SATURN 2019

DOMAIN DRIVEN DESIGN OF BIG DATA SYSTEMS

Babu Samuel

SIEMENS Healthineers

Carnegie Mellon University
Software Engineering Institute

DOMAIN DRIVEN DESIGN OF BIG DATA SYSTEMS

Big Data Design (Domain Driven Approach)

Architecturally Significant Requirements for Big Data Systems

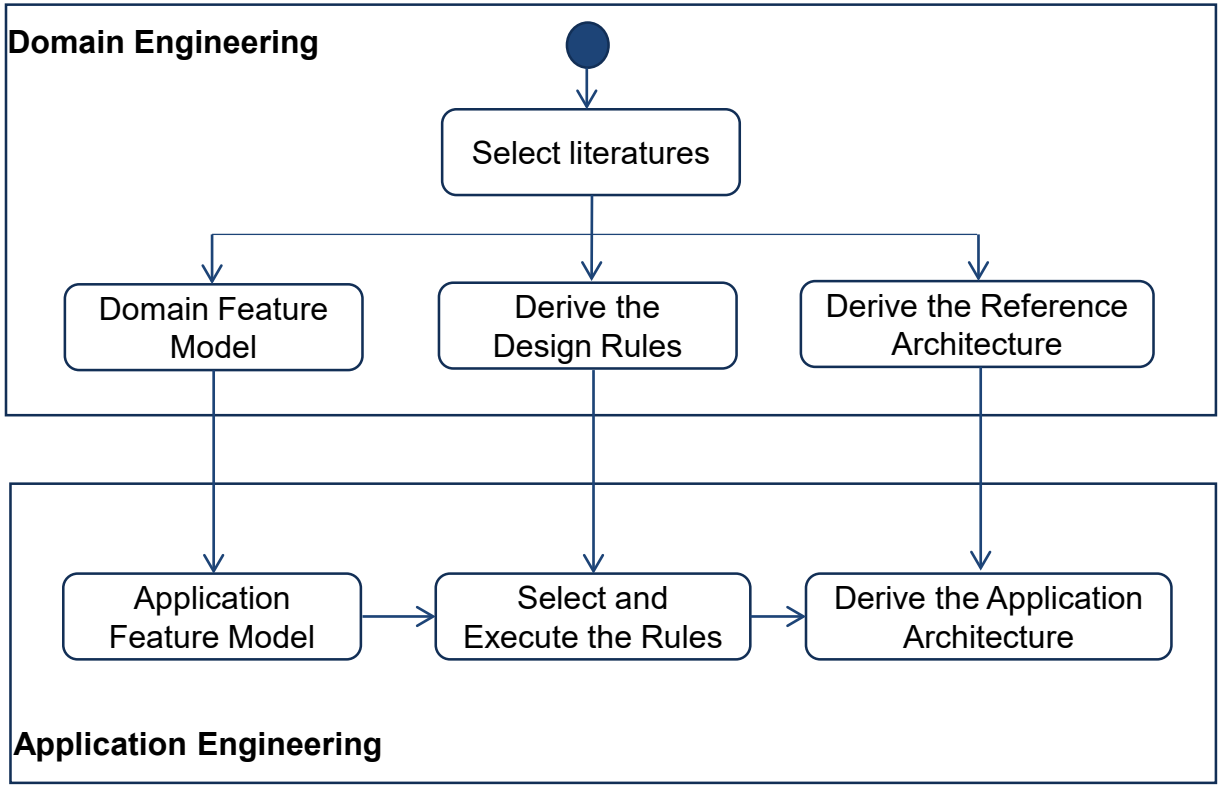
Conclusion

Big Data Design (Domain Driven Approach)

Based on

- Feature model (defines common and variant features of a domain)
- Design Rules (requirements and use-cases)
- Non-functional requirements (quality attributes)

Big Data Design (Domain Driven Approach)



Reference Architecture

Reference Architecture is an **abstraction** of ‘real’ architectures.

