Návrh digitální platformy pro zpracování dat z bezpečnostních incidentů

TARZAN

2017-06-05

Cíle projektu

- Zpracování různých typů dat z různých zdrojů
- Potenciálně velké množství dat
- Spolupráce analytiků
- Inteligentní analýza

Cíle projektu

- Zpracování různých typů dat z různých zdrojů
  Otevřený systém
- Potenciálně velké množství dat
  Škálovatelný systém
- Spolupráce analytiků
  Víceuživatelský systém, interaktivní analýza
- Inteligentní analýza
  Machine Learning

Klient-server aplikace, úložiště a výpočty v klusteru, jednotlivé funkce implementovatelné v libovolném jazyce či jako wrapper nad existujícími nástroji, volně integrované.

Investigation Framework

Collection

- Triage
- Sampling
- Selective Acquisition
- Intelligent Acquisition

Preservation

- Unstructured Data
- Structured Data
- Distributed Storage
- Indexing

Reduction

- Filtering
- Extraction
- Compression
- Transformation

Examination

- Enrich
- Label
- Search
- Clustering
- Outliers
- Correlation

Analysis

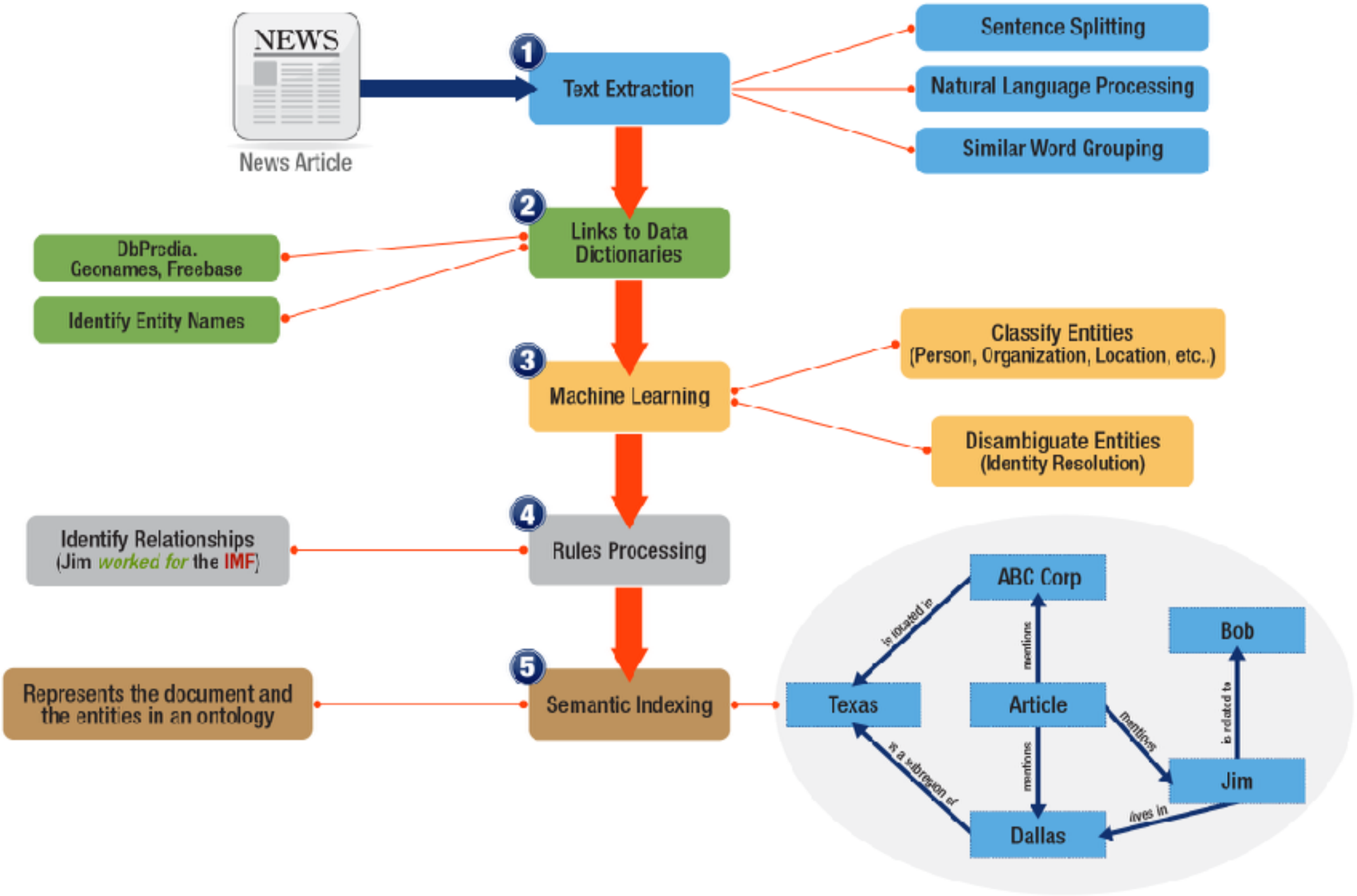
- Data Mining
- Semantic Analysis
- Ontology
- Drilling
- Functional
- Temporal

