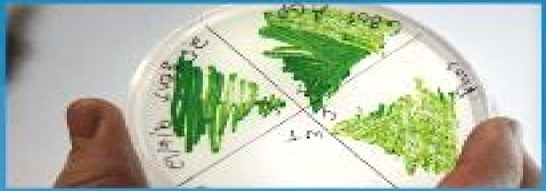




Transportation Data Science at NREL

Adam Duran, Kenneth Kelly, Caleb Phillips

5/22/2018

NREL is a national laboratory of the U.S. Department of Energy, Office of Energy Efficiency and Renewable Energy, operated by the Alliance for Sustainable Energy, LLC.

Why do you need a Big Data factory?

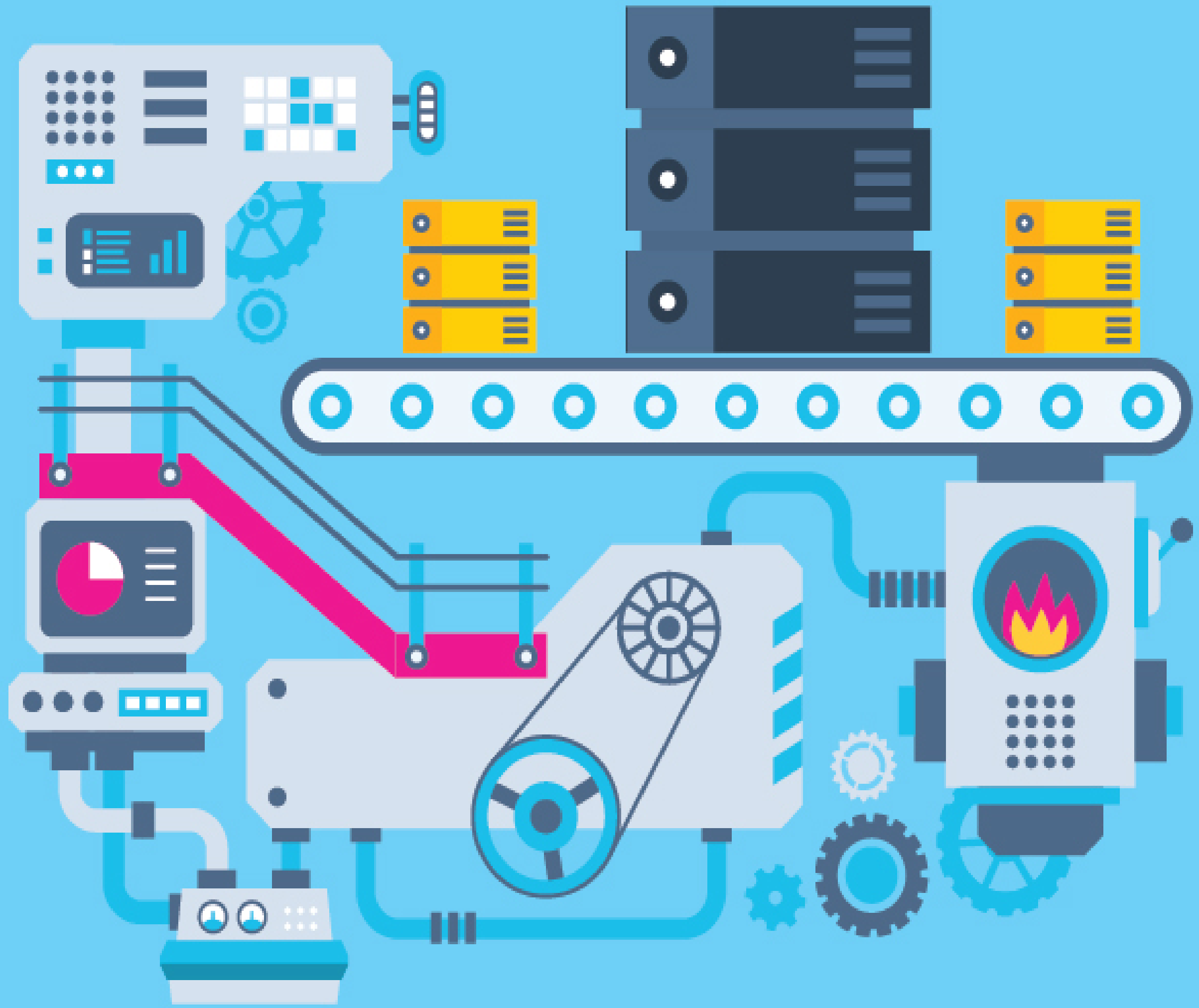


Image - http://www.cmsitservices.com/blog/why-do-you-need-a-big-data-factory/

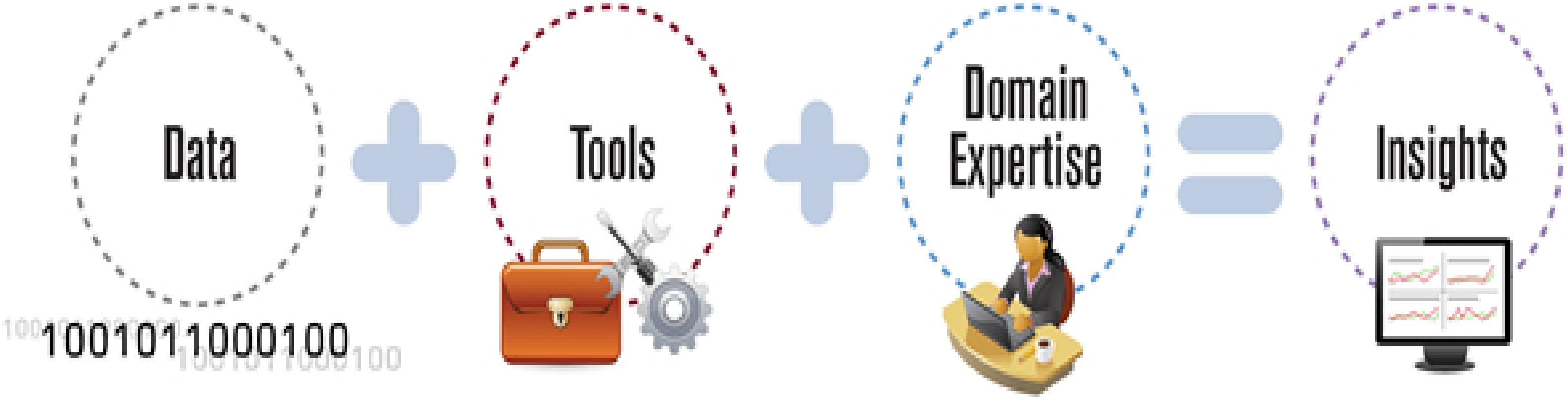


Image - https://tdwi.org/articles/2017/05/30/future-of-big-data-is-cloudy-and-bumpy.aspx

How Big is Big? The 5 Vs of Big Data

Do you have big data?

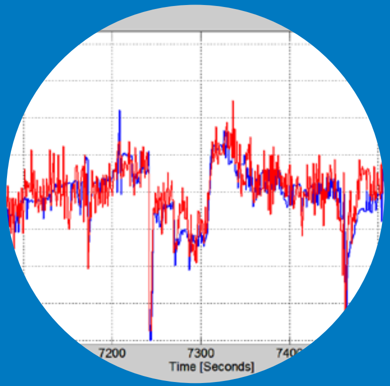
Volume – how big?
Variety – what type and nature?
Velocity – how fast does it arrive?