According to Muller,

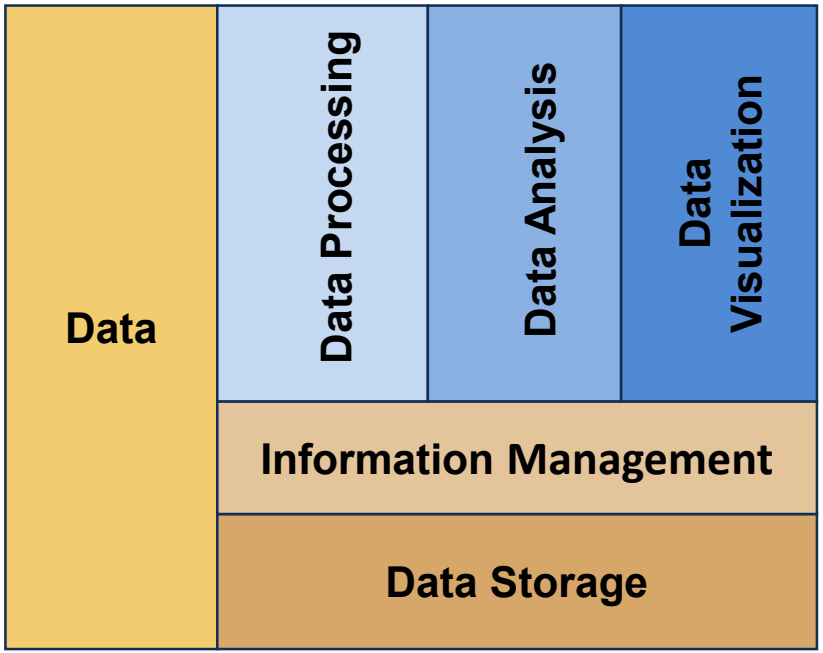
“architects can use a reference architecture as **guidance to create a concrete architecture** for their organization, business context and technology”

Reference Architecture

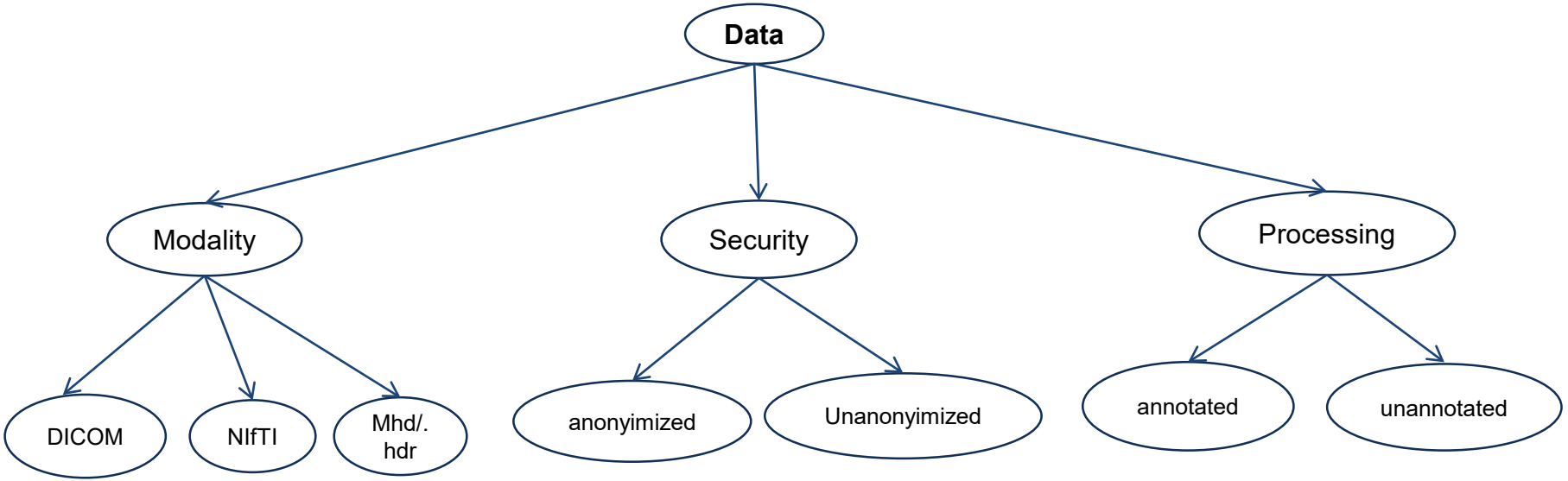
Important Observations:

