A Modern Approach to Situation Awareness:

The Ultimate Challenge for Event Processing

Ralf Mueller - Oracle
Christoph Brandt – TU Darmstadt

Dieter Gawlick
Adel Ghoneimy
Kenny Gross
Zhen Liu

ORACLE

Safe Harbor Statement

The following is intended to outline our general product direction. It is intended for information purposes only, and may not be incorporated into any contract. It is not a commitment to deliver any material, code, or functionality, and should not be relied upon in making purchasing decisions. The development, release, and timing of any features or functionality described for timing of any features or functionality described for Oracle's products remains at the sole discretion of Oracle.

Agenda – Part 1

- Situation Awareness (SA)
  - Motivation
  - From the OODA-loop to KIDS
- High quality event detection in Big Data streams
  - Requirements
  - MSET as the technology of choice for multi-dimensional time series
  - Demo
- IDP – Intelligent Data Pre-processing
  - Architecture
  - Temporal property graphs as core technology to organize data
- Summary of part 1

Agenda – Part 2

- Qualitative Data Processing Problem and Solution: Proposed Approach
- Graph Transformation: Examples
- Triple Graph Transformation: Examples & Demo
- Some theoretical Results & Available Tools
- Q&A

Agenda – Part 1

- Situation Awareness (SA)
  - Motivation
  - From the OODA-loop to KIDS
- High quality event detection in Big Data streams
  - Requirements
  - MSET as the technology of choice for multi-dimensional time series
  - Demo
- IDP – Intelligent Data Pre-processing
  - Architecture
  - Temporal property graphs as core technology to organize data
- Summary of part 1

What is Situation Awareness (SA)?

http://en.wikipedia.org/wiki/Situation_awareness *

"…is the perception of environmental elements and events with respect to time or space, the comprehension of their meaning (situation), and the projection of their future status."

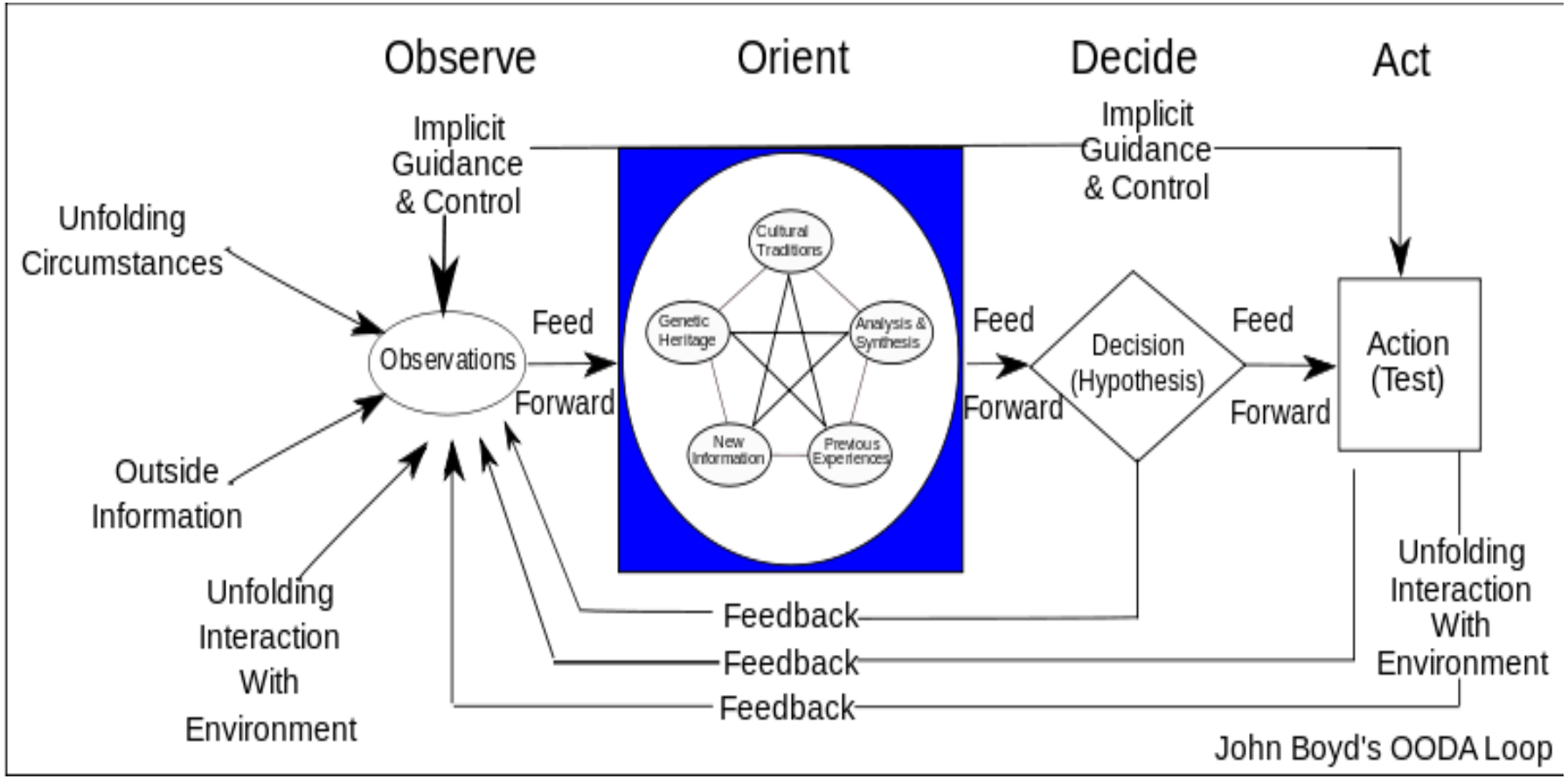
"…has been recognized as a critical, yet often elusive, foundation for successful decision-making across a broad range of complex and dynamic systems"

* as of May, 6th, 2019

Why do We Care?

- A practically unlimited number of applications have to deal with Situation Awareness (SA):
  - **Internet of Things (IoT),** Cloud Services Management, Human Capital Management (HCM),Customer Relationship Management (CRM), Enterprise Resource Planning (ERP), Supply Chain Management (SCM), Power Grid Management, Bug Management (SW/HW), Flight Control, airplane surveillance, patient care, medical Research, Military **(SA started here)**, and many, many more
  - Practically **ANY** critical application needs to deal with SA
- A comparison
  - The **Cloud** is in the process to fundamentally change/improve **operations**
  - **Situation Awareness** will fundamentally change/improve **functionality**

The OODA Loop – John Boyd (1976)



