A Unified Data Analytics Framework for Time-Series Data

Sundeep Tuteja, PACCAR

- Engineering Solutions team, has held multiple positions at PACCAR for over 12 years
- Prior Experience
  - Belcan Engineering Group, Engineering tool development for diesel engine data analytics
  - Howard Hughes Medical Institute, software tools to assist with computational neuroscience research.
- Master’s degree in electrical engineering from North Carolina State University and a bachelor's degree in instrumentation engineering from University of Mumbai.

MathWorks
AUTOMOTIVE CONFERENCE 2024
North America

A Unified Data Analytics Framework For Time-Series Data

Sundeep Tuteja

PACCAR Inc

(He/Him)

MathWorks

PACCAR Technical Center

- ✓ Established in 1982
- ✓ Located on 375 acres
- ✓ 210,000 sq. ft. of Laboratories
- ✓ 1.6 mile High Speed Test Track
- ✓ 1.5 mile Durability Test Track

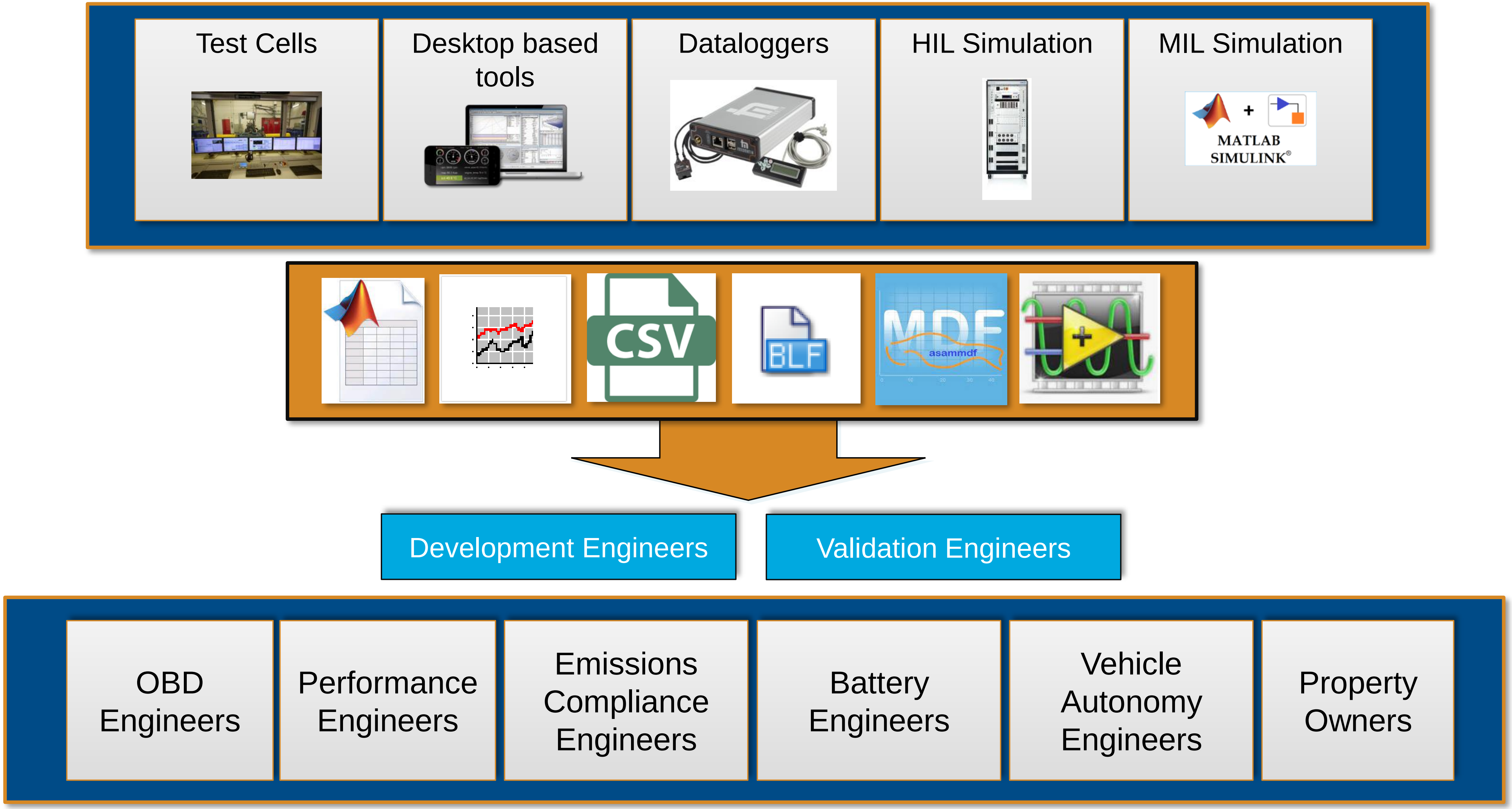
KENWORTH
Peterbilt
DAF
PACCAR MEXICO
PACCAR Australia
LEYLAND TRUCKS
PACCAR POWERTRAIN
PACCAR INFORMATION TECHNOLOGY
PACCAR INNOVATION CENTER
PacLease
PACCAR PARTS
DYNACRAFT A PACCAR COMPANY
PACCAR FINANCIAL
PACCAR WINCH DIVISION BRADEN CARCO Gearmatic
PACCAR TECHNICAL CENTER

- **Founded as Seattle Car Manufacturing Company in 1905**
- **Headquarters: Bellevue, WA**
- **Major manufacturer of heavy-duty and medium-duty trucks under the Kenworth, Peterbilt, and DAF nameplates**
- **Approximately 31000 employees**