Presentation

- Visualization
- Reporting

http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=7740422

Zpracování nestrukturovaných dat



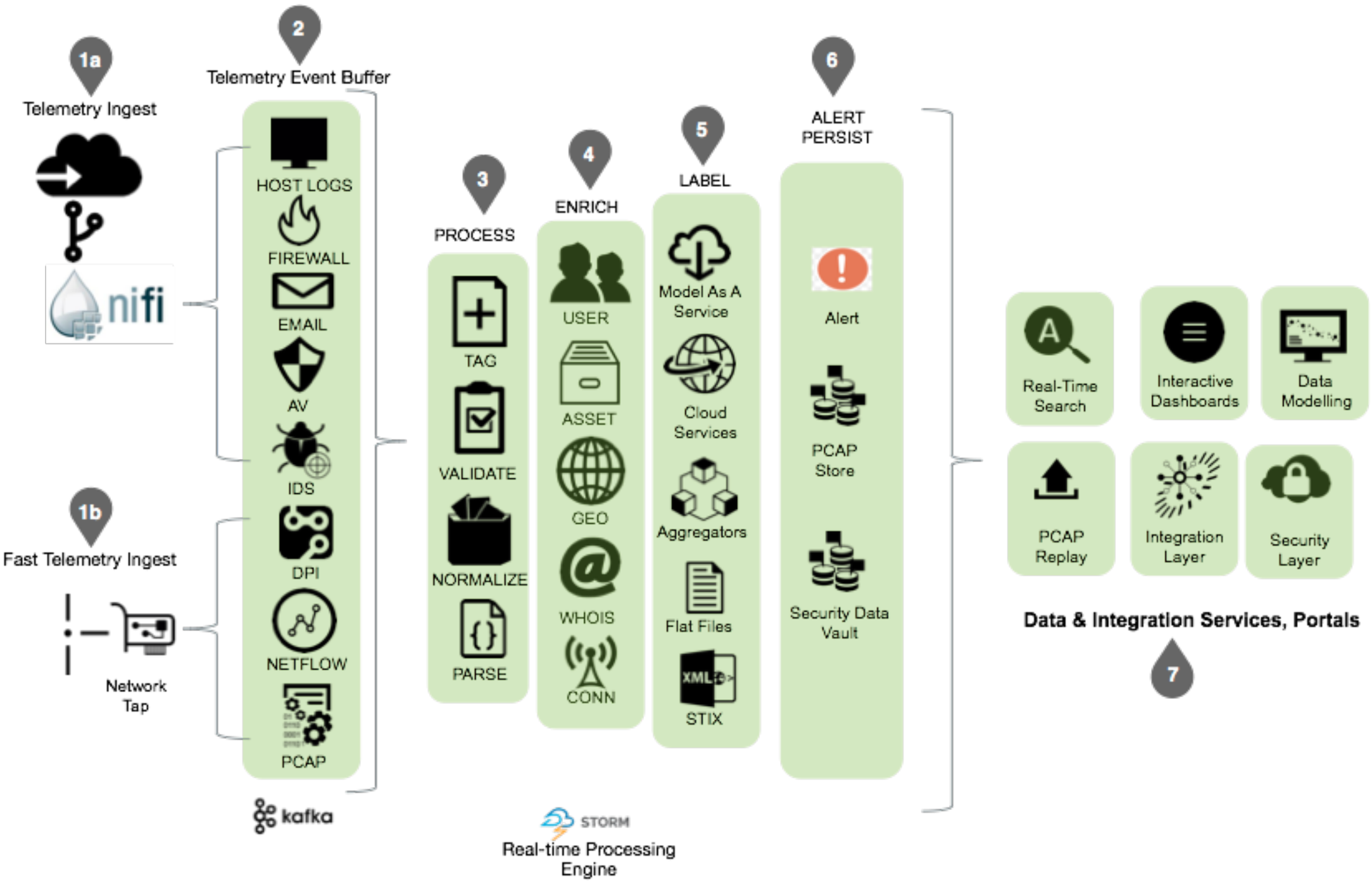
Existující nástroje

- E-Detective, NetDetector, NetIntercept, NetSleuth, Network Miner, Netfox Detective, …

