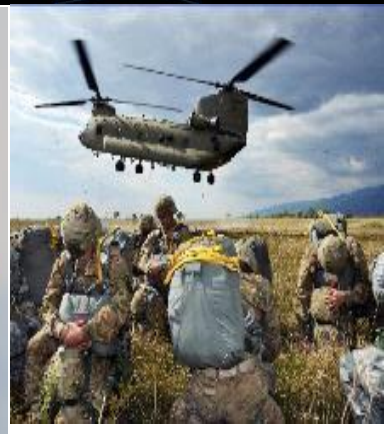
Enabling Analysis of Large-Scale Maintenance Data through Innovative Techniques

Engineered Resilient Systems Track
NDIA Systems and Mission Engineering Conference 2018
October 24, 2018

ERDC: Maria Seale, Andrew Strelzoff, Alicia Ruvinsky, Daniel Martinez, Amanda Hines, Grace Nabholz, Joshua Church, Stone Abdullah, Glenn Bond, Wesley Brewer, Owen Eslinger; Technical Directors: David Richards, Rob Wallace, Cary Butler, Dave Stuart, Dharhas Pothina
AMRDEC: Daniel Wade, Andrew Wilson, Nathan Rigoni



ITL INFORMATION TECHNOLOGY LABORATORY

AMRDEC

ERS ENGINEERED RESILIENT SYSTEMS DEPARTMENT OF DEFENSE

Topics

- Introduction
- Big Data Management
- Sensor Data Reconstruction
- Automated Logbook Labeling
- Integrated Data Engineering for Automated Labeling (IDEAL)

Distribution A: Approved for public release

Introduction

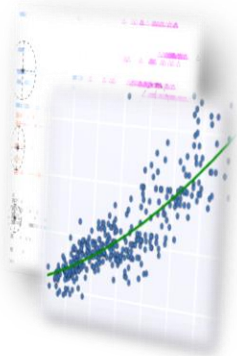
Motivation:

- Reduce maintenance costs
- Increase vehicle and component useful life
- Improve analysis and reporting
- Influence future platform designs

Challenges:

- Technological infeasibility of managing massive data sets
- Isolated nature of different types of maintenance data
- Infeasibility of applying manual analysis techniques over large, dispersed data