From https://en.wikipedia.org/wiki/OODA_loop *

- John Boyd (US Air Force Colonel) established SA as a discipline

- Decision Making occurs in recurring cycle of Observe-Orient-Decide-Act

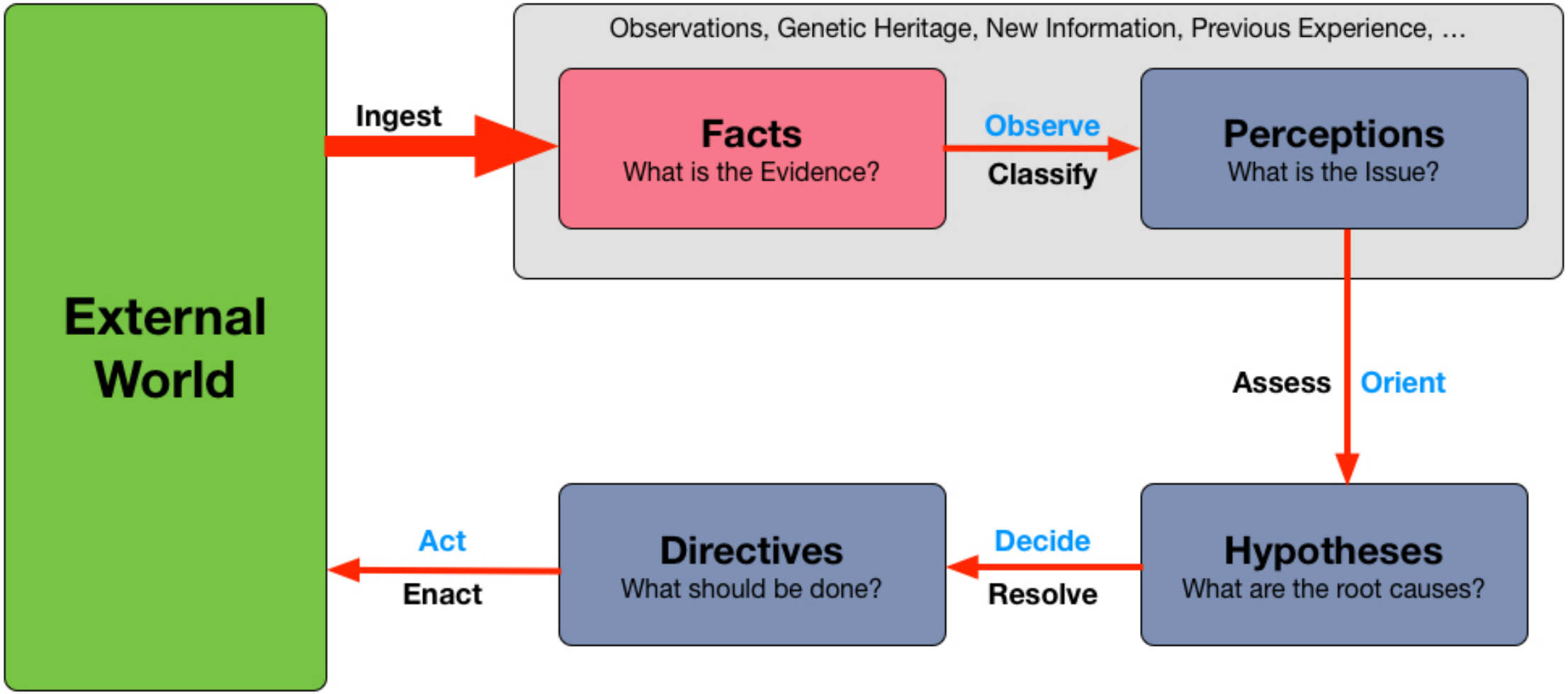
* as of May, 30th 2019

A Lot has Happened Since 1976

- IT has matured; we experienced the emergence of
  - Database management
  - Big Data
  - Cloud Computing
  - Machine Learning/AI
  - … many, many other things
- Pervasive, Ubiquitous and Mobile Computing
  - IoT (Internet of Things)
  - RFID
  - Smart Phones, Tablets, Watches,…
- **Today, we live in a data driven world**
- **We need to map the OODA loop/concepts to modern IT**
  - Capturing and interpreting an evolving state represented by multi-dimensional time series is one of the most important underlying technologies

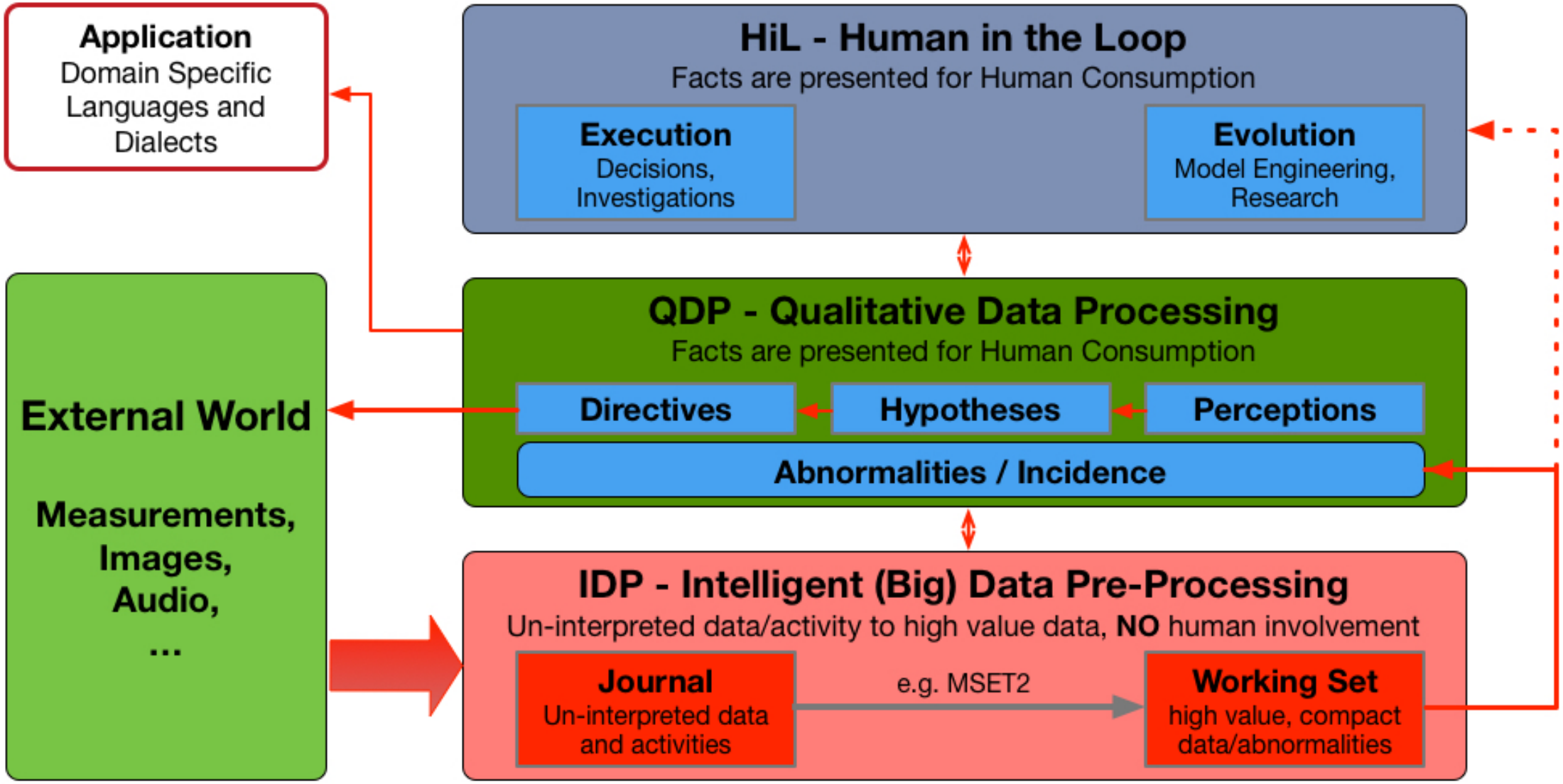
KIDS*: Mapping the OODA loop into IT Technology

A comprehensive view based on Data, Knowledge, and Processes



* KIDS: Knowledge Intensive Data-processing Systems

KIDS – Process/Data Flow



Some of the Major Challenges

- ✓ Ingest and store incoming (sensor) data
- ✓ Prepare data for high speed real time processing as well as batch processing
- ✓ Identify abnormal conditions (events) with lowest possible false positive/negative as well as high sensitivity and minimal resource consumption
- ✓ Create an incident for each event and organize these incidents as they relate to each other
- ✓ Inform users (domain experts) about incidents in their domain language
- ✓ Allow domain experts to investigate with their domain language the underlying data
  - Humans are unable to absorb Big Data directly
- ✓ Provide full provenance for data, knowledge, and processes

Agenda – Part 1

- Situation Awareness (SA)
  - Motivation
  - From the OODA-loop to KIDS
- High quality event detection in Big Data streams
  - Requirements
  - MSET as the technology of choice for multi-dimensional time series
  - Demo
- IDP – Intelligent Data Pre-processing
  - Architecture
  - Temporal property graphs as core technology to organize data
- Summary of part 1

Requirements

Requirement	Functional	Operational
High-Speed data ingest	• Supervision of incoming data • Support for any data type • **Provenance**	• Performance • Scalable • Reliable • Secure
Abnormal condition detection	• Lowest false-alarm, missed-alarm • High sensitivity • Minimal programming and maintenance effort - 'Cleanup' of sensor data for easier follow-up processing • Explainable	• Performance • Scalable • 'Cleanup' of sensor data for faster follow-up processing

Will Focus on Sensor Data Processing (IoT)

- State of the art:
  - Preprocessing typically involves only the ingestion of data, detection/filtering of obvious “outliers” that are deemed to have no prognostic significance….everything else is up to users

- What is missing?
  - Real time determination of abnormal conditions
    - Human are unable to ask the right question at the right time
    - Popular ML technology has high false positive/negative, low sensitivity
    - No provenance and tamper-proofing
    - Investigations are difficult, expensive, and not reliable or conceptually not possible due to non-deterministic algorithms
  - Transformation to compact, high quality data – Domain Jargon (Fachsprache)
    - Low quality data leads to low quality output
      - Un-condensed (high volume) data lead to high resource consumption, high false/ missed alarm probabilities

What are the Options Detecting Events?