Agenda

- Data collection and analysis scenarios
- Industry challenges
- Introducing Dashboard2
- Working with disparate data sources
- Working with remote data sources
- Automated report generation
- MathWorks tools and capabilities used
- Future capabilities
- The MathWorks tool ecosystem at PACCAR
- Acknowledgements

Data Collection And Analysis Scenarios


Test Cells
Desktop based tools
Dataloggers
HIL Simulation
MIL Simulation
MATLAB
SIMULINK
CSV
BLF
MDF
asammdf
Development Engineers
Validation Engineers
OBD Engineers
Performance Engineers
Emissions Compliance Engineers
Battery Engineers
Vehicle Autonomy Engineers
Property Owners

Challenges

- Multiple data formats and tool vendors
- Heterogenous datasets
- Data collection imperfections
- Interactive on-demand analysis
- Automated regular analysis
- Identification of data sets of interest

Dashboard2

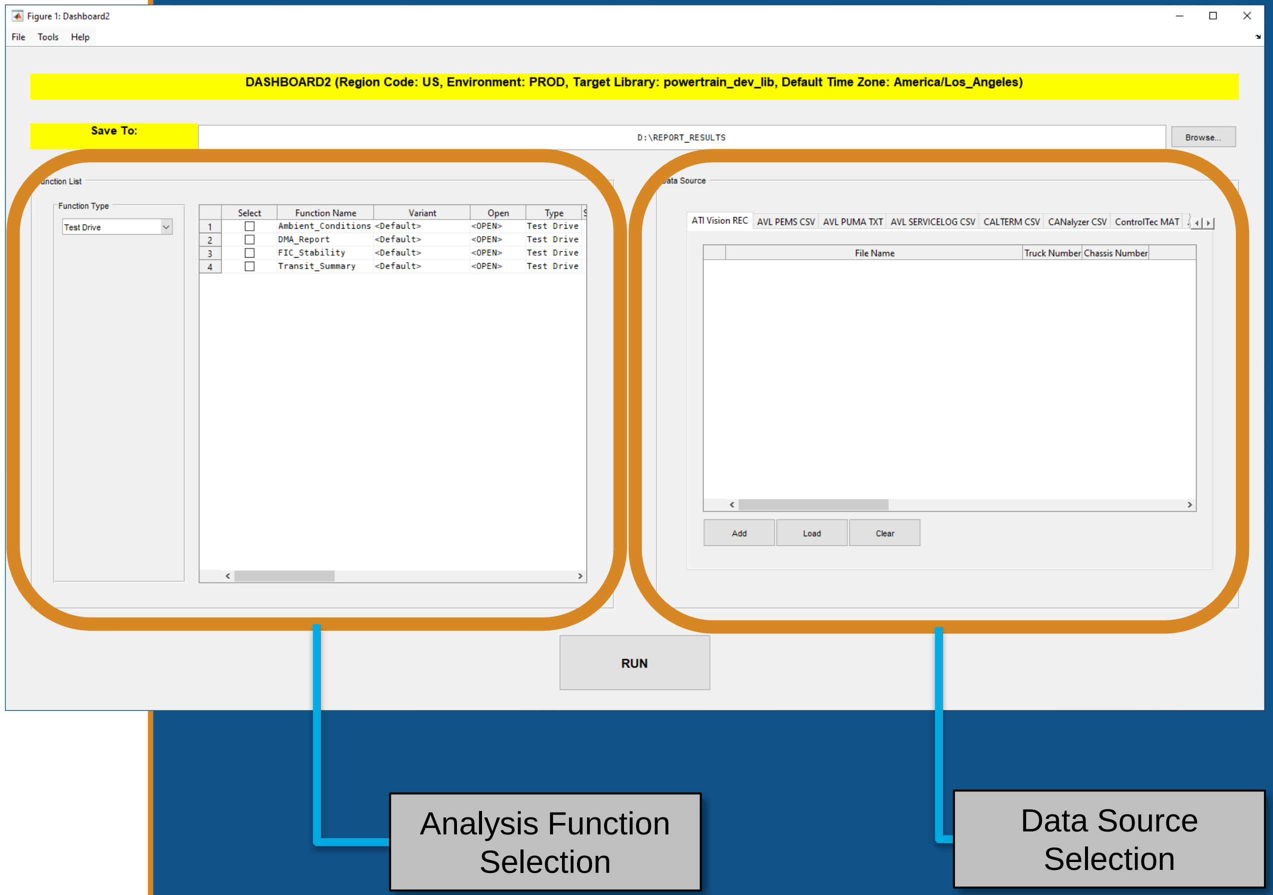


Figure 1: Dashboard2
File
Tools
Help
Browse...
Function List
Function Type
Test Drive
Data Source
ATI Vision REC
AVL PEMS CSV
AVL PUMA TXT
AVL SERVICELOG CSV
CALTERM CSV
CANalyzer CSV
ControlTec MAT
File Name
Truck Number
Chassis Number
Add
Load
Clear
RUN
Analysis Function Selection
Data Source Selection

DASHBOARD2 (Region Code: US, Environment: PROD, Target Library: powertrain_dev_lib, Default Time Zone: America/Los_Angeles)

Save To: D:\REPORT_RESULTS

	Select	Function Name	Variant	Open	Type
1	☐	Ambient_Conditions	<Default>	<OPEN>	Test Drive
2	☐	DMA_Report	<Default>	<OPEN>	Test Drive
3	☐	FIC_Stability	<Default>	<OPEN>	Test Drive
4	☐	Transit_Summary	<Default>	<OPEN>	Test Drive

- Unified data analytics tool supporting 16 file formats
- Configurable analysis parameters (signal alias lists, units, sampling rates)
- Analysis functions reside in a separate namespace, for independent development
- Interactive operation and command line operation
- Includes utilities for format independent visualization, merging, cleaning, querying, and time aligning data

