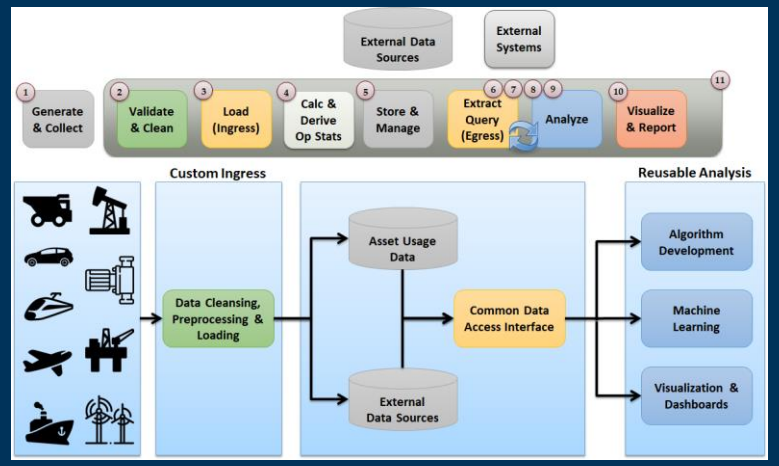
HBM Prenscia Technology Day

Virtual Seminar Series



VePRO -A Program Overview –Scalable CBM

Presented by Dr. Mark A. Pompetzki
Director – Engineering Solutions
Prenscia Engineering Solutions

Session #4: **Vehicle Fleet Data Analytics**

VePRO - A Program Overview – Scalable CBM

Abstract

The U.S. Army is working to improve Reliability, Availability and Maintainability (RAM) as well as durability for its tactical wheeled vehicle fleet while reducing operating costs through the use of Reliability Centered Maintenance (RCM), Condition Based Maintenance (CBM) and Health and Usage Monitoring Systems (HUMS).

The Vehicle Performance, Reliability & Operations - Analysis (VePRO-A) program focuses on the scalability of a HUMS and CBM software system with the ultimate goal of reducing cost and extending equipment life. The objective of this program is to configure scalable and robust software system components along with the end-to-end integrated system for deployment to broaden the understanding of operational usage severity and deterioration as it relates to RAM, cost, readiness, durability, etc. The VePRO-A system approach demonstrates a progressive increase in the ability to analyze operational usage data with a program goal of managing up to 20,000 vehicles and a scalable solution that can be used as a baseline towards achieving the U.S. Army's long-term desired goal of monitoring up to 80,000 Tactical Wheeled Vehicles (TWVs). Overall system architecture and features will be discussed and presented.

About the presenter, Dr. Mark A. Pompetzki

Dr. Pompetzki has over thirty years of experience in fatigue life evaluation and test, data acquisition, duty cycle development, Health and Usage Monitoring Systems (HUMS) and analytics. He has over twenty-five technical publications on durability, fatigue applications and HUMS and is a member of two SAE committees related to fatigue and reliability.

He is the Director of Prenscia Engineering Solutions and HBM nCode Federal LLC. He has been with HBM nCode / HBK for over 20 years and has led multiple projects on Health and Usage Monitoring Systems (HUMS), Condition Based Maintenance (CBM) and fatigue evaluation of components and systems.

Mark holds a Ph.D. in Civil Engineering from University of Waterloo, Canada.

HBK
HOTTINGER BRÜEL & KJÆR

VePRO-A
Program Overview

Vehicle Performance, Reliability & Operations - Analysis
November 2020

A modular, flexible & scalable solution that can be applied to enable enterprise level data management & analysis of assets

Mark Pompetzki
Iain Dodds

HBM nCode Federal LLC

Prenscia
Engineering Solutions

HBM

Brüel & Kjær

VePRO: Vehicle Performance, Reliability & Operations - Analysis

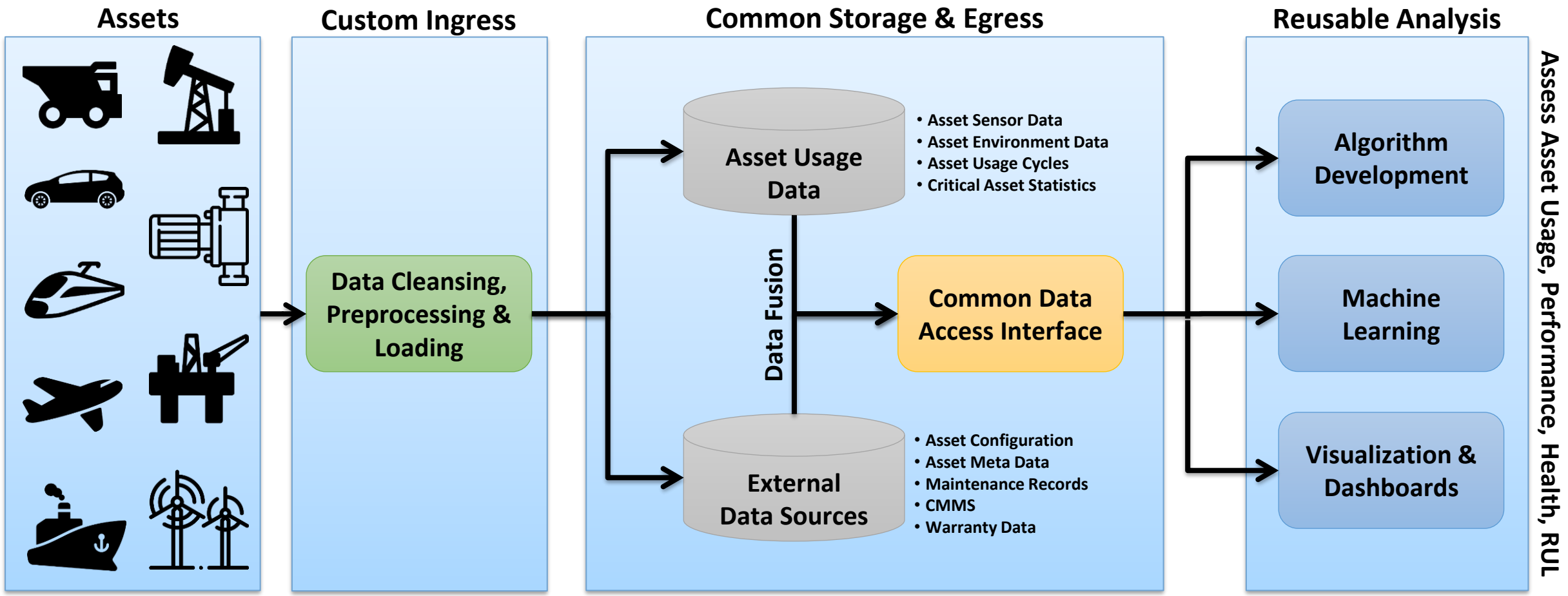
Program Scope:

- Provide an **end-to-end framework & components** for handling and processing CBM datasets
  - Allows end users to focus on using the data and extracting value rather than data handling
  - **Manage and analyze** operational usage data for up to 20,000 assets
- Focus on tasks including **data handling, cleaning, storage, analysis and reporting / visualization**
- Deliver **scalable enterprise level** capabilities
- Provide yearly incremental improvements and capabilities to a complete end-to end solution

Goal: Efficiently Manage, Process, Analysis, Refine & Contain

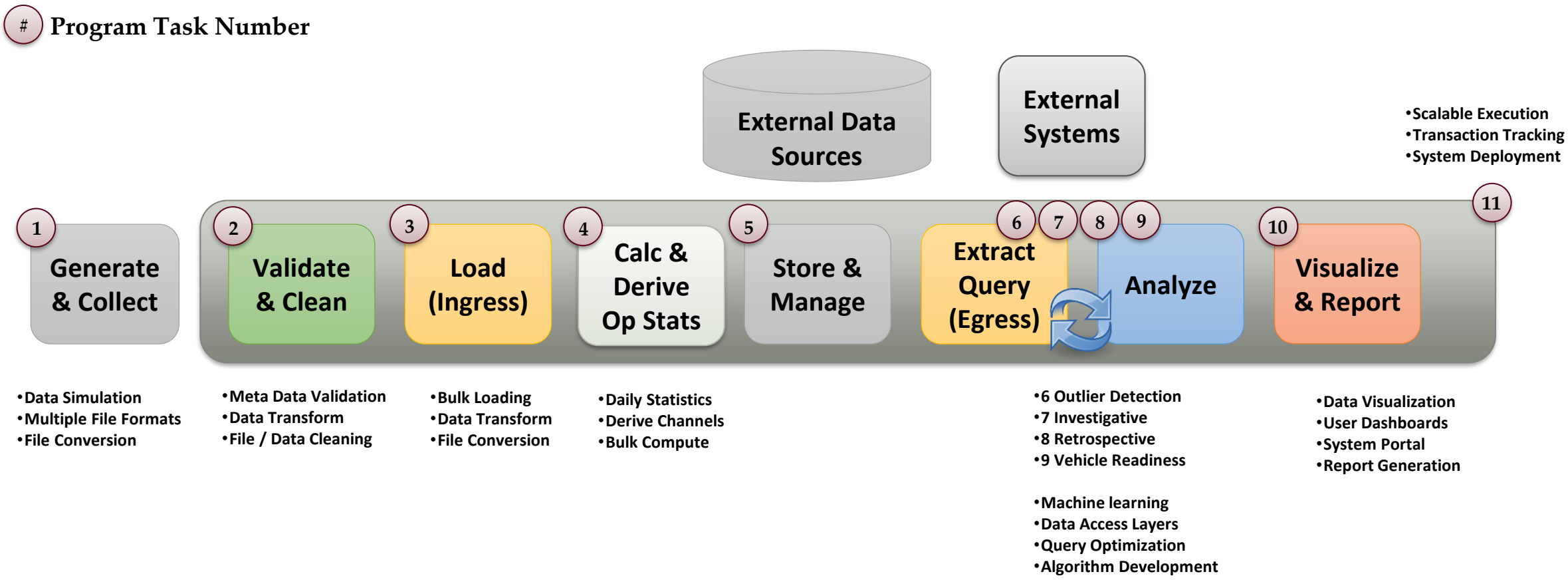
- 20,000 vehicles, 40 channels per vehicle, 2 hours of usage per day for 3 years
- ~ 11 million data files
- ~ 100TB of searchable time series data
- > “four trillion” timestamped data points

Custom, Common & Reusable Solutions



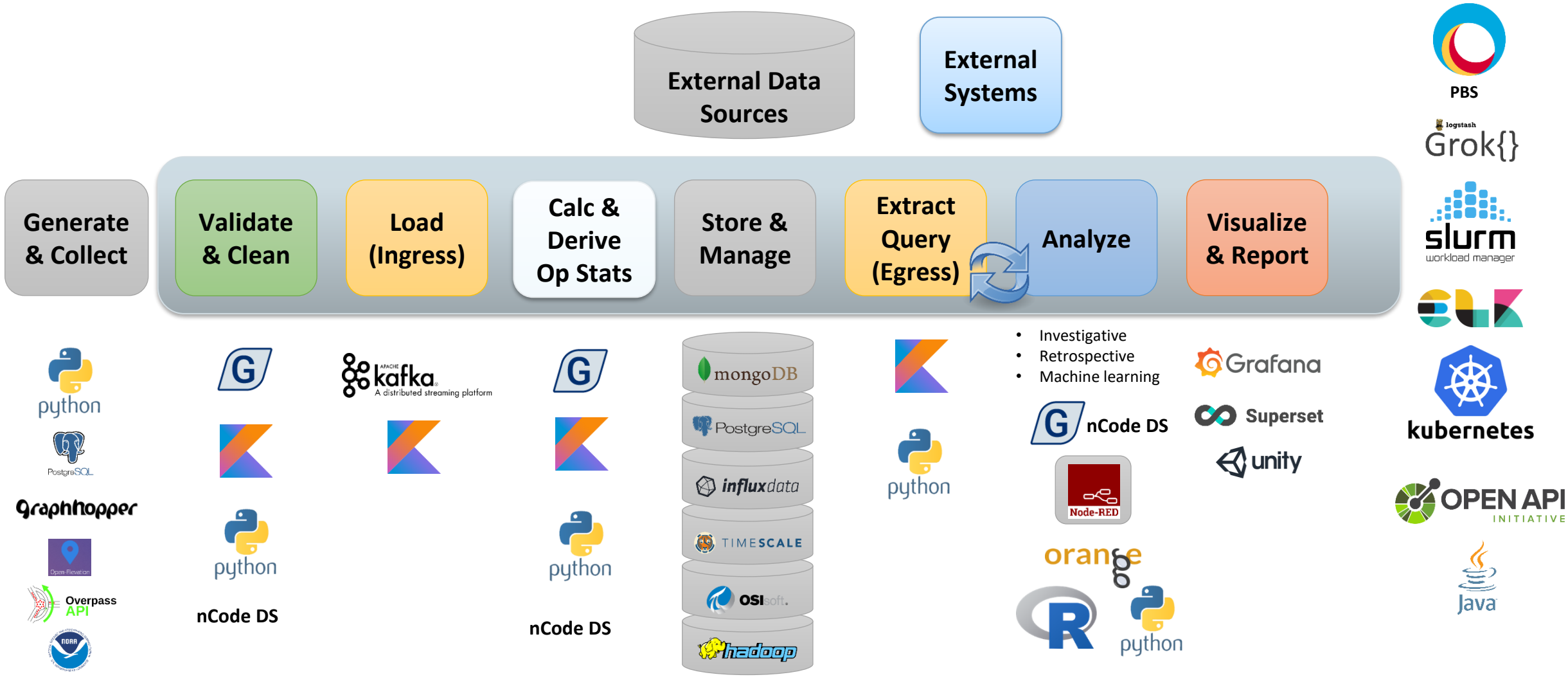
Enhancing the 3 D2D's at scale
Data-to-Discovery, Data-to-Decisions, Data-to-Dollars

