Design of Big Data Reference Architectures for Use Cases in the Insurance Sector

Vladimir Elvov, Bachelor's Thesis – Final Presentation, 05.03.2018

Chair of Software Engineering for Business Information Systems (sebis)
Faculty of Informatics
Technische Universität München
wwwmatthes.in.tum.de

Outline


Motivation
Approach & Research Questions
Use Cases
Big Data Architecture
Evaluation & Conclusion

Motivation

2015: Only 5 out of 30 DAX companies have Big Data applications deployed*

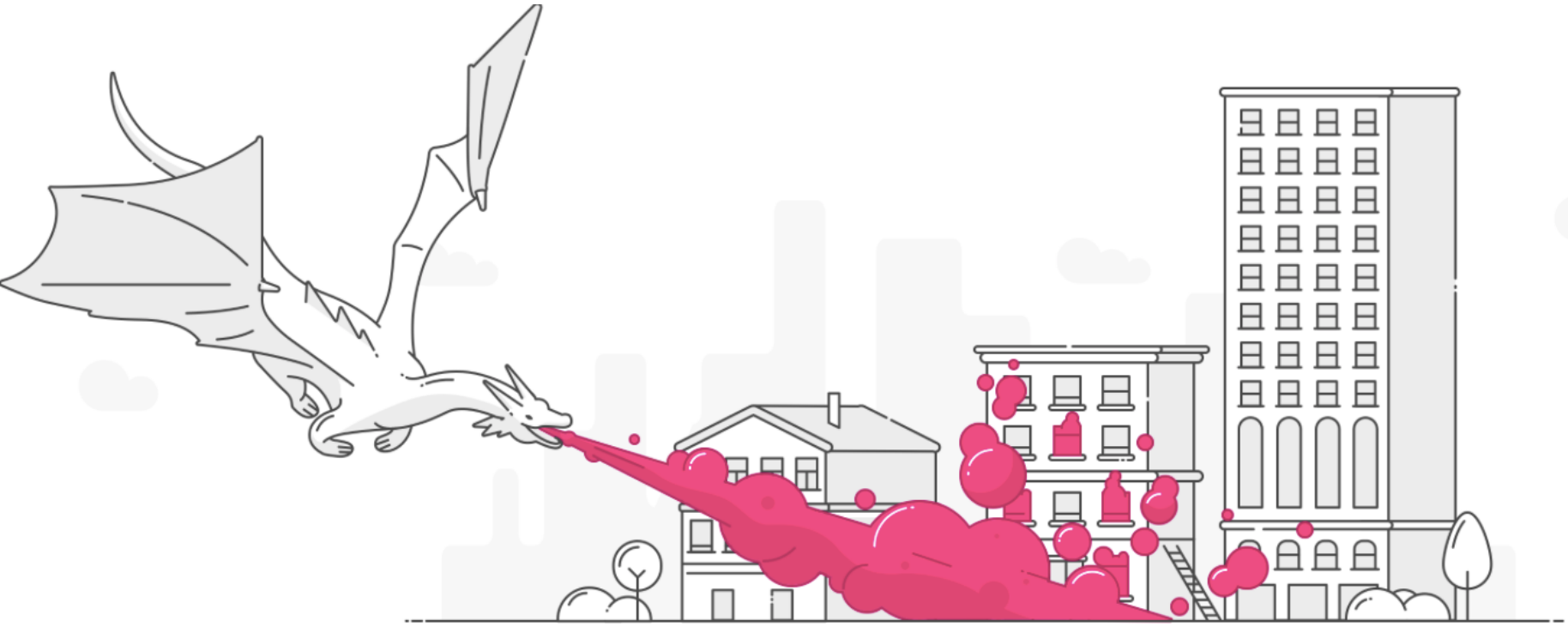
Data in silos, products too complex – **few insights about customers**

90 seconds to get insured – **3 minutes** to get paid

*) Matthes, F. & Kazman, E.

Motivation: Insurtech in action

Lemonade



Approach & Research Questions

RQs

RQ1

What are possible Big Data Use Cases in the insurance sector and which ones do have the highest potential?

RQ2

Which requirements have to be fulfilled in order to implement these Use Cases?

RQ3

What does a Big Data Reference Architecture look like in order to operationalize the Use Cases?