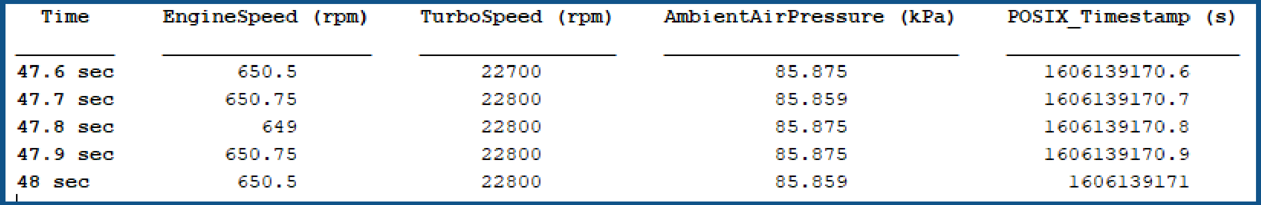
Dashboard2 Stages

LOADING

Data files are “skimmed” for high level information (headers, metadata, identifier names, starting timestamp) and an object is instantiated for each file.

PROCESSING

Required signals are extracted and processed into a uniform data representation consisting of MATLAB timetables. POSIX timestamps and source file identifiers are added, and the data is cleaned if applicable (utilizes parallelization)



```
   Time        EngineSpeed (rpm)    TurboSpeed (rpm)    AmbientAirPressure (kPa)    POSIX_Timestamp (s)
  ________     _________________    ________________    ________________________    ___________________

  47.6 sec           650.5               22700                   85.875                1606139170.6
  47.7 sec          650.75               22800                   85.859                1606139170.7
  47.8 sec             649               22800                   85.875                1606139170.8
  47.9 sec          650.75               22800                   85.875                1606139170.9
  48 sec             650.5               22800                   85.859                  1606139171
```

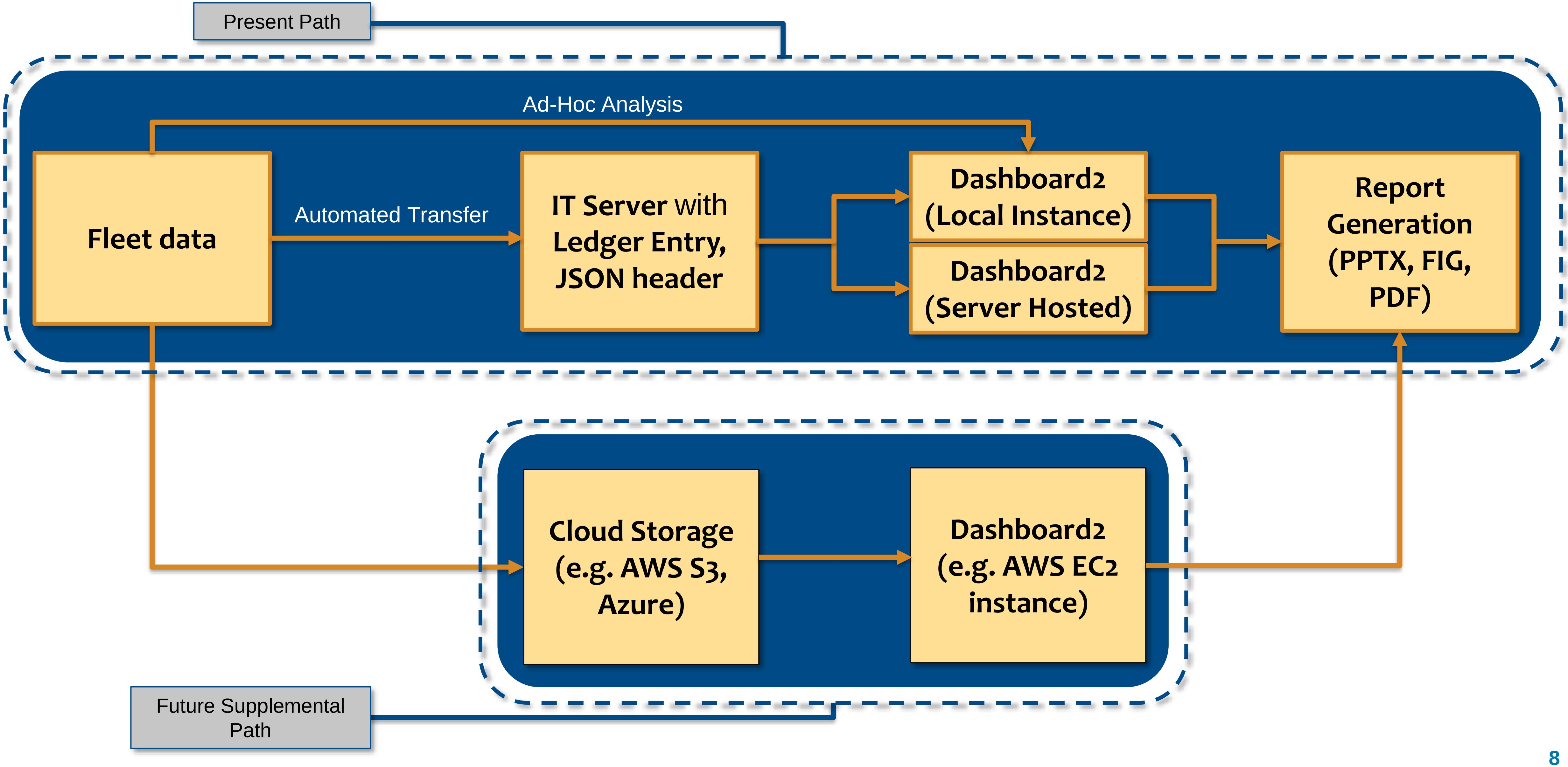
MERGING

If applicable, the extracted data store is merged and passed on to the following stage. Data can be processed sequentially (file by file) or all at once depending on data volumes and reporting requirements