VePRO Data Management Landscape



A modular, flexible & scalable solution that can be applied to enable enterprise level data management & analysis of assets

VePRO Technology Landscape

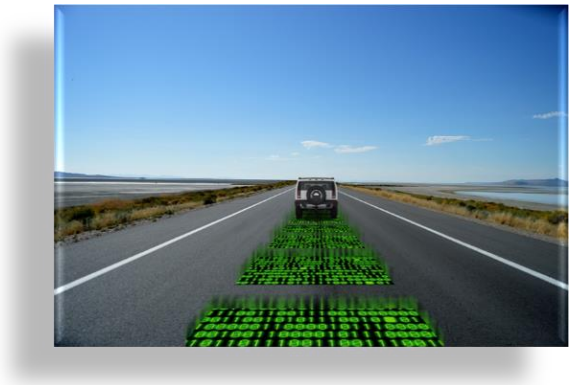


STRATEGY

1) Implement best in class open source solutions
2) Integrate leading commercial products where appropriate

Task 1 – Generation of Representative Data Set

- Vehicle Data Simulator
  - Data simulator generates usable & representative daily data
  - Provided volume data input to test/prove system scale
  - Can be used as a ‘digital twin’ to assist in forecasting activities (What if scenarios)



- Real Field Data

Summary of AMSAA Scrubbed Dataset	
Total Dataset Size	67.8 GB
Total Folders	6,816
Collection Date Range	2/24/2012 - 8/31/2014
Total # CDF Files	286,883
VINs	2,282
Models	78
Families	8
Collection Box Types	2
Unique Days of Data	713 out of 919

US Army does not have 20,000 vehicle instrumented, hence the need for both data types

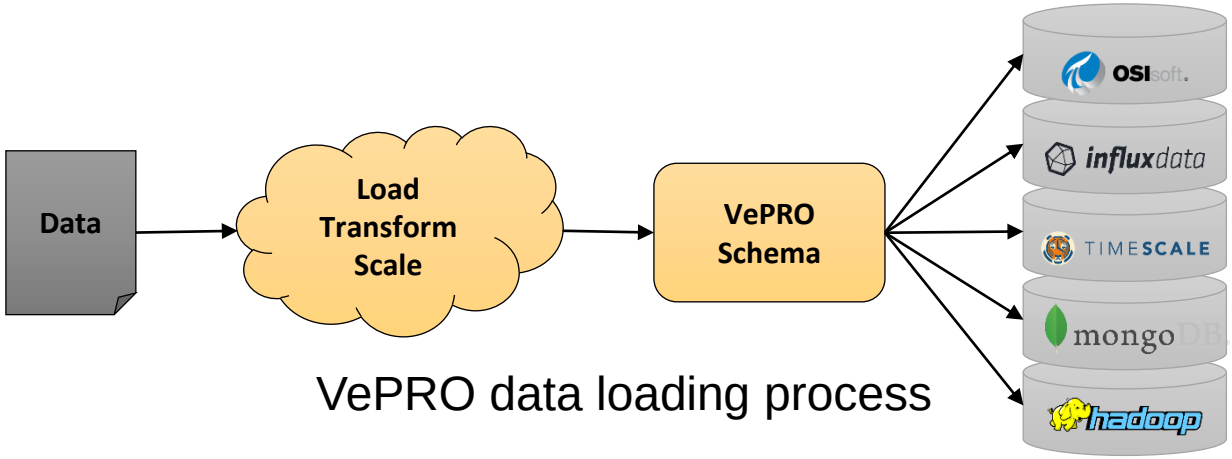
Task 2 - Data Validation & Cleaning

- Automated methods to assess field data file quality and fix known file issues
  - Common, automated & consistent cleaning process for all data files
  - Volume of data means limited manual intervention

  - Year 1 - Validate & fix files, meta data
  - Year 2 - Identifying, tagging and/or cleaning of anomalies, bad data and missing data

The value of the downstream processes relies on good quality data

Task 3 – Data Loading



VePRO data loading process

	VePRO Initial Investigation	VePRO Year 1	VePRO Year 2
Threads	1	6	6
Calc Stats (files/min)	-	206	1032
Index & Load (files/min)	2.31	172	999

Efficient data load is key to maximize throughput in 24hr window

Task 4 – Calculate & Derive Data

Types of Calculations

- Basic Channel Stats
  - Max, Min, Mean, StdDev
- Derived Channels & Features
  - New channel – values for every timestamp
  - Idle on/off
- Duration Calculation
  - Time at a condition
  - Example – Time at idle
- Multi Channel Formula / Stats
  - Example – Remaining oil life
- Rollup Stats
  - Daily / Monthly / Yearly Totals
  - Example – Total time at idle

Where & When in Process to Calculate

- Per database load
- As parallel streaming data
- Post database load
- Where is data coming from and going to

How

- Optimize data egress / ingress
- Optimize compute algorithms
- Parallel calculation execution

Calculation method grouping, data location & process methods drives design of scalable computational components

Task 5 – Data Store

Benchmarking

Key Characteristics

- Performance
- Optimized Schema Design for Ingress & Egress
- System Integration
- Total Cost of Ownership