Variability – are their inconsistencies?
Veracity – challenging assure quality?

Big data and machine learning challenges exist **across all industries**

Do you need a truck?

Machine Learning ❤ Big Data

Managing Big Data – Transportation Data Sources, Types, and Volumes



Timeseries

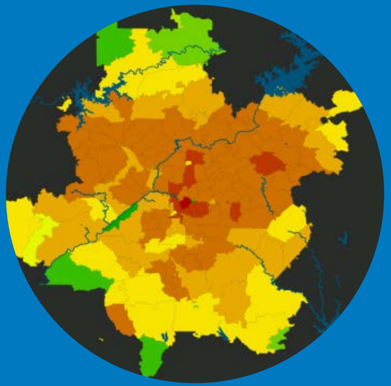
- 1 Hz CAN/OBD and Instrument Data
- Fuel Rates
- Vehicle Speed
- Engine and Emissions Parameters

Geospatial

- 1Hz GPS Data
- Latitude
- Longitude
- Elevation
- Heading

Categorical

- Vehicle Classifiers for Sorting Results
- Weight Class
- Transmission
- Fuel
- Body



Supplemental

- Road Networks
- Infrastructure
- Solar Exposure
- Climate and Temperature

Structured vs. Unstructured Data

Structured:

- Traditional Databases (SQL)

Semi-Structured:

- XML
- JSON

Unstructured:

- Text
- Images
- Audio

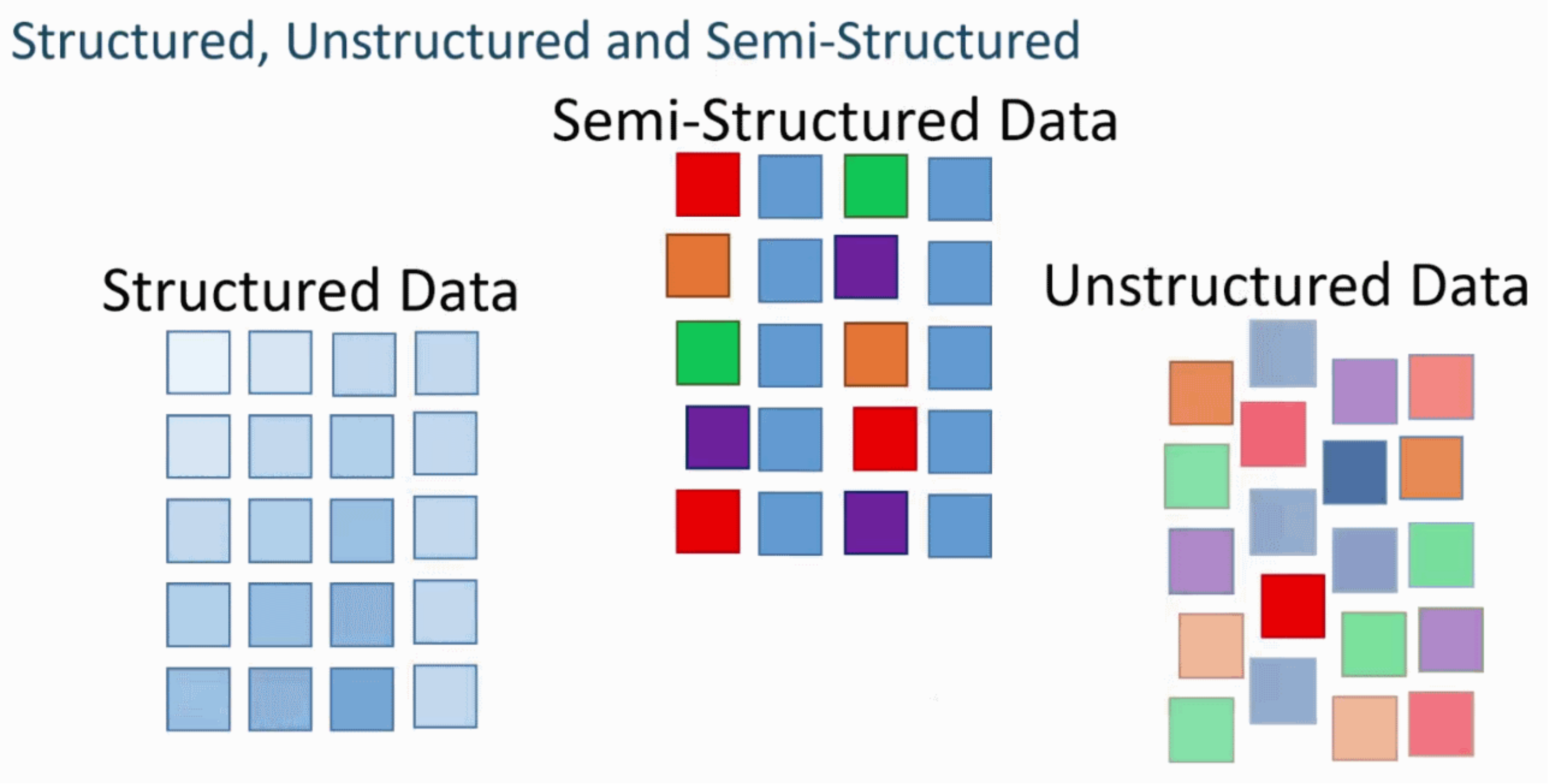


Image - https://e-skillsbusinesstoolbox.webnode.nl/big-data/voorbij-de-hype/

What do we do with all of this data? Enter the Data Scientist!