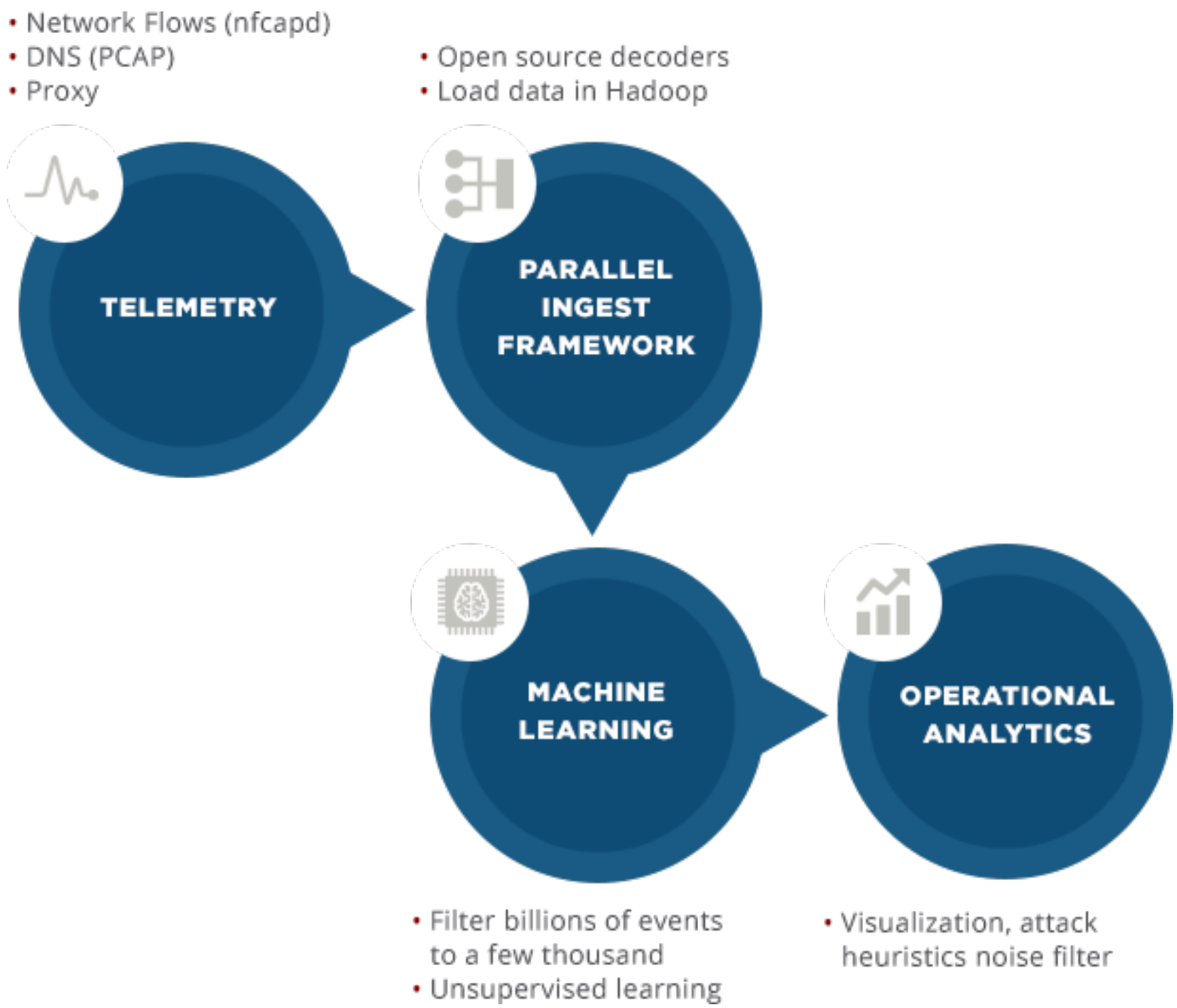
- XPLICO
  http://www.xplico.org/

- GRR
  https://github.com/google/grr
- Apache METRON
  http://metron.apache.org/
- APACHE SPOT
  http://spot.incubator.apache.org/

Apache METRON



Apache SPOT


• Network Flows (nfcapd)
• DNS (PCAP)
• Proxy
TELEMETRY
• Open source decoders
• Load data in Hadoop
PARALLEL INGEST FRAMEWORK
MACHINE LEARNING
OPERATIONAL ANALYTICS
• Filter billions of events to a few thousand
• Unsupervised learning
• Visualization, attack heuristics noise filter