	Notes	Open Source	HPC Friendly	Time Series Data	Meta Data	TS Statistical	Derived Data	Packaged Analytics
influxdata	• Open Source (TICK Stack) • Cluster Solution is **not** open source • Fast ingress / egress • SQL like syntax	Y/N	TBD	Y	N	Y	Y	N
TIMESCALE	• Open Source • 100% SQL database	Y	TBD	Y	Y	Y	Y	N
OPENTSDB	• Open Source • Runs on Hadoop and HBase • Scales to millions of writes per second • Add capacity by adding nodes	Y	TBD	Y	N	Y	Y	N
hadoop	• Open Source • Distributed and indexed file based data model	Y	Y	Y	Y	Y	Y	N
OSIsoft.	• Commercial Tool • Integrated analysis engine • Well established industry Tool • Only runs on Windows Server	N	N	Y	Y	Y	Y	Y
	• GPU database solution • Improved performance for aggregations, sorts, and grouping operations	Y/N	TBD	Y	Y	Y	Y	N

Efficient, flexible, scalable data store

Task 6 – Identification of Vehicle Outliers

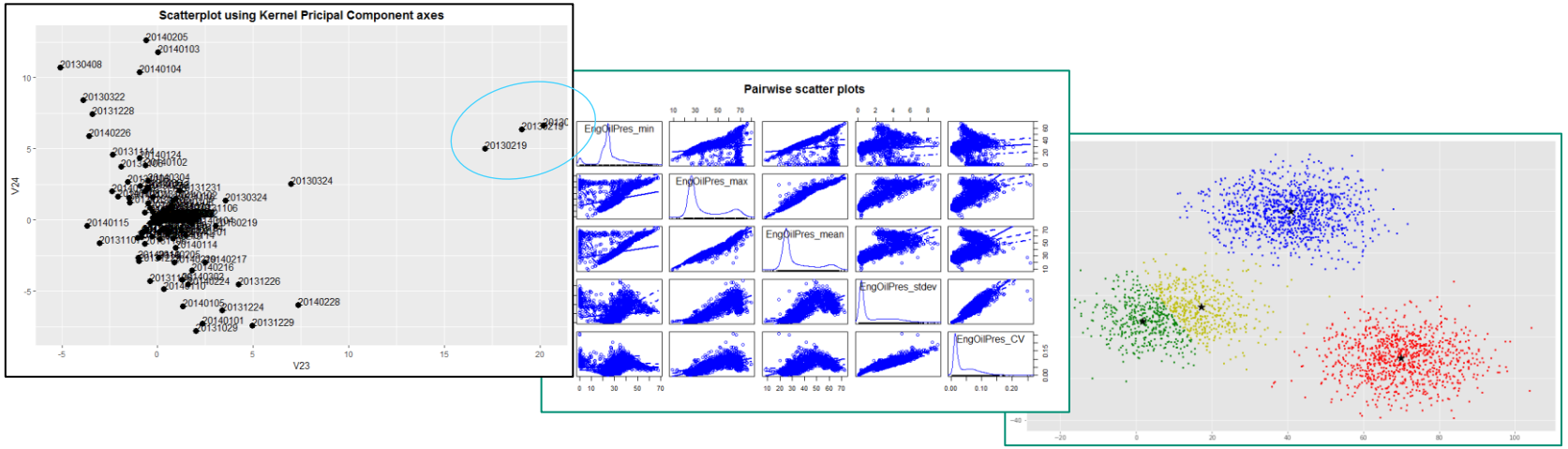
Which one is the outlier?

Strategy

- Feature based identification
  - Classification:
- Long/short term failure
- FWG
- Data driven identification

Methods

- Dimension Reduction Methods
- Distance Based Methods
- Clustering Methods
- Density Based Methods
- Time-indexed Methods



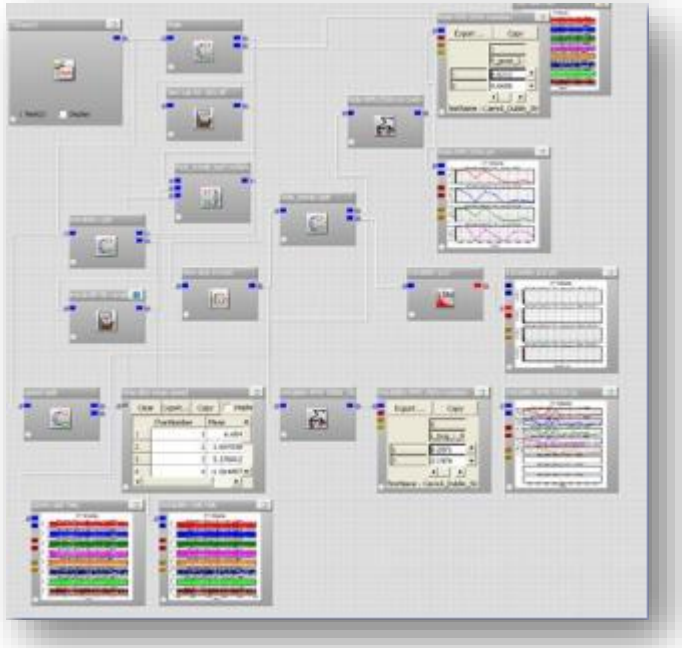
Identify vehicles with health concerns based on data trends