The Analytics Process

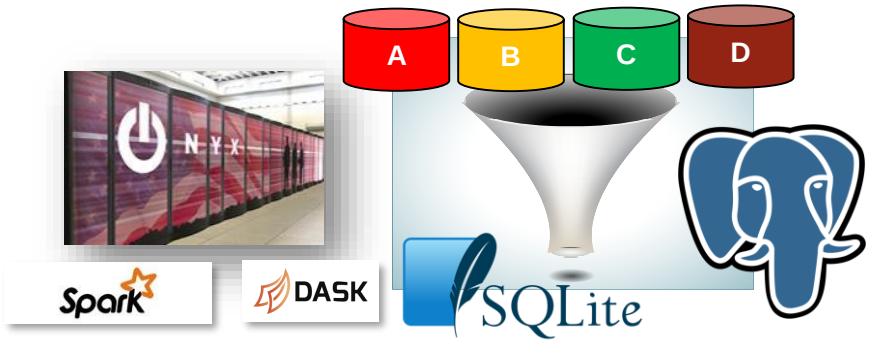


TB-Scale Datasets → **HPC Ecosystem** → **Data Analytics** → **Deployment**

Distribution A: Approved for public release

Research Areas

Big Data Management



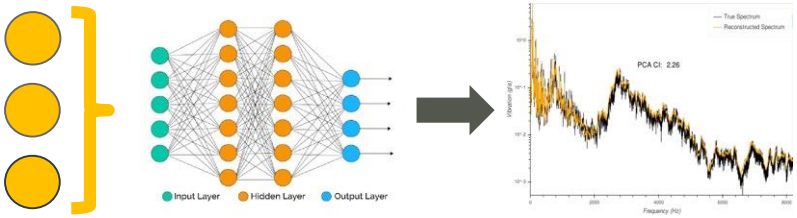
Natural Language Processing

Natural Language Text Fields

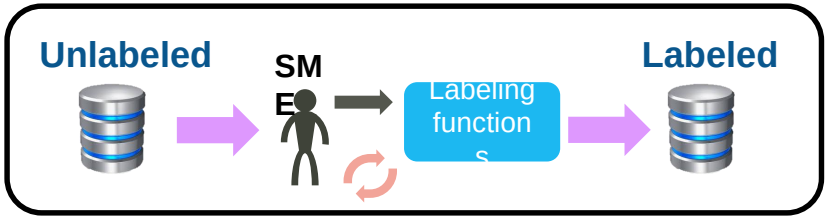
Labels to predict

NARR	CORR_NARR	SCD1	SCD2	RFG
WATER LEAK CHECK REQUIRED FOR REPLACEME...	INSP OK	C	N	02A01A
OPERATIONAL CHECK OF WINDSHIELD ANTI-ICE ...	MOC COMPLETE	C	N	02A01A
PMD	INSP OK	S	N	02_PMD
2ND 9-11 HR TQ CHECK REQ ON RED M/R DAM...	COMPLETED PASSED AT 1320 IN LBS	C	N	05A01Z01
INSP A020 - PERFORM M/R SPINDLE LUGS, M/R ...	INSP OK COMPLETED WITH PMD	S	N	05A01I
PMD	COMPLETED	S	N	02_PMD
PMD	PMD COMPLETED 4MAY SEE PMD E...	S	N	02_PMD
INSP A100 - #1 ENGINE HISTORY RECORDER REA...	COMPLETED	S	N	04A068
INSP A101 - #2 ENGINE HISTORY RECORDER REA...	COMPLETED	S	N	04A068
MED #2 FAILCAUTION LIGHT ILLUMINATED 3 TI...	CHECK FOUND OK COULD NOT DU...	U	U	08817
		U	H	04A06

Deep Learning & Machine Learning



Automated Labeling



Big Data Management
Why?

Enable decision makers to have timely access to complete information sets for data reporting and analysis

- All sensor data, logbook maintenance data and high value component data reside together

Capability to host, manage, and query extremely large data sets in an HPC environment

Distribution A: Approved for public release

Big Data Management
How?

Use HPC resources and tools designed for large-scale distributed data access and management

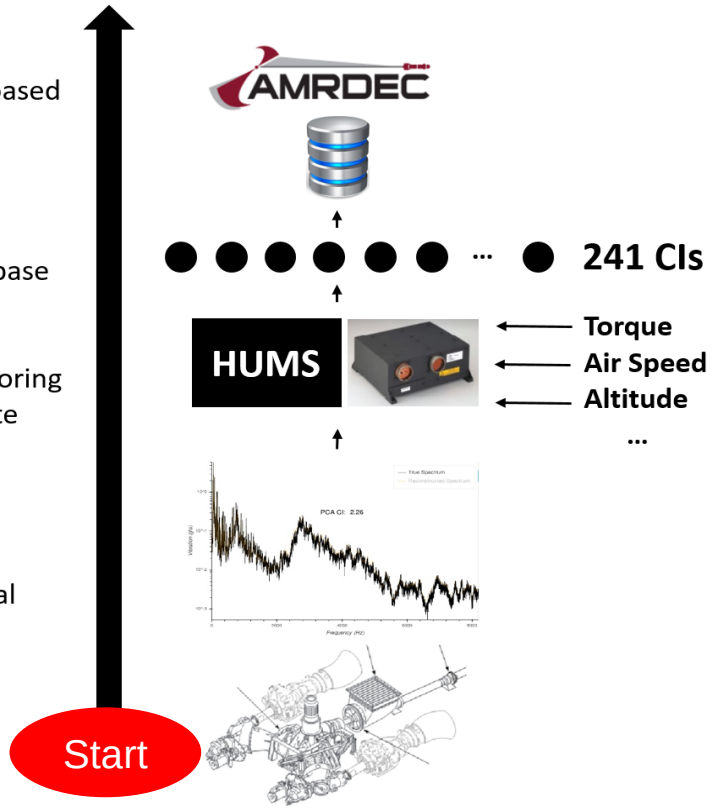
- Dedicated storage on HPC
- Workflow management
- High-performance open-source tools for ML
- Scalable database technologies
- *Demonstrated query over 1.3M sensor logs in 25 minutes*

Distribution A: Approved for public release

Sensor Data Reconstruction

Why?