- CQL's (Continuous Query Languages)
  - Too complex, expensive, time consuming to program and to maintain
- Machine Learning
  - NNs (Neural Networks), SVMs (Support Vector Machines)
    - Not deterministic, cannot support provenance
  - MSET (Multivariate State Estimation Technique)
    - Fully deterministic
    - MSET is focused on multi-dimensional time series

The MSET Technology

- The Multivariate State Estimation Technique (**MSET**) is a non-linear, non-parametric regression modeling method that was originally developed in the 1990's for prognostic anomaly detection in nuclear power plants, commercial aviation and business critical applications.

- Oracle was the first company to pull MSET-type statistical ML into enterprise servers, engineered systems and DB clusters
  → ***Autonomous Systems***

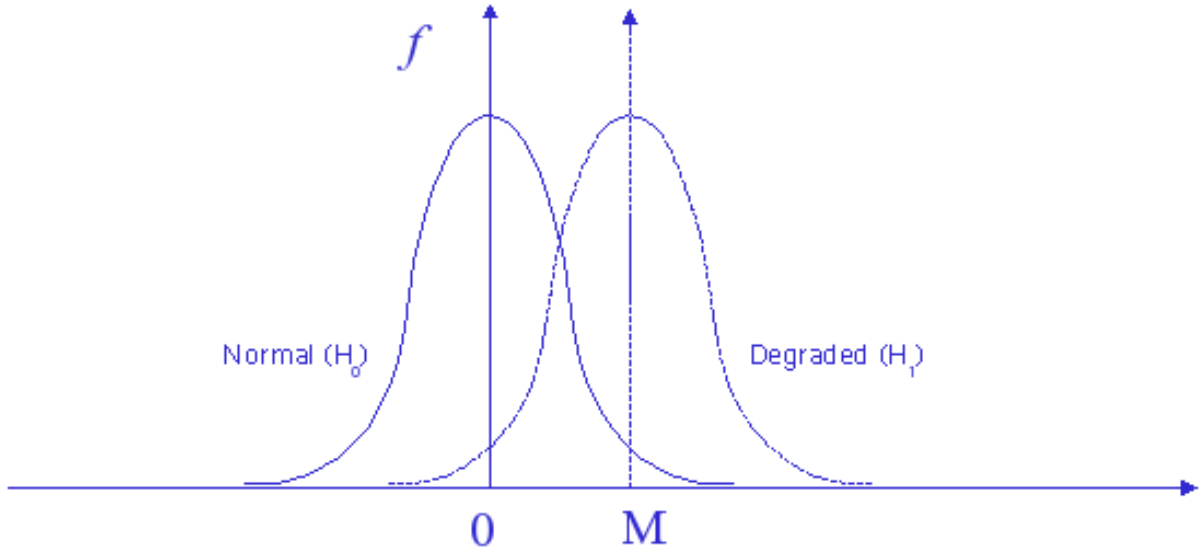
- MSET is perfect for detecting subtle anomalies while attaining high sensitivity in noisy or even chaotic process metrics, but with ultra-low false-alarm and missed-alarm probabilities making MSET an ideal ML algorithm for IoT real-time prognostic applications

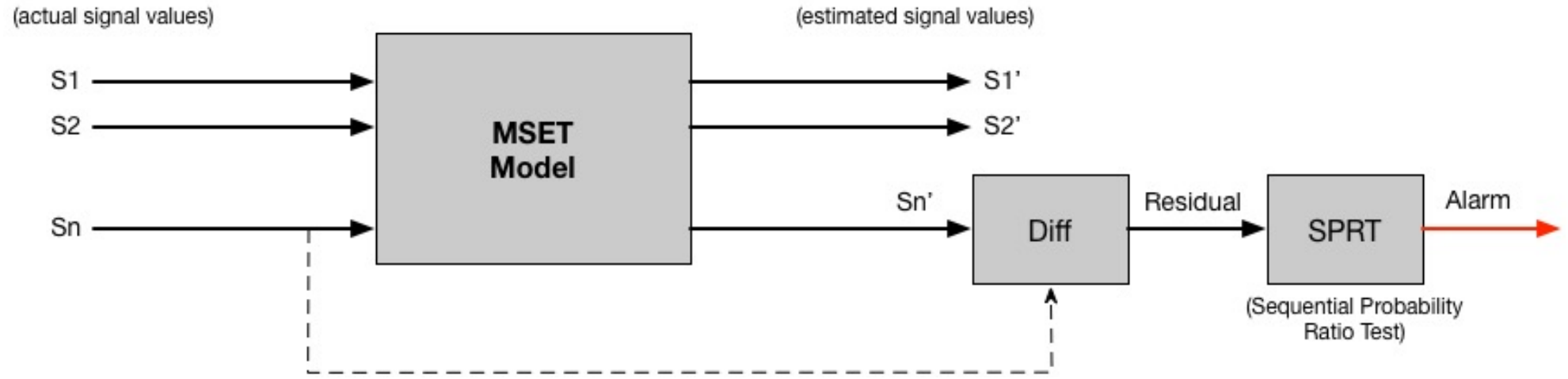
Background: Prognostic Algorithmic Innovations

- Sequential Probability Ratio Test (**SPRT**) For Stationary Time Series
  - Advanced pattern recognition technique for high sensitivity, high reliability sensor and equipment operability surveillance
  - Developers proved in refereed journals that the SPRT provides the earliest mathematically possible annunciation of a subtle fault in noisy process variables.
  - Crucial capability: Ultra-low and separately specifiable false-alarm and missed-alarm probabilities
- Multivariate State Estimation Technique (**MSET**) For Dynamic Time Series
  - Online model-based fault detection and identification (Supervised ML)
  - MSET predicts in real time what each process metric should be on the basis of learned correlations among all process variables
  - MSET incorporates the SPRT to monitor the residuals between the actual observations and the estimates MSET predicts on the basis of the correlated variables.

