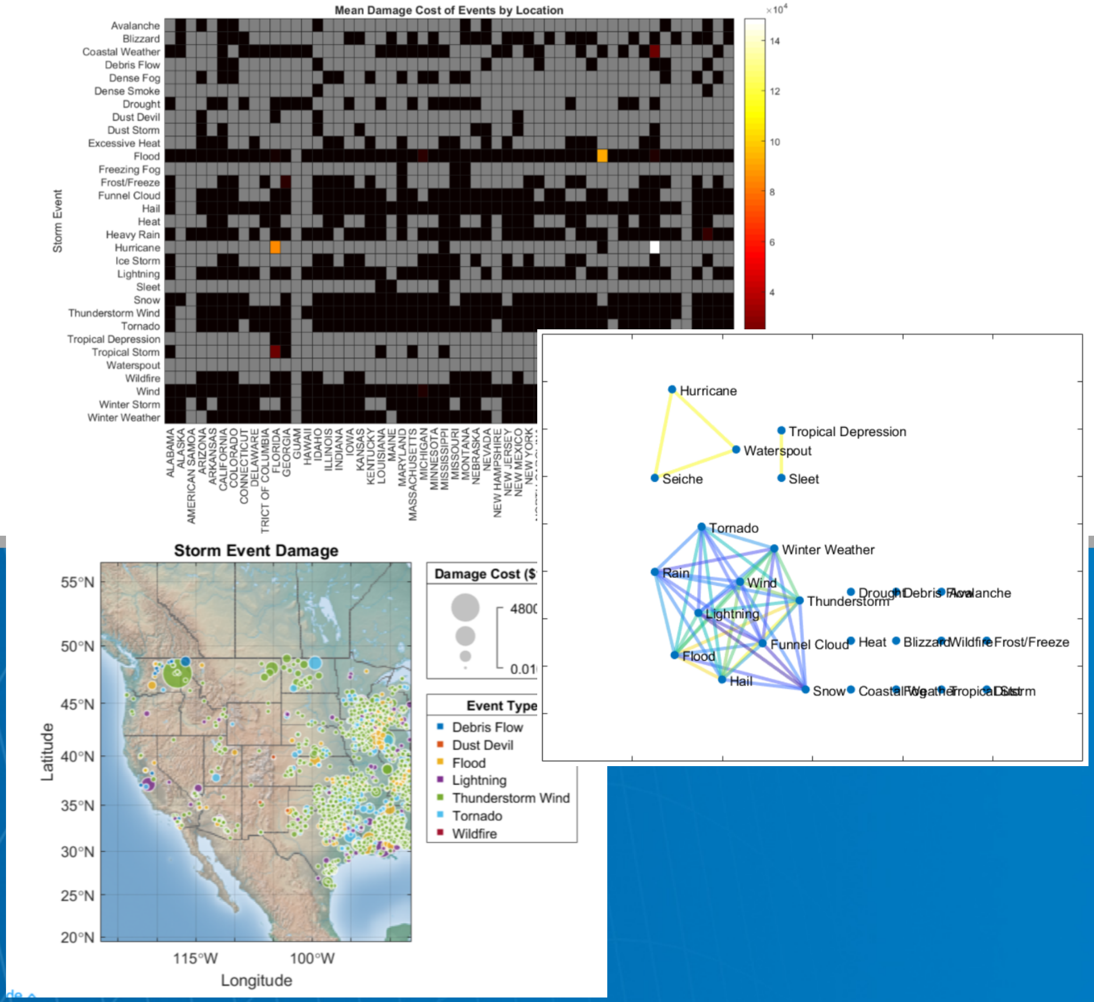
Data Analytics and Machine Learning With MATLAB



Sumit Tandon
Senior Customer Success Manager, MathWorks, Inc.
Sumit.tandon@mathworks.com

Why MATLAB?

Focus on solving your problems

Productive environment
tuned for engineering and scientific work

Ready to use
with toolboxes that work out of the box

Ready to run on production systems without rewriting code

Reliable
entrusted to send a spacecraft to Pluto, create certified code for medical devices

Execution speed
with optimized code that leverages GPUs, clusters, and clouds

Complementary, Interactive, Self-paced MATLAB Tutorials

Ideal for new users or a refresher

MATLAB Onramp

Get started quickly with the basics of MATLAB.

