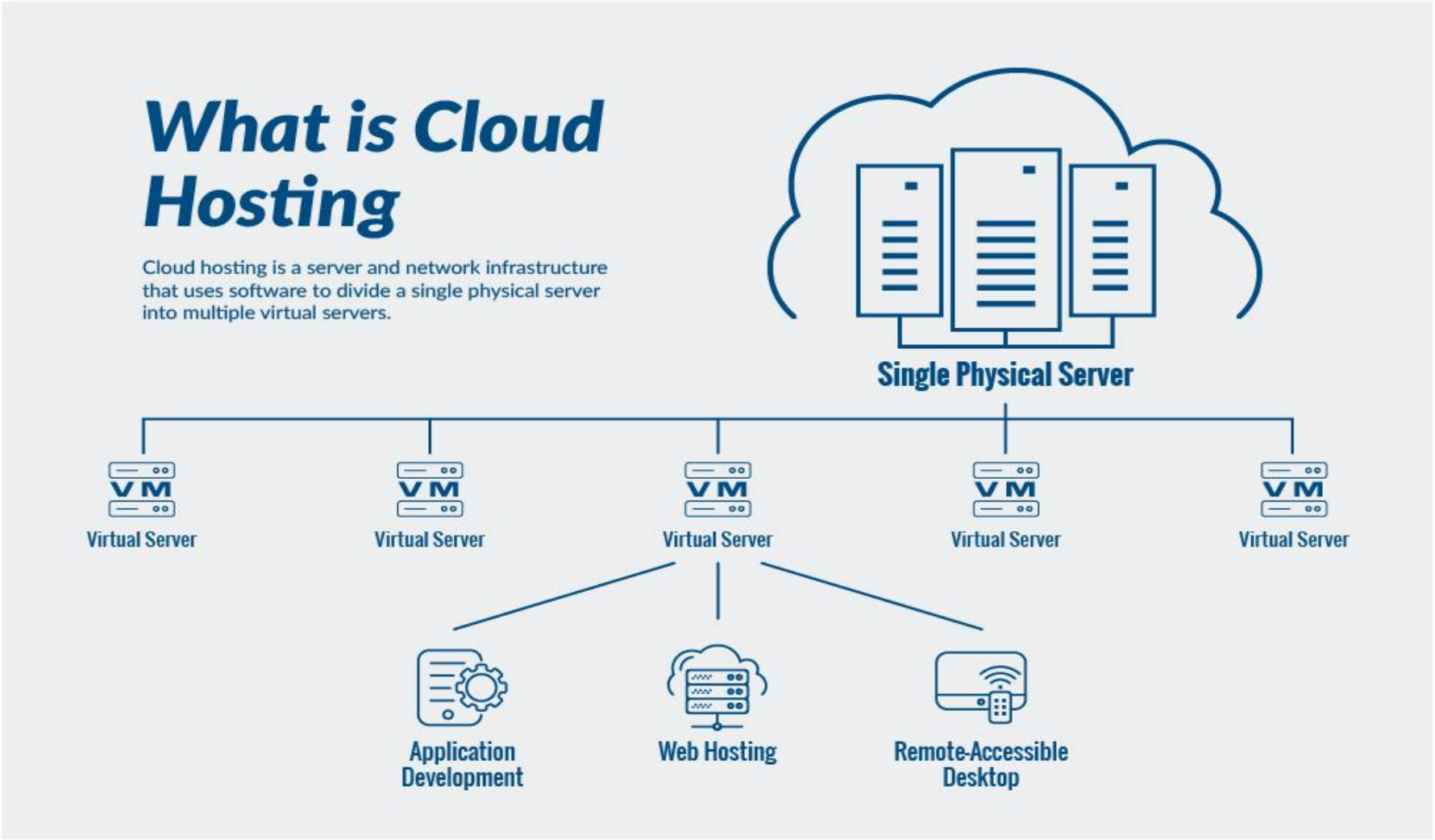
Blockchain Pillars



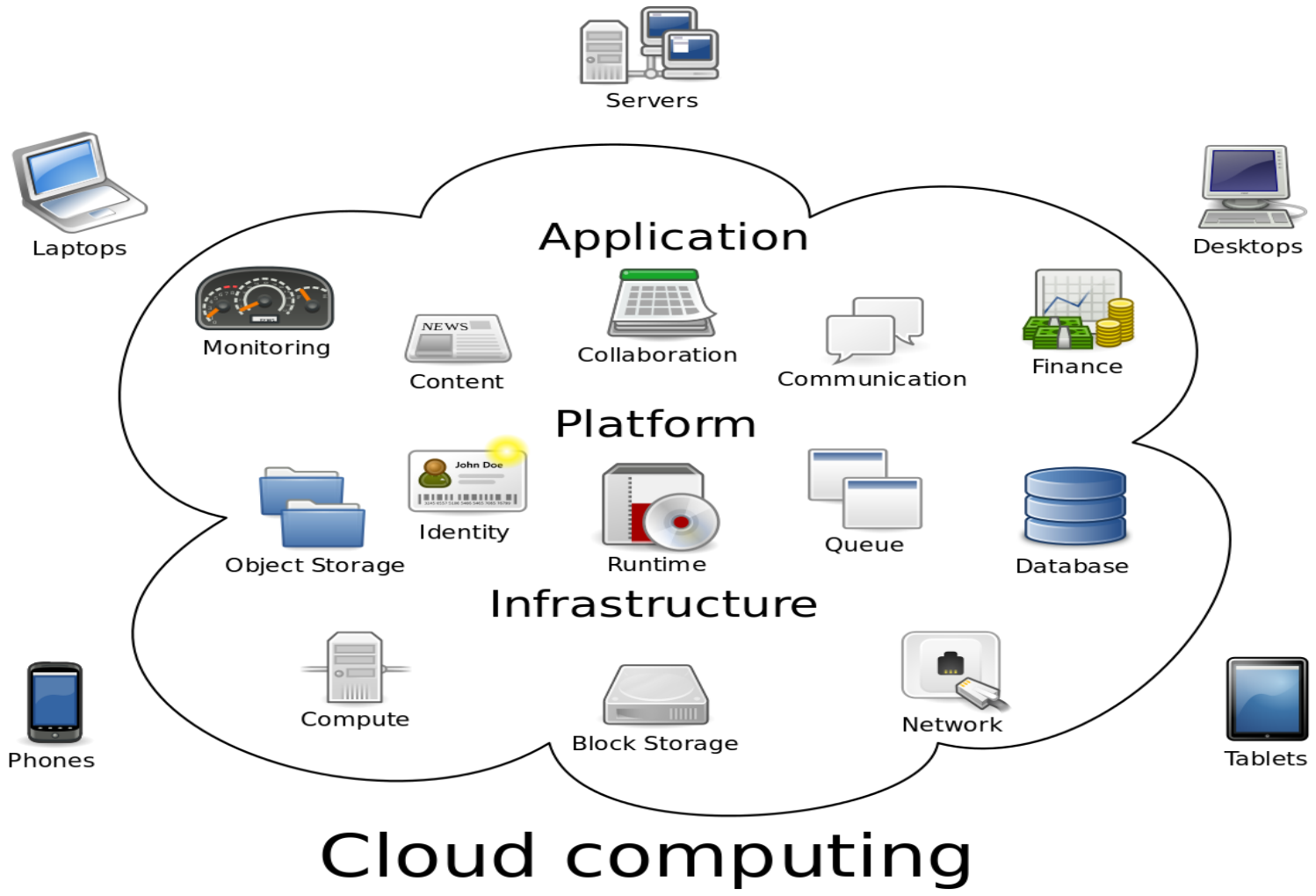

BLOCKCHAIN
Glorious Insight
Smart Contracts
Digital Currency
Securities
Record Keeping
Digital Rights
Wagers
Escrow
E-commerce
Global Payments
Remittance
P2P Lending
Microfinance
Equity
Private Markets
Debt
Crowdfunding
Derivatives
Healthcare
Title Records
Ownership
Voting
Intellectual Property