SPRT vs. MSET

- **SPRT:**
  - Works with non-Gaussian noise signals (accommodate any measurement noise)
  - Sequential binary-hypothesis test compares likelihood of observations coming from reference distribution (H_0) vs. degraded distribution (H_1)
  - Empirically learns reference distribution H_0 and use as baseline for system
- **MSET:**

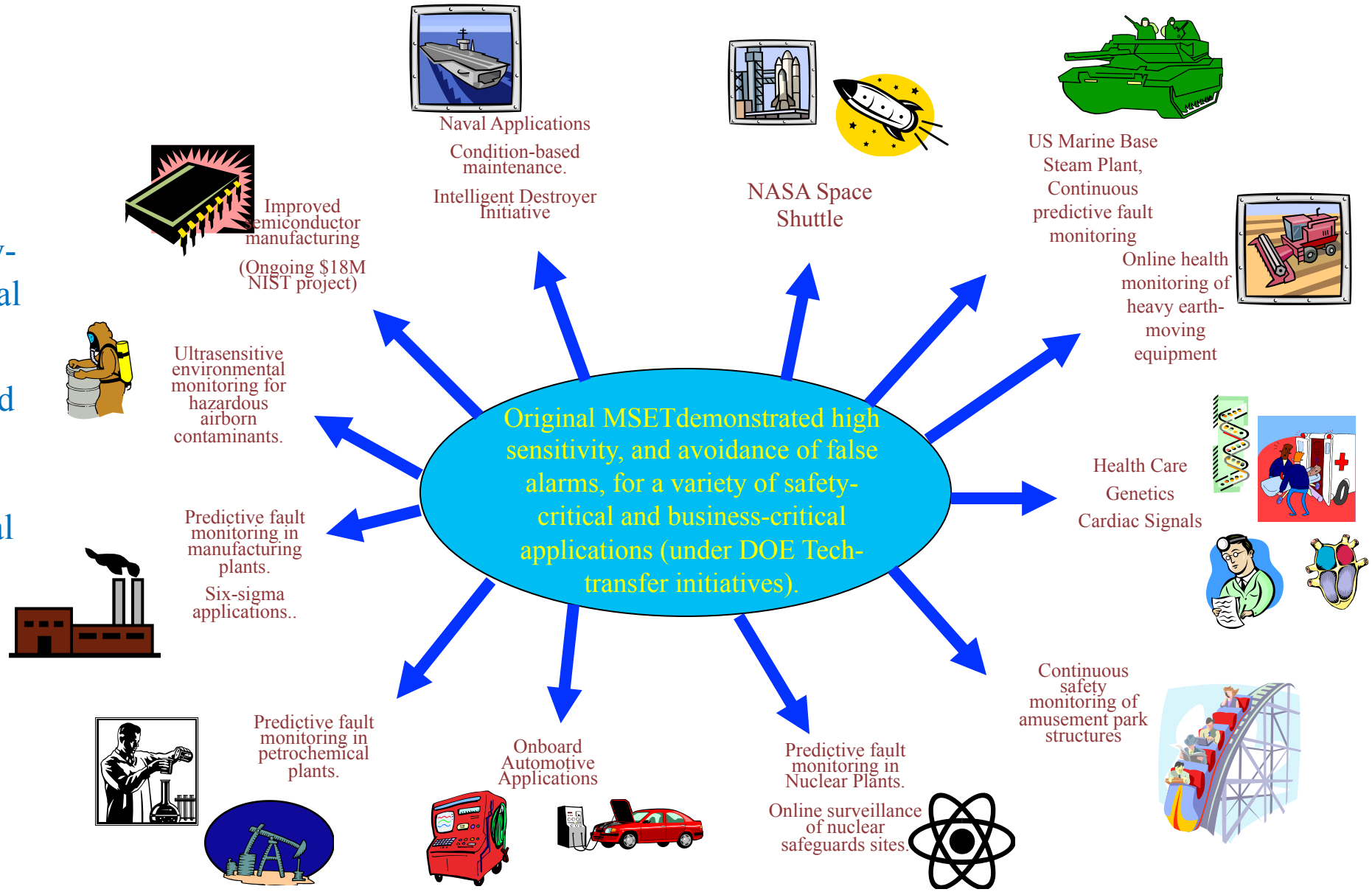




MSET Background

Original MSET (1998) is mature and in use for prognostics in many safety-critical and business-critical industries.

Oracle MSET2 inherits and improves the value proposition for real-time prognostics for IoT optimal predictive maintenance of critical assets.



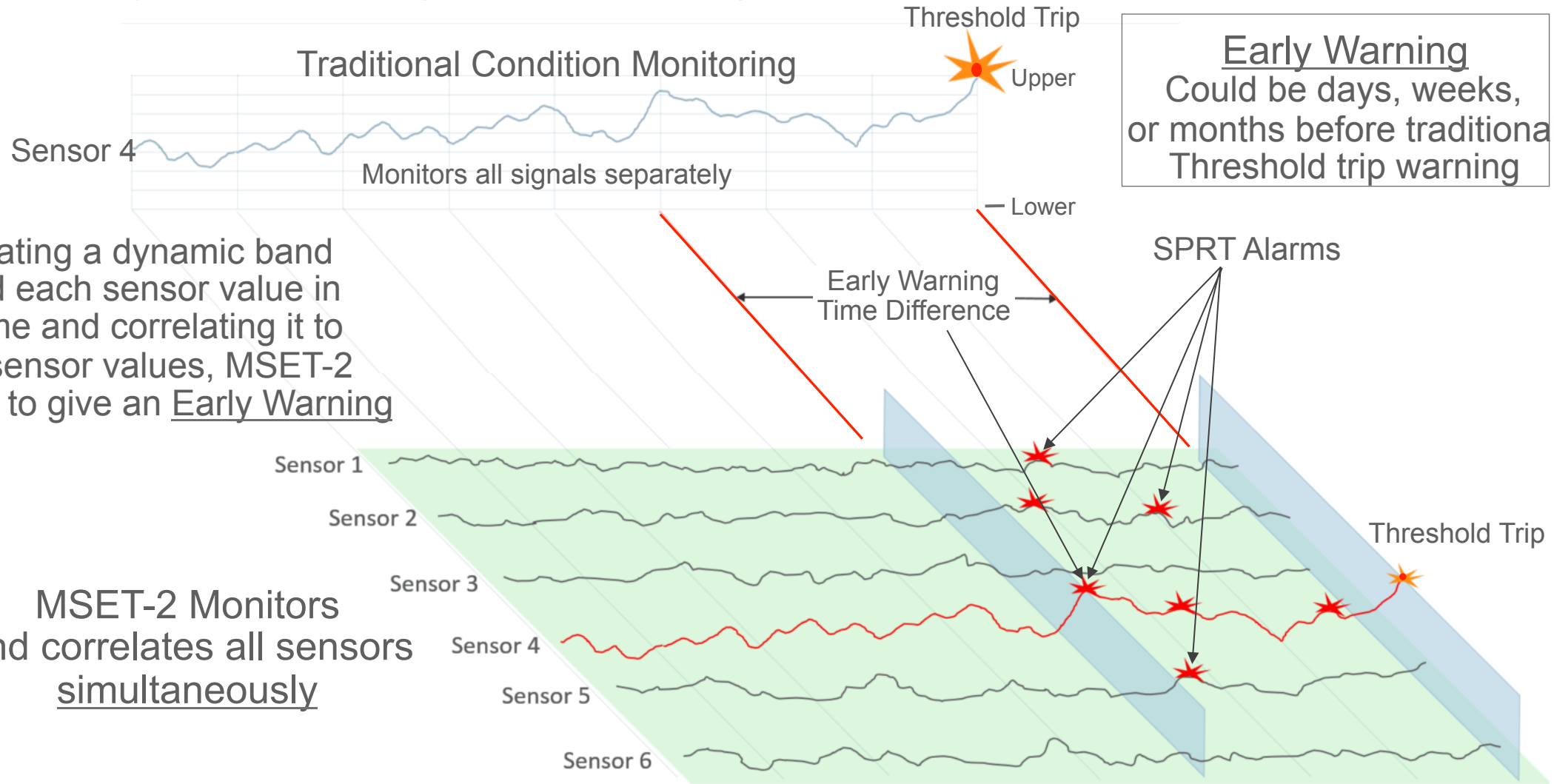
The ‘Magic’ of MSET

- Disambiguation between faulty sensor readings, vs anomalies in monitored processes/systems
  - Sensors most often have smaller Mean-Time-Between-Failures than the assets the sensors are supposed to protect
  - Sensor disturbances (de-calibration bias, intermittent stuck-at faults, change-of-gain drifts, spikiness) are primary cause of false-alarms, missed-alarms in many IoT use cases
  - MSET2 flags faulty sensors for recalibration at next opportunity
    - Swaps in a highly accurate “virtual sensor”, no need to shut down critical assets from sensor-de-calibration events
- Identify abnormalities of assets
  - Describes deviations from the norm
  - Computes quantitative Remaining Useful Life (RUL) - recommended optimal remedial actions