ANALYZING

The data store representation is then analyzed to generate an interactive FIG file report. Analysis types include transient analysis and aggregate analysis

Data Pipeline



Present Path

Ad-Hoc Analysis

Fleet data

Automated Transfer

IT Server with **Ledger Entry, JSON header**

Dashboard2 (Local Instance)

Dashboard2 (Server Hosted)

Report Generation (PPTX, FIG, PDF)

Cloud Storage (e.g. AWS S3, Azure)

Dashboard2 (e.g. AWS EC2 instance)

Future Supplemental Path

Working with Disparate Data Sources

Different Data Sources Utilizing Different System Clocks

Reference Signal Selection

Shift Value

Automatic Alignment

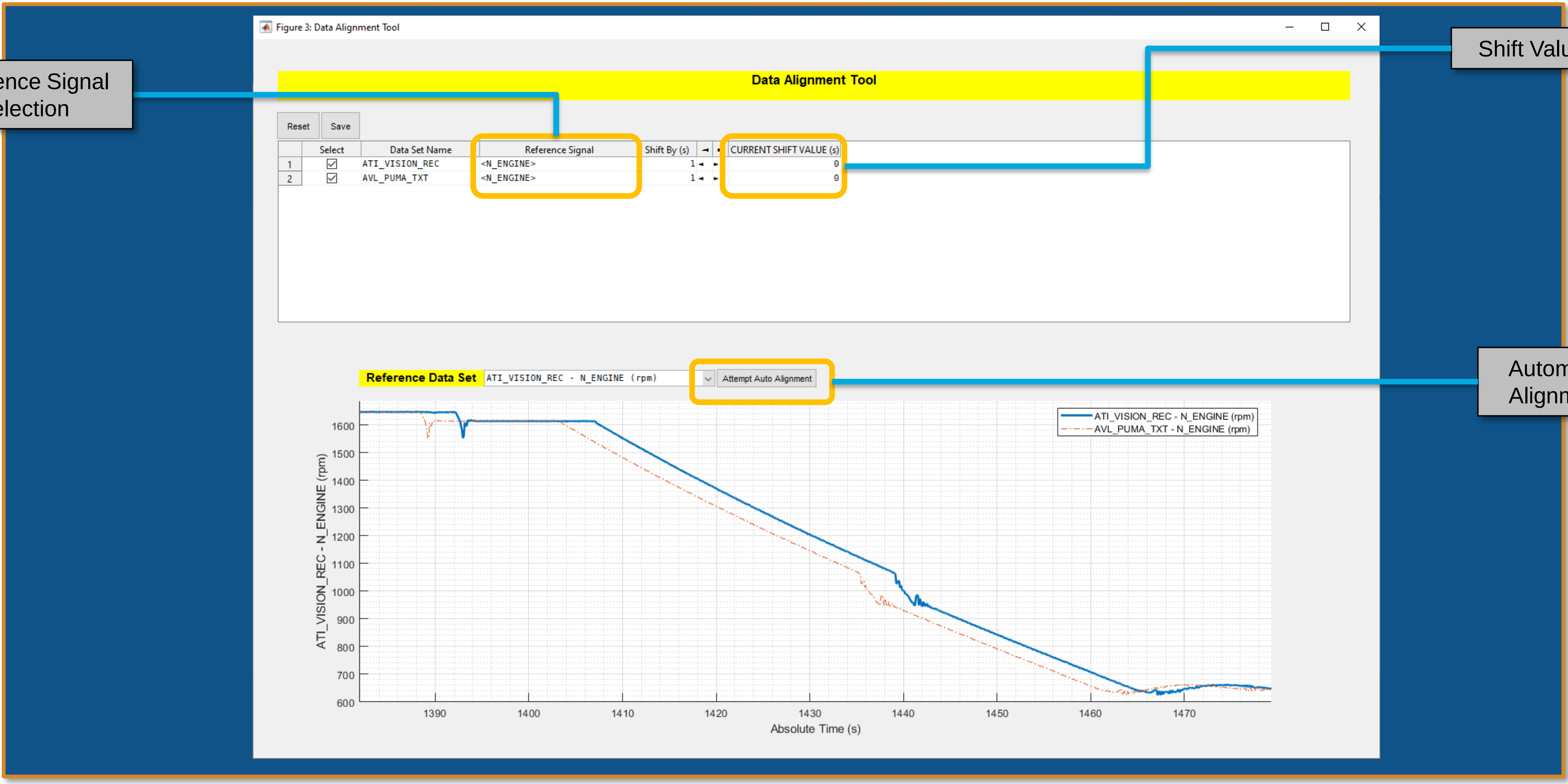


Figure 3: Data Alignment Tool
Data Alignment Tool
Reset
Save
Attempt Auto Alignment
ATI_VISION_REC - N_ENGINE (rpm)
AVL_PUMA_TXT - N_ENGINE (rpm)
ATI_VISION_REC - N_ENGINE (rpm)
1600
1500
1400
1300
1200
1100
1000
900
800
700
600
1390
1400
1410
1420
1430
1440
1450
1460
1470
Absolute Time (s)

	Select	Data Set Name	Reference Signal	Shift By (s)	◄	►	CURRENT SHIFT VALUE (s)
1	☑	ATI_VISION_REC	<N_ENGINE>	1	◄	►	0
2	☑	AVL_PUMA_TXT	<N_ENGINE>	1	◄	►	0

Reference Data Set ATI_VISION_REC - N_ENGINE (rpm)