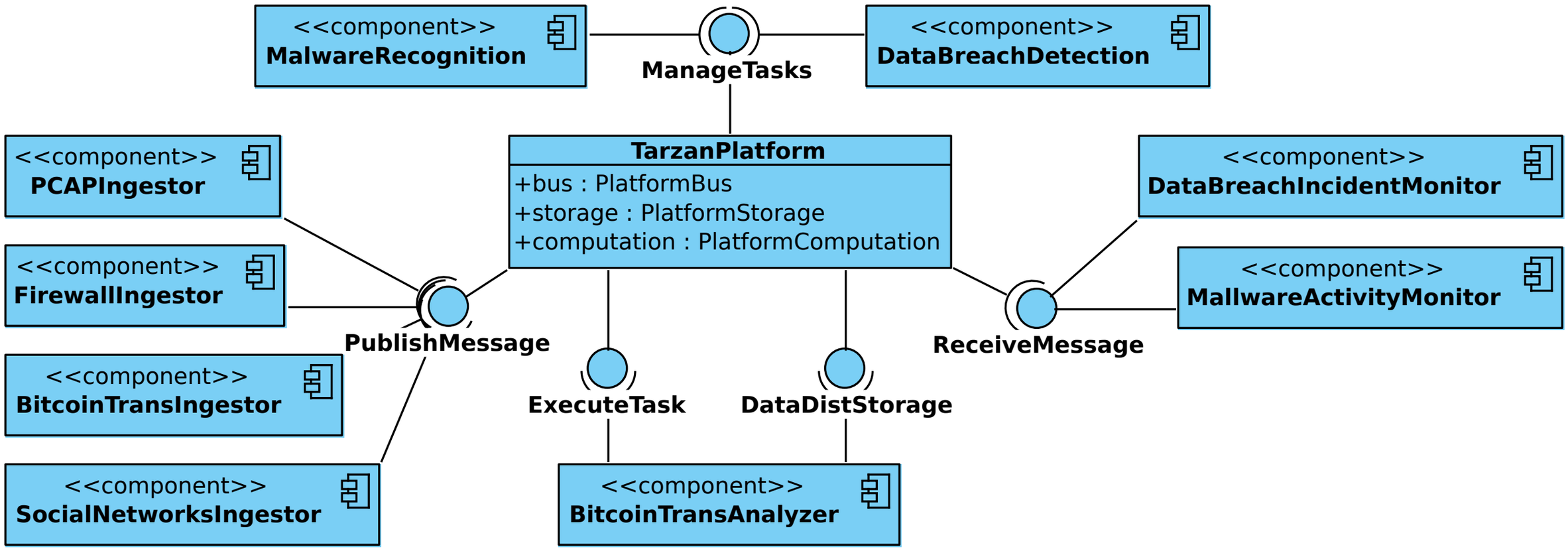