BLOCKCHAIN
VERIFIED
ADDRESS
WALLET
ICO
NETWORK
TRUST
CRYPTOCURRENCY
MINING
%
TRANSACTION
SMART CONTRACTS

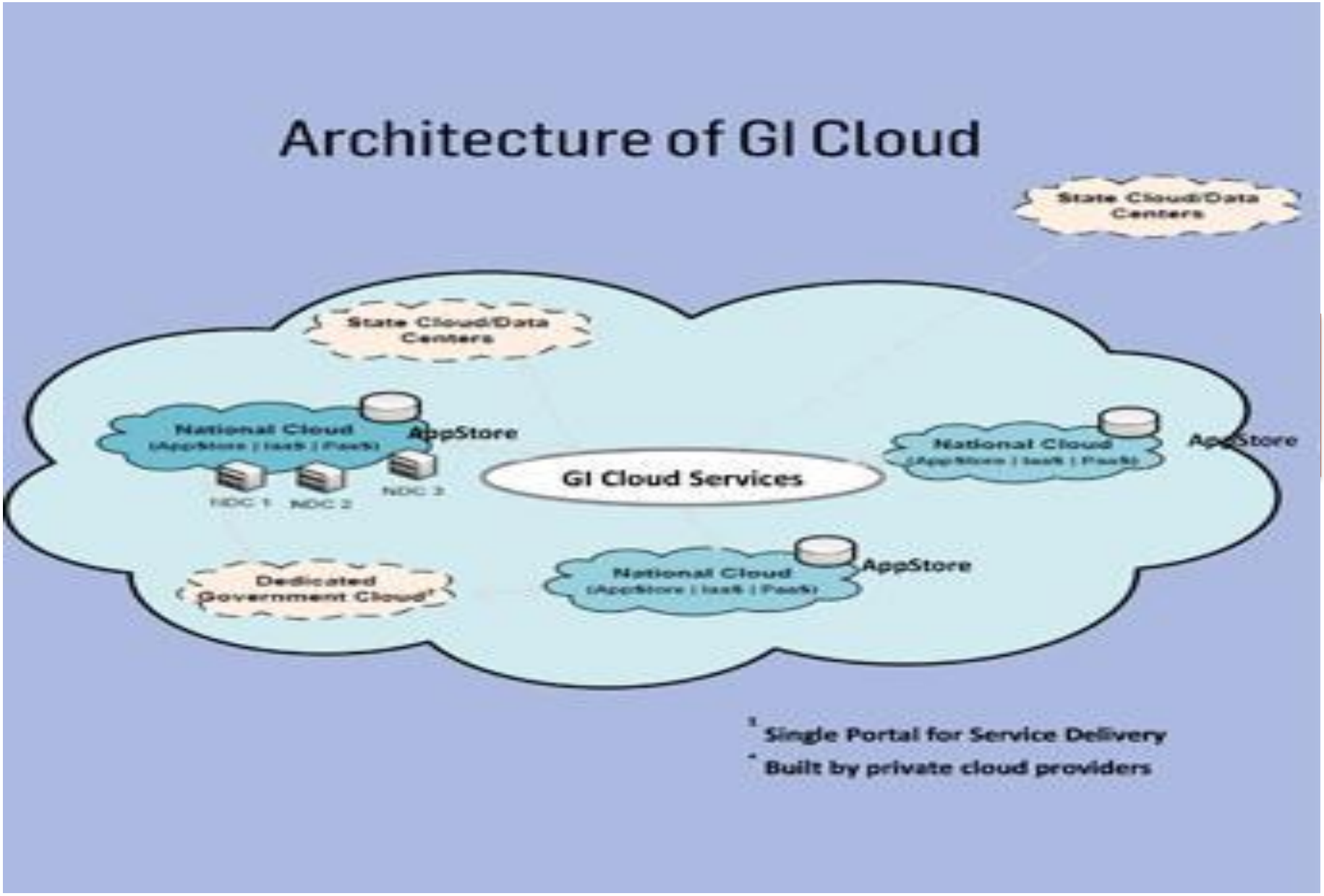
Cloud Computing



Types of Cloud Computing