Working with Disparate Data Sources

Different Data Sources Utilizing Different System Clocks

Reference Signal Selection

Shift Value

Automatic Alignment

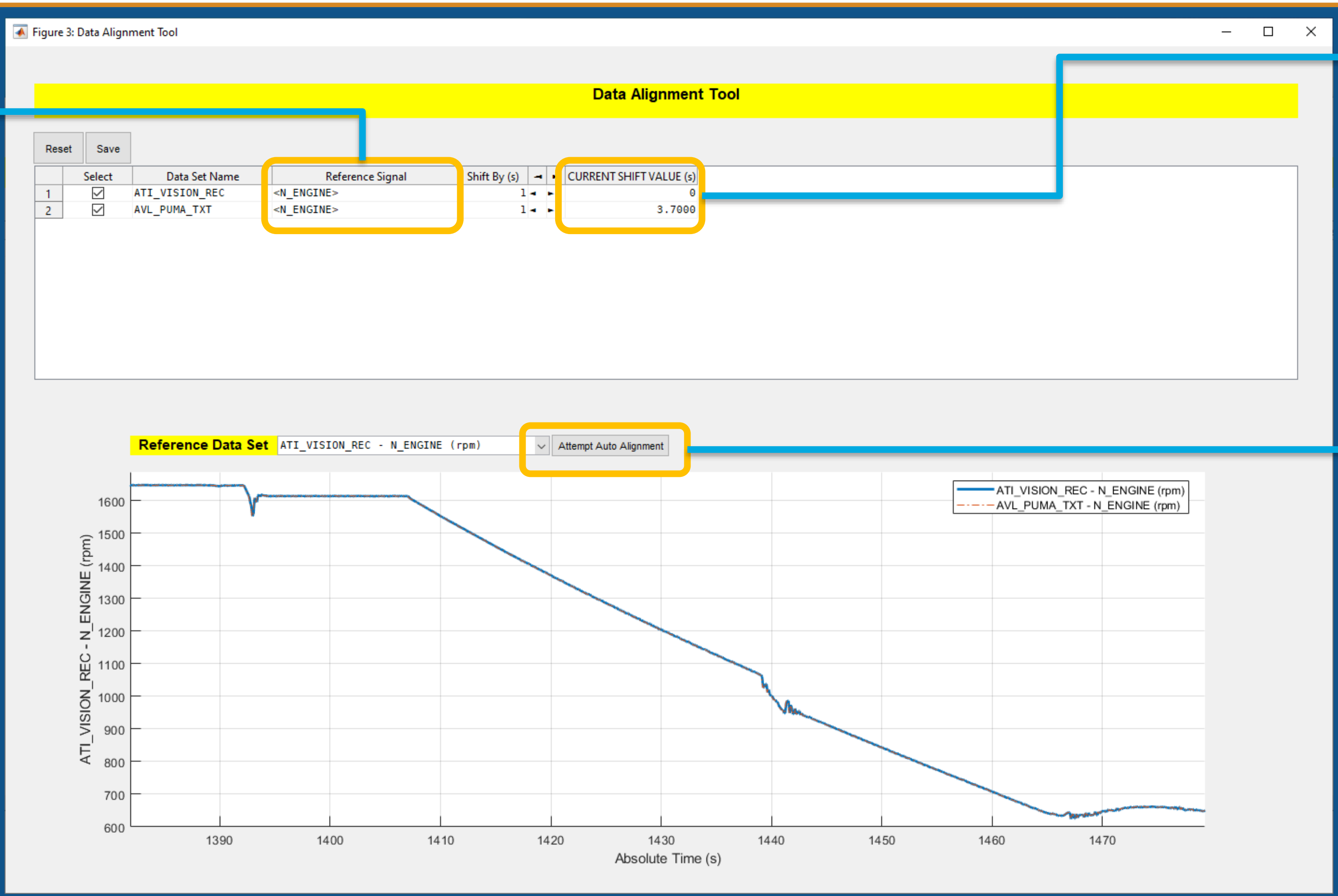


Figure 3: Data Alignment Tool

Data Alignment Tool

	Select	Data Set Name	Reference Signal	Shift By (s)	◄	►	CURRENT SHIFT VALUE (s)
1	☑	ATI_VISION_REC	<N_ENGINE>	1	◄	►	0
2	☑	AVL_PUMA_TXT	<N_ENGINE>	1	◄	►	3.7000

Reference Data Set ATI_VISION_REC - N_ENGINE (rpm)

Reset
Save
Attempt Auto Alignment
ATI_VISION_REC - N_ENGINE (rpm)
AVL_PUMA_TXT - N_ENGINE (rpm)
ATI_VISION_REC - N_ENGINE (rpm)
600
700
800
900
1000
1100
1200
1300
1400
1500
1600
1390
1400
1410
1420
1430
1440
1450
1460
1470
Absolute Time (s)

Working with Remote Data Sources

- Large numbers of data files are streamed daily from vehicles in the field and archived in remote locations
- High level metadata is automatically extracted and entered into a file ledger (SQL Server Database), and queried using this utility