The ‘Magic’ of MSET continued

- Data quality improvements
  - **UnQuantize** – Automatically identifies low-resolution “quantized” transducers
    - Transforms low-resolution input signals into high-accuracy output signals, prior to consumption by MSET2 for prognostics
  - **Imputation** – for missing data replacement
    - Not interpolation! A “blind spot” filled in with interpolation is still a blind spot (in terms of prognostics)
    - Oracle’s optimal missing-value imputation (MVI) fills in missing values with highly-accurate estimates based on cross correlation with other non-missing values
  - **ARP** - Re-synchronization of measurements out of phase due to (common) clock-sync disparities in measurement instrumentation
- Tamper-proof provenance certification for original raw time series data
- Data reduction, and compression
  - Optimal Memory Vectorization
    - To represent the data with a minimal set
  - Statistical Compression - Uses Oracle’s “Zeno’s Circular File” compression methodology

Early Warning Advantage

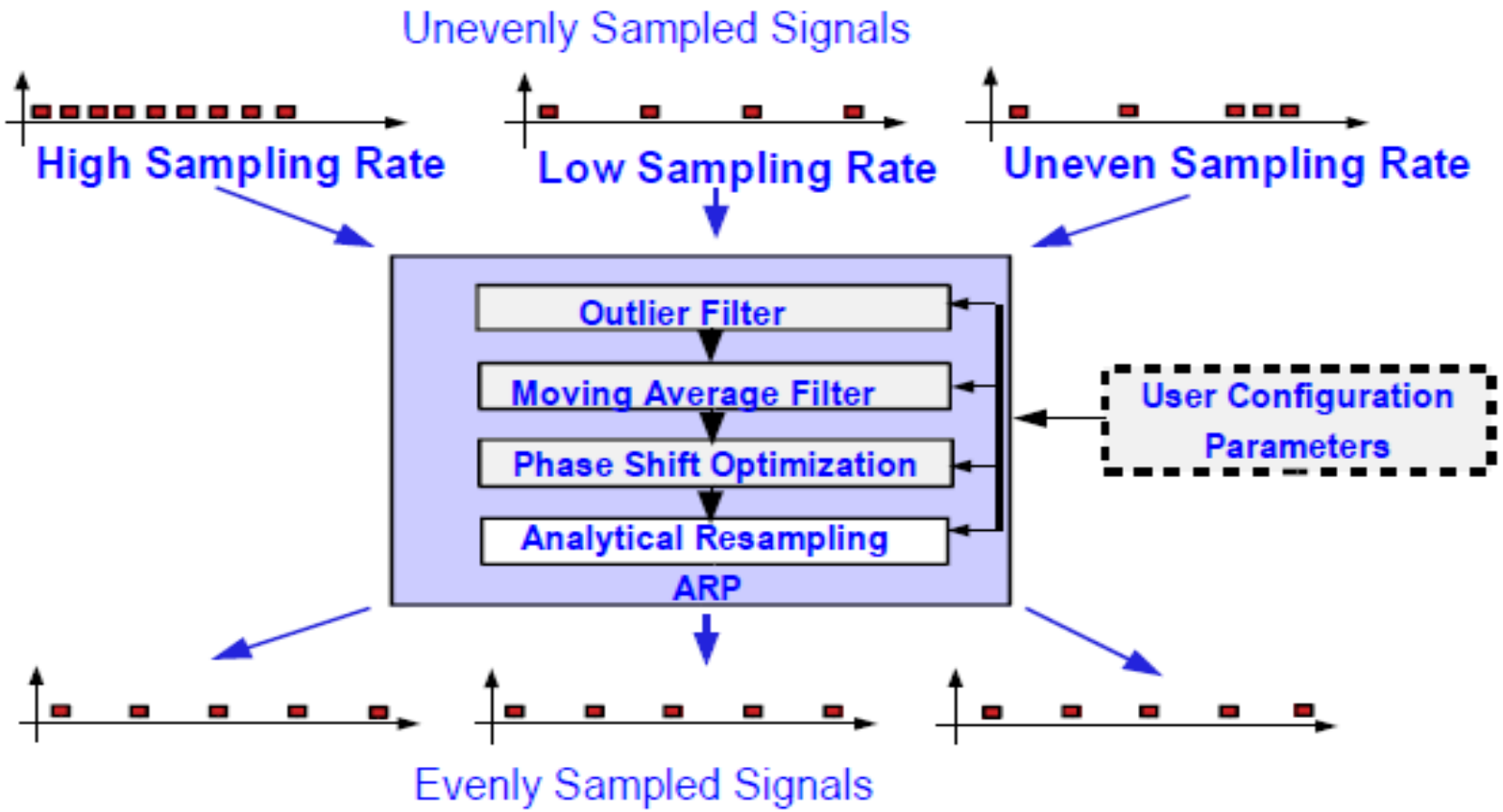


Early Warning
Could be days, weeks, or months before traditional Threshold trip warning

By creating a dynamic band around each sensor value in real time and correlating it to other sensor values, MSET-2 is able to give an Early Warning

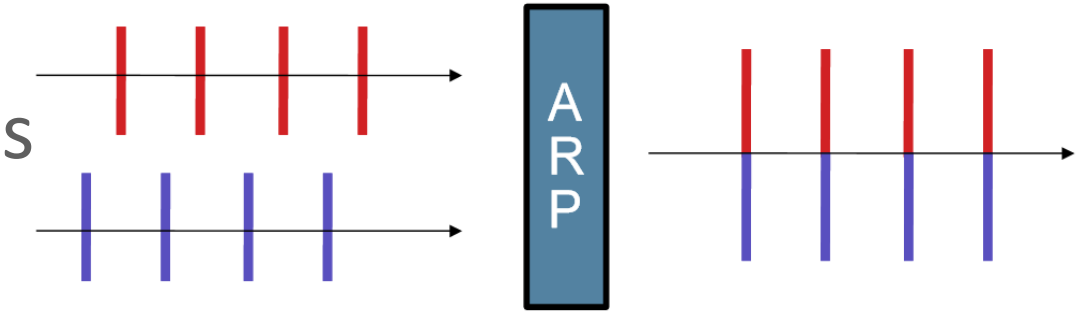
MSET-2 Monitors and correlates all sensors simultaneously

ARP:Analytical Resampling Process



- ARP: Essential for Multi-Signal Diagnostics/Prognostics
  - Challenges for multi-sensor diagnostics/prognostics, plus ARP solutions to those challenges, summarized next slide

Analytical Resampling Process



Essential for Prognostic Machine-Learning Algorithmics

Telemetry streams originate with differing sampling rates:

- ✓ APR uses interpolation-based up-sampling/down-sampling methods to generate uniform sampling intervals for all telemetry time series
- ✓ It is very commonly the case that the four primary sources of telemetry signatures to be used in prognostic security originate from clocks that are significantly out of sync

Asynchronies from Clock Mismatch Issues:

- ✓ Clock mismatch issues will cause almost all time-series Machine Learning algorithms to fail prognostic functional requirement criteria
- ✓ ARP performs real-time “adaptive empirical synchronization”

MSET, Summary

- Developed in response to a request from the NCR (Nuclear Regulatory Commission)
  - Too many false positives
- Challenge: Find an unknown number of abnormal conditions
  - Training set: Measurement of a running system without any issues
  - Test set: Measurements of a faulty system
- The result:
  - MSET found all issues without any false positives
  - Competing approaches; e.g., neural networks, found none: Neural networks were banned from nuclear power plants
- **Kenny Gross** is the lead inventor/developer of the MSET technology
  - Kenny has significantly enriched MSET, Kenny has 200+ patents and published 200+ papers
  - Kenny Gross: https://labs.oracle.com/pls/apex/f?p=LABS:bio:0:2256#ct07tabcontent1 and https://sites.google.com/site/kennycgross/
  - Kenny is a member of the KIDS team