1. Core of a big data architecture
   - Parallel batch processing engine (Hadoop MapReduce)
   - Distributed File system (HDFC)
   - NoSQL database (Hbase)
2. Other components
   - Querying engine
   - Predictive Analytics engine
   - Statistical Analysis / ML engine
   - Data importing engine
   - Real time / complex event processing engine
3. Architecture principles
   - Loose coupling
   - Cloud computing
   - Scalability
4. Best practices
   - Data pipeline approach
5. More consensus about the hardware and software than the principles and best practices

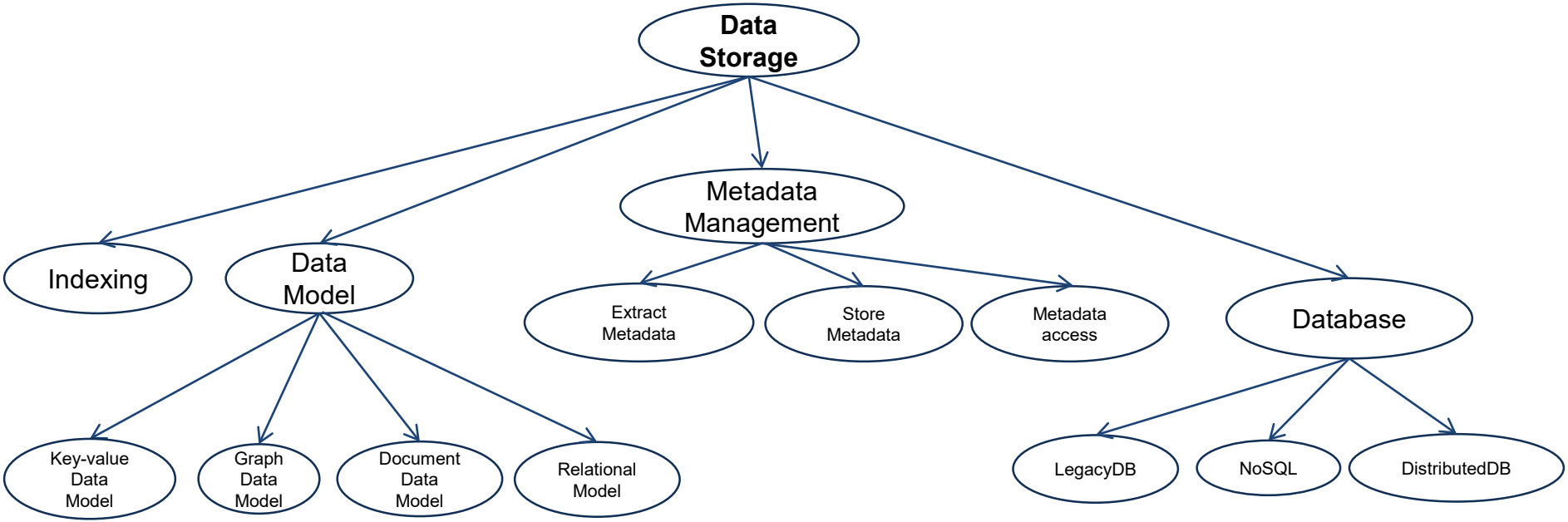
Big Data Design (Domain Driven Approach)



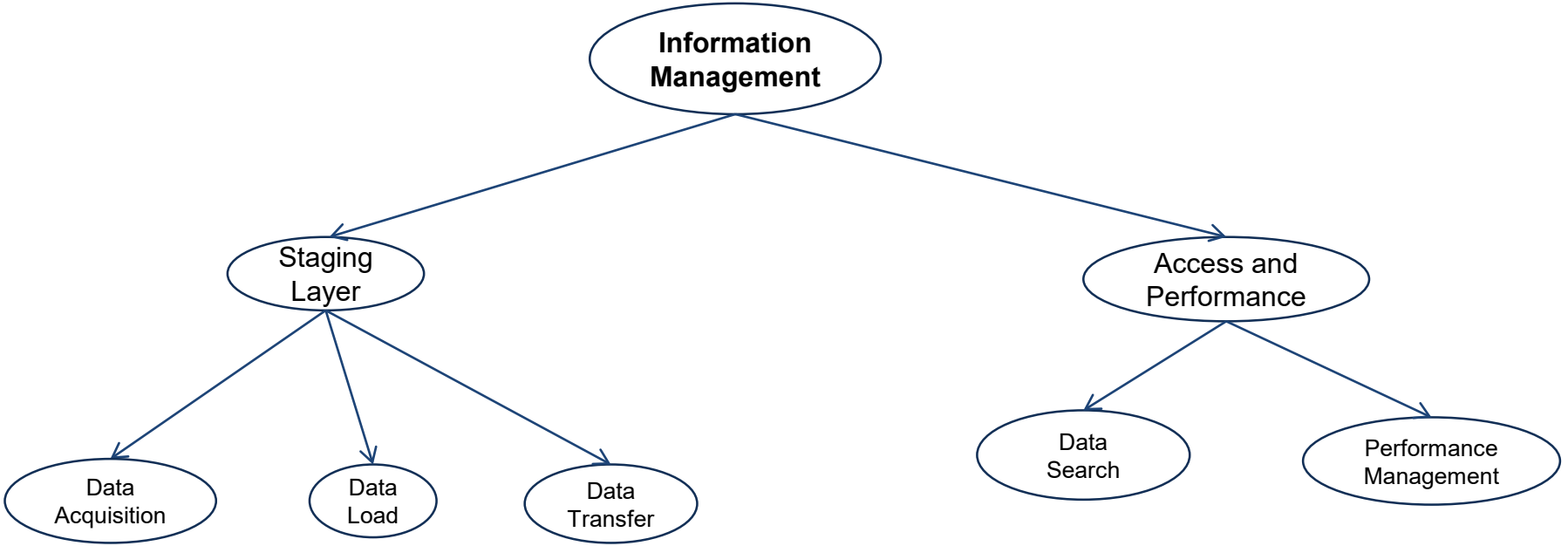
Big Data Design (Domain Driven Approach)



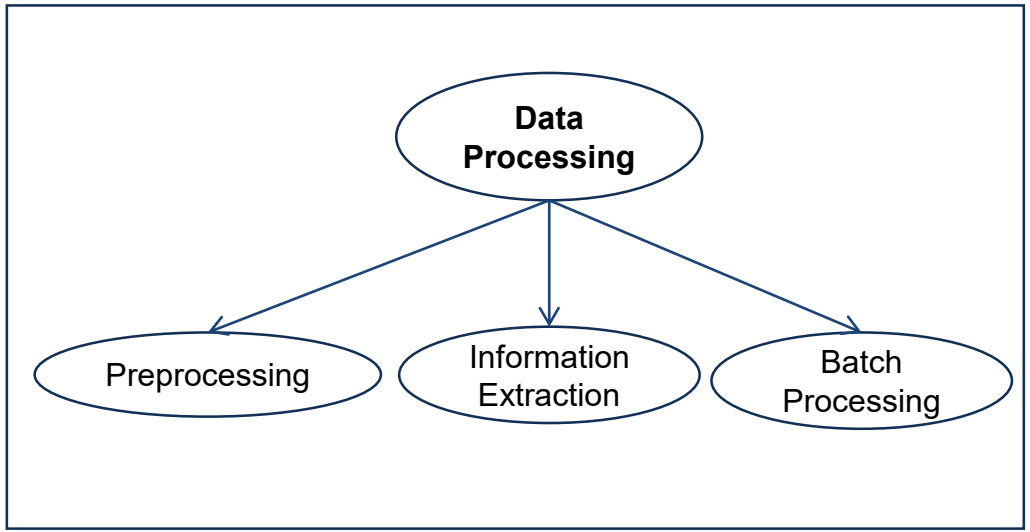
Big Data Design (Domain Driven Approach)

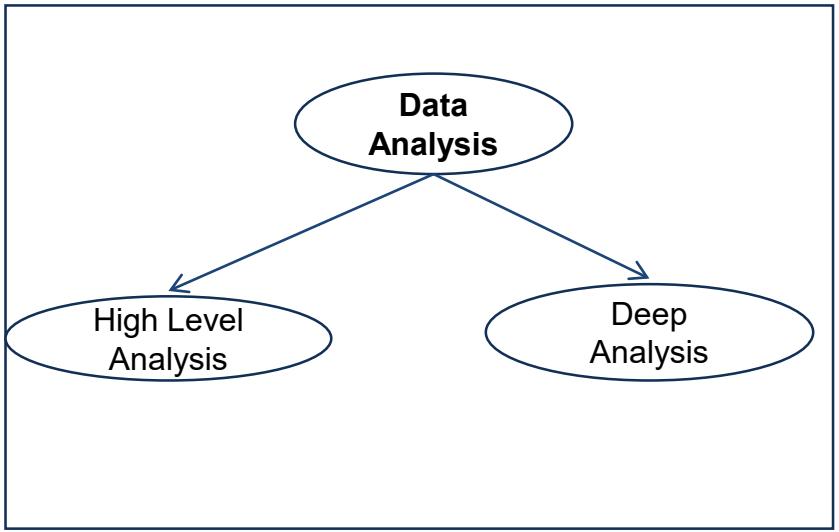


Big Data Design (Domain Driven Approach)

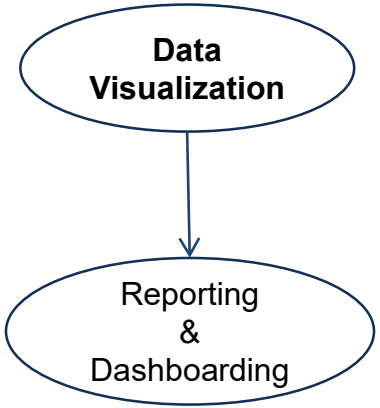


Big Data Design (Domain Driven Approach)





Big Data Design (Domain Driven Approach)



Design Rule Modeling

According to Design Rule Definition Language

“**if <feature> is selected then [action] <feature> on node [name]**”

DR1: If ***MULTIPLE DATA TYPES*** is selected then load ***METADATA MANAGEMENT*** on component DATA STORAGE

DR2: If ***KEY-VALUE DATA MODEL*** is selected then load ***NoSQL*** on component DATA STORAGE

DR3: If ***STREAMING*** is selected then load ***DATA ACQUISITION*** on component INFORMATION MANAGEMENT

Architecturally Significant Requirements for Big Data Systems

- The **Five Vs** - **V**olume, **V**elocity, **V**ariety, **V**eracity and **V**alue
- **Hyper-scalability** - able to support an exponential growth in computing requests even though the available resources only grow linearly
- **Resilience** – need to identify attacks and react to those circumstances
- **Privacy and Security** – ability to reliably control, monitor and audit who can perform what actions on the resources

Conclusion

- Select the **feature model** which covers the common and variant features of a broad set of different applications
- Integrate the **design rules** in the overall process in deriving a feasible big data architecture
- Use the **reference architecture** for deriving concrete application architectures

Serial No.	List of Papers to derive the reference architecture of Big Data Systems
1	B. Geerdink, “A Reference Architecture for Big Data Solutions”
2	C. Ballard et al., Information Governance Principles and Practices for a Big Data Landscape. IBM Redbooks, 2014
3	D. Chapelle, “Big Data & Analytics Reference Architecture.” An Oracle White Paper (2013)
4	M. Maier, A. Serebrenik, and I.T.P. Vanderfeesten, “Towards a Big Data Reference Architecture.” (2013)
5	NIST Big Data PWG, Draft NIST Big Data Interoperability Framework: Volume 6, Reference Architecture (2014)
6	N. Marz, and J. Warren, “Big Data: Principles and best practices of scalable realtime data systems.” Manning Publications Co. (2015)
7	Oracle, Information Management and Big Data A Reference Architecture, An Oracle White Paper, February (2013)
8	P. Pääkkönen, and D. Pakkala, “Reference Architecture and Classification of Technologies, Products and Services for Big Data Systems.” Big Data Research (2015)
9	S. Soares, “Big Data Governance.” Information Asset, LLC (2012)