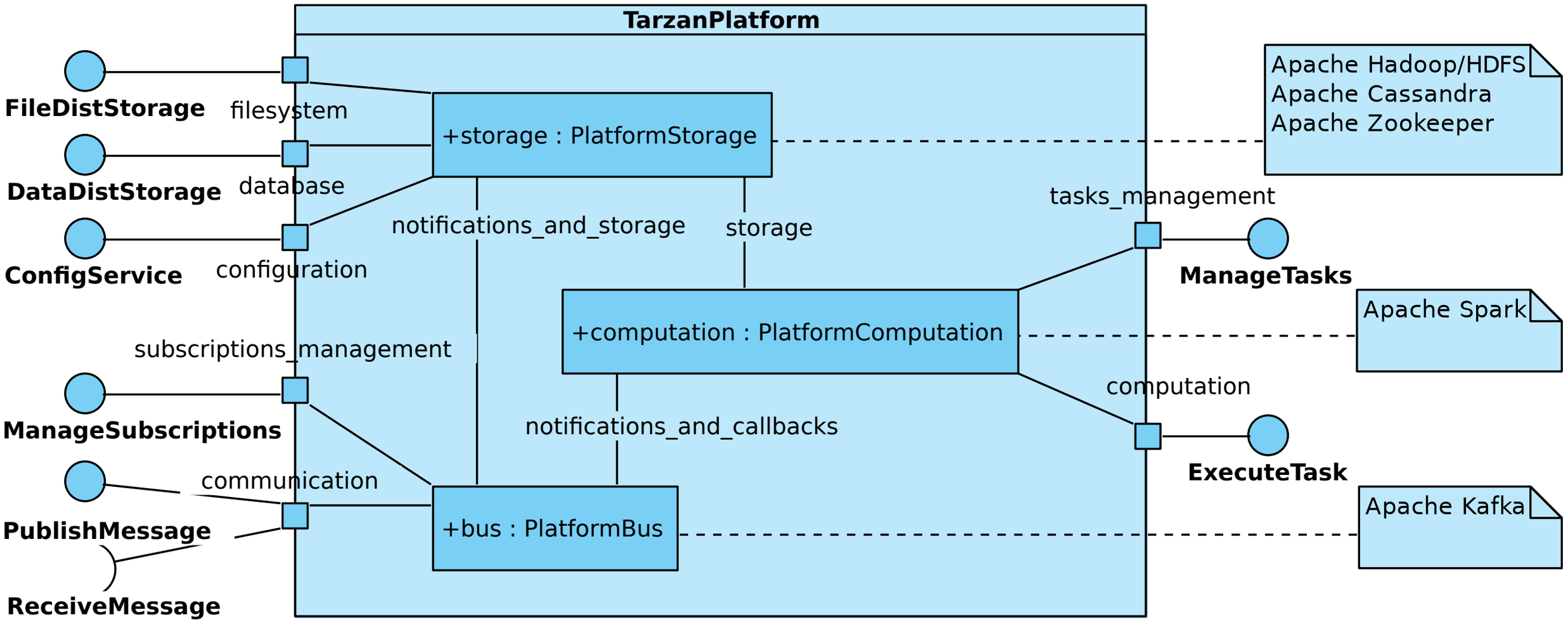