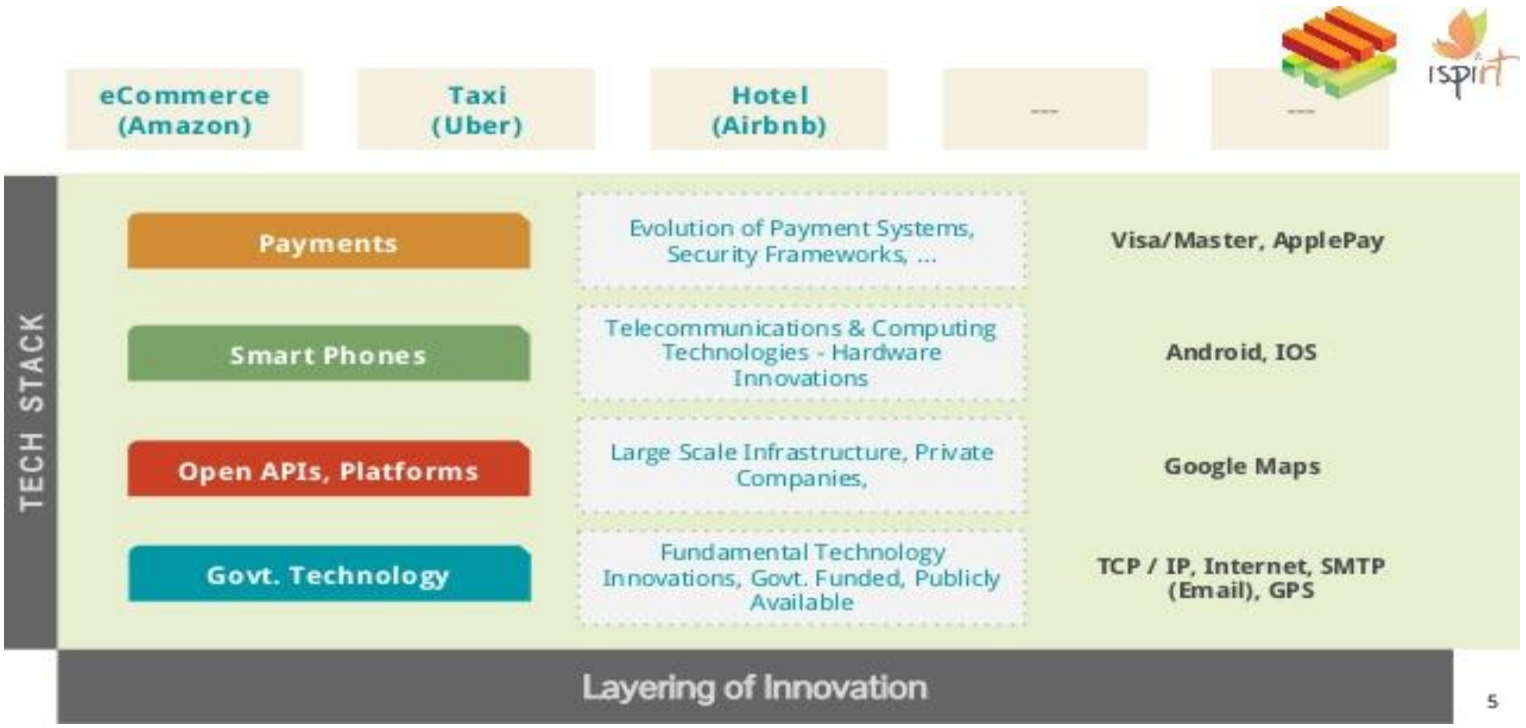
Architecture of GI Cloud



Thank you

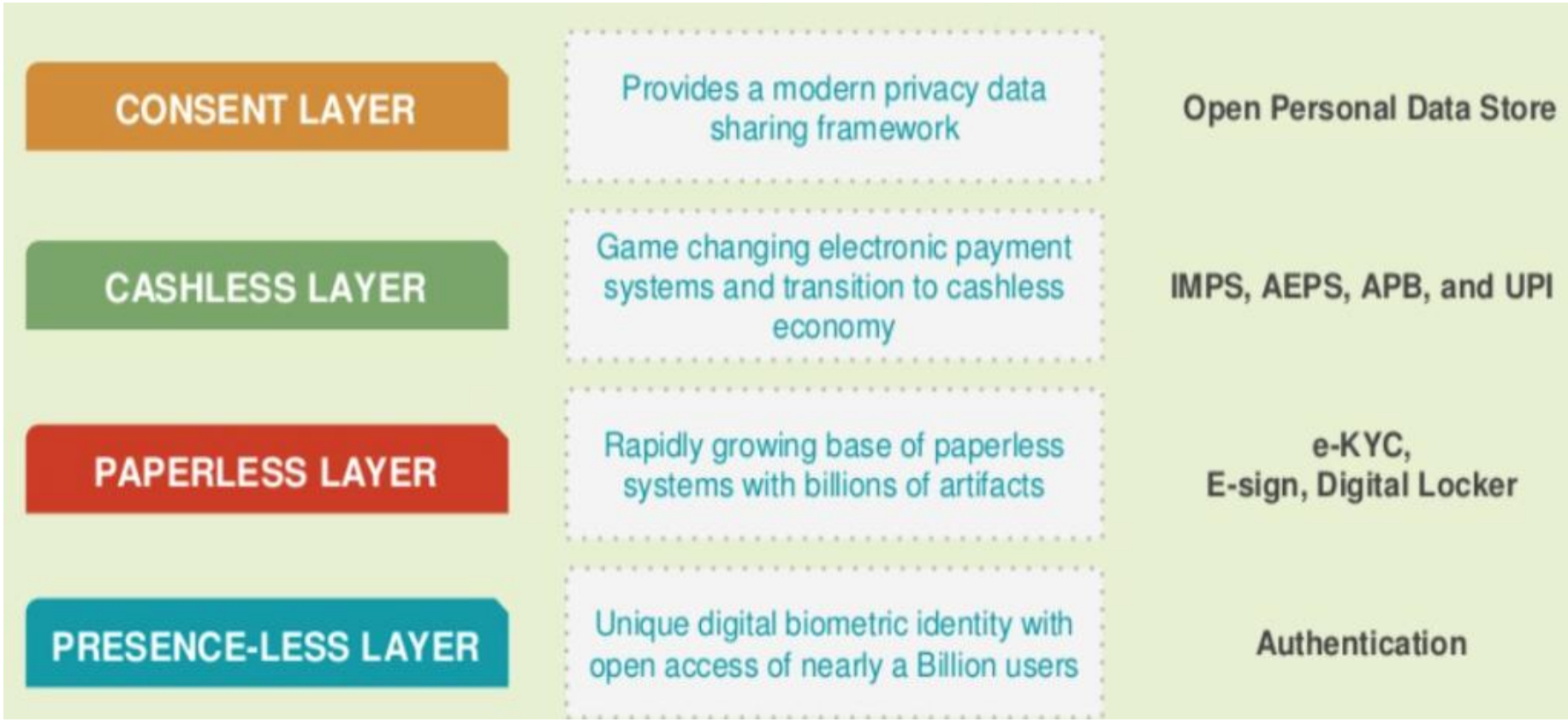
Srinivas.bhoosarapu@gmail.com
Mobile : 9154035731