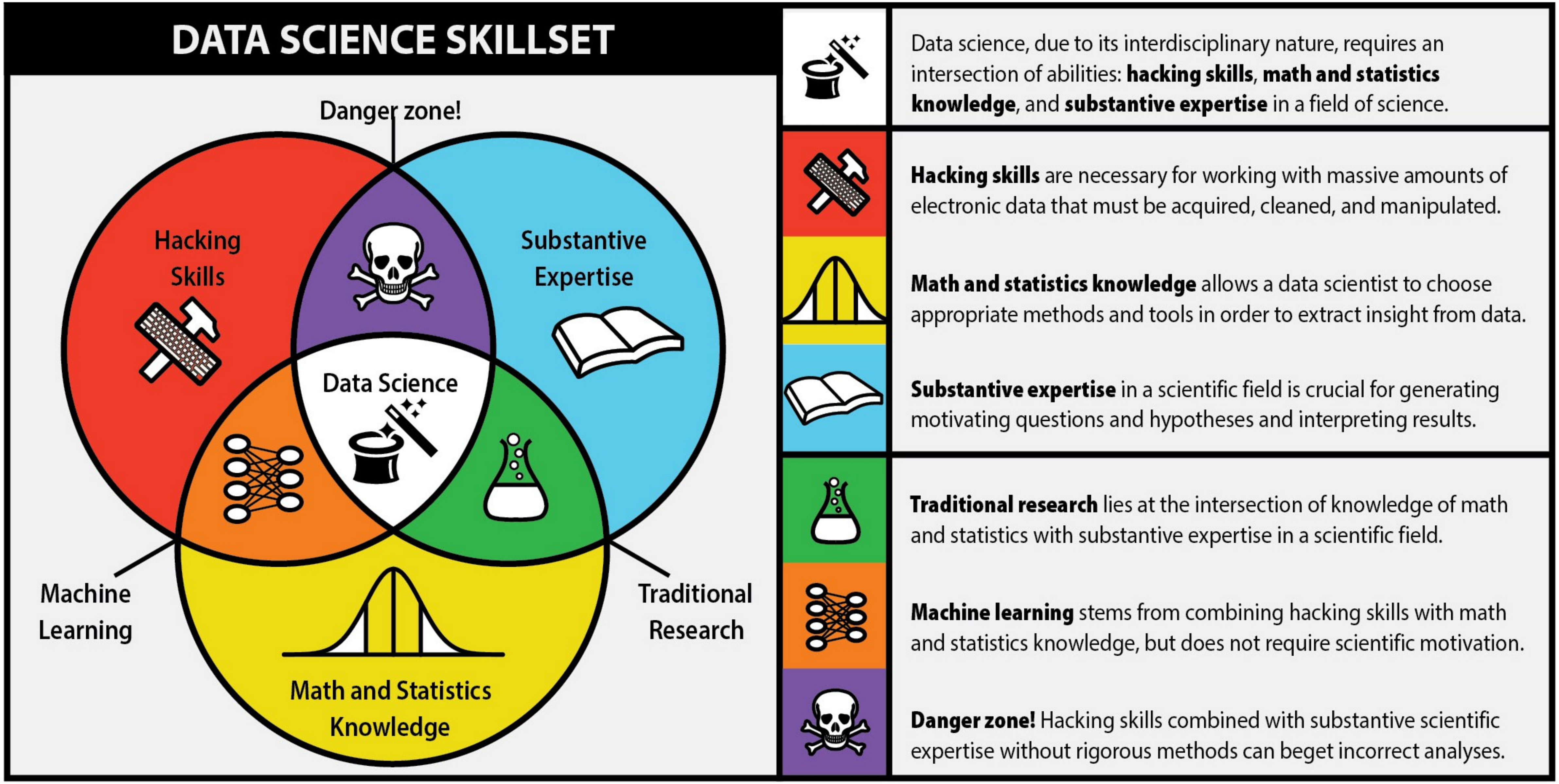


Image - http://berkeleysciencereview.com/how-to-become-a-data-scientist-before-you-graduate/

NREL Computational Sciences / ESIF

Computational Sciences Center

- HPC Systems and Operations
- **Data Analysis and Visualization**
- Simulation and Optimization
- Algorithms and Fluid Dynamics

ESIF High Performance Computing (HPC) Center

NREL Data Resource Landscape

Established

- Peregrine
  - Parallel File system
  - Mass Storage
  - Visualization

- Hitachi Storage
- Relational Database Servers
- Timeseries Cluster
- ESIF Data Repository
- Data Relays
- APIs & Web services
- Invites external collaborators

NREL Data Resource Landscape

Established

- Peregrine
  - Parallel File system
  - Mass Storage
  - Visualization

- Hitachi Storage
- Relational Database Servers
- Timeseries Cluster
- ESIF Data Repository
- Data Relays
- APIs & Web services
- Invites external collaborators