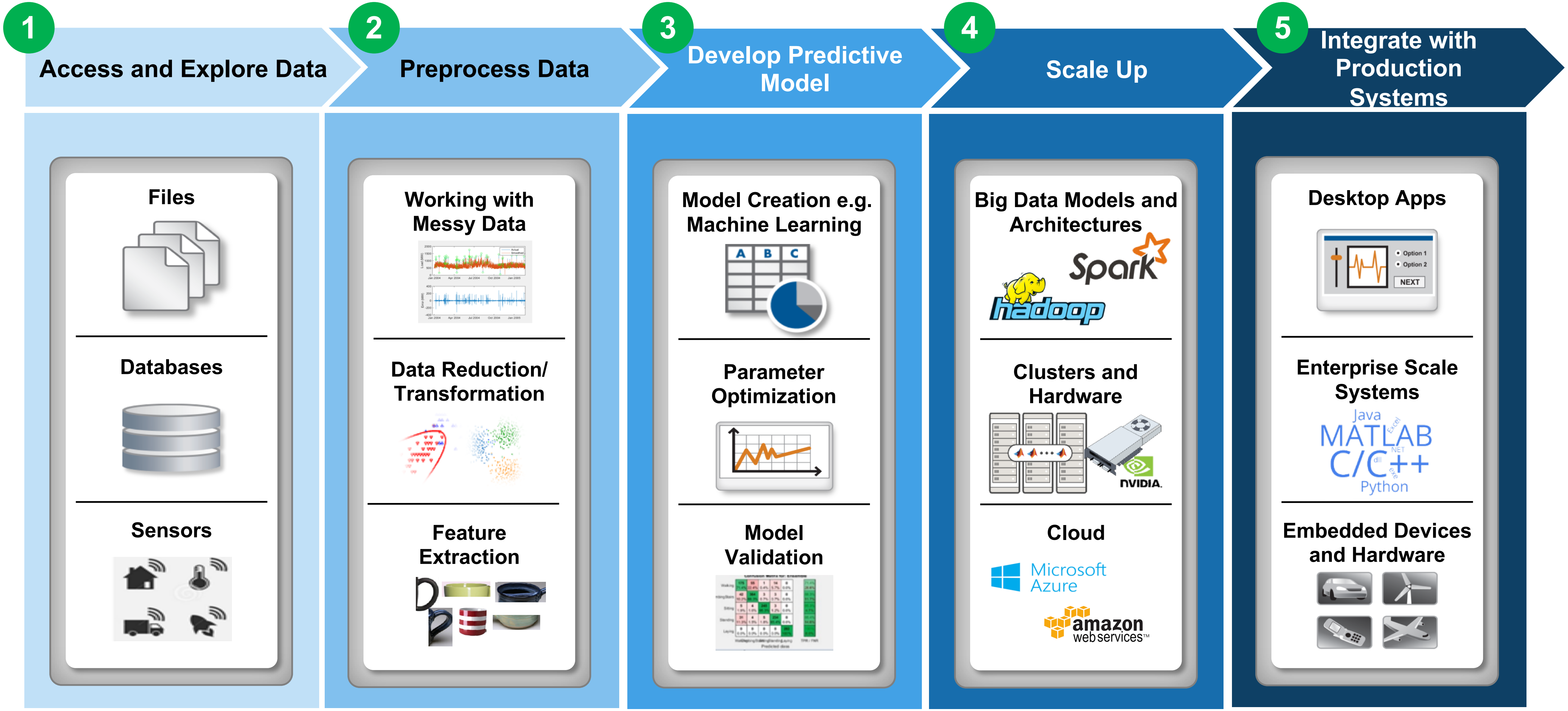
Launch

Deep Learning Onramp

Get started quickly using deep learning methods to perform image recognition.