India Stack

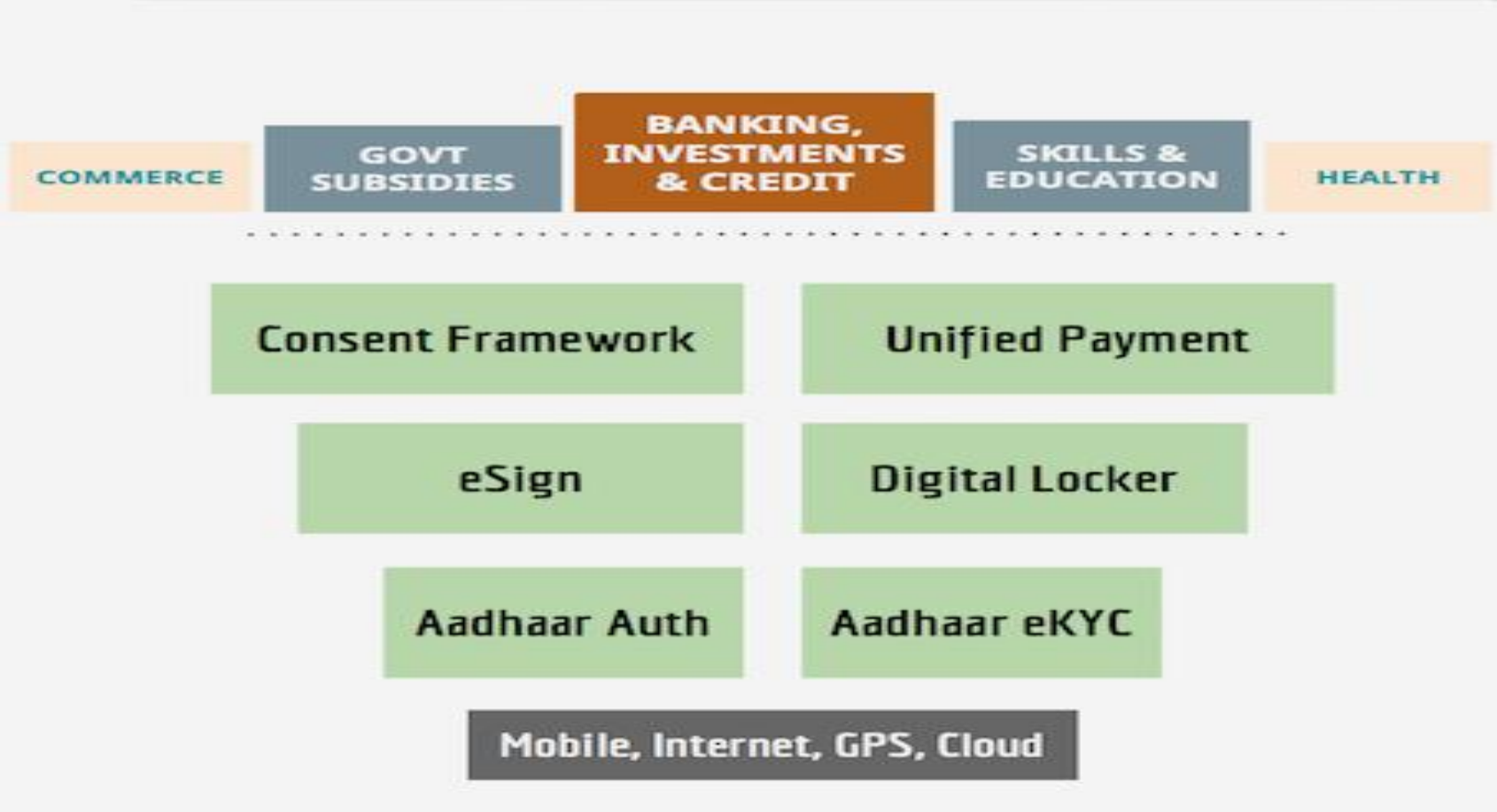

eCommerce (Amazon)
Taxi (Uber)
Hotel (Airbnb)

TECH STACK
Payments
Evolution of Payment Systems, Security Frameworks, ...
Visa/Master, ApplePay
Smart Phones
Telecommunications & Computing Technologies - Hardware Innovations
Android, IOS
Open APIs, Platforms
Large Scale Infrastructure, Private Companies,
Google Maps
Govt. Technology
Fundamental Technology Innovations, Govt. Funded, Publicly Available
TCP / IP, Internet, SMTP (Email), GPS
Layering of Innovation
5

India Stack


CONSENT LAYER
Provides a modern privacy data sharing framework
Open Personal Data Store
CASHLESS LAYER
Game changing electronic payment systems and transition to cashless economy
IMPS, AEPS, APB, and UPI
PAPERLESS LAYER
Rapidly growing base of paperless systems with billions of artifacts
e-KYC,
E-sign, Digital Locker
PRESENCE-LESS LAYER
Unique digital biometric identity with open access of nearly a Billion users
Authentication

India Stack


COMMERCE
GOVT SUBSIDIES
BANKING, INVESTMENTS & CREDIT