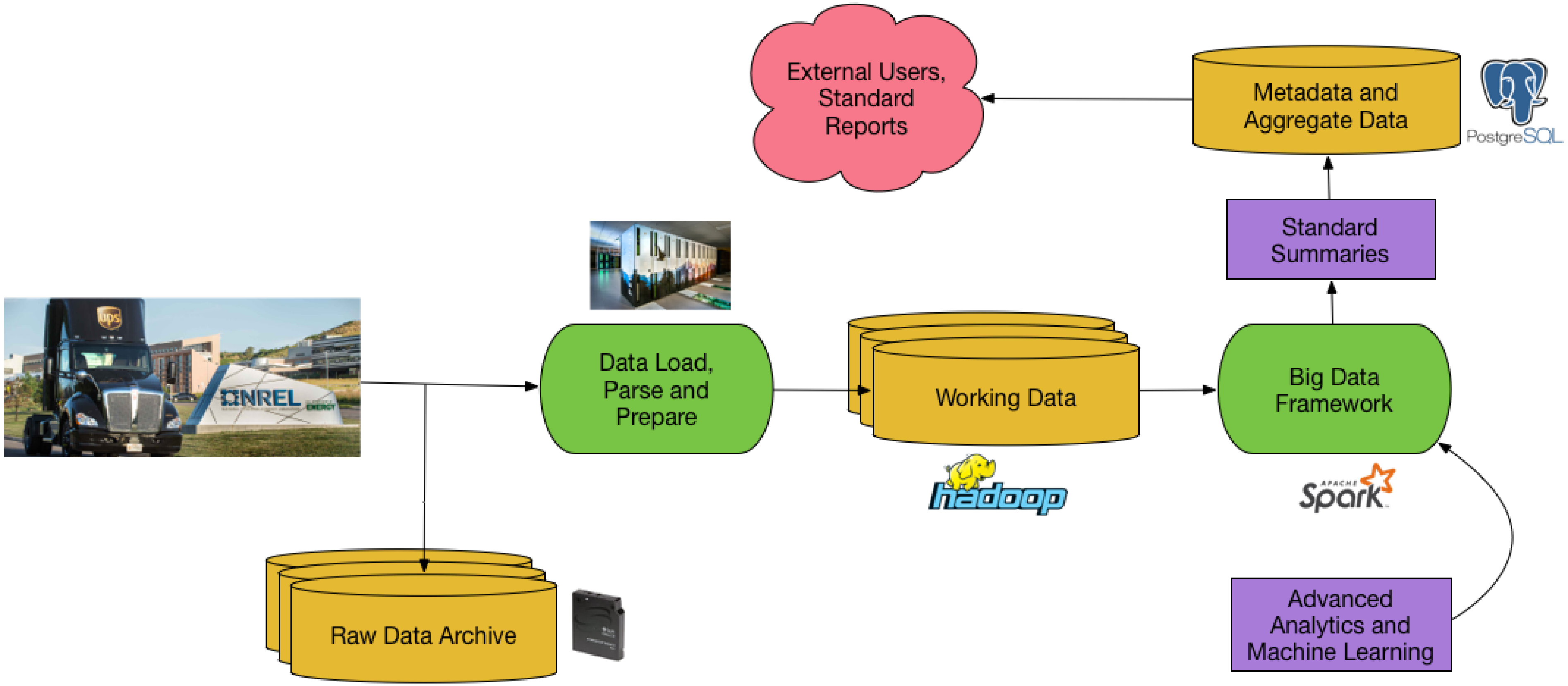
Emerging

- Sparkplug
  - Openstack
  - Spark
  - Hadoop
  - Kafka
- Scalable Attached (Object) Storage
- Peregrine 2 (August!)
  - HPC -> Big Data
- Scalable Relational Databases

Cloud Compatibility Allows Arbitrary Scalability

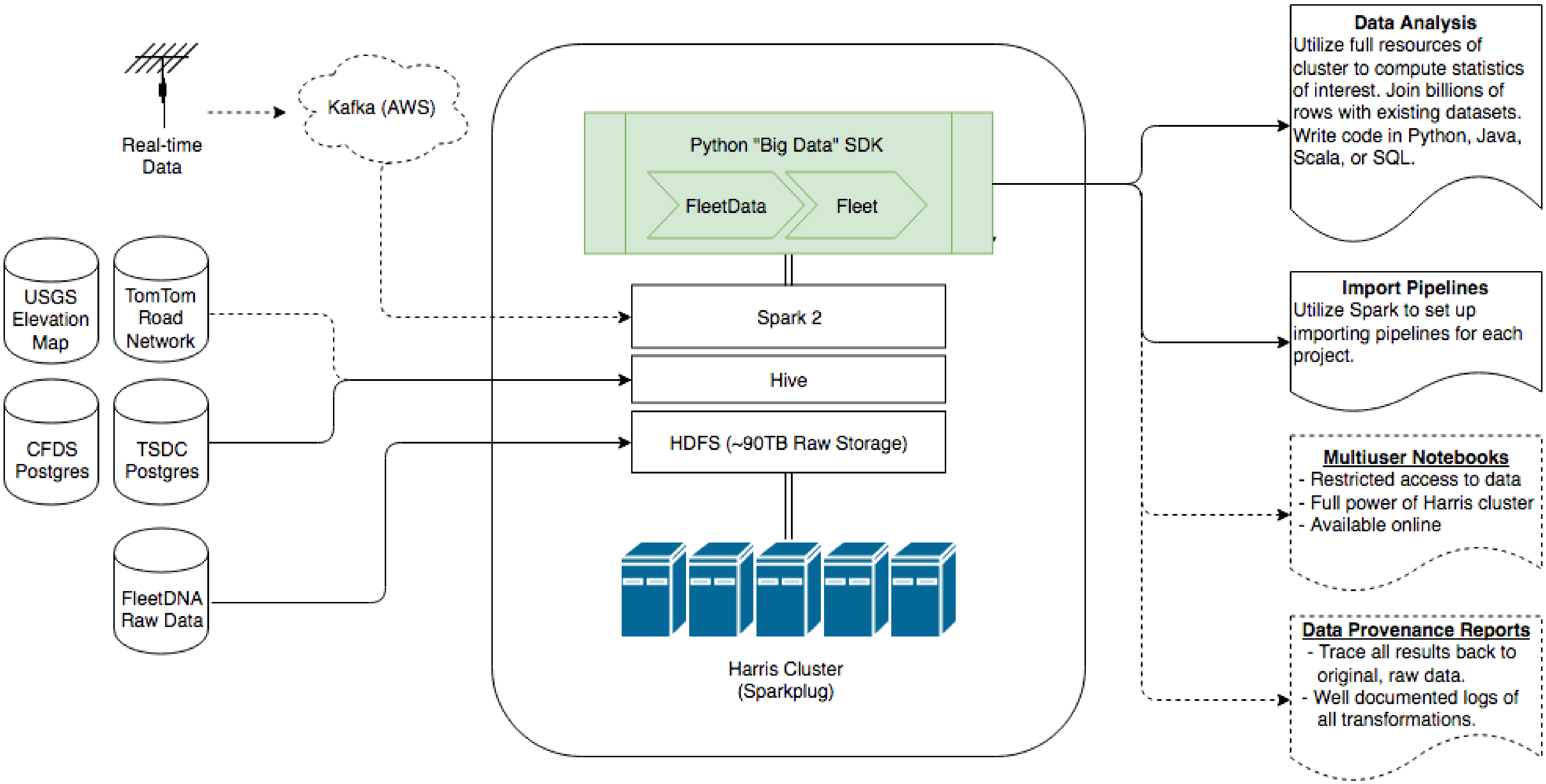
- **Amazon and Competitors offer Hundreds of Services**
- **Increasing adoption by large companies**
- **NREL approach: cloud-replicable infrastructure**
- **Key services:**
  - **S3 – Scalable Object-based Storage**
  - **EC2 – Scalable Compute**
  - **Lambda – Pay per 'function' execution**
  - **Marketplace Gateway -- Monetize data access**