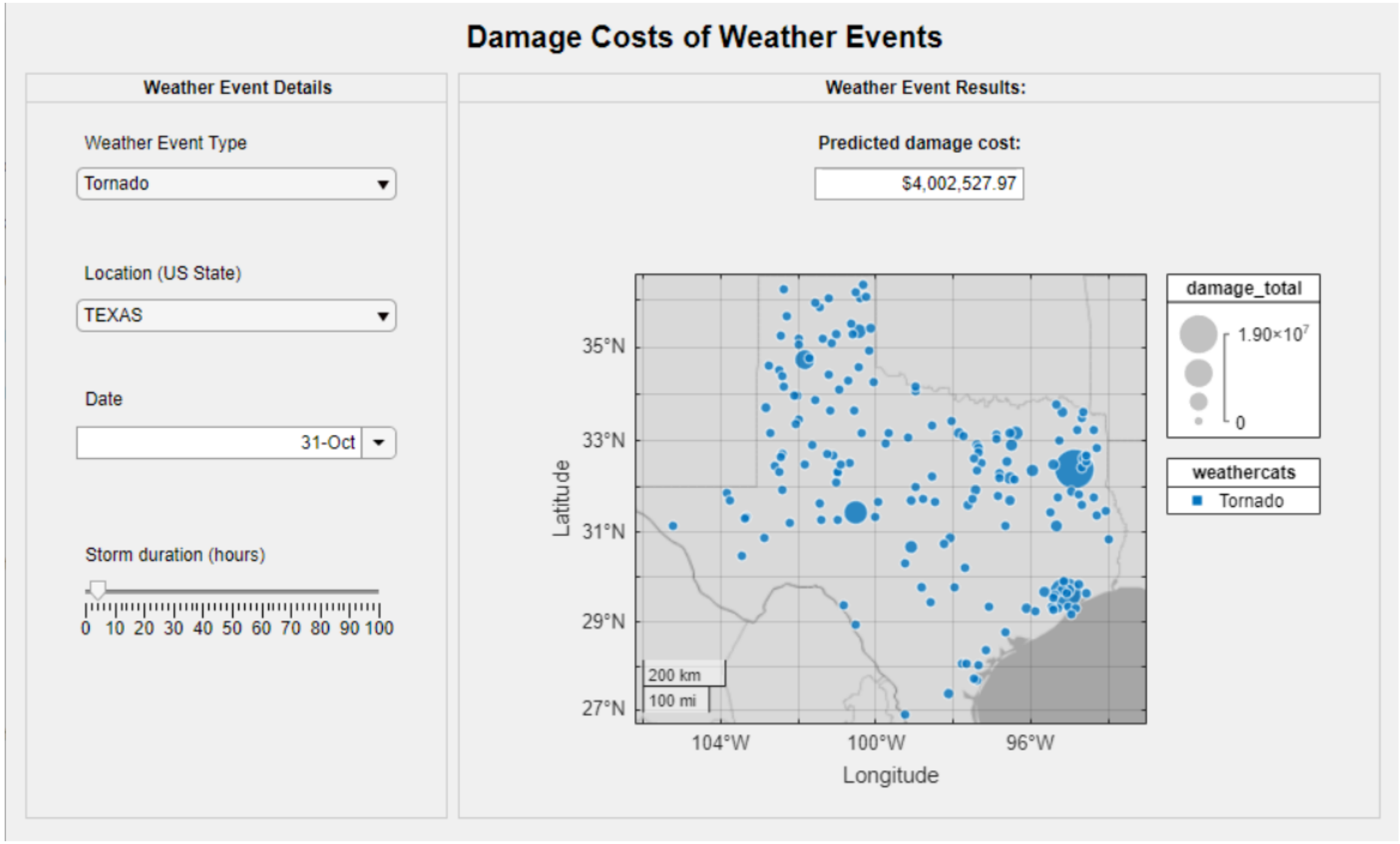
Launch

Data Analytics Workflow

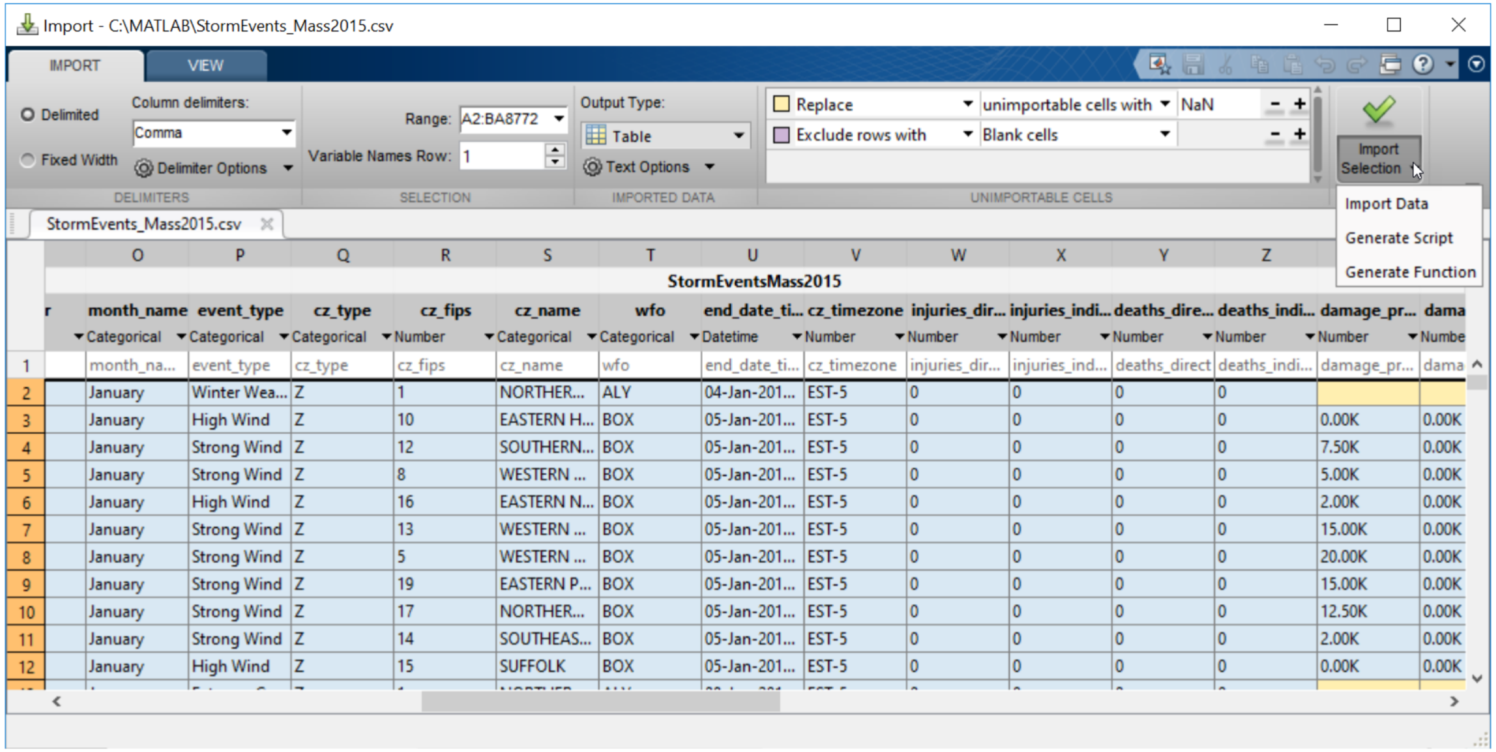


Demo: Predict Damage Cost of Weather Events

- Use historical weather events data from 1980-2017
- Preprocess data
- Develop prediction model based on event type, location, time/month/year
- Predict damage value to prepare for future