HBM nCode Federal LLC

HBK HOTTINGER BRÜEL & KJÆR

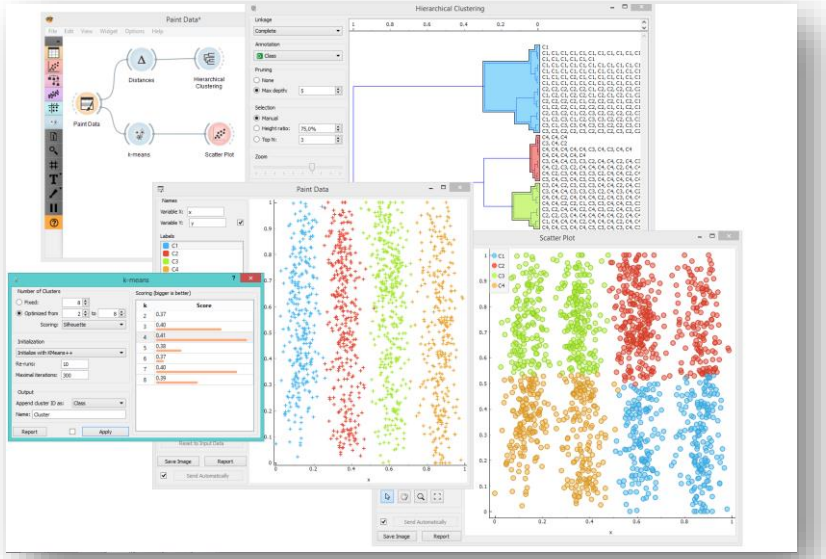
Task 7 – Interactive Analysis Domains

Interactive Data Analysis



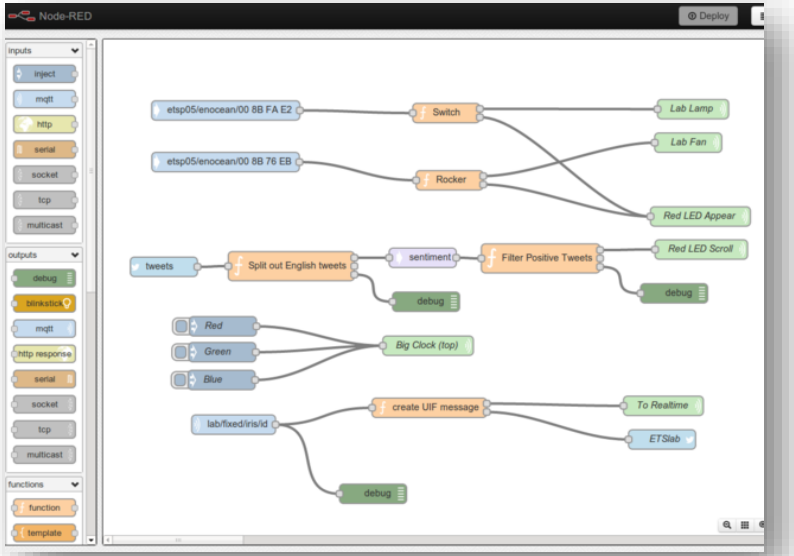
GlyphWorks nCode DS

Data Science Analysis



orange

IOT Data Streaming Analysis



Node-RED

Integration with multiple domains for Investigative Analysis

HBM nCode Federal LLC

HBK HOTTINGER BRÜEL & KJÆR

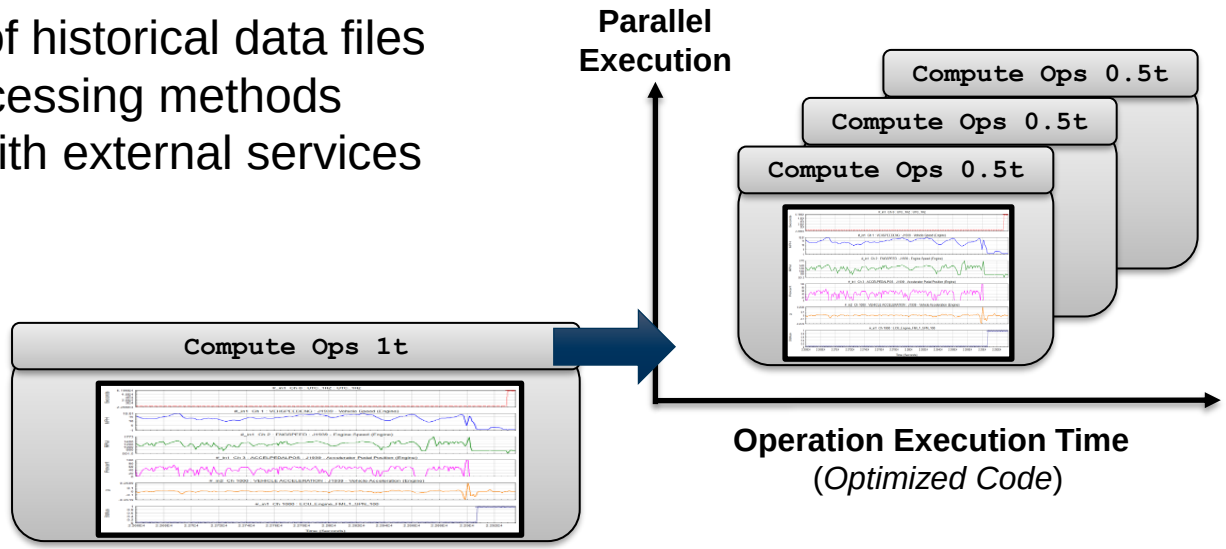
Task 8 – Retrospective Analysis

- Summary / Objective
  - Calculate and store new derived data for a new methodology from all historical measured data for an entire vehicle platform
  - Focus on calculation performance and throughput.

- Challenges
  - Execute analysis processes against a defined set of historical data files
  - Processing of large data sets requires efficient processing methods
  - Tradition HPC operations typically don't interface with external services (Web Services / Databases)
  - Data volume and throughput

Strategy

- Algorithm Optimization
- Scale through Parallel Execution

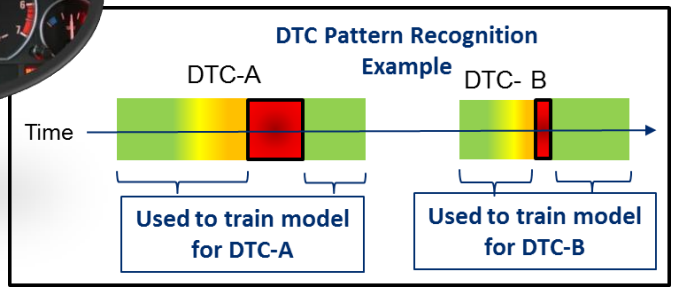


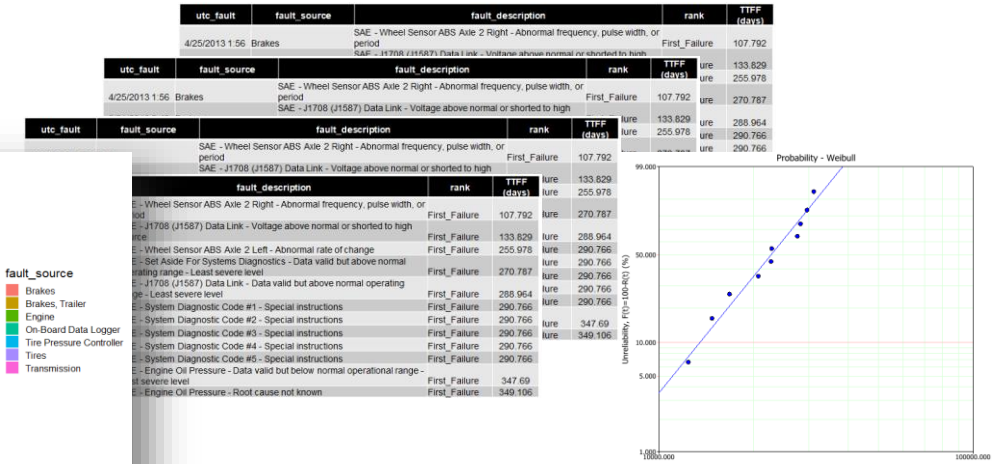
Apply new algorithms methods to large historic datasets