Methodology

Literature review

14 Interviews

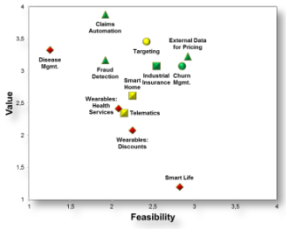
Literature review

Literature review

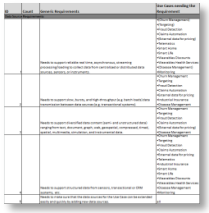
6 Interviews

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

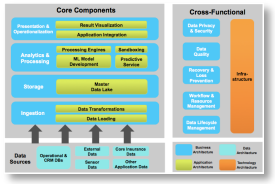
Results



Use Cases



Requirements



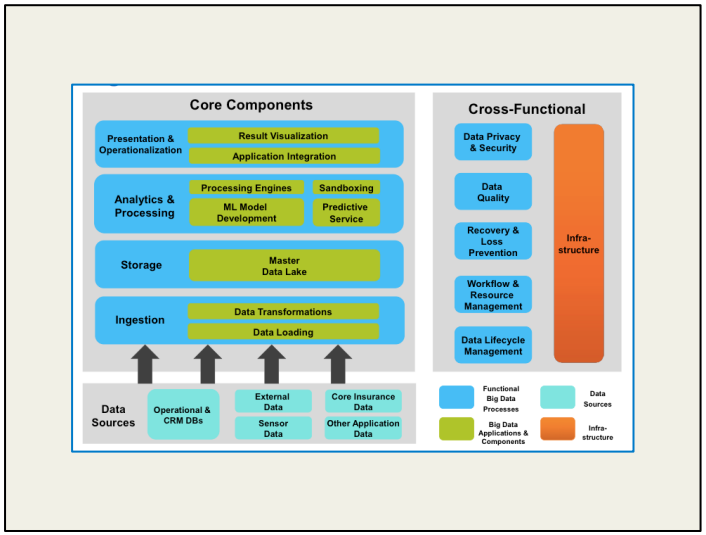
Reference Architecture

Use Cases

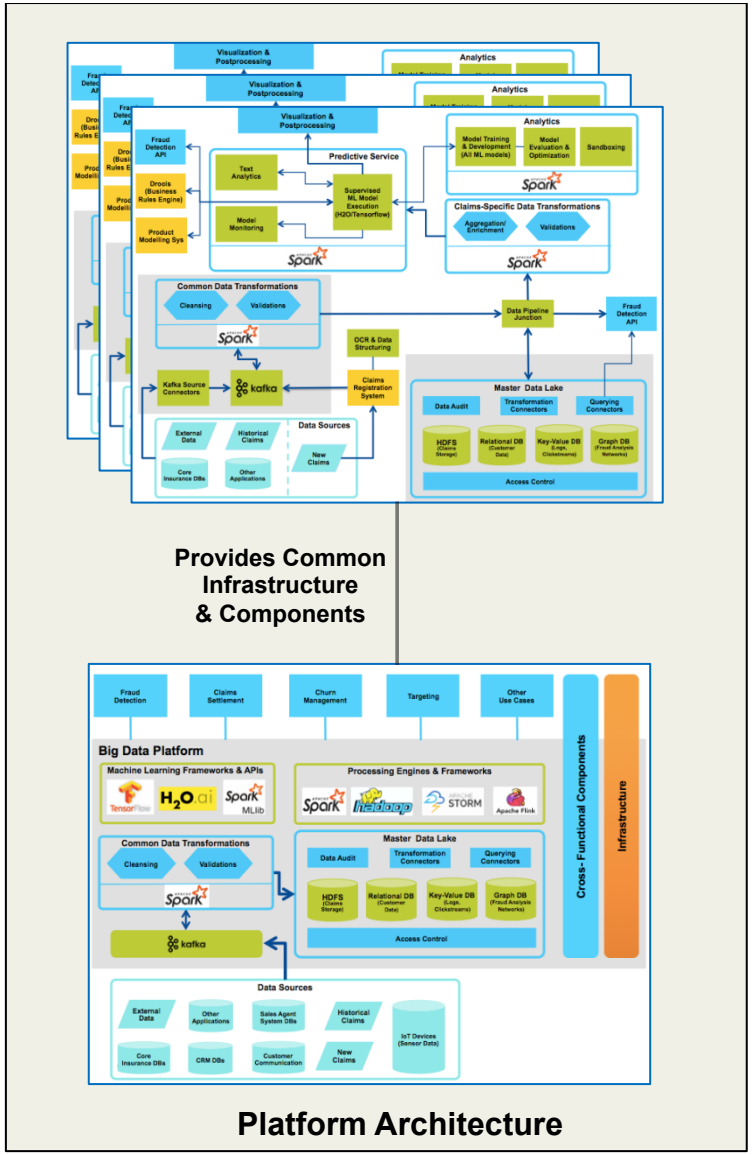
1. **Customer Analytics**
   1) Churn Detection and Management
   2) Targeting
2. **Internal Processes**
   1) Fraud Detection
   2) Claims Automation
   3) External Data for optimized Pricing and Risk Assessment
   4) Analysis of the Enterprise Architecture and Business Processes based on Monitoring Data
3. **IoT in P&C**
   1) Telematics
   2) Industrial Insurance
   3) Smart Home
4. **Smart Health & Smart Life**
   1) Health Insurance based on Wearables Data (Discounts)
   2) Health Services based on Wearables Data
   3) Disease Management
   4) Sensor-based Services in Life Insurance