SKILLS & EDUCATION
HEALTH
Consent Framework
Unified Payment
eSign
Digital Locker
Aadhaar Auth
Aadhaar eKYC
Mobile, Internet, GPS, Cloud

Case Study – Power Sector

Vulnerabilities in the Value Chain

Power sector is vulnerable to both short-term and long-term disruptions, e.g.:
- Unauthorized access to control systems causes outages, overloads or other damages
- Malicious data transmission causes unintended system behavior
- Meter tampering causes huge financial losses due to replacement
- Theft of personally identifiable information reveals usage patterns, home occupancy, etc.

Generation Vulnerabilities

- Weaknesses in GenCos' IT systems
- SCADA vulnerabilities: Weak authentication, backdoors, ladder logic

Transmission Vulnerabilities

- D-DOS attack on smart grids
- Malicious data injection
- Attacks on controllers (SCADA, PLCs)

Distribution Vulnerabilities

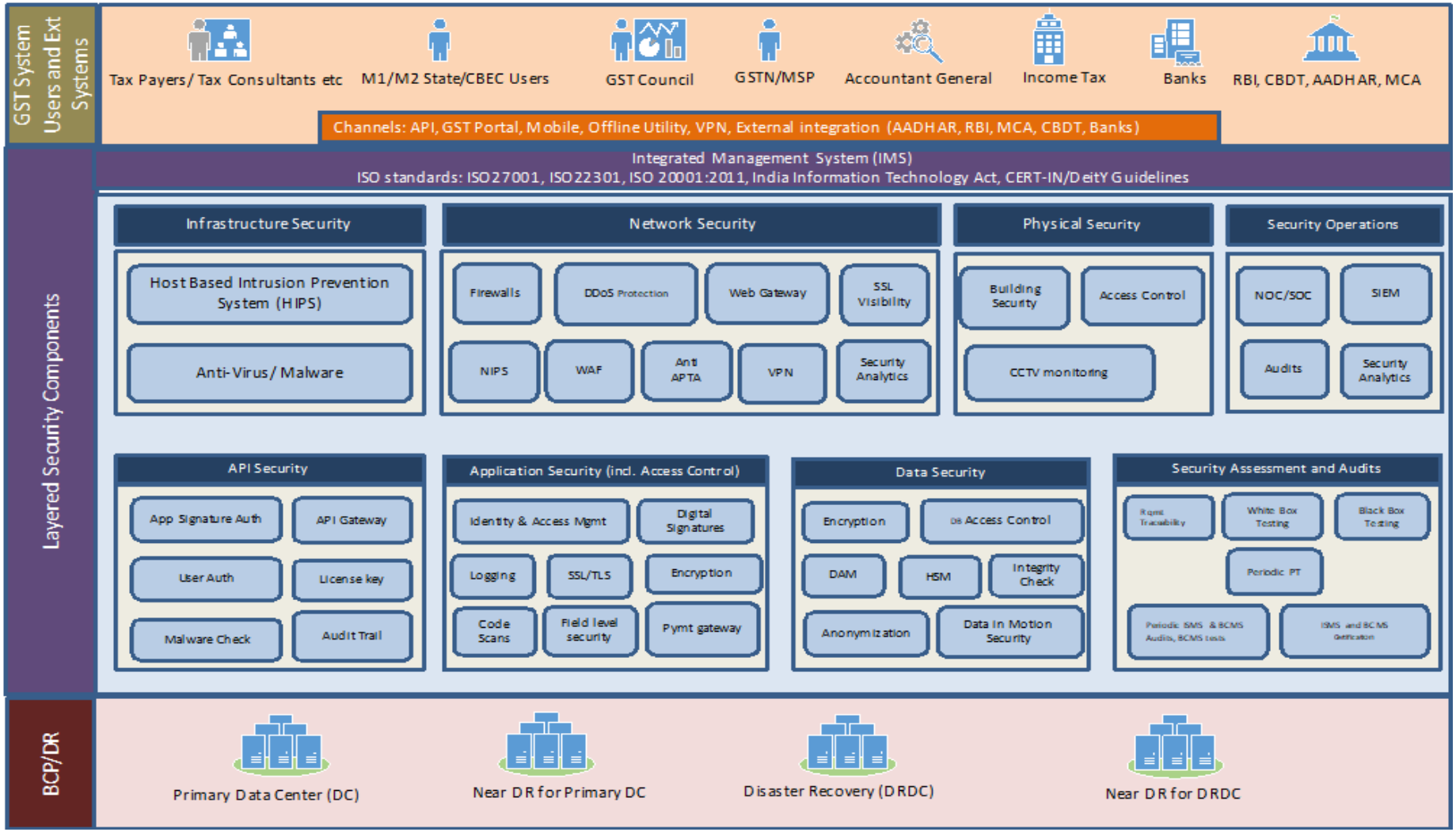
- Network Operating Centre impersonation
- Smart Meter tampering through unauthorized control