Demo

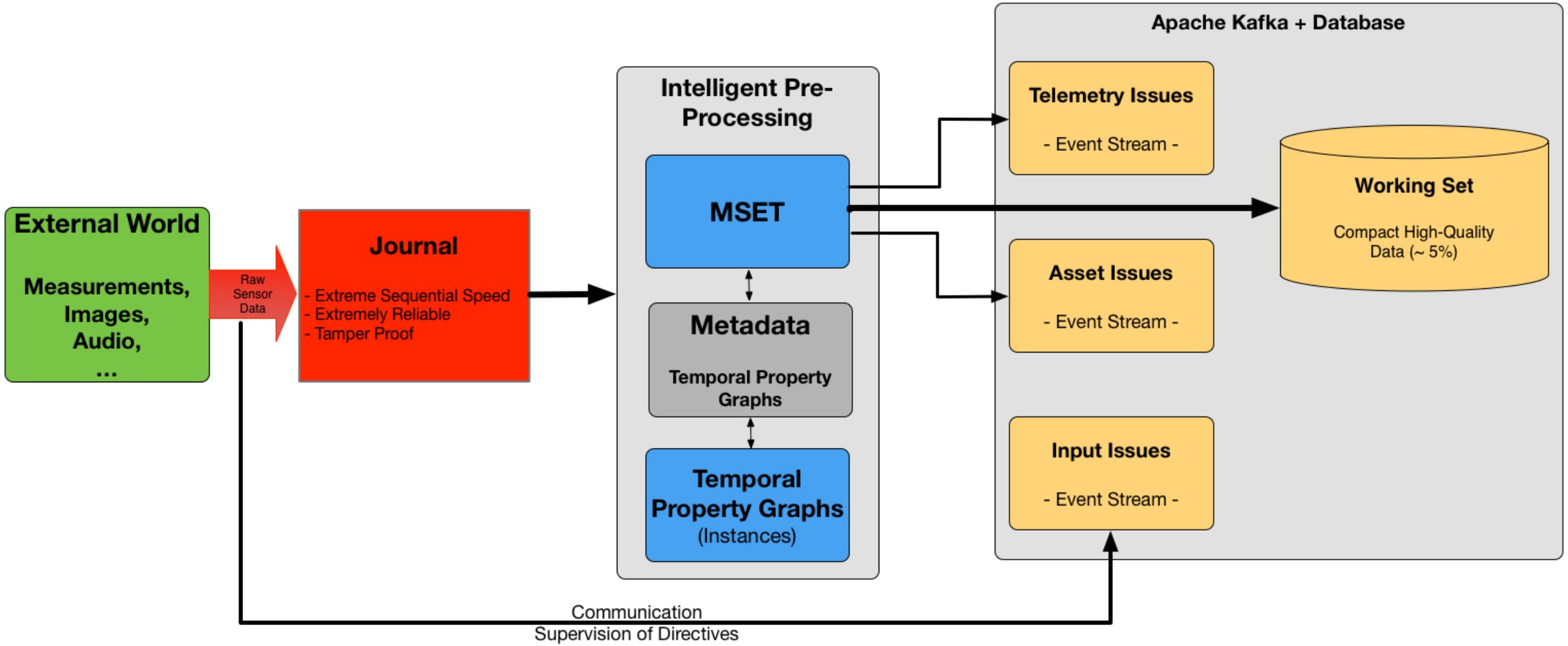
Agenda – Part 1

- Situation Awareness (SA)
  - Motivation
  - From the OODA-loop to KIDS
- High quality event detection in Big Data streams
  - Requirements
  - MSET as the technology of choice for multi-dimensional time series
  - Demo
- IDP – Intelligent Data Pre-processing
  - Architecture
  - Temporal property graphs as core technology to organize data
- Summary of part 1

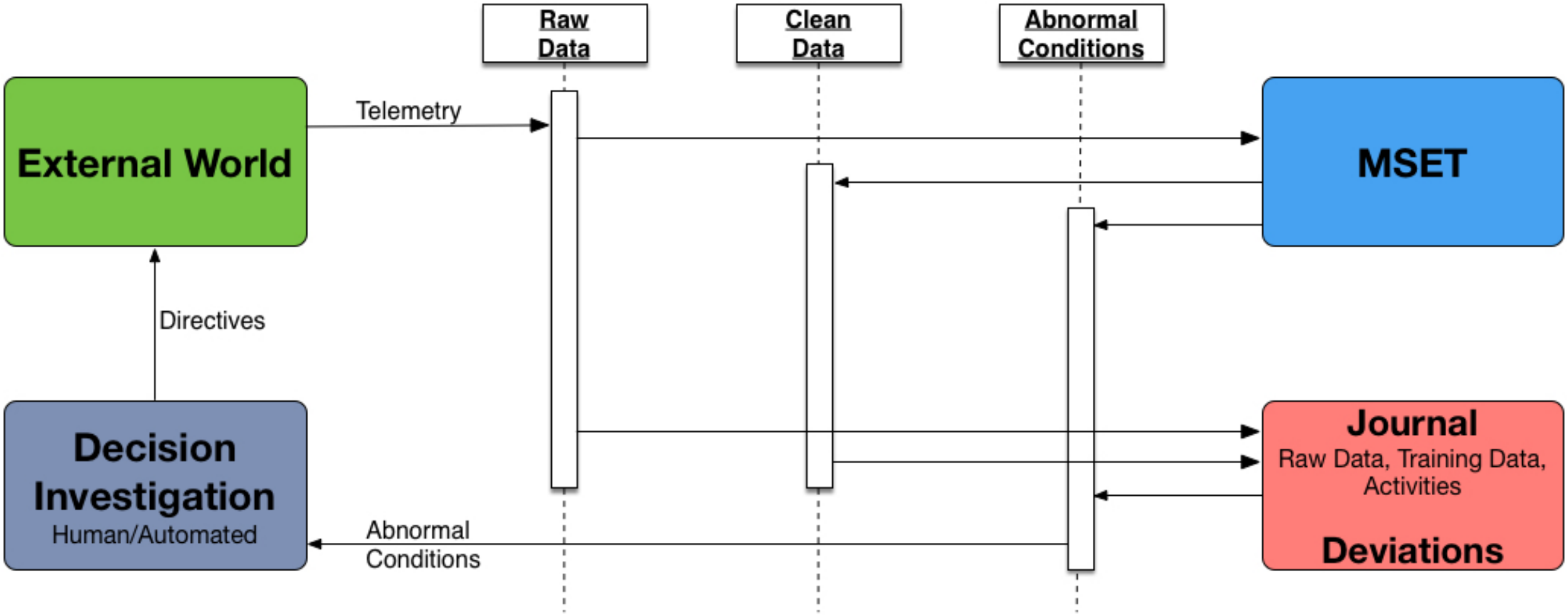
IDP – Intelligent Data Pre-processing Functionality

- Capturing and Supervising all incoming data, access and change
  - This is a tamper-proof journal providing direct access
  - Incoming data is not interpreted
  - (Arrival of) data will be checked according to directives though
- Determining abnormal conditions with provably minimal false-alarm & missed-alarm
  - MSET uniquely disambiguates between sensor issues versus issues in monitored assets (fundamental challenge for Situation Awareness) and creates actionable event notifications
- Transforming data with MSET
  - This creates high quality data with significantly reduced data volume
  - Effectively removes “dark data” (random noise associated with sensor measurements), culls out the vectors that optimally represent the underlying structure of the time series
  - Follow-up processing gets improved quality and reduced resource consumption
- Providing operational characteristics for mission critical applications
  - Scalability, performance, security, tamper-proof provenance - deterministic, and more