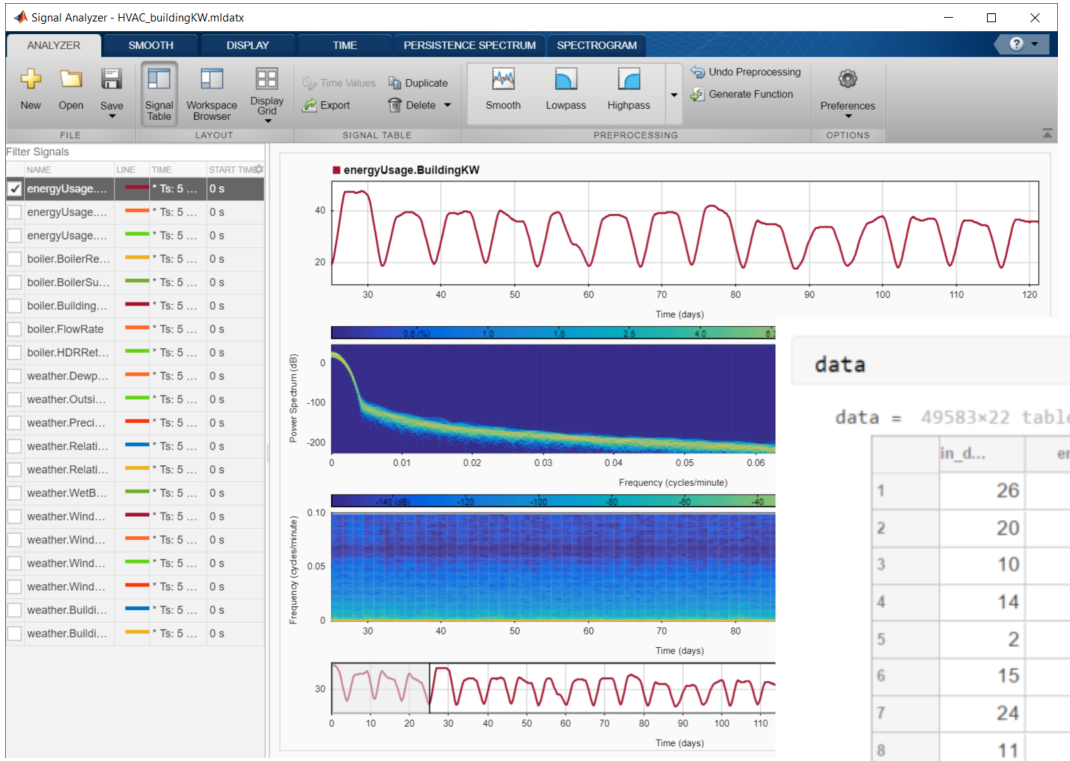
Access data from many sources



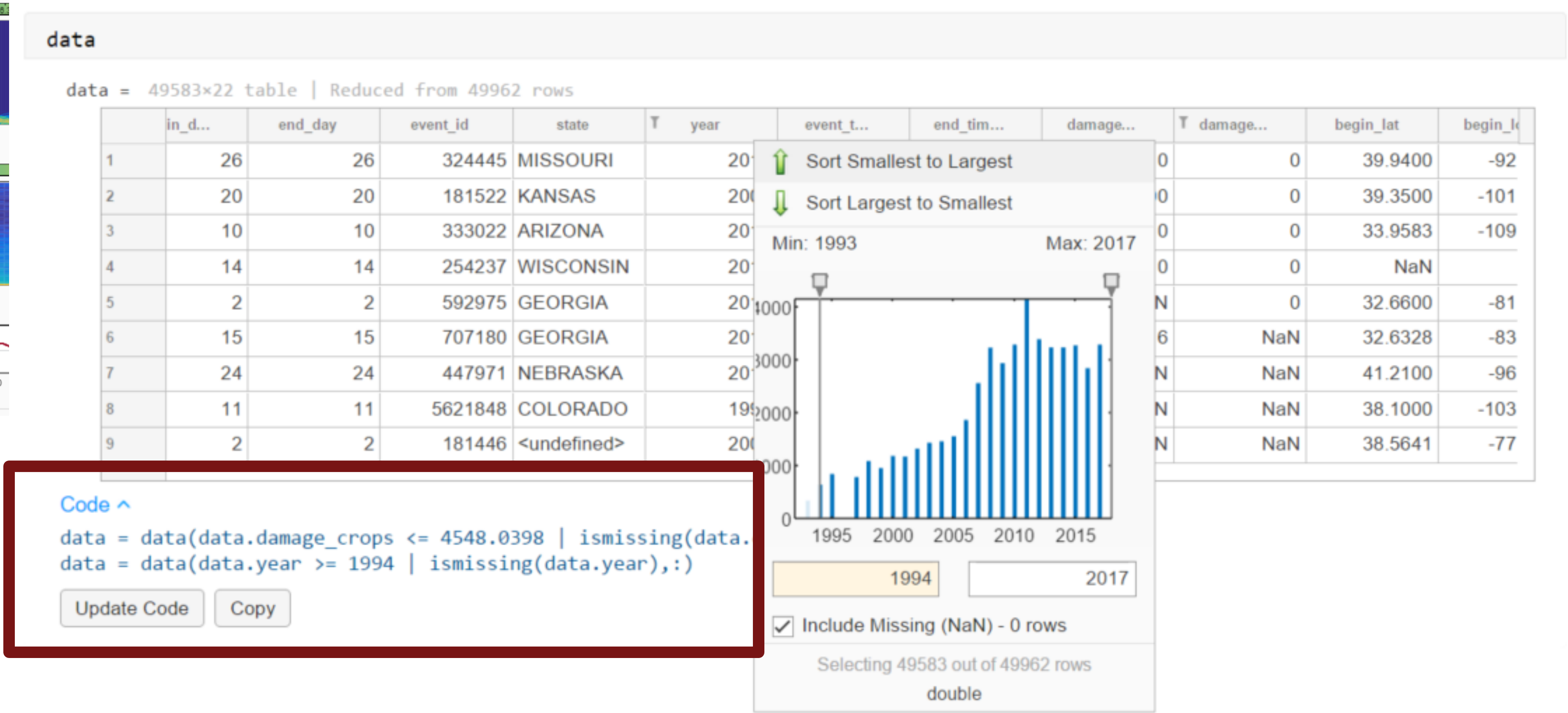
2

Preprocess Data

Spend less time cleaning data




```
data = sortrows(data);
data = fillmissing(data,'linear');
data = smoothdata(data);
```



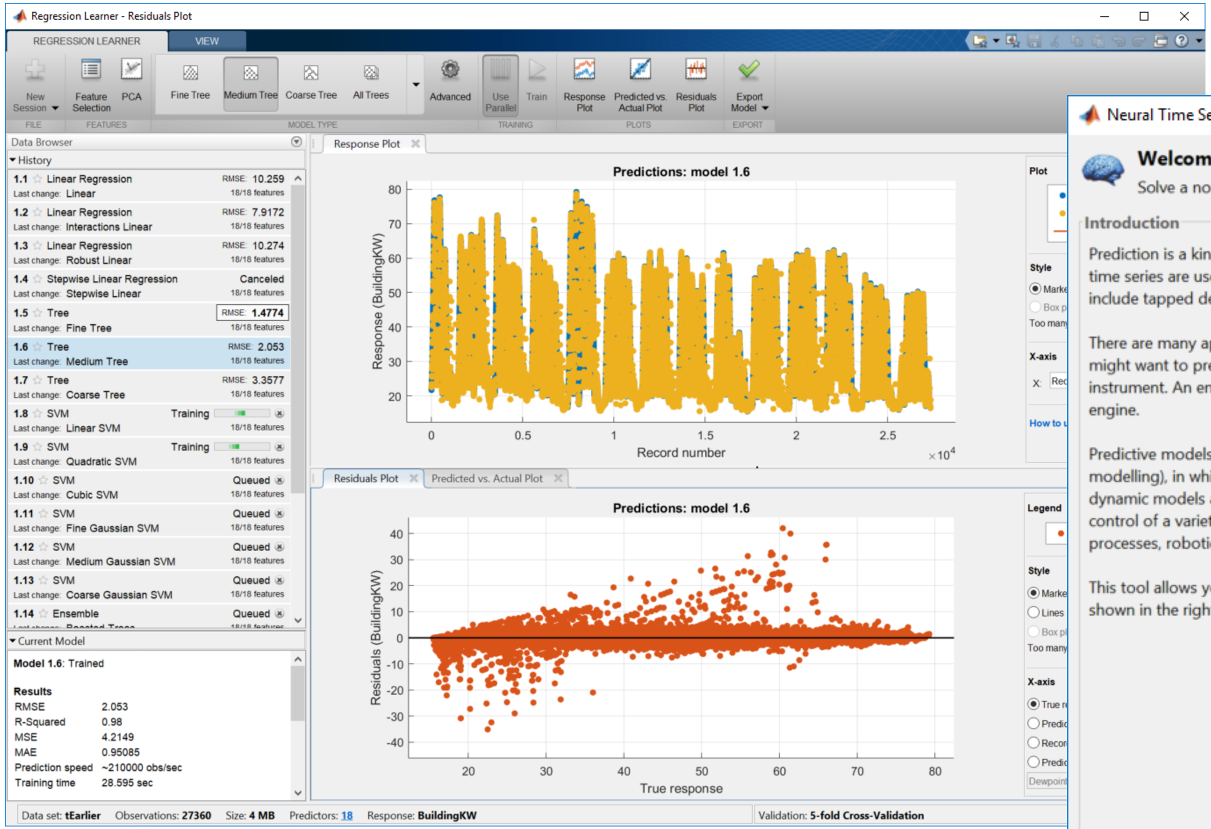
	in_d...	end_day	event_id	state	year	event_t...	end_tim...	damage...	damage...	begin_lat	begin_l
1	26	26	324445	MISSOURI	20			0	0	39.9400	-92
2	20	20	181522	KANSAS	20			0	0	39.3500	-101
3	10	10	333022	ARIZONA	20			0	0	33.9583	-109
4	14	14	254237	WISCONSIN	20			0	0	NaN	
5	2	2	592975	GEORGIA	20			N	0	32.6600	-81
6	15	15	707180	GEORGIA	20			6	NaN	32.6328	-83
7	24	24	447971	NEBRASKA	20			N	NaN	41.2100	-96
8	11	11	5621848	COLORADO	19			N	NaN	38.1000	-103
9	2	2	181446	<undefined>	20			N	NaN	38.5641	-77