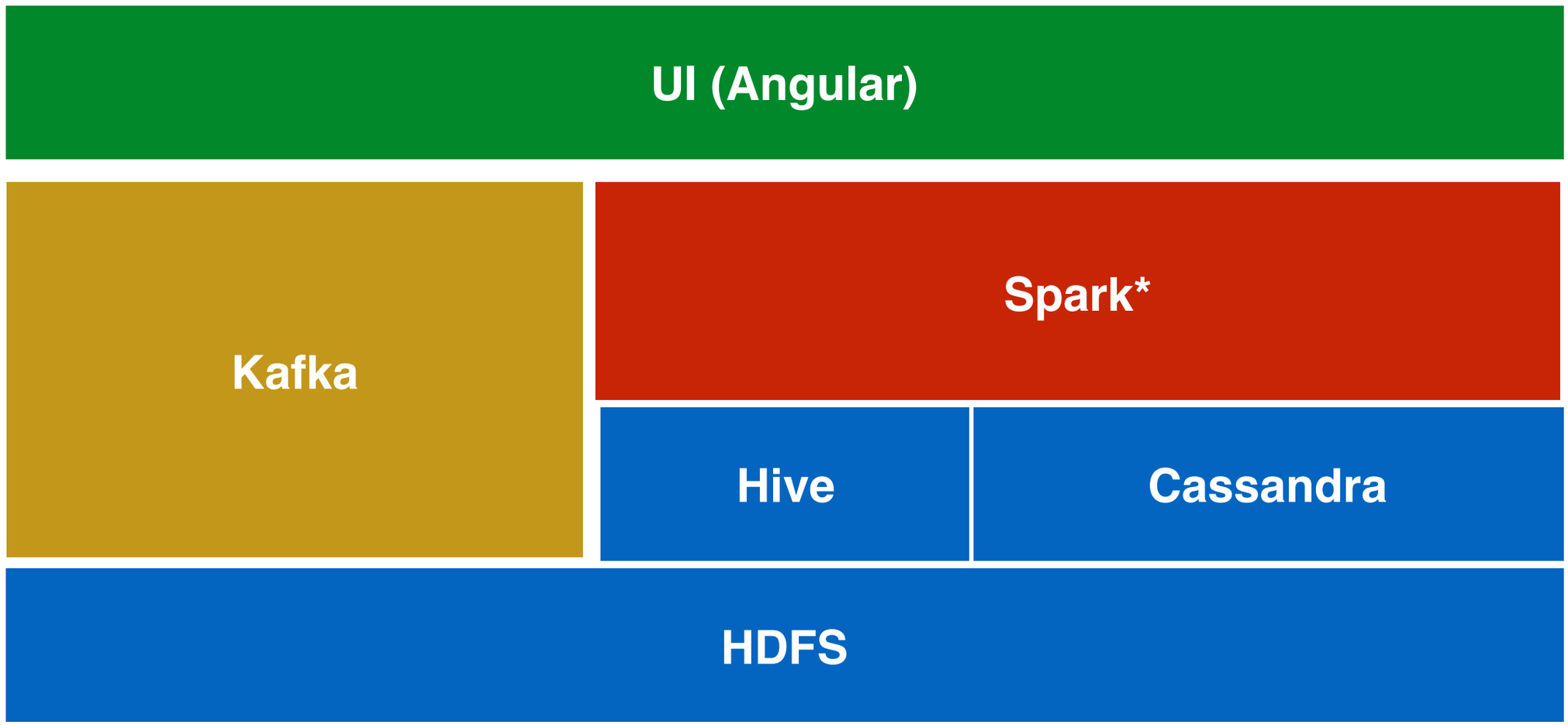