Big IIOT Data – For transportation systems

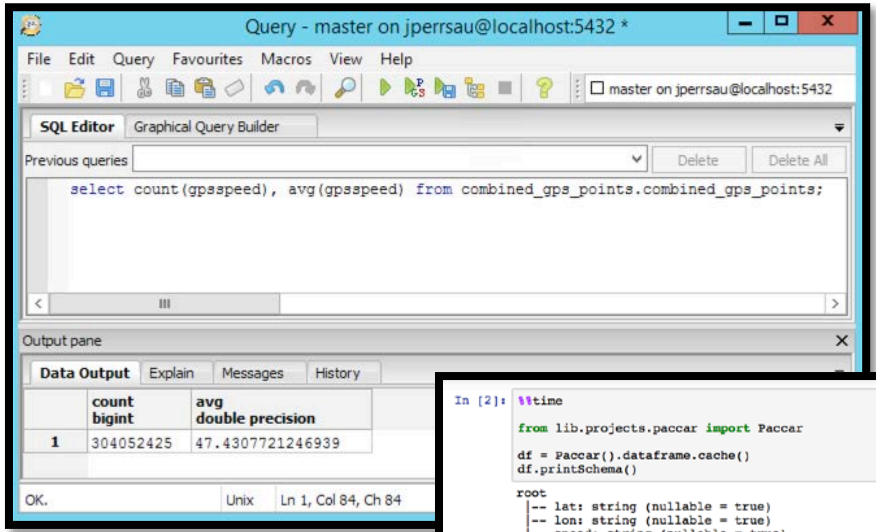


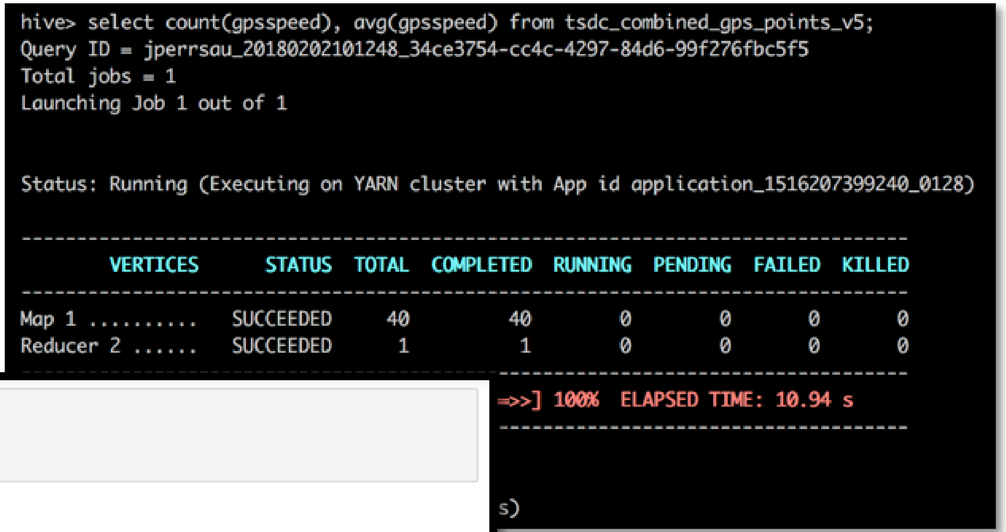
- Fully compatible with cloud services – industry standard technology
- Can process streaming (high velocity) or offline (high volume) data
- Designed for petabyte-scale (or bigger) datasets]
- Can support traditional HPC or Big-Data use cases
- Promotes collaboration with external users