Task 9 – Vehicle Readiness based on Reliability

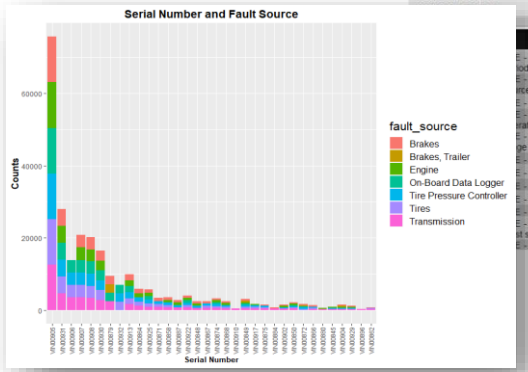
- Summary / Objective
  - Identify vehicles with readiness issues based on
    - Diagnostic Trouble Codes (DTCs)
    - Maintenance history
    - Vehicle usage data
- Challenges
  - Approach & algorithm development
  - Merging and analysis of data from multiple data sources (DTCs, time series, maintenance, etc.)
  - Data volume and throughput

Strategy





utc_fault	fault_source	fault_description	rank	TTFF (days)
4/25/2013 1:56	Brakes	SAE - Wheel Sensor ABS Axle 2 Right - Abnormal frequency, pulse width, or period	First_Failure	107.792
		SAE - J1708 (J1587) Data Link - Voltage above normal or shorted to high	First_Failure	133.829
		SAE - Wheel Sensor ABS Axle 2 Left - Abnormal rate of change	First_Failure	255.978
		SAE - Set Aside For Systems Diagnostics - Data valid but above normal operating range - Least severe level	First_Failure	270.787
		SAE - J1708 (J1587) Data Link - Data valid but above normal operating range - Least severe level	First_Failure	288.964
		SAE - System Diagnostic Code #1 - Special instructions	First_Failure	290.766
		SAE - System Diagnostic Code #2 - Special instructions	First_Failure	290.766
		SAE - System Diagnostic Code #3 - Special instructions	First_Failure	347.69
		SAE - System Diagnostic Code #4 - Special instructions	First_Failure	290.766
		SAE - System Diagnostic Code #5 - Special instructions	First_Failure	290.766
		SAE - Engine Oil Pressure - Data valid but below normal operational range - Least severe level	First_Failure	347.69
		SAE - Engine Oil Pressure - Root cause not known	First_Failure	349.106



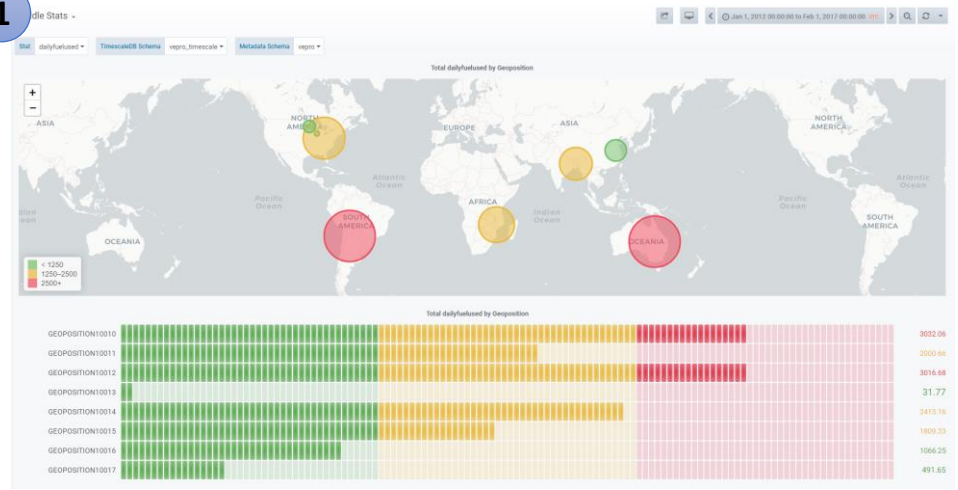
Identify vehicle readiness based on maintenance & usage data