Fig. 3. An example of external application components utilizing the TARZAN Platform (the ingestors on the left side are feed data to the platform, computation tasks and an application on the top and bottom are processing the data, and the monitors on right side are passing results to clients).

Architektura



TARZAN Technologie



*https://zdatainc.com/2014/09/apache-storm-apache-spark/

Případová studie

Analýza PCAP souborů a extrakce “zajímavých” objektů (Network Miner):

- Identifikace komunikujících stran
- Extrakce souborů z HTTP, FTP, SMB, SMTP, POP3 a IMAP
- Enhancement (GEO IP)
- Uživatelská jména
- DNS dotazy
- Identifikace OS, webového prohlížeče
- Decapsulation GRE, 802.1Q, PPPoE, VXLAN, OpenFlow, SOCKS, MPLS and EoMPLS

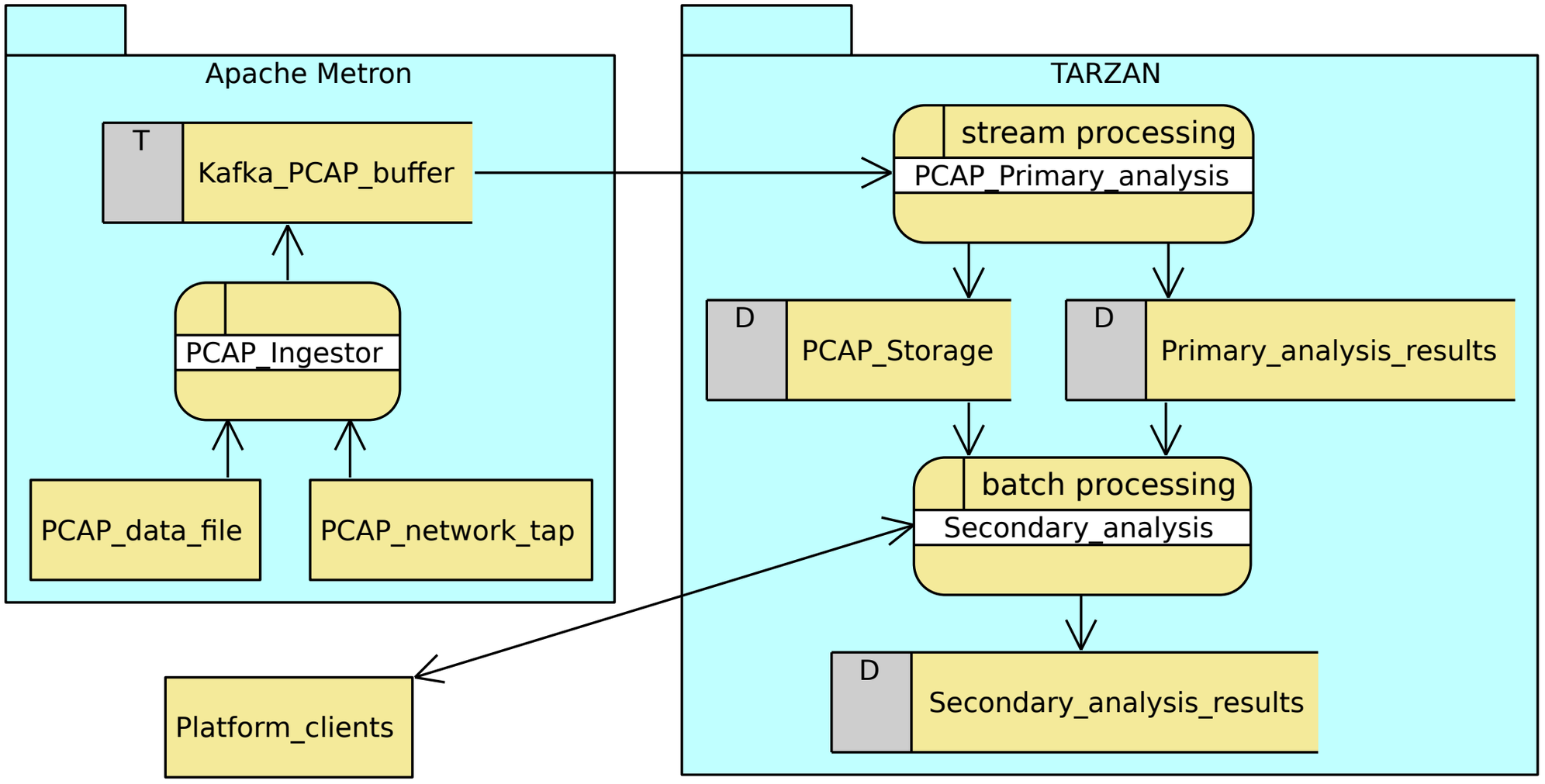


Fig. 4. Architecture of the PCAP Analysis tool with data-flows (including processes, data storages, and external data sources and entities).

Závěr

- Provedena analýza problému, vytvořen přehled v současnosti dostupných řešení.
- Současný stav 2017-06:
- Vytvořen návrh systému dle specifických požadavků.
- Zahájeny práce na proof-of-concept nástroje pro analýzu PCAP souborů - demonstrace 2017-09.

- TARZAN je fork projektu Apache SPOT (proč ne Metron?)
- Prototyp bude k dispozici 2017-12.
- 2018+ rozšiřování platformy a integrace dalších datových zdrojů a analytických modulů

References