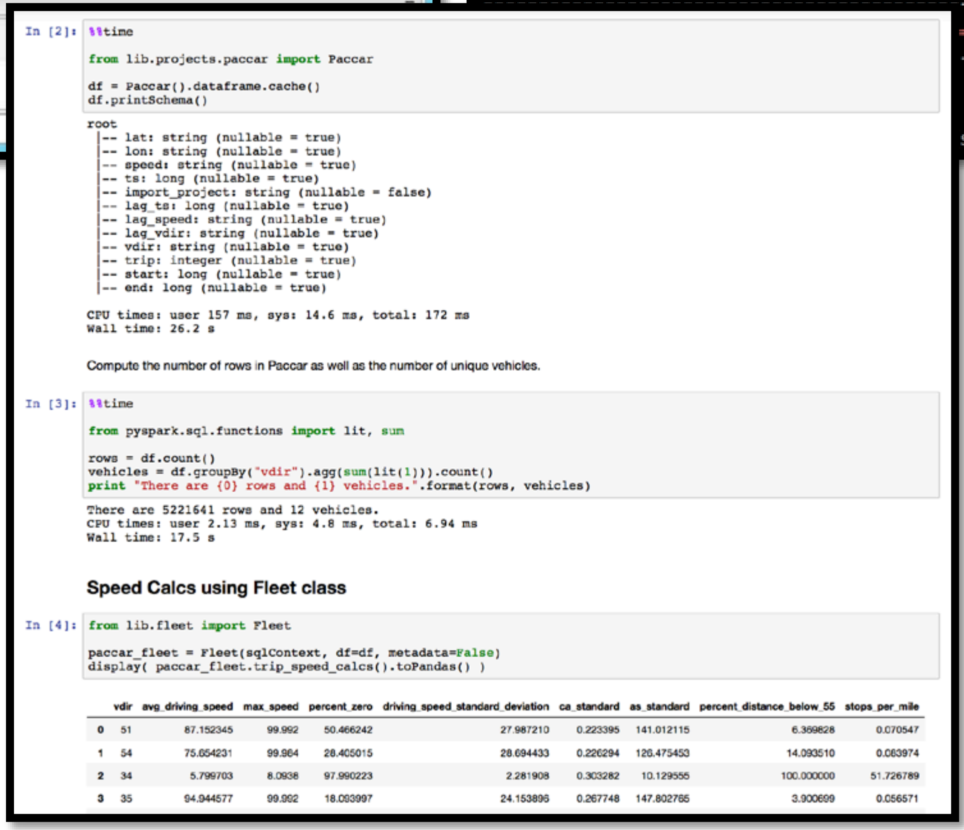
Current Status: 'Big Data SDK' for Transportation Data



‘Big Data SDK’ for Transportation Data







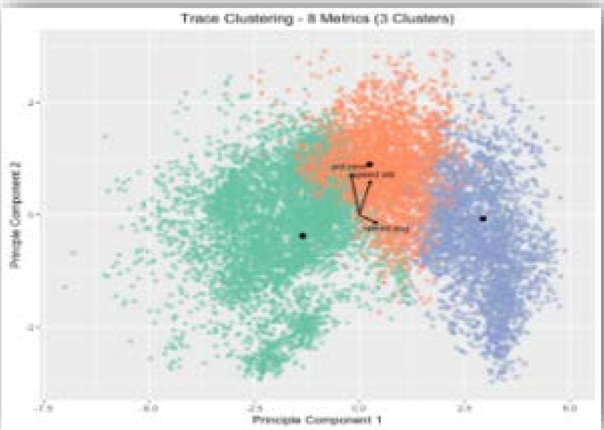
Primary Interface:

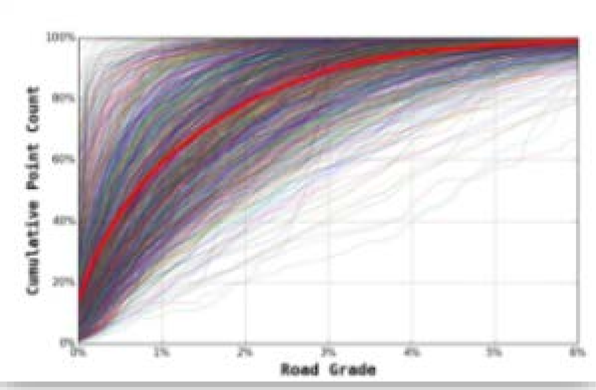
Or, if you prefer:

Big Data Applications – Smart Classification – GHG Phase 2

Scientific Approach & Accomplishment

- NREL Fleet DNA data and analytic expertise provided information crucial to EPA's development of Phase II GHG and fuel efficiency standards for medium- and heavy-duty vehicles.
- NREL segmented vocational vehicle drive-cycle characteristics into multi-dimensional operating groups—including urban, mixed urban, and highway driving conditions—to develop a series of transient drive cycles with weighting factors representative of the acceleration rates, speed distributions, and idle times seen in real-world commercial vehicle driving.
- NREL applied map-matching techniques with USGS elevation data and then weighted the profiles using freight activity data.
- Statistically representative highway segments were identified for on-road testing, and road grade profiles were incorporated into EPA certification cycles.





This work tapped into Fleet DNA data, fused with national road network and freight activity data using NREL's Peregrine high-performance computing system.

Significance & Impact

- Analysis of Fleet DNA vocational vehicle data helped EPA ensure Phase II GHG regulations are more representative of real-world driving conditions.