Other vulnerabilities

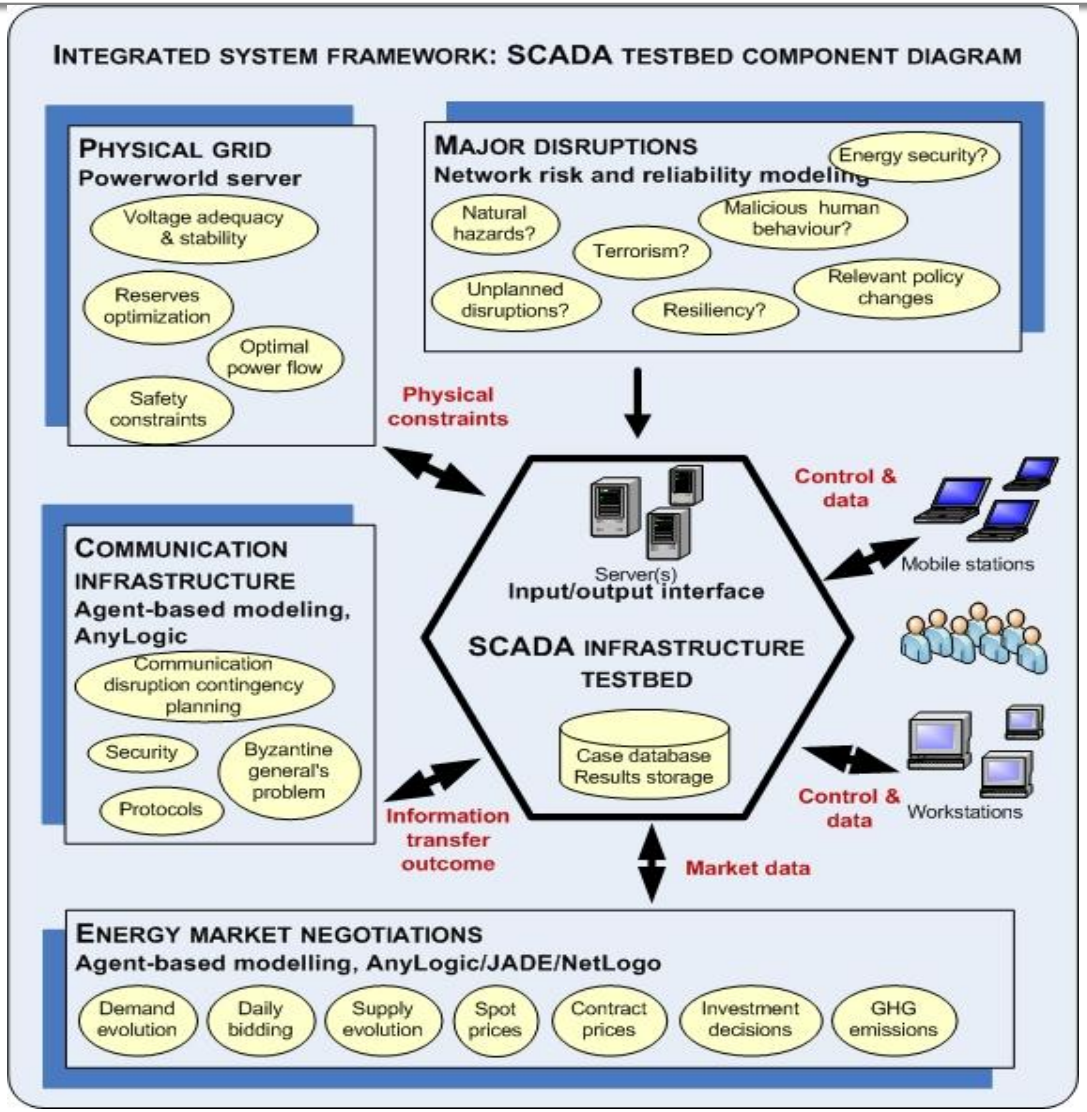
- **Telemetry** (data connectivity) systems have little to no security protocols
- **Consumer data** can potentially be stolen from Smart Grids and put to malicious use
- **Zero-day threats** due to gaps in network zoning, default passwords, dated patch updates.