Architecture

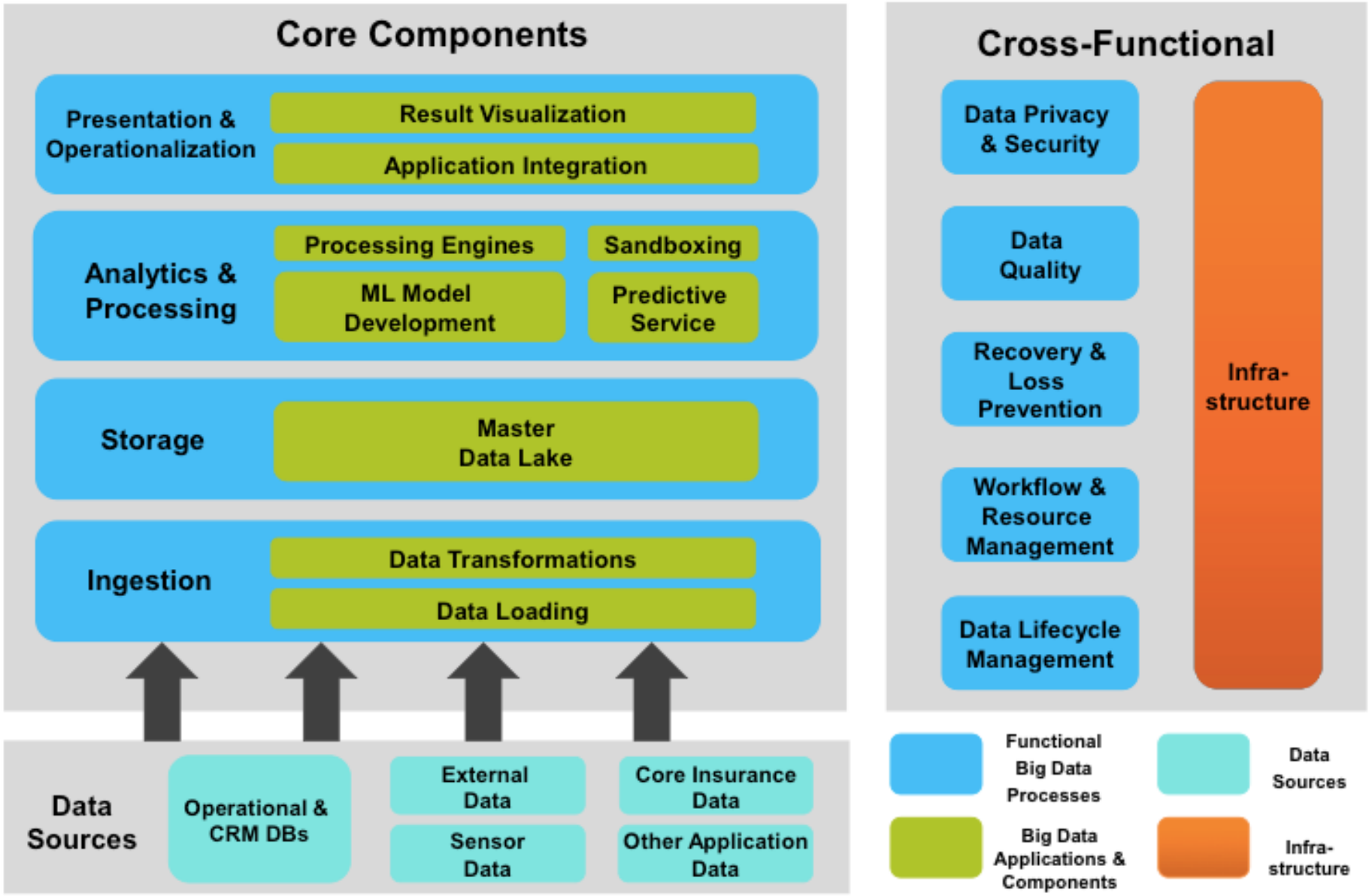
Reference Architecture



Solution Architecture for Use Cases

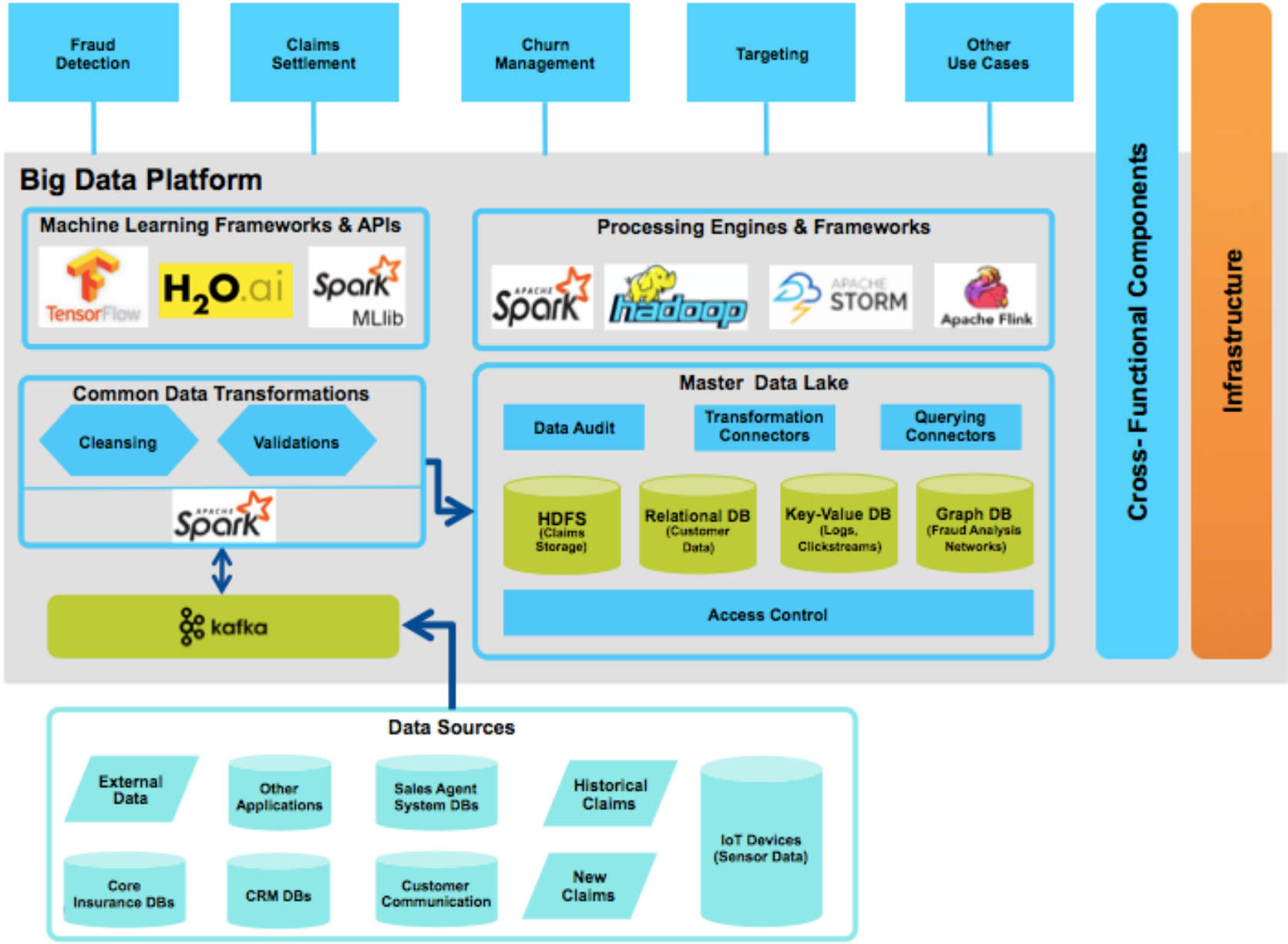


Big Data Reference Architecture

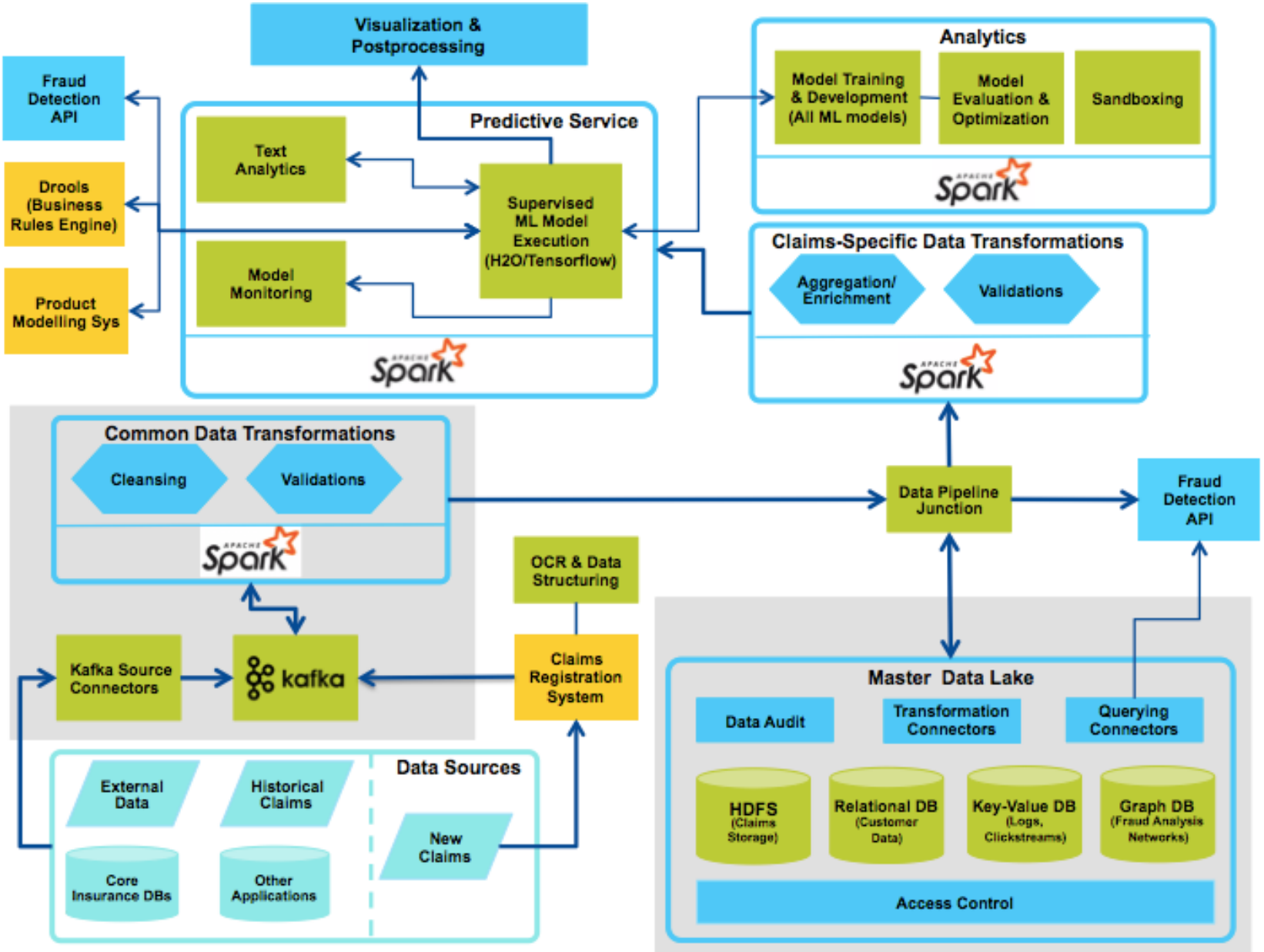


One Big Data Platform to rule them all, one Kafka to find them, one Data Pipeline to bring them all and in the Data Lake bind them

Big Data Platform Architecture



Solution Architecture for Claims Settlement



Evaluation & Conclusion

- **Expert Interviews – general feedback:**
  - ❖ All required capabilities for Use Cases offered by Reference Architecture
  - ❖ Solution Architecture is a good example for using the Reference Architecture
  - ❖ Additional components suggested:
    - ➢ Platform Approach
    - ➢ Separate ML model training and execution areas

Expert Interview Partners

Google
- **Cloud Solution Architect**

Microsoft
- **Cloud Solution Architect**

cloudera
- **Chief Architect**

Allianz
- **Senior Enterprise Architect**
- **Program Lead Data Lake**
- **Senior Data Scientist**

Any Questions?

Sources (Selection)

Matthes, F. & Kazman, R. (2015): *Demystifying Big Data Adoption: Beyond IT Fashion and Relative Advantage.*

Google & Bain&Company (2016): *Digitalisierung der Versicherungswirtschaft: Die 18-Milliarden-Chance.*

Marr, B. (2015): *Using SMART Big Data, Analytics and Metrics To Make Better Business Decisions and Improve Performance.*

National Institute on Standards and Technologies (2015): *NIST Big Data Interoperability Framework: Volume 6, Reference Architecture.*

National Institute on Standards and Technologies (2015): *NIST Big Data Interoperability Framework: Volume 5, Architectures Whitepaper Survey*

Fox, G. & Chang, W. (2015): *Big Data Use Cases and Requirements.*

Clarke, R. & Libarikian, A. (2014): *Unleashing the value of advanced analytics in insurance.*

Digital McKinsey (2017): *Digital disruption in insurance: Cutting through the noise.*

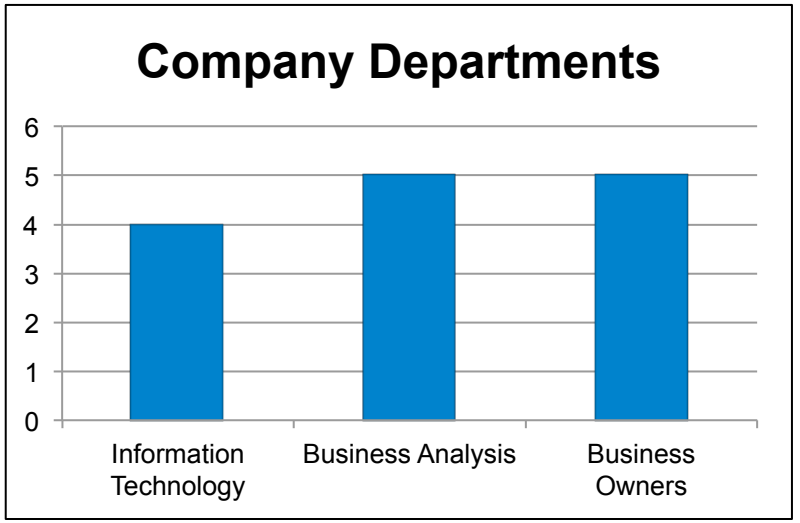
https://bigdatawg.nist.gov/

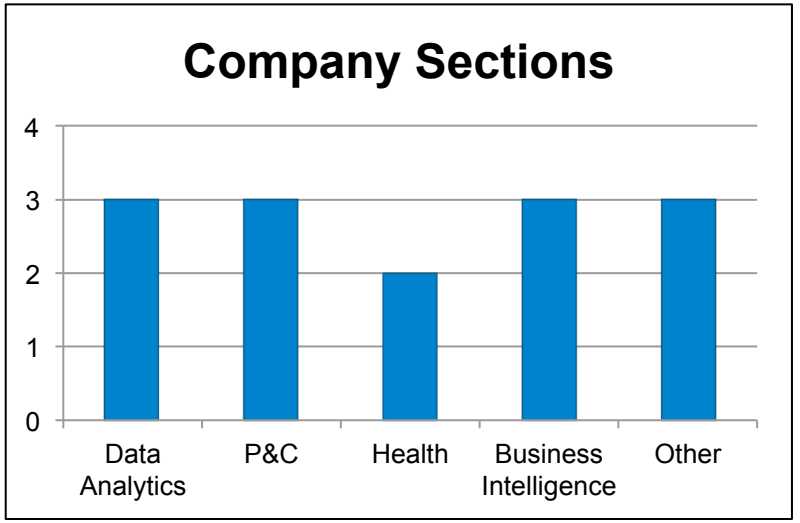
Backup slides

Use Case Evaluation Methodology

- **14** Interview partners
- 6 questions per 13 Use Cases
- **8 weeks** for interviews and result analysis
- **Questions**:
  - Use Case's Business Value on a scale from 1 to 4
  - Use Case's Feasibility on a scale from 1 to 4
  - Possible Risks in Implementation
  - Possible cooperation partners

➢ Results show average scores for these categories





Requirements Analysis

- Common template for all Use Cases
- NIST-based
- Analysis of Use Case specific requirements
- All relevant aspects covered:
  - Data sources
  - The 4 V's
  - Data processing & analysis
  - Security & Privacy
  - Business requirements

- 33 generic requirements in total
- Each requirement mapped against the Use Case that needs it

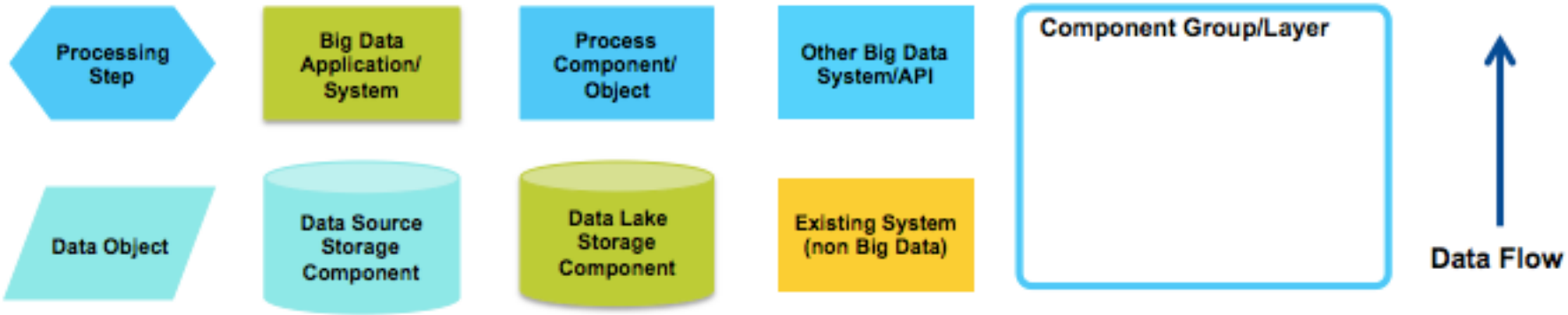
Requirements Template

Use Case Title		
Description		
Big Data Characteristics	Data Source	
	Volume	
	Velocity	
	Variety	
Big Data Science	Veracity and Data Quality	
	Visualization	
	Data Types	
	Data Analytics	
Security and Privacy	Personally Identifiable Information (PII) used?	
	Highly sensitive data used?	
	Governance, Compliance & Audit	
Organizational & Business Requirements	Knowhow	
	External Partners	
	Other business challenges	
Other Big Data Challenges		

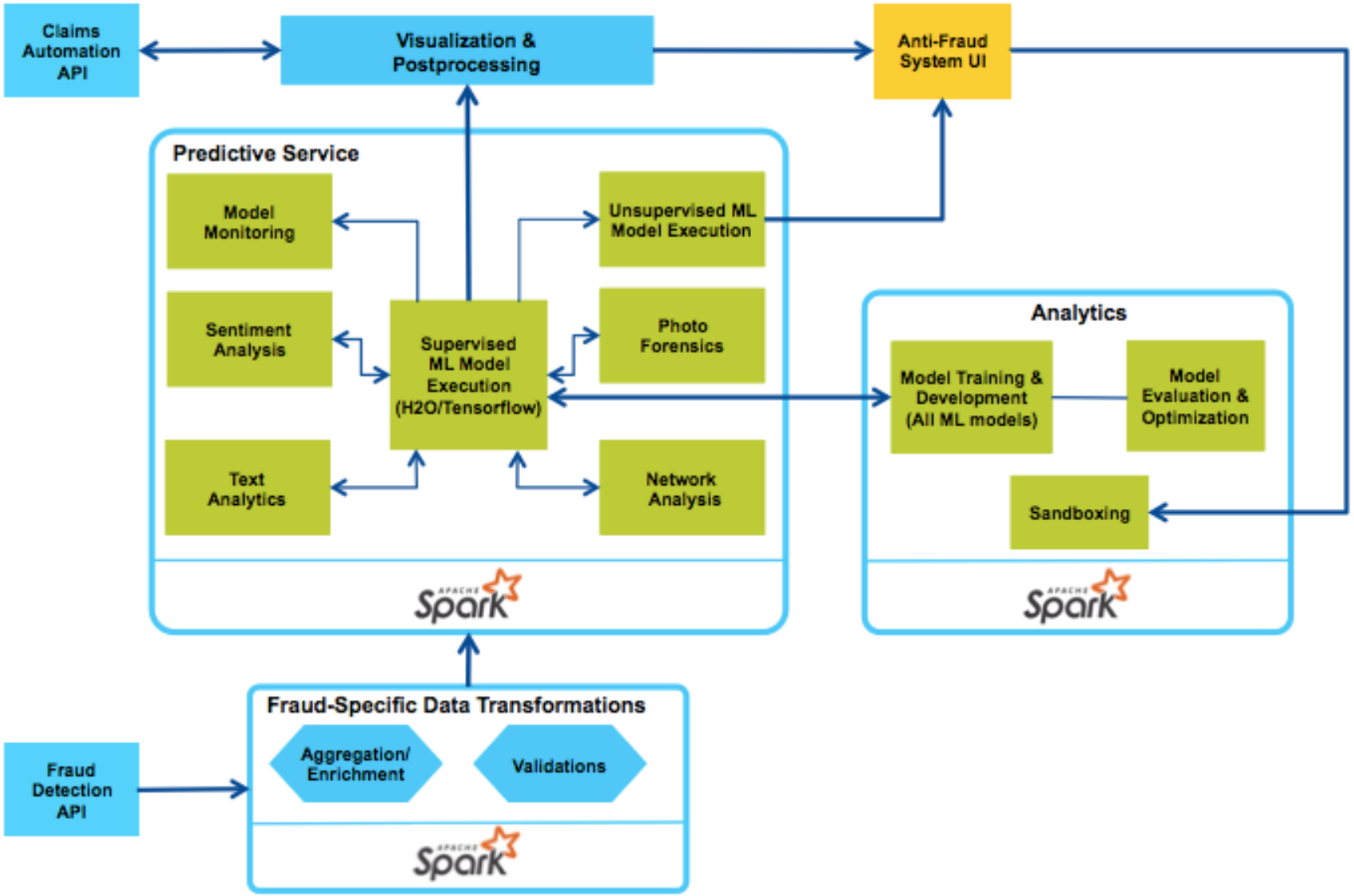
Generic Requirements

ID	Count	Generic Requirements	Use Cases needing the Requirement
Data Source Requirements			
1		Needs to support reliable real time, asynchronous, streaming processing/loading to collect data from centralized or distributed data sources, sensors, or instruments.	•(Churn Management) •(Targeting) •Fraud Detection •Claims Automation •(External data for pricing) •Telematics •Smart Home •Smart Life •Wearables Discounts •Wearables Health Services •(Disease Management) •Monitoring
2		Needs to support slow, bursty, and high-throughput (e.g. batch loads) data transmission between data sources (e.g. transactional systems).	•Churn Management •Targeting •Fraud Detection •Claims Automation •External data for pricing •Industrial Insurance •Disease Management
3		Needs to support diversified data content (semi- and unstructured data) ranging from text, document, graph, web, geospatial, compressed, timed, spatial, multimedia, simulation, and instrumental data.	•Churn Management •Targeting •Fraud Detection •Claims Automation •(External data for pricing) •Disease Management
			•Churn Management

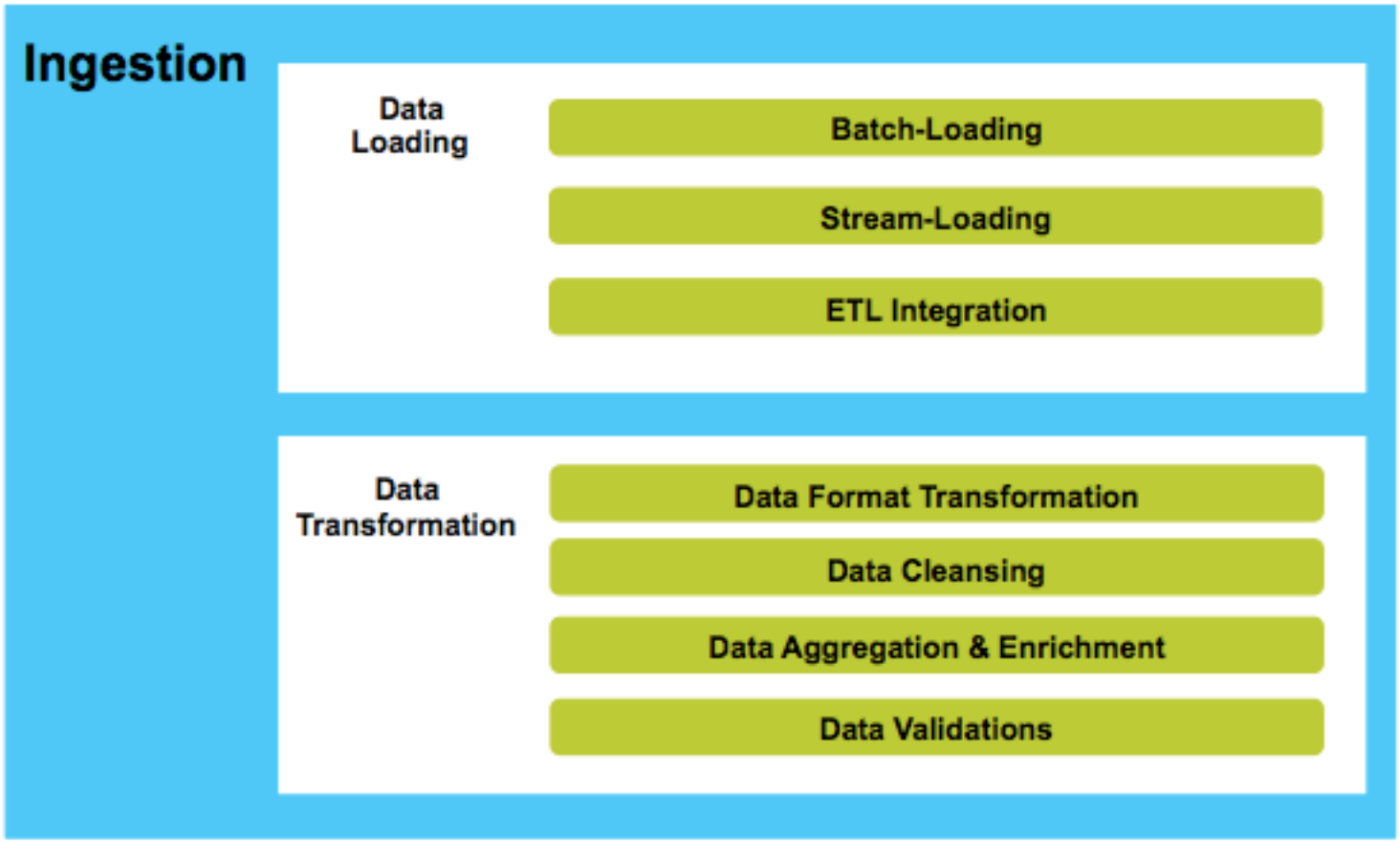
Notation Definition for Solution Architecture



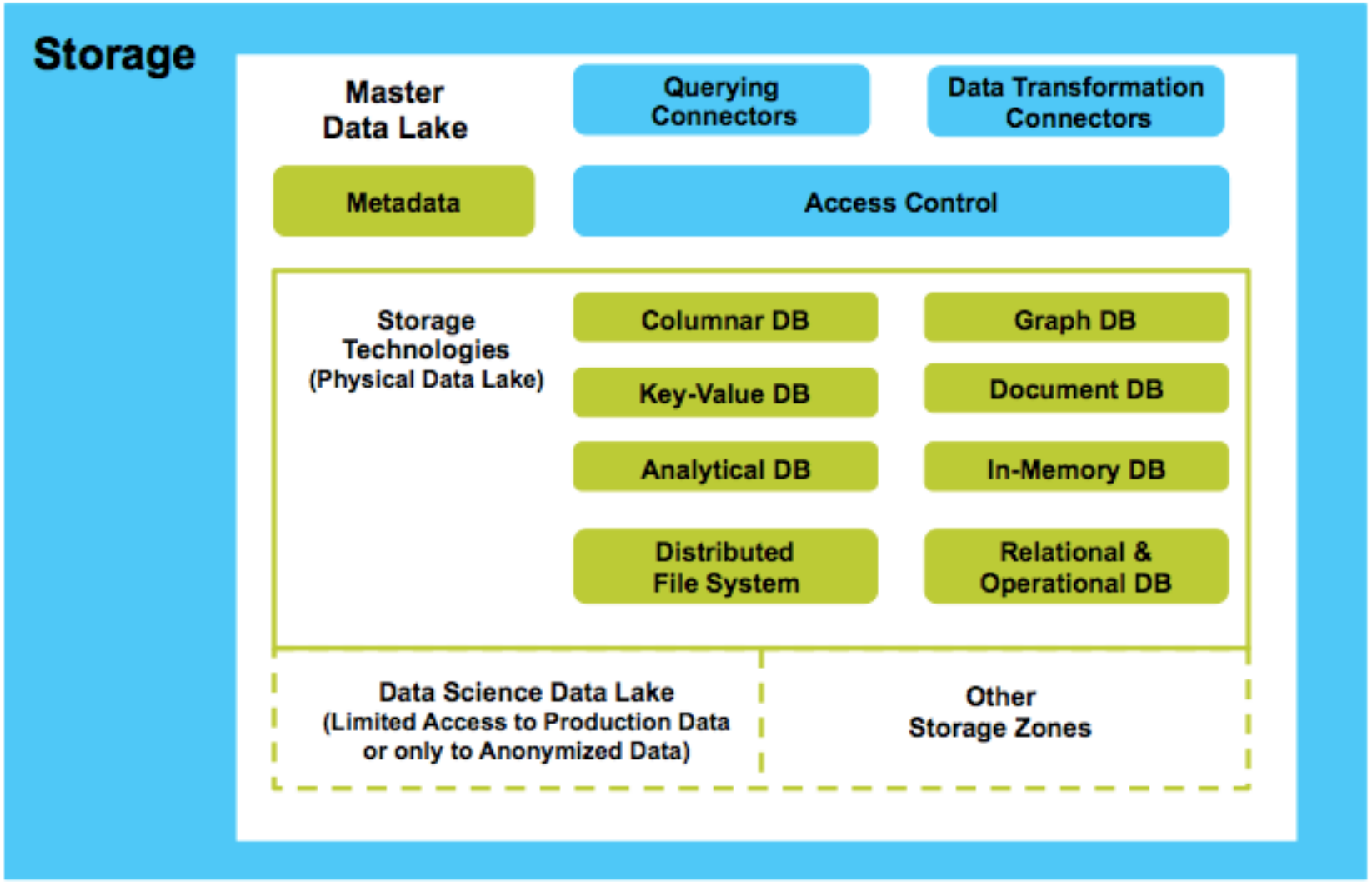
Solutions Architecture for Fraud Detection



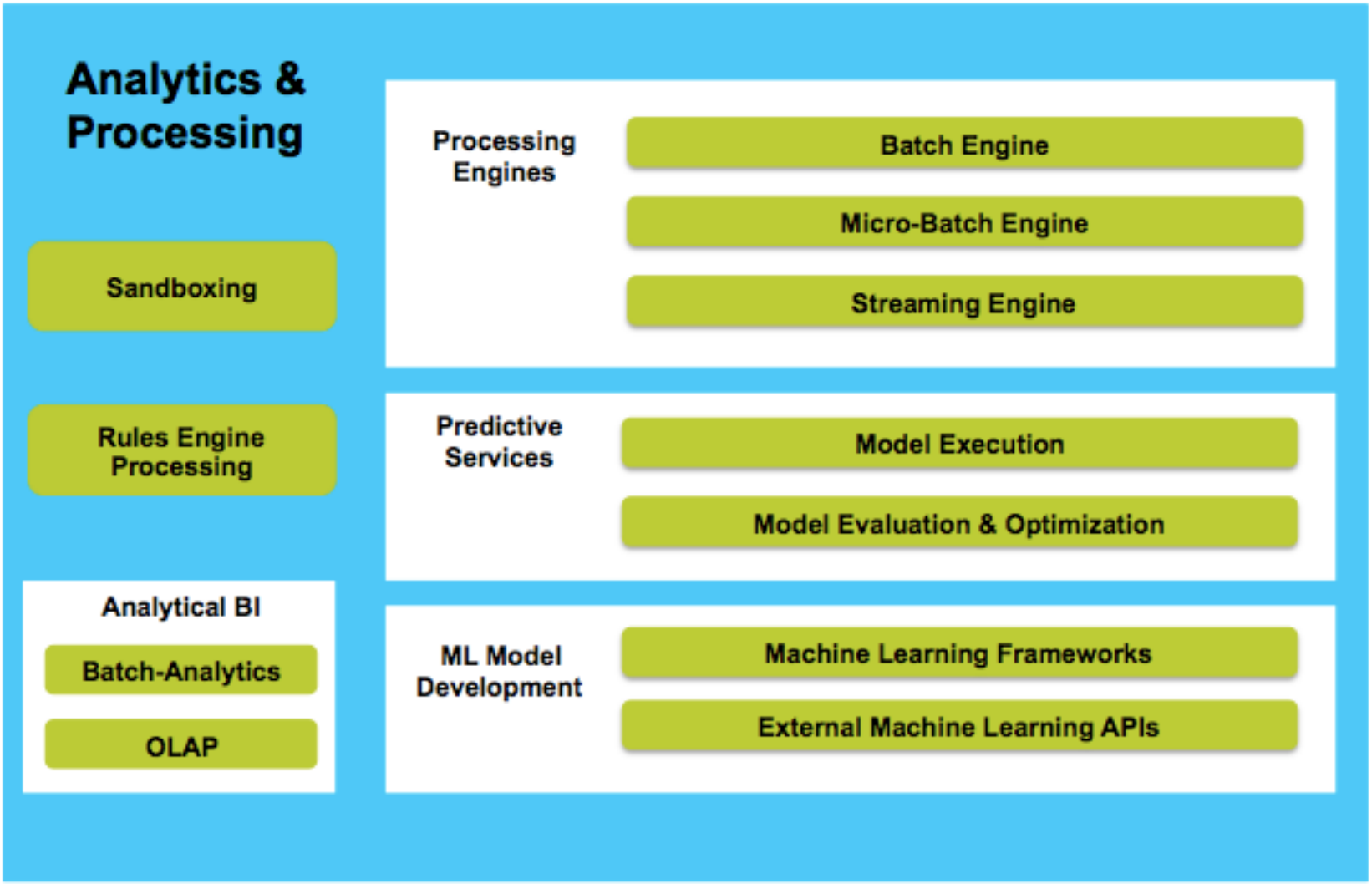
Level 2 – Ingestion



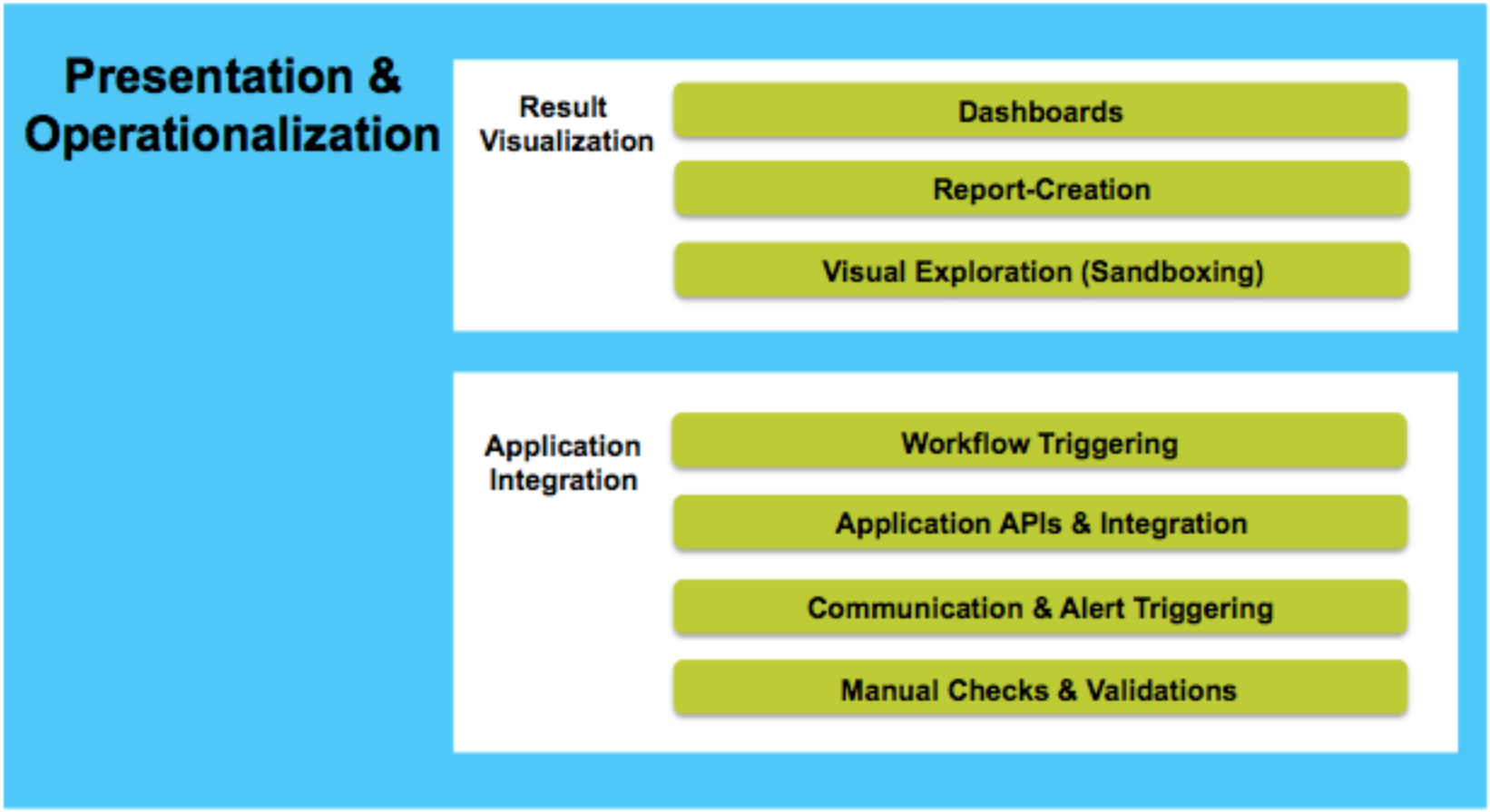
Level 2 – Storage



Level 2 – Analytics & Processing



Level 2 – Visualization & Postprocessing



Level 2 – Cross-Functional Components