- [1] M. Qi, Y. Liu, L. Lu, J. Liu, M. Li, Big data management in digital forensics, in: 17th IEEE Int. Conf. Comput. Sci. Eng. CSE 2014 - Jointly with 13th IEEE Int. Conf. Ubiquitous Comput. Commun. IUCC 2014, 13th Int. Symp. Pervasive Syst. Algorithms, a, 2015: pp. 238–243. doi:10.1109/CSE.2014.74.
- [2] J. Lee, S. Un, Digital forensics as a service: A case study of forensic indexed search, in: Int. Conf. ICT Converg., 2012: pp. 499–503. doi:10.1109/ICTC.2012.6387185.
- [3] J. Rrushi, P. a Nelson, Big Data Computing for Digital Forensics on Industrial Control Systems, Inf. Reuse Integr. (IRI), 2015 IEEE Int. Conf. (2015) 593–608. doi:10.1109/IRI.2015.94.
- [4] M. Qi, Y. Liu, L. Lu, J. Liu, M. Li, Big Data Management in Digital Forensics, 2014 IEEE 17th Int. Conf. Comput. Sci. Eng. (2014) 238–243. doi:10.1109/CSE.2014.74.
- [5] A. Guarino, Digital forensics as a big data challenge, in: ISSE 2013 Secur. Electron. Bus. Process., 2013: pp. 197–203. doi:10.1007/978-3-658-03371-2_17.
- [6] S. Zawoad, R. Hasan, Digital Forensics in the Age of Big Data : Challenges , Approaches , and Opportunities, 2015 IEEE 17th Int. Conf. High Perform. Comput. Commun. (HPCC), 2015 IEEE 7th Int. Symp. Cybersp. Saf. Secur. (CSS), 2015 IEEE 12th Int. Conf Embed. Softw. Syst. (2015) 1320–1325. doi:10.1109/HPCC-CSS-ICESS.2015.305.
- [7] A. Guarino, Digital forensics as a big data challenge, in: ISSE 2013 Secur. Electron. Bus. Process., 2013: pp. 197–203. doi:10.1007/978-3-658-03371-2_17.