- Vibrational spectra must be submitted to Aviation Engineering Directorate (AED) for life extension
- Only a few rotorcraft store vibrational spectra
- Therefore, a predictive model is required



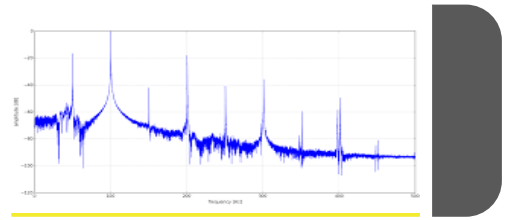
Distribution A: Approved for public release

Sensor Data Reconstruction

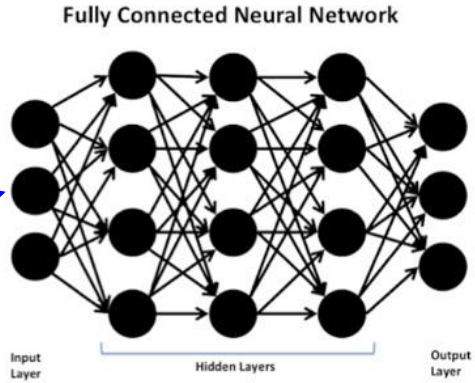
How?

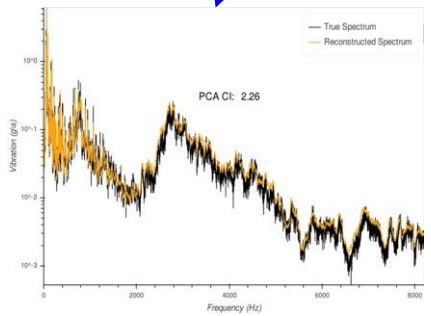
Fully Connected Neural Net that learns the mapping between CIs and the raw sensor data

RAW Data



Condition Indicators (CIs)





Automated Logbook Labeling
Why?

Convert aviation maintenance data to engineering reliability data

- Increase analyst-scored data to 25%
- Provide machine-labeled data for remaining 75%

> 35M total records

~ 3.5M scored

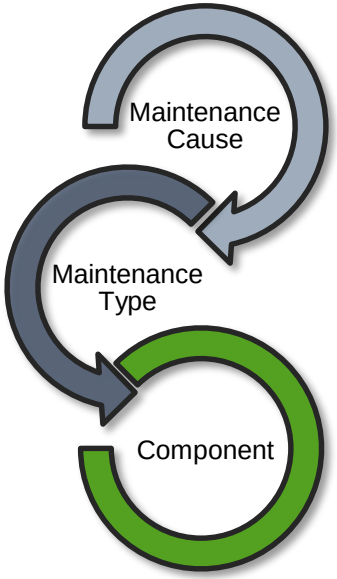
Distribution A: Approved for public release

Automated Logbook Labeling *How?*

Natural Language Processing and Machine Learning techniques:

- Word2Vec
- Distributed Random Forest algorithm
- Learning Using Privileged Information strategy