IDP - Intelligent Pre-Processing Architecture



The IDP Data Flow

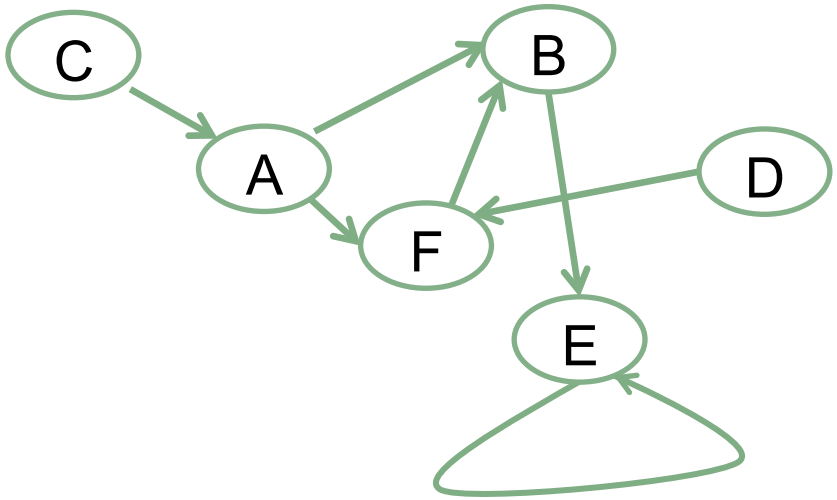


The Meta-Data

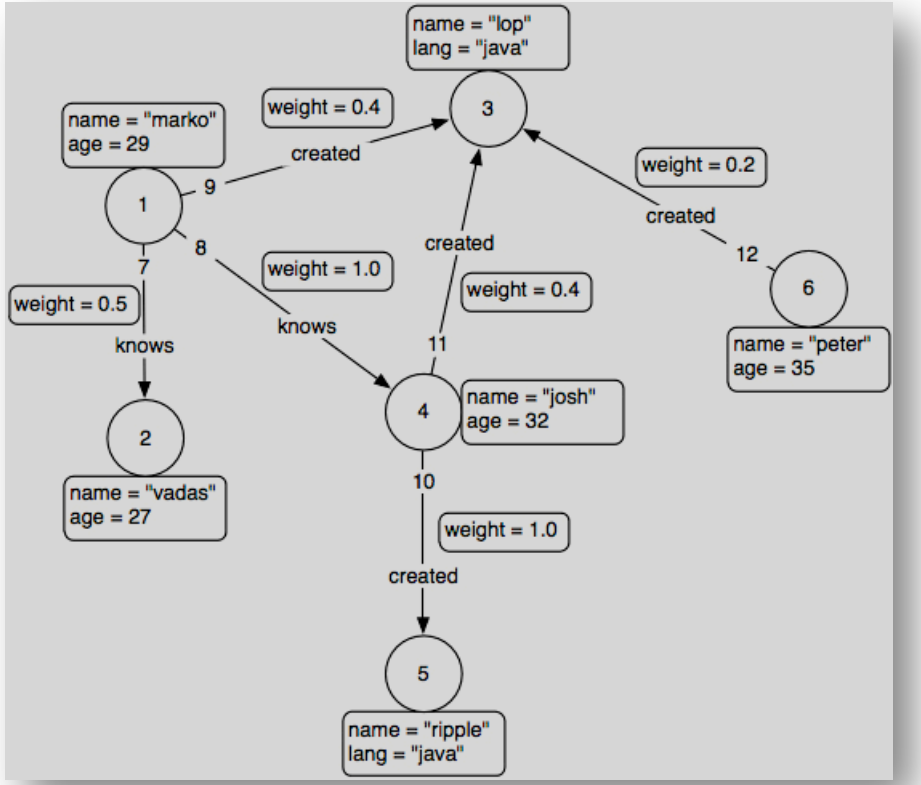
- The temporal status of objects and their relationship
  - Assets
  - Sensors
  - Incidents – the abnormal conditions/events
- Everything on type and instance level
  - Types – to capture the concepts and ontologies
  - Instances – to capture the individual assets and their status
- Everything will be captured with temporal property graph technology
  - Most flexible way to capture objects and their dynamic relationships
  - Provides full provenance with snapshot isolation
- Major use:
  - Understanding sensor data in their (temporal) context
    - For real-time processing – activation of models with the proper data
    - For off-line processing – investigation, analysis, and model development and evolution
  - Optimal clustering of data

Overview of Graph

- What is a graph?
  - A set of vertices and edges (with optional properties)
  - A graph is simply **linked data**
- Why do we care?
  - Graphs are everywhere
    - Road networks, power grids, biological networks
    - Social networks/Social Web (Facebook, Linkedin, Twitter, Baidu,,…)
    - Knowledge graphs (RDF, OWL, JSON-LD etc.)
  - Many public, open data initiatives use graphs
  - Graphs are intuitive and flexible
    - Easy to navigate, easy to form a path, natural to visualize
    - Do not require a predefined schema



Property Graph Data Model



- A set of vertices (or nodes)
  - each vertex has a unique identifier.
  - each vertex has a set of in/out edges.
  - each vertex has a collection of **key-value** properties.
- A set of edges
  - each edge has a unique identifier.
  - each edge has a head/tail vertex.
  - each edge has a label denoting type of relationship between two vertices.
  - each edge has a collection of **key-value** properties.
- Blueprints Java APIs
- Implementations
  - Oracle, Neo4j, Titan, InfiniteGraph, Dex, Sail, MongoDB …
- A property graph can be modeled as an RDF Graph

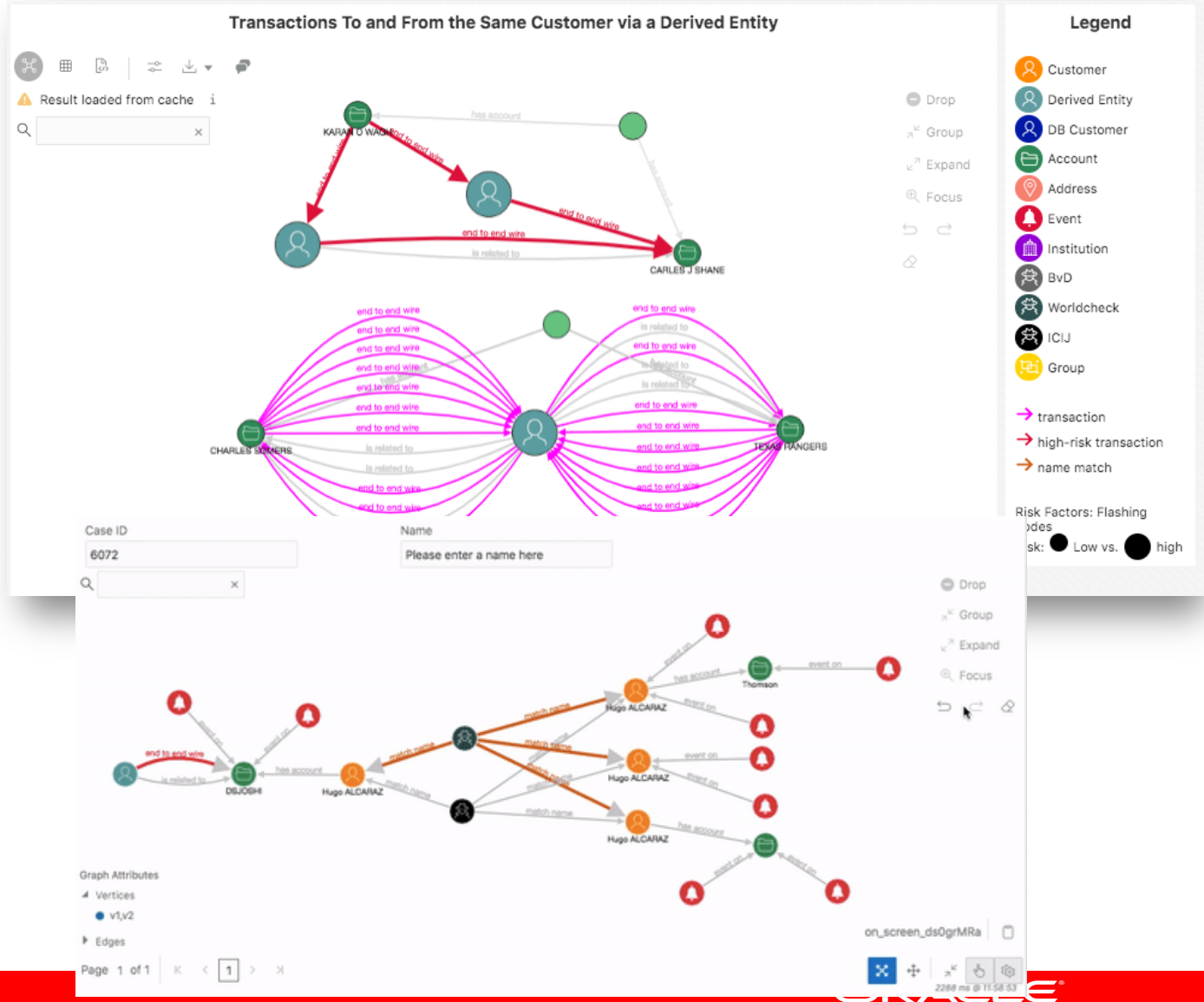
https://github.com/tinkerpop/blueprints/wiki/Property-Graph-Model

What do Temporal Property Graphs Provide?