3

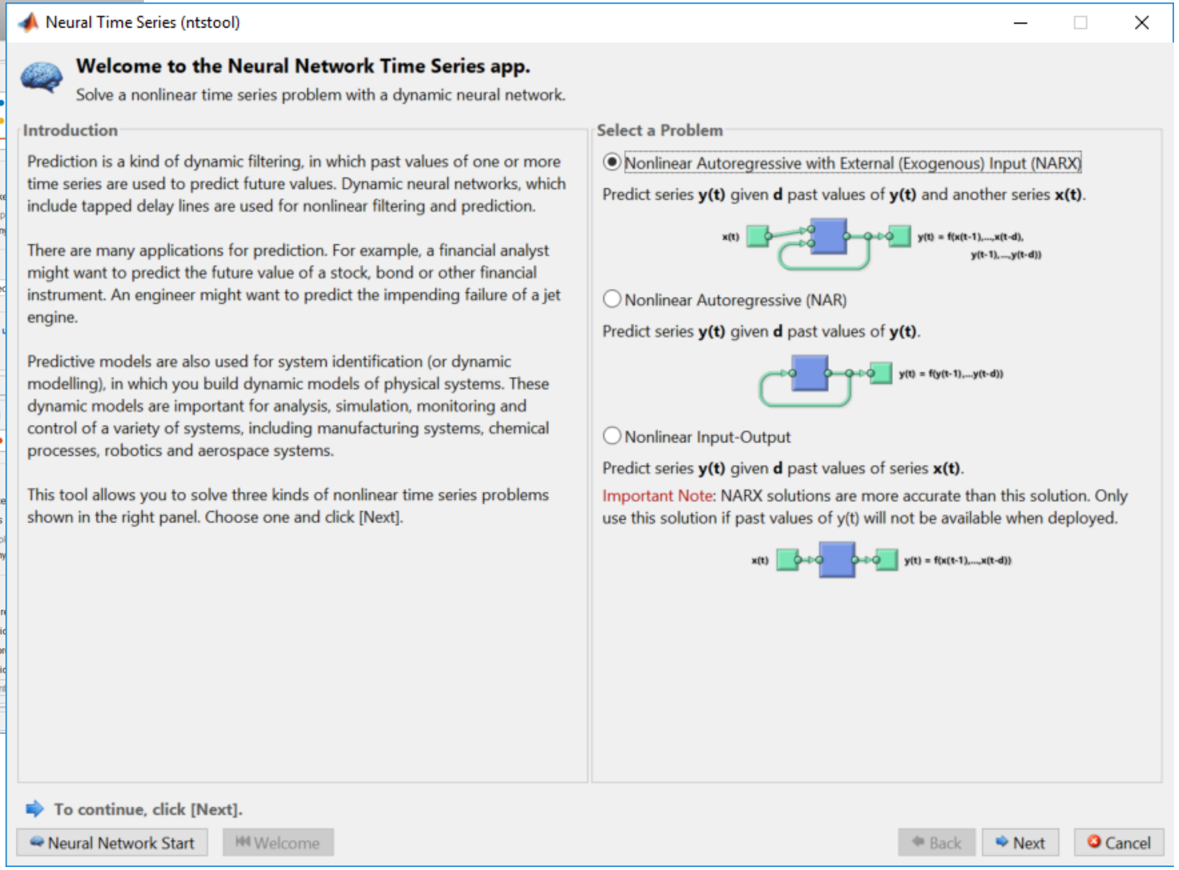
Develop Predictive Models

Get started easily with advanced techniques

Classification
Regression



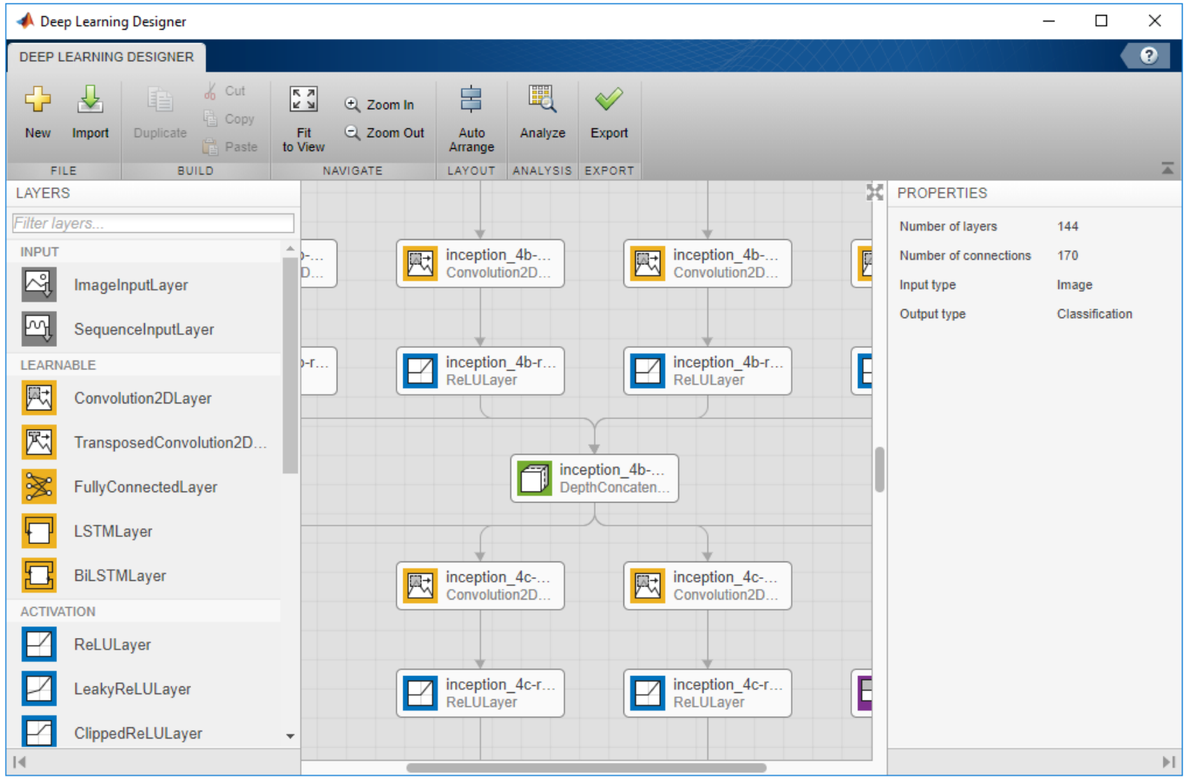
Deep Learning



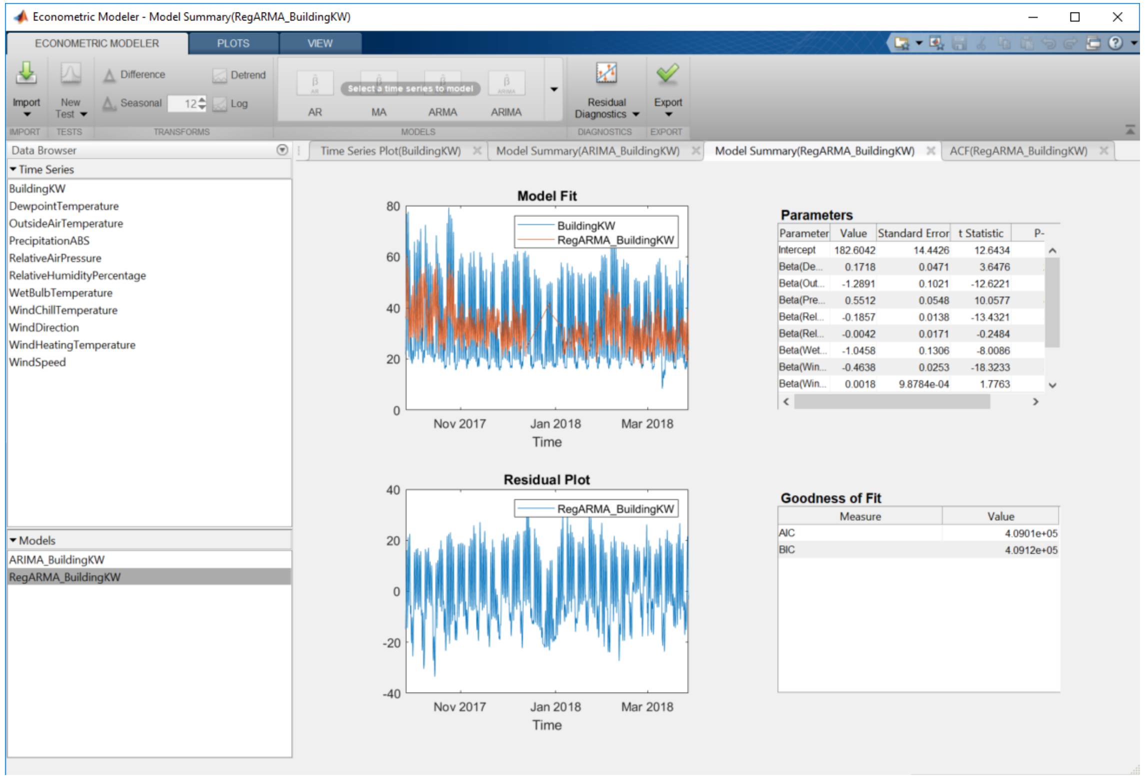
Explore different types of models

3 Develop Predictive Models

Neural Networks

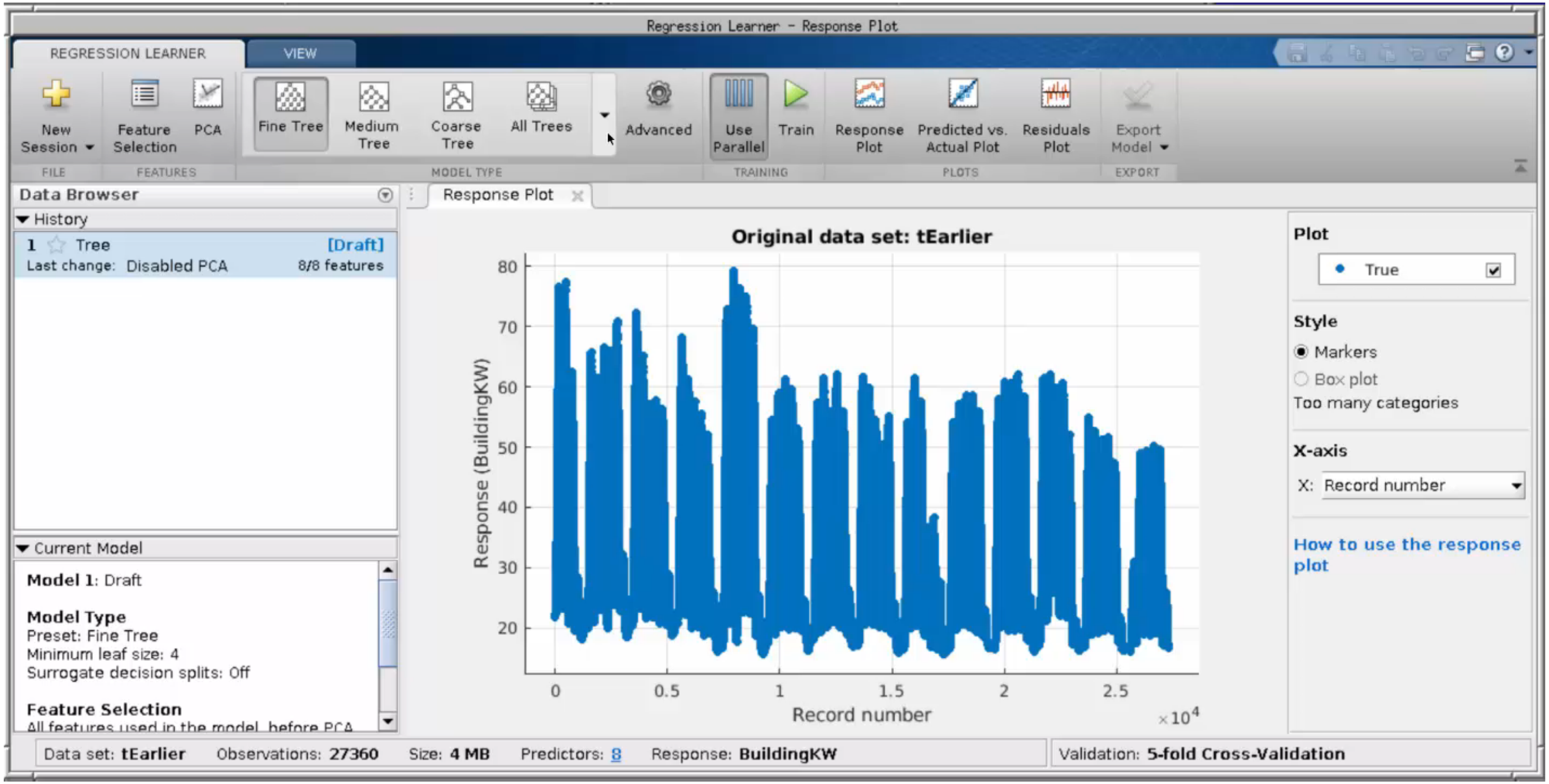


Time Series Models



Try many algorithms in parallel and validate results

3 Develop Predictive Models



4

Scale Up

Scale to big data using the same code