Electric Terrorism: % of attacks by grid components targeted (1994-2004)

Case Study - GSTN

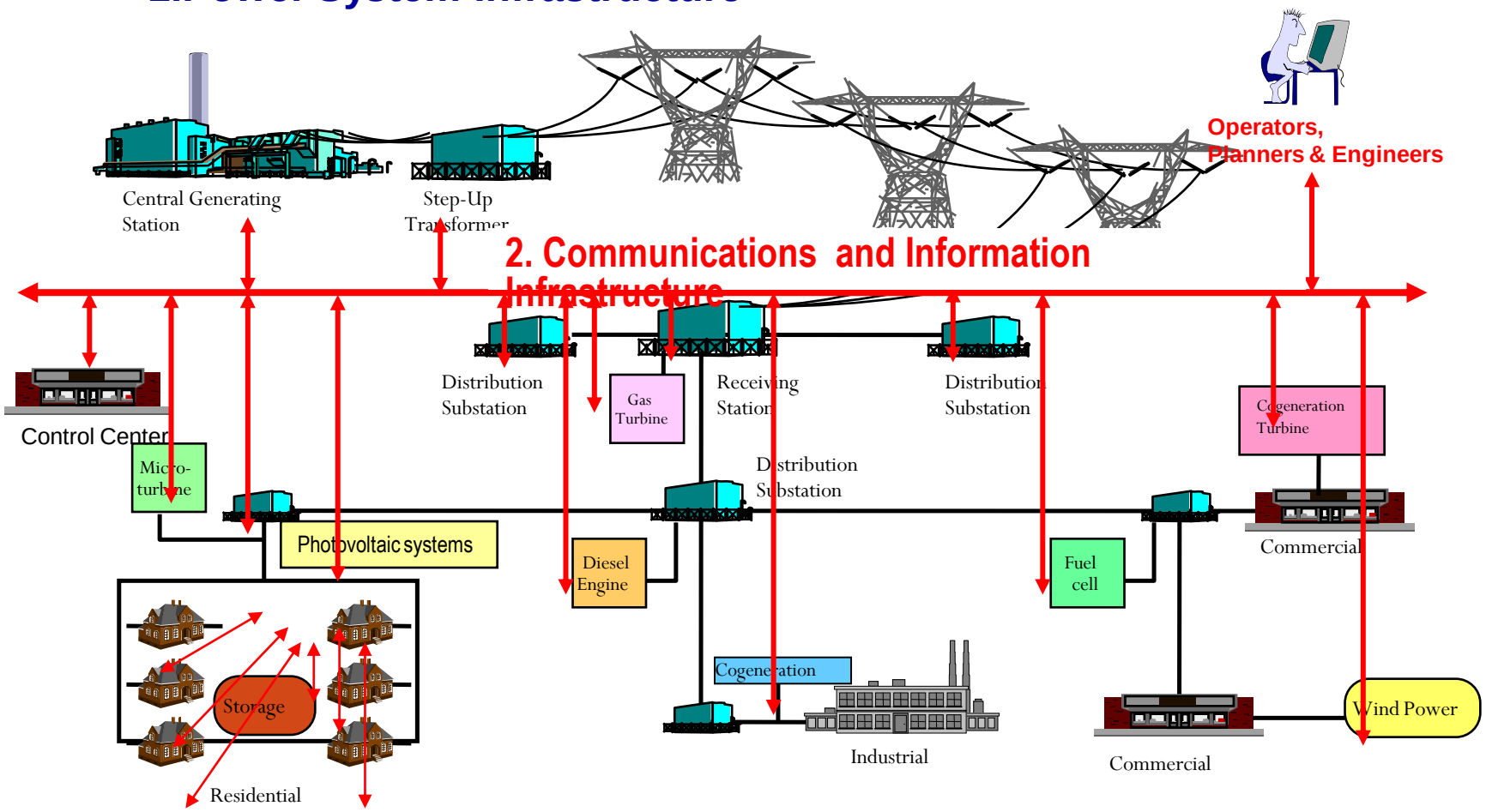


Case Study in ISMS

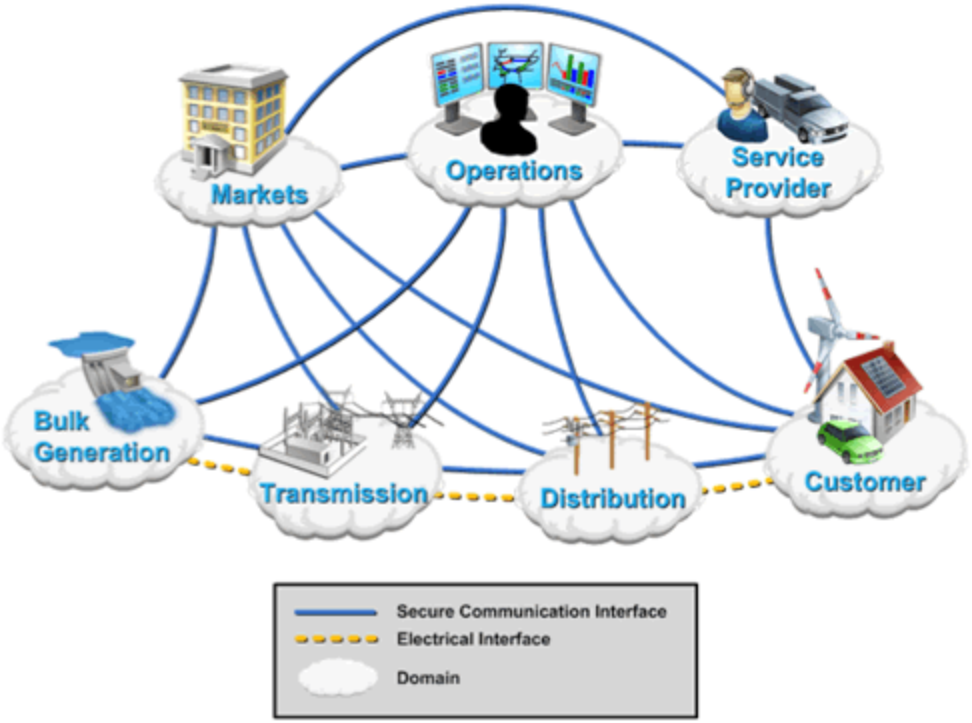


To maintain power system reliability, need to manage both the Power System Infrastructure and its supporting Information Infrastructure