HBM nCode Federal LLC

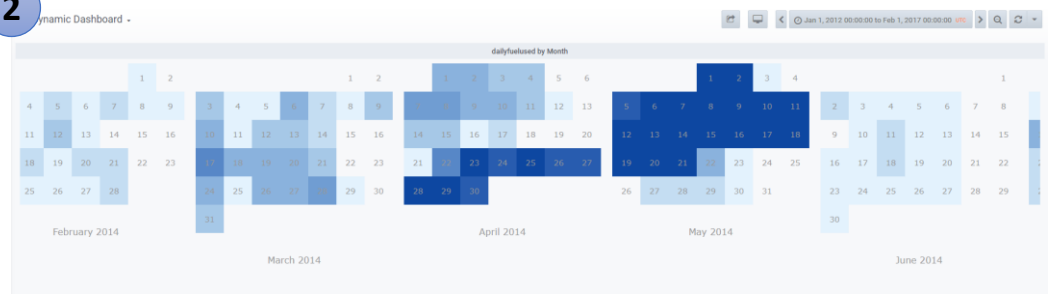
HBK HOTTINGER BRÜEL & KJÆR

Task 10 – Drilldown Dashboards - Sample

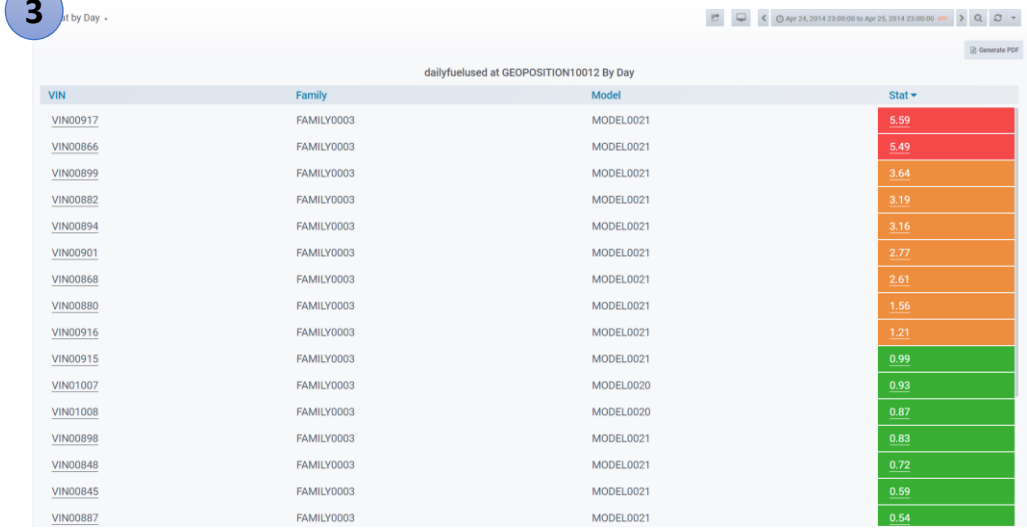
1



2



3



dailyfuelused at GEOPOSITION10012 By Day

VIN	Family	Model	Stat
VIN00917	FAMILY0003	MODEL0021	5.59
VIN00866	FAMILY0003	MODEL0021	5.49
VIN00899	FAMILY0003	MODEL0021	3.64
VIN00882	FAMILY0003	MODEL0021	3.19
VIN00894	FAMILY0003	MODEL0021	3.16
VIN00901	FAMILY0003	MODEL0021	2.77
VIN00868	FAMILY0003	MODEL0021	2.61
VIN00880	FAMILY0003	MODEL0021	1.56
VIN00916	FAMILY0003	MODEL0021	1.21
VIN00915	FAMILY0003	MODEL0021	0.99
VIN01007	FAMILY0003	MODEL0020	0.93
VIN01008	FAMILY0003	MODEL0020	0.87
VIN00898	FAMILY0003	MODEL0021	0.83
VIN00848	FAMILY0003	MODEL0021	0.72
VIN00845	FAMILY0003	MODEL0021	0.59
VIN00887	FAMILY0003	MODEL0021	0.54

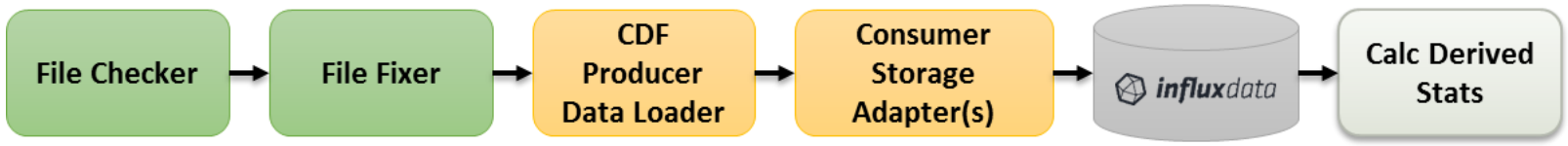
4

High level aggregated statistics to asset specific time series data in < 5 clicks

HBM nCode Federal LLC

HBK HOTTINGER BRÜEL & KJÆR

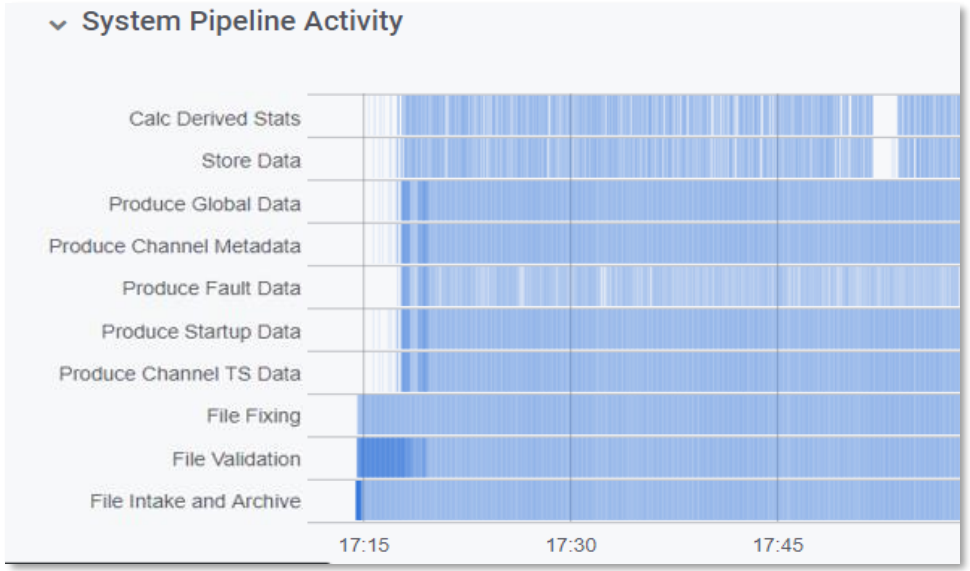
Task 11 - End To End Dynamic Process Control



Elastic End-End Thread Management Example

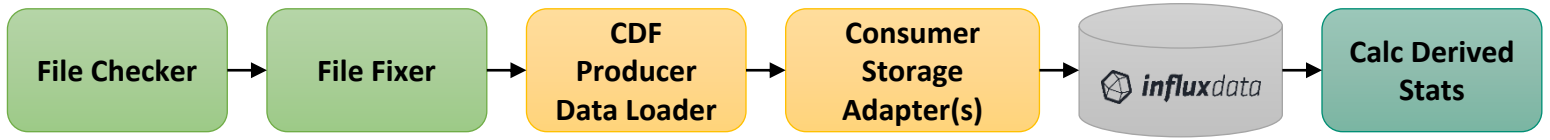
MODEL0021 (9287 files) on single VM (16 threads)

A. No elasticity control, single thread for each process
B. No elastic control, best guess fixed thread counts
C. Elastic control, system controlled thread management