- Associations between incoming data and their sensors and assets provides rich context information
  - The context consists of
    - Instances and their relations
      - Status of sensors and assets
    - Types and their relations
      - Full description of functionality and ontology
  - Temporality captures the evolution of the context
    - Which sensor was connected to which asset when the measurement was done?
    - What was the status of the sensors and assets at the time of the measurement?
    - What was the functionality and ontology of the assets and sensors when the measurement was done?
    - What is the meaning of the sensor data when the measurement was done?
  - Allows (temporal) enrichment of sensor data with minimal reference

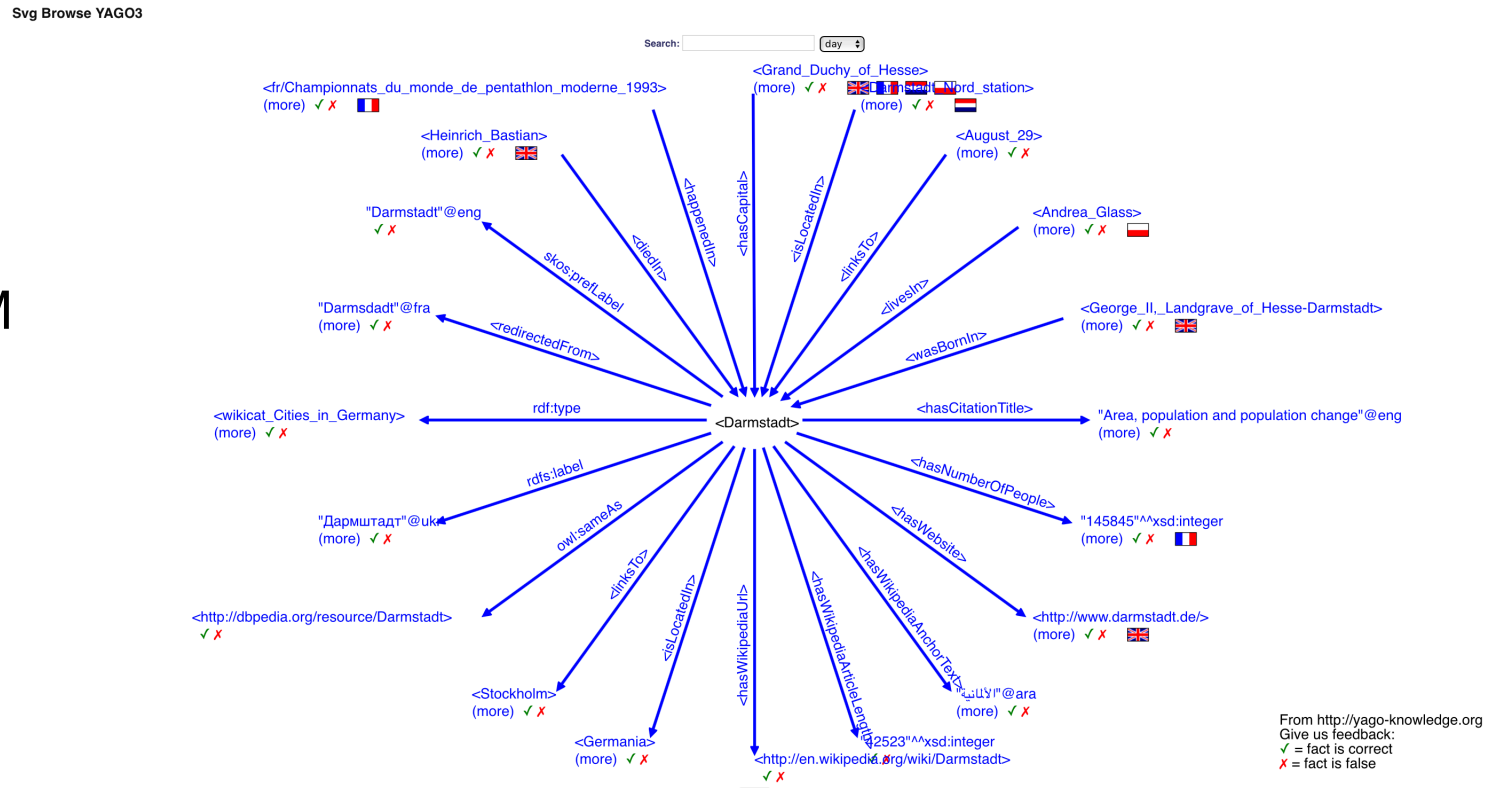
Example: Oracle Financial Crime & Compliance Studio

- Catch predicate criminals
  - Illicit gains (e.g. drugs, trafficking)
  - Money laundering, rapid movement of funds
- Captures relations between customers, accounts, events, transactions
- Entity Linking and Visualization
- Case Correlation



Example: YAGO *

- Created and maintained by MPI Saarbruecken & Telecom Paris Tech University
- Combination of data from Wikipedia, Wordnet & Geonames
- Contains >17M entities and >150M facts
- Is anchored in time and space by attaching a temporal and spatial dimension to many of its facts and entities
- Confirmed accuracy of 95%



* https://www.mpi-inf.mpg.de/departments/databases-and-information-systems/research/yago-naga/yago/

IDP – the Interface

- Training sets
  - The quality of the training data makes or breaks the quality of the interpretation of the facts; given a trainings set:
    - MSET2 will guarantee best and earliest possible abnormal condition detection
      - Mathematical proven
      - Empirically confirmed with many implementations in many dense-sensor IoT prognostic use cases
    - MSET2 requires minimal resource consumption
      - Empirically confirmed with many Big Data use cases
  - Oracle can derive high-quality training data from its customer base *
- Temporal property graphs
  - Fully declarative interface and analysis tools for developer, rich sets of visualization tools for end users
    - Reduces development and maintenance effort
    - Increases quality and operational characteristics

* Oracle does NOT have access to customer data though

Agenda – Part 1

- Situation Awareness (SA)
  - Motivation
  - From the OODA-loop to KIDS
- High quality event detection in Big Data streams
  - Requirements
  - MSET as the technology of choice for multi-dimensional time series
  - Demo
- IDP – Intelligent Data Pre-processing
  - Architecture
  - Temporal property graphs as core technology to organize data
- Summary of part 1

KIDS Contribution to SA

- KIDS provides guidance for the development and evolution of applications and system components.
  - ✓ Provides applications with guidance how to structure them to support SA.
  - ✓ Offers new components; e.g., IDP, QDP
  - ✓ Promotes a synergetic approach to the design and interaction of these components.
  - ✓ Provides system components with guidance of missing functions and operational characteristics.

Status

- MSET exists and is matured
  - Can be used by sophisticated teams
- Situation Awareness
  - Has been around for a while and is heavily studied
  - There is increasing demand from customers
  - Many major applications attempt to deal with it, but
    - Lack of a solid model based on modern IT technology
    - Difficult and expensive development, evolution without architecture
    - Static (and much reduced) functionality
- We have an architecture and many critical pieces
  - We will continue to identify gaps and intend to close them
  - We intend to use the KIDS technology to evolve existing products

Agenda – Part 2

- Qualitative Data Processing Problem and Solution: Proposed Approach
- Graph Transformation: Examples
- Triple Graph Transformation: Examples & Demo
- Some theoretical Results & Available Tools
- Q&A

Enjoy The Break