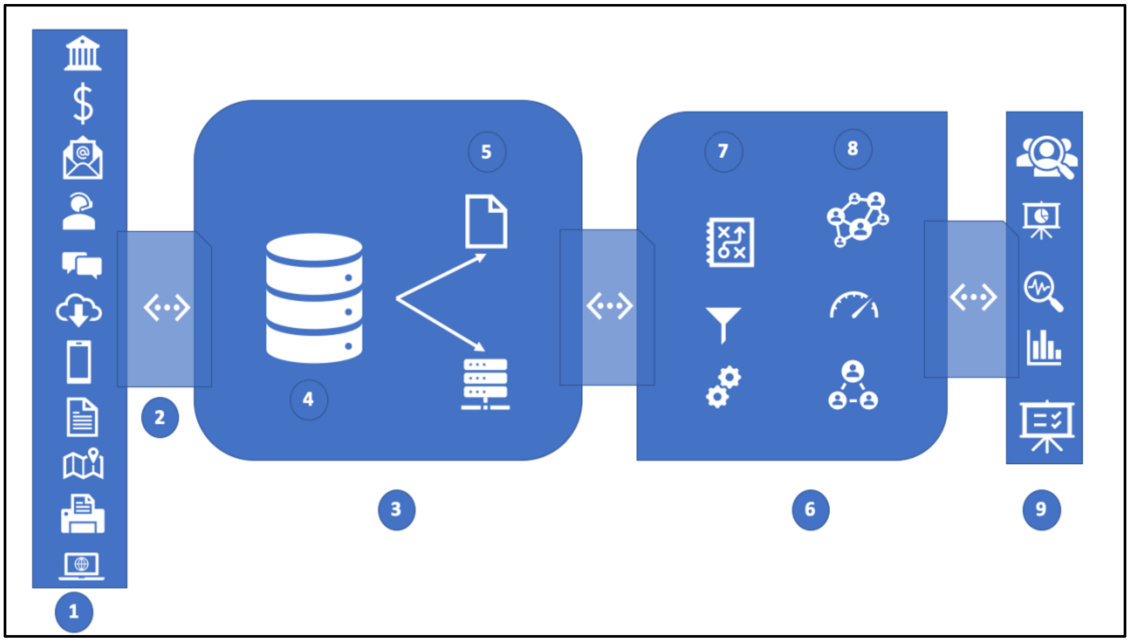
Surveillance System Architecture

Following is an attempt to outline a surveillance system architecture using advanced analytics and big data paradigms.



The prominent blocks of the architecture are as follows:

1. Data sources:
    a. Internal Sources:
        i. Structured:
            1. Transaction data
            2. Behavioral data (Traders HR data, Entity/workstation network data)
            3. Print logs
            4. Badge Access logs
            5. Download logs
            6. Browsing data
        ii. Unstructured:
            1. Email
            2. IM chat
            3. Voice call transcripts
    b. External Sources
        i. Structured:
            1. Financial results
            2. Market data
            3. Geolocation data

            4. Alerts from third party compliance platform
        ii. Unstructured:
            1. Market news
            2. Social media (Blogs, twitter, discussion forums)
            3. Financial filings

2. Data Ingestion:
    a. Apache Kafka
    b. ETL tools (SSIS/SSRS, Apache Airflow)
    c. API (SEC, EDGAR, etc.)
3. Big Data and archival store
4. HDFS based big datastore (Cloud based or On-premise)
5. Cache and analytics database
    a. Graph Database (Neo4j) as primary
    b. NoSQL DB (MongoDB) as secondary
    c. In memory (Redis) as cache
6. Real time analytics machine
7. Data pre-processing
    a. Data filtering
    b. Third party data enrichment
    c. Apache Spark
8. Data analytics processing
    a. Rule based analytics
    b. Lexicon based analytics
    c. Exploratory processing (Hive, MapReduce)
        i. Trading floor communication visualization
    d. Predictive Analytics (Spark MlLib)
    e. Behavioral Analytics
        i. Pattern recognition (Clustering-kNN, Hierarchical clustering)
            1. Quote stuffing
            2. Dumping
        ii. Anomaly detection (LSTM, RNN)
            1. Layering and spoofing detection
            2. Large, unusual volume detection
9. Realtime surveillance Alerts
    a. Visualizations graphs
    b. Reports
    c. Dashboards
    d. Holistic behavior profiling