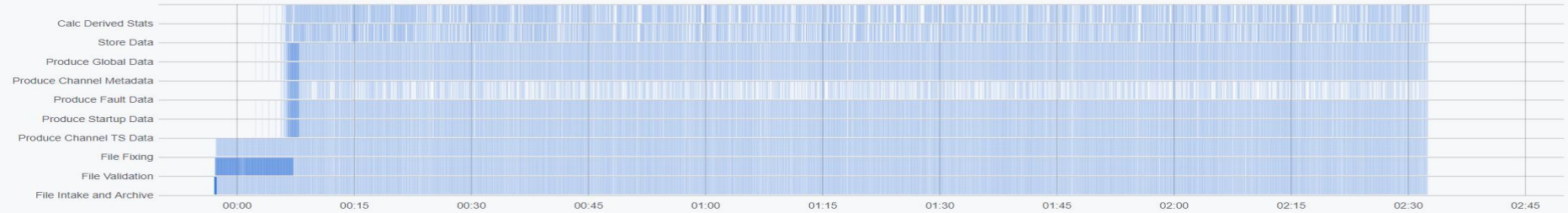


Scalable, optimized & automated process control

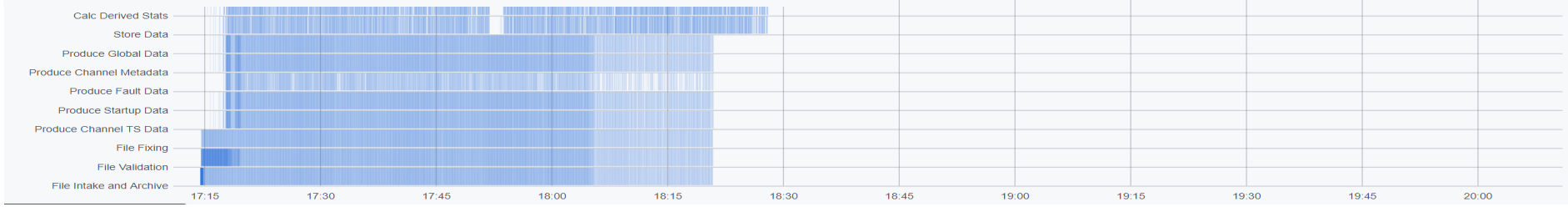
Task 11 – End-to-End Process Control



A. No elasticity control, single thread for each process — **Total Time: 2:35:30**



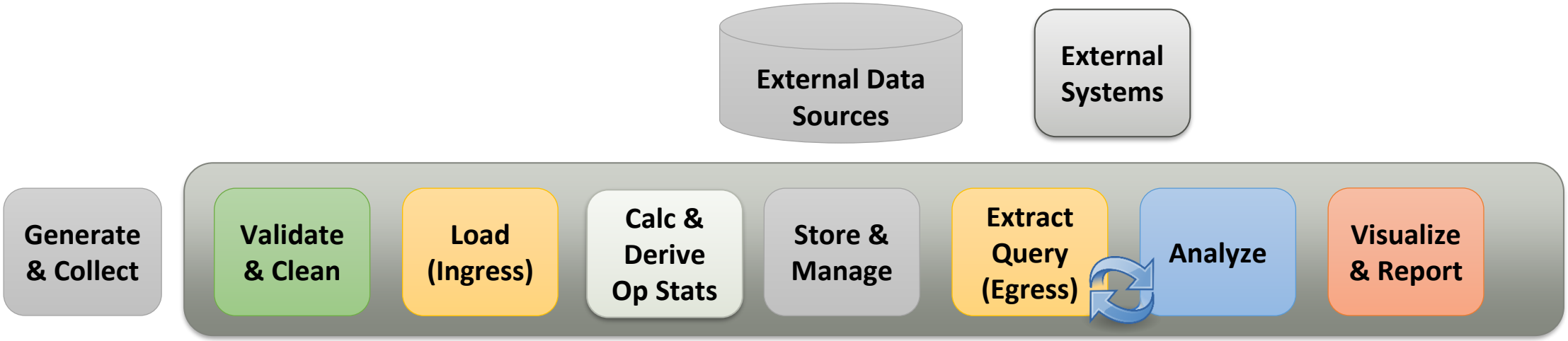
B. No elastic control, best guess fixed thread counts — **Total Time: 73:05**



C. Elastic control, system controlled thread management — **Total Time: 63:10**

MODEL0021 (9287 files) on single VM (16 threads)

Summary and Q&A



A modular, flexible & scalable solution that can be applied to enable enterprise level data management & analysis of assets

Thank You

HBM nCode Federal LLC

HBK
HOTTINGER BRÜEL & KJÆR

Q&A

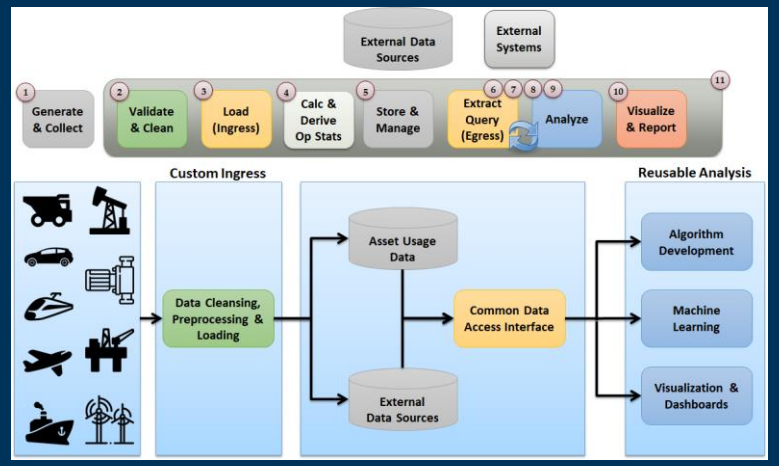
- Muhammad Afzal (Omnicell, USA)
  - I understand the strategy of data gathering, storing etc. What's the outcome in terms of fleet life improvement?
    - *Fleet life improvement is a result of the analysis performed on the data that's is stored and managed within the VePRO system. The VePRO system was created to perform these types of analysis at scale and elevate the analyst from the task of data wrangling. We have various customer that are at different stages in the maturity curve of fleet life predictions and view the VePRO architecture as an enabling tool to achieve their CBM visions.*
  - Are you using Tableau for data visualization?
    - *We have explored and used various visualization tools within the VePRO system. Examples include Grafana, Tableau, and even Unity Virtual Reality (VR) type environments. So yes, it is possible to structure the queries from the VePRO system to drive Tableau charts and reports.*

- Peter Bebbington (Brainpool Ltd, UK)
  - Do you use Apache Kakfa for the brokage of data between Assets sources and Ingress?
    - *Kafka is the core message bus that communicates data and execution commands between the various components of the VePRO system. If the asset can stream data directly to the VePRO system Kafka can be used as the entry point to catch the incoming data streams.*
  - Did you look at Kdb+ as a time series db option?
    - *Yes Kbd+ is a tool that we have on our radar to be assessed for possible future integrations.*
  - Since you are using Kubernettes, are you using kubeflow for managing the ML pipeline?
    - *We have primarily used Kubernetes as a distribution bundling mechanism for the various VePRO components. It also allows us to scale very easily within a cloud based environment. We have not evaluated Kubeflow as of yet to support our Machine Learning pipeline.*

HBK
HOTTINGER BRÜEL & KJÆR

HBM Prenscia Technology Day

Virtual Seminar Series



VePRO -A Program Overview –Scalable CBM

Presented by Dr. Mark A. Pompetzki
Director – Engineering Solutions
Prenscia Engineering Solutions

Session #4: **Vehicle Fleet Data Analytics**