Computer
Learning
Algorithm for
Records
Evaluation



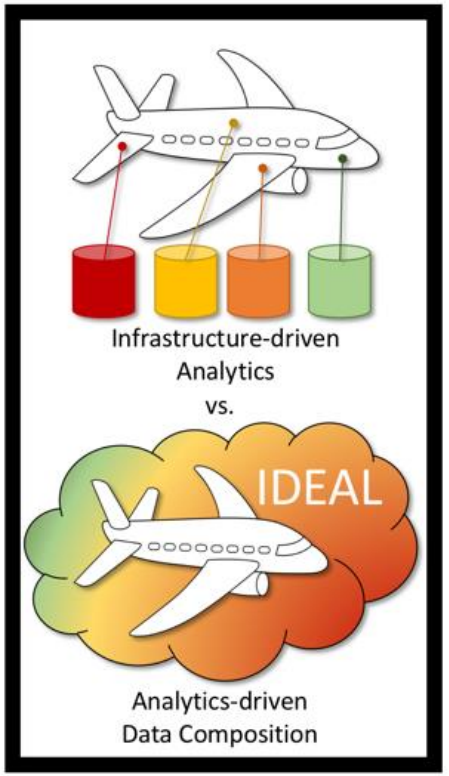
> 90% per record accuracy

Distribution A: Approved for public release

Integrated Data Engineering for Automated Labeling (IDEAL)

Why?

- **Enable holistic analysis of platforms from disparate data sets**
- **Create large, labeled data sets for supervised learning**
  - Virtual integration of data sets to support autonomous labeling
  - Discovery of currently unrecognized events through data synthesis
  - Massive, labeled data sets made available for algorithm development

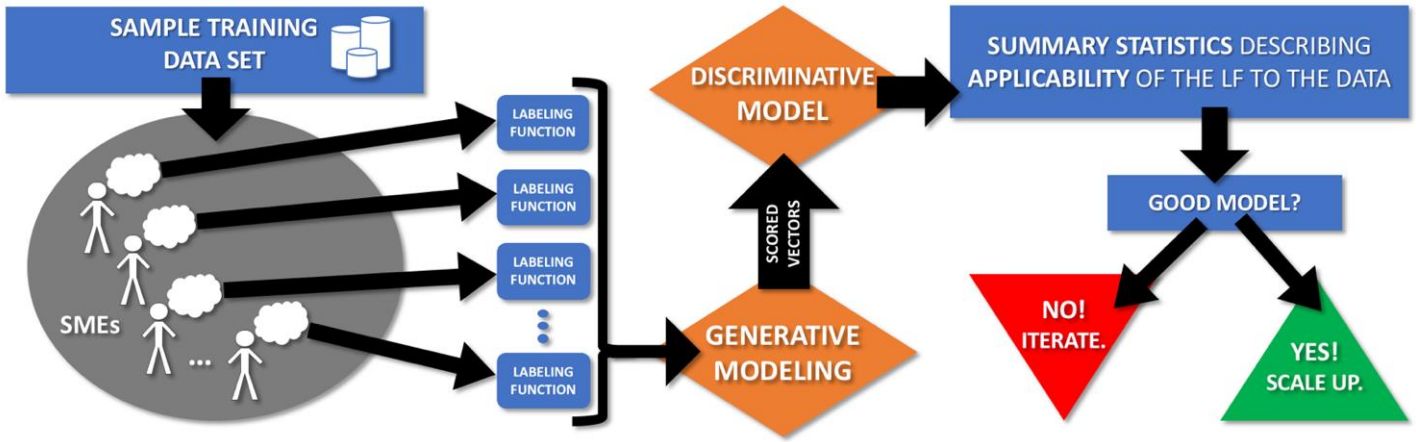


Integrated Data Engineering for Automated Labeling (IDEAL)
How?

Data Programming Paradigm

Autonomously label engineered data compilations from integrated, repurposed data

- Labeling functions produced by SMEs
- Generative model trained
- Label probability distribution produced by discriminative model
- Label accuracies assessed



Summary

- **Managing big data requires a different approach, with special tools and techniques**
- **Machine Learning can derive new insights from data that can lead to more effective decisions**
- **ERDC ITL and AMRDEC have demonstrated the ability to manage and analyze large-scale maintenance data**

Distribution A: Approved for public release

Contact

Maria Seale
US Army Engineer Research and Development Center
Information Technology Laboratory