One file

Access Data

```
measured = readtable('PumpData.csv');
measured = table2timetable(measured);
```

Preprocess Data

Select data of interest

```
measured = measured(timerange(seconds(1),seconds(2)),'Speed')
```

Work with missing data

```
measured = fillmissing(measured,'linear');
```

Calculate statistics

```
m = mean(measured.Speed);
s = std(measured.Speed);
```

One hundred files

Access Data

```
measured = datastore('PumpData*.csv');
measured = tall(measured);
measured = table2timetable(measured);
```

Preprocess Data

Select data of interest

```
measured = measured(timerange(seconds(1),seconds(2)),'Speed')
```

Work with missing data

```
measured = fillmissing(measured,'linear');
```

Calculate statistics

```
m = mean(measured.Speed);
s = std(measured.Speed);
```

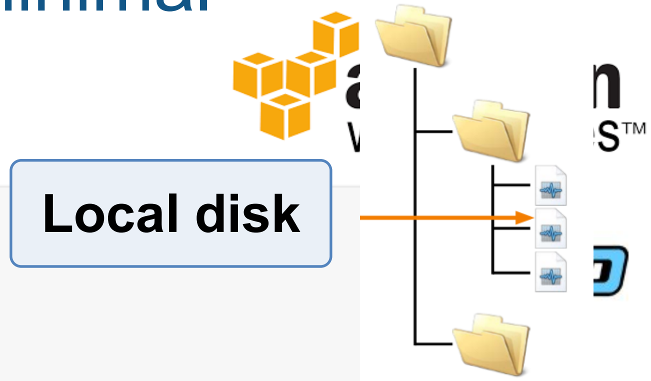
```
[m,s] = gather(m,s);
```

Access data from anywhere with minimal changes

```
setenv('AWS_ACCESS_KEY_ID', 'ACCESS_KEY_ID')
setenv('AWS_SECRET_ACCESS_KEY', 'ACCESS_KEY')
```