Primary Query Parameters

Serial Number Selection

Header Queries

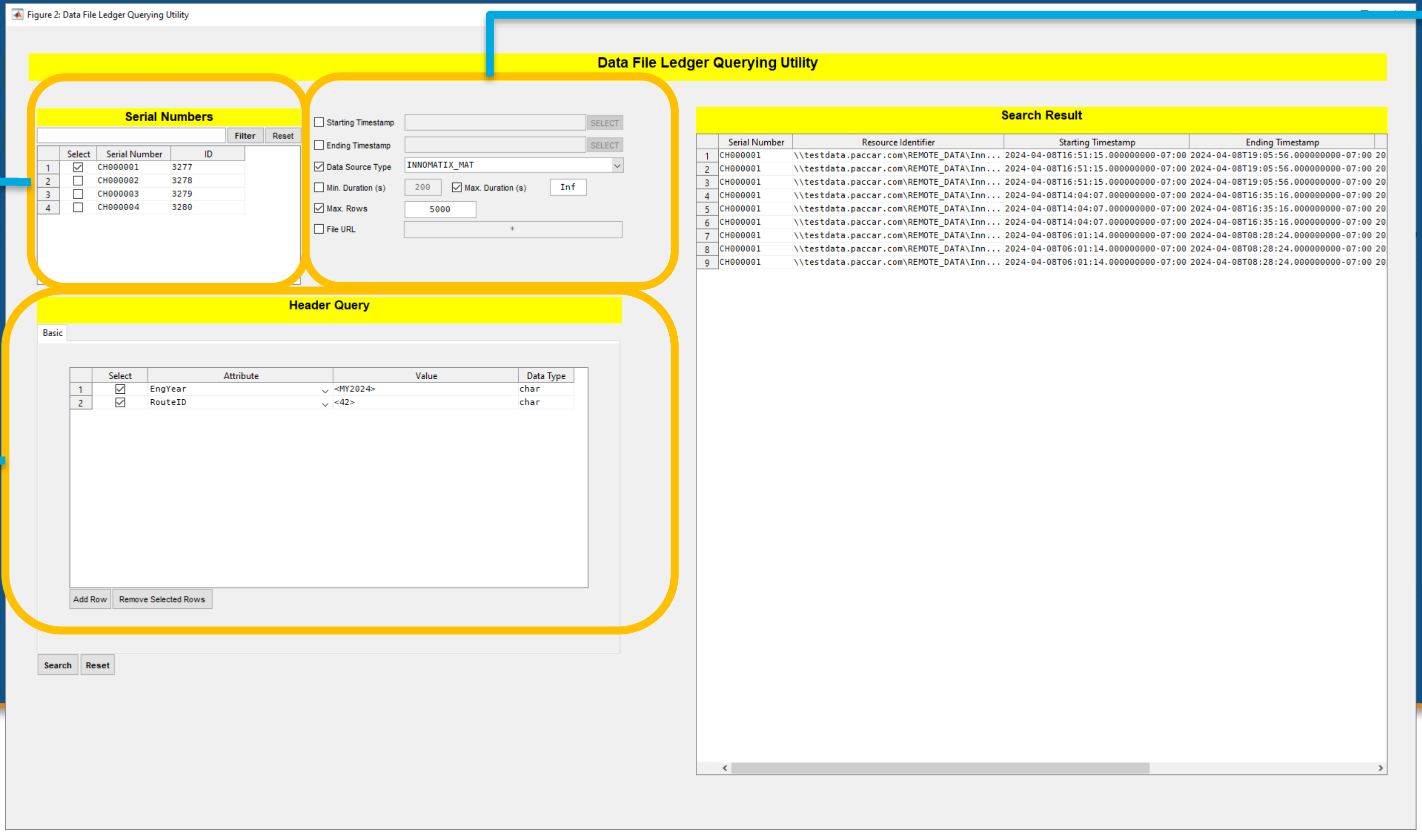


Figure 2: Data File Ledger Querying Utility

Data File Ledger Querying Utility

Serial Numbers

	Select	Serial Number	ID
1	☑	CH000001	3277
2	☐	CH000002	3278
3	☐	CH000003	3279
4	☐	CH000004	3280

☐ Starting Timestamp

☐ Ending Timestamp

☑ Data Source Type: INNOMATIX_MAT

☐ Min. Duration (s): 200

☑ Max. Duration (s): Inf

☑ Max. Rows: 5000

☐ File URL: *

Header Query

	Select	Attribute	Value	Data Type
1	☑	EngYear	<MY2024>	char
2	☑	RouteID	<42>	char

Search Result

	Serial Number	Resource Identifier	Starting Timestamp	Ending Timestamp	
1	CH000001	\\testdata.paccar.com\REMOTE_DATA\Inn...	2024-04-08T16:51:15.000000000-07:00	2024-04-08T19:05:56.000000000-07:00	20
2	CH000001	\\testdata.paccar.com\REMOTE_DATA\Inn...	2024-04-08T16:51:15.000000000-07:00	2024-04-08T19:05:56.000000000-07:00	20
3	CH000001	\\testdata.paccar.com\REMOTE_DATA\Inn...	2024-04-08T16:51:15.000000000-07:00	2024-04-08T19:05:56.000000000-07:00	20
4	CH000001	\\testdata.paccar.com\REMOTE_DATA\Inn...	2024-04-08T14:04:07.000000000-07:00	2024-04-08T16:35:16.000000000-07:00	20
5	CH000001	\\testdata.paccar.com\REMOTE_DATA\Inn...	2024-04-08T14:04:07.000000000-07:00	2024-04-08T16:35:16.000000000-07:00	20
6	CH000001	\\testdata.paccar.com\REMOTE_DATA\Inn...	2024-04-08T14:04:07.000000000-07:00	2024-04-08T16:35:16.000000000-07:00	20
7	CH000001	\\testdata.paccar.com\REMOTE_DATA\Inn...	2024-04-08T06:01:14.000000000-07:00	2024-04-08T08:28:24.000000000-07:00	20
8	CH000001	\\testdata.paccar.com\REMOTE_DATA\Inn...	2024-04-08T06:01:14.000000000-07:00	2024-04-08T08:28:24.000000000-07:00	20
9	CH000001	\\testdata.paccar.com\REMOTE_DATA\Inn...	2024-04-08T06:01:14.000000000-07:00	2024-04-08T08:28:24.000000000-07:00	20

Automated Report Generation

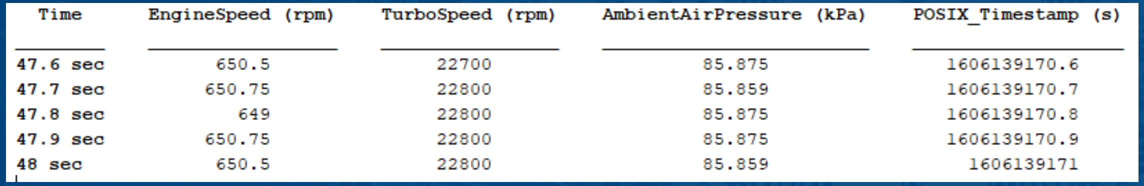
Done in conjunction with the previously described Data File Ledger

```
Dashboard2_US_OBD_DEV.exe Action runAutomatedReport FunctionName <AnalysisFunctionName>
StartDateVec "[2023 10 1 0 0 0]" StopDateVec "[2023 10 2 23 59 59.999]" SerialNumberList
<ChassisNumberList> VariantName <VariantName> EmailList <EmailList>
```

- **Action runAutomatedReport:** Required to run the automated report
- **FunctionName <AnalysisFunctionName>:** Which function is to be run
- **StartDateVec "[2023 10 1 0 0 0]":** The starting timestamp [year month day hours minutes seconds], in the local system time zone
- **StopDateVec "[2023 10 1 23 59 59.999]":** The ending timestamp [year month day hours minutes seconds], in the local system time zone
- **SerialNumberList <ChassisNumberList>:** A comma separated list of chassis numbers of interest
- **VariantName <VariantName>:** Specifies the signal list variant to be utilized for this execution
- **EmailList <EmailList>:** Comma separated list of email addresses

Backend

- Utilizing POSIX timestamps as a uniform representation of timestamps with a high level of precision, using MATLAB’s “posixtime” functions

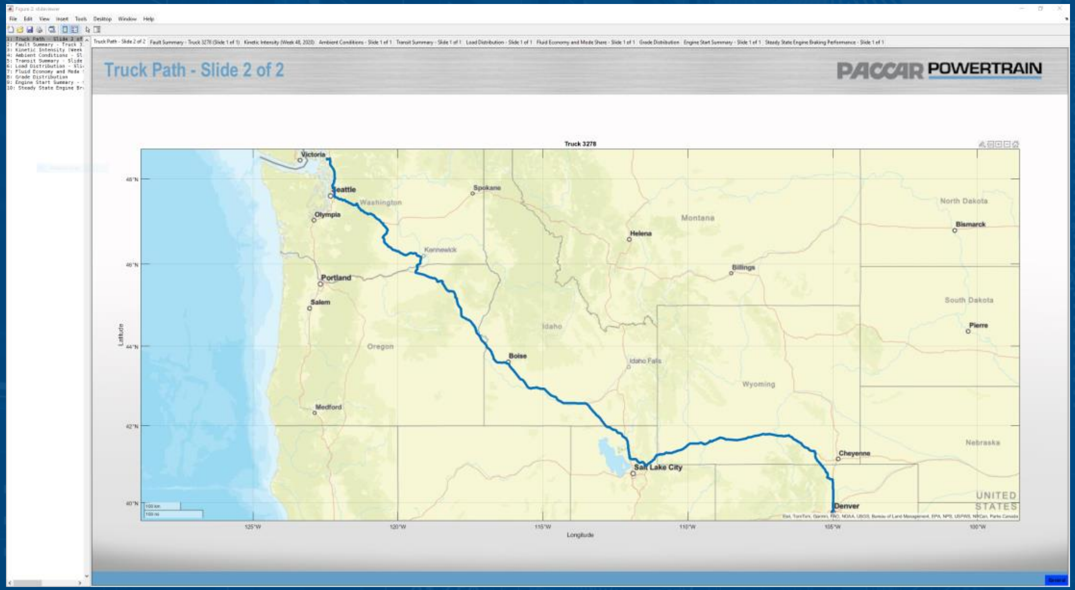


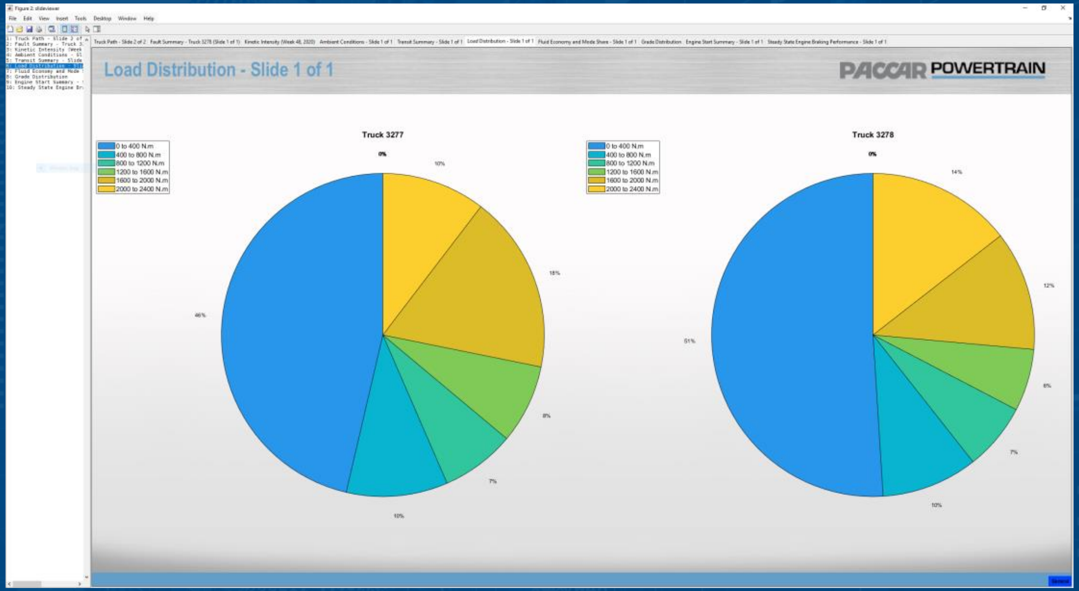
Time	EngineSpeed (rpm)	TurboSpeed (rpm)	AmbientAirPressure (kPa)	POSIX_Timestamp (s)
47.6 sec	650.5	22700	85.875	1606139170.6
47.7 sec	650.75	22800	85.859	1606139170.7
47.8 sec	649	22800	85.875	1606139170.8
47.9 sec	650.75	22800	85.875	1606139170.9
48 sec	650.5	22800	85.859	1606139171

- A specific reader function designed for each data format to ensure that only the data applicable for a specific report is ever loaded, thus allowing one data collection cycle to serve multiple stakeholders
  - MATLAB’s “load” and “matfile” functions: Permit loading only the necessary labels from a MAT file
  - MATLAB’s “readtable” function: Permits the selection of specific columns in a delimited ASCII file (like CSV)

- Parallel computing allows processing of multiple files simultaneously, utilizing upto 32 workers

Example Reports

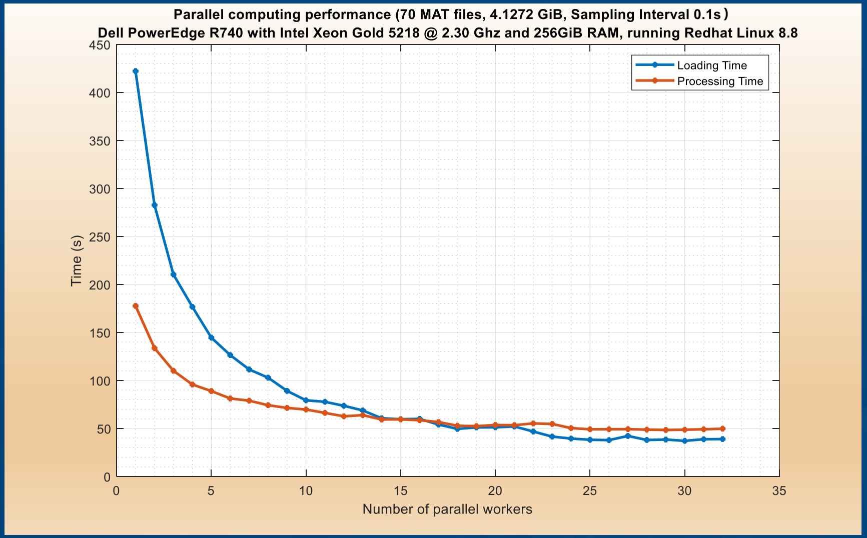

Truck Path - Slide 2 of 2
PACCAR POWERTRAIN
Truck 3278
Victoria
Seattle
Olympia
Washington
Spokane
Kennewick
Portland
Salem
Oregon
Boise
Idaho
Idaho Falls
Medford
Montana
Helena
Billings
Wyoming
Salt Lake City
Cheyenne
Denver
North Dakota
Bismarck
South Dakota
Pierre
Nebraska
UNITED STATES
Latitude
Longitude


Load Distribution - Slide 1 of 1
PACCAR POWERTRAIN
Truck 3277
Truck 3278
0 to 400 N.m
400 to 800 N.m
800 to 1200 N.m
1200 to 1600 N.m
1600 to 2000 N.m
2000 to 2400 N.m

- Reports are generated in a tabbed MATLAB FIG file by default
- Distributable, viewable from the compiled tool without MATLAB
- All of MATLAB’s interactive plotting and annotation tools are available within this window
- Data traceability can be incorporated

MathWorks Tools and Capabilities Used

- **Parallel Computing Toolbox**: Utilization of up to 32 workers for improved data processing speeds



- **Database Toolbox:** Cross platform connectivity to SQL Server
- **MATLAB Interfaces to Java:** Used for Java’s built-in cryptography routines
- **Vehicle Network Toolbox:** Adds support for MF4 files and Vector CAN DBC files
- **MATLAB tables and timetables:** Form the basis of the abstraction layer provided by Dashboard2
- **MATLAB Compiler:** Compiles Dashboard2 targeted towards different stakeholders

Future Capabilities

- Integration with cloud storage
- Archiving of processed data in a queryable form in a cloud storage area, such that it can serve all stakeholders
- Scaling with the MATLAB Parallel Cluster, interfacing with Simulink based tools for generating Simulated data
- Availability of reports through a “Software as a Service” (SaaS) model

The MathWorks tool ecosystem at PACCAR

- Software release tools to prepare memory images for programming various ECUs
- Vehicle simulation tools for optimizing calibration work
- Rapid controls prototyping tools and frameworks utilizing MATLAB Coder, Simulink Coder, and Embedded Coder

MathWorks
AUTOMOTIVE CONFERENCE 2024
North America

Thank you

Acknowledgements

- Our supervisors Mike Brennan and Rob Orlowski for encouraging us to present at this event
- The MathWorks team's Nishant Singh and Veronica Ma for inviting us
- Our colleagues for helping make this project cover a wide variety of areas
  - Mark Woodland
  - Matt Swart
  - Maryia Shautsova
  - James Castner
- All of our users that provided valuable feedback towards the improvement of the tool and our IT department for supporting infrastructure needs

MathWorks®

© 2024 The MathWorks, Inc. MATLAB and Simulink are registered trademarks of The MathWorks, Inc. See *mathworks.com/trademarks* for a list of additional trademarks. Other product or brand names may be trademarks or registered trademarks of their respective holders.