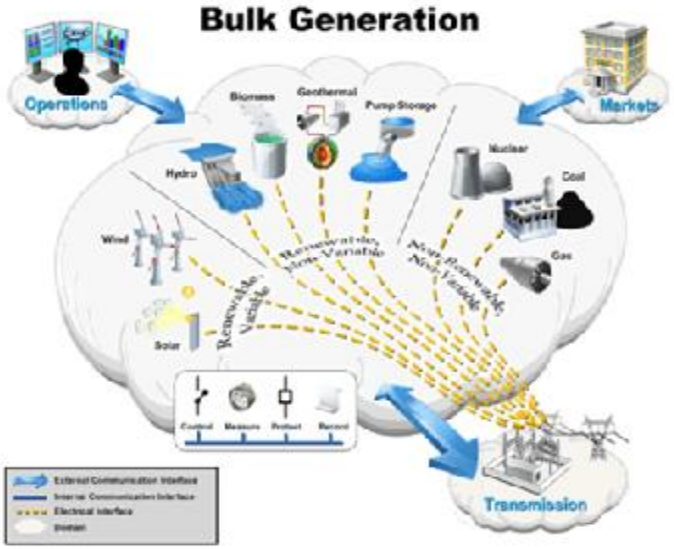
1.Power System Infrastructure

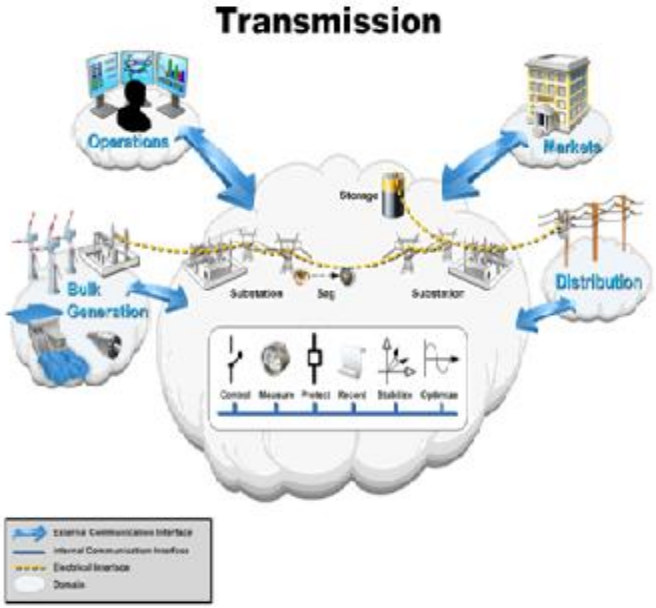


Case Study in ISMS

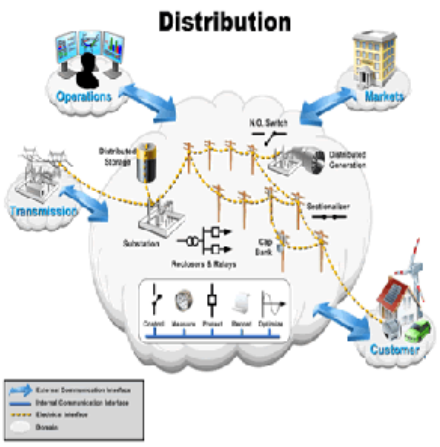


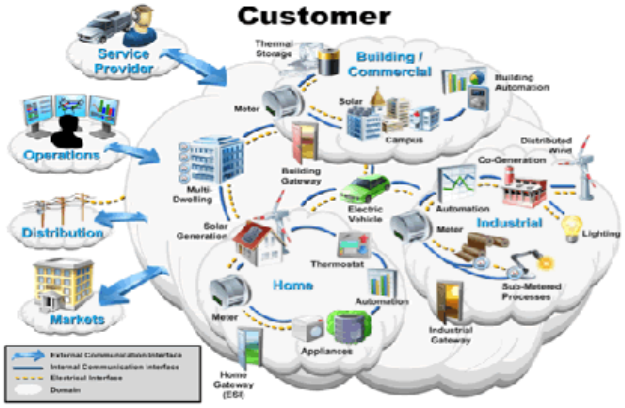
ISMS - Best Practices





ISMS - Best Practices





Thank you

Srinivas.bhoosarapu@gmail.com
Mobile : 9154035731

Thank you