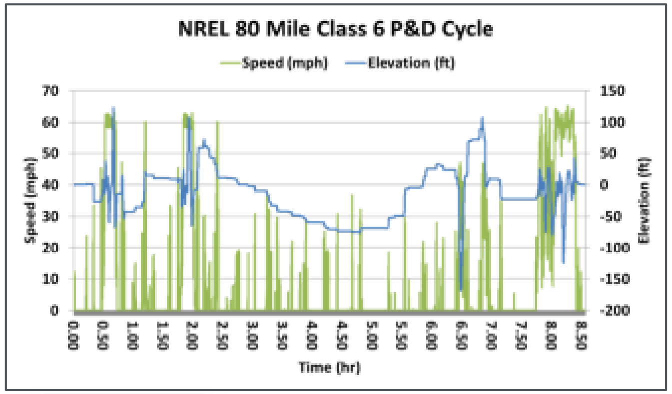
Big Data Applications – Model Based Design – Range Extended EV

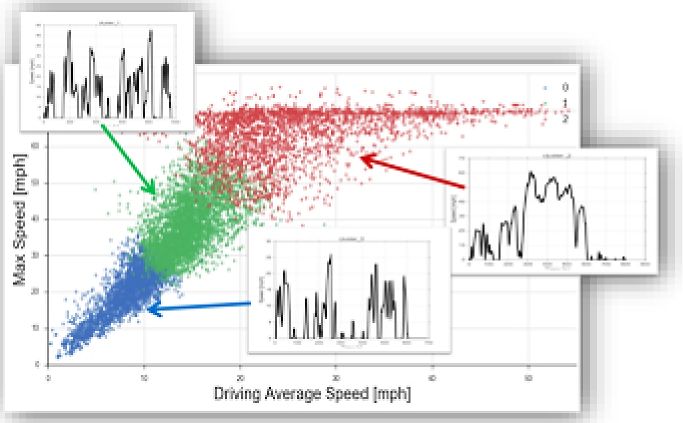
Scientific Approach & Accomplishment

- Leveraging Fleet DNA data to characterize real-world duty cycles from urban delivery vehicles, NREL applied the k-medioid clustering algorithm to segment in-use driving profiles into operational modes and developed representative drive cycles for various modes using the DRIVE tool.
- NREL developed analytical methods to incorporate other parameters, such as road grade, idle time, and key status into the drive cycles.
- NREL's drive cycles are being used to size drivetrain components and optimize energy storage control strategies to meet performance requirements and validate performance relative to program objectives.

Significance & Impact

- This work was conducted as part of two industry partnerships under DOE FOAs led by ***Cummins and Robert Bosch*** to develop commercially viable, range-extended EVs for urban delivery applications targeting a 50% efficiency improvement.





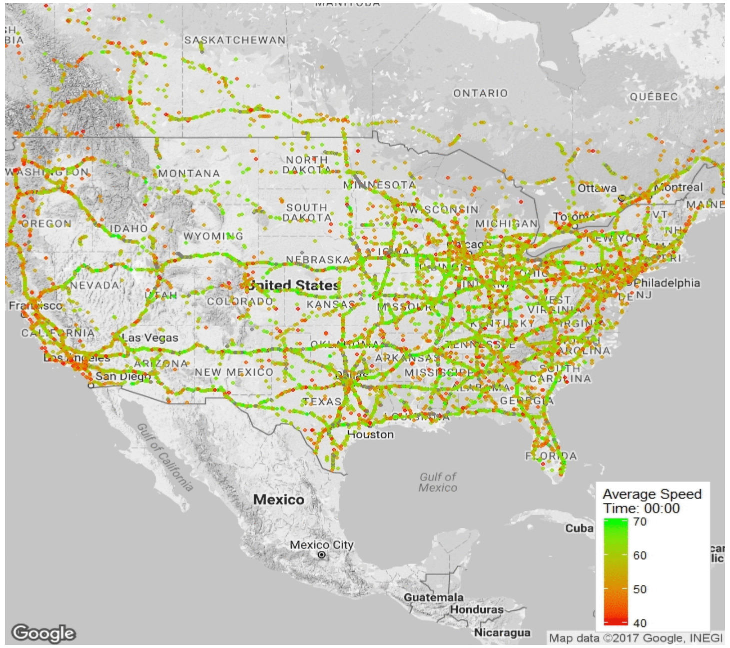
*NREL-developed representative drive cycles are used by **Cummins and Bosch** in powertrain optimization and performance evaluations.*

Big Data Applications – Forecasting – National Scale Platooning Potential

Scientific Approach & Accomplishment

- NREL analyzed fuel-savings data from six independent platooning studies conducted between 2013 and 2016 with Class 8 tractor trailers, including four independent track test studies, wind tunnel results from LLNL, and CFD simulations from Denso.
- NREL followed up track testing efforts with large scale (50k+ vehicles) evaluating real world potential for platooning on US roadways.

Platooning reduces aerodynamic drag by decreasing the driving distance between vehicles.



Significance & Impact

- NREL evaluation and analysis have characterized platooning performance under a range of speeds, loads, and following distances, including reduced benefits at very close following distances.
- NREL platooning data and analysis are being used in an ***ARPAe NEXTCAR project with Purdue, Cummins, and Peloton*** to develop next-generation adaptive platooning technologies and in other research efforts at ***LLNL, LBNL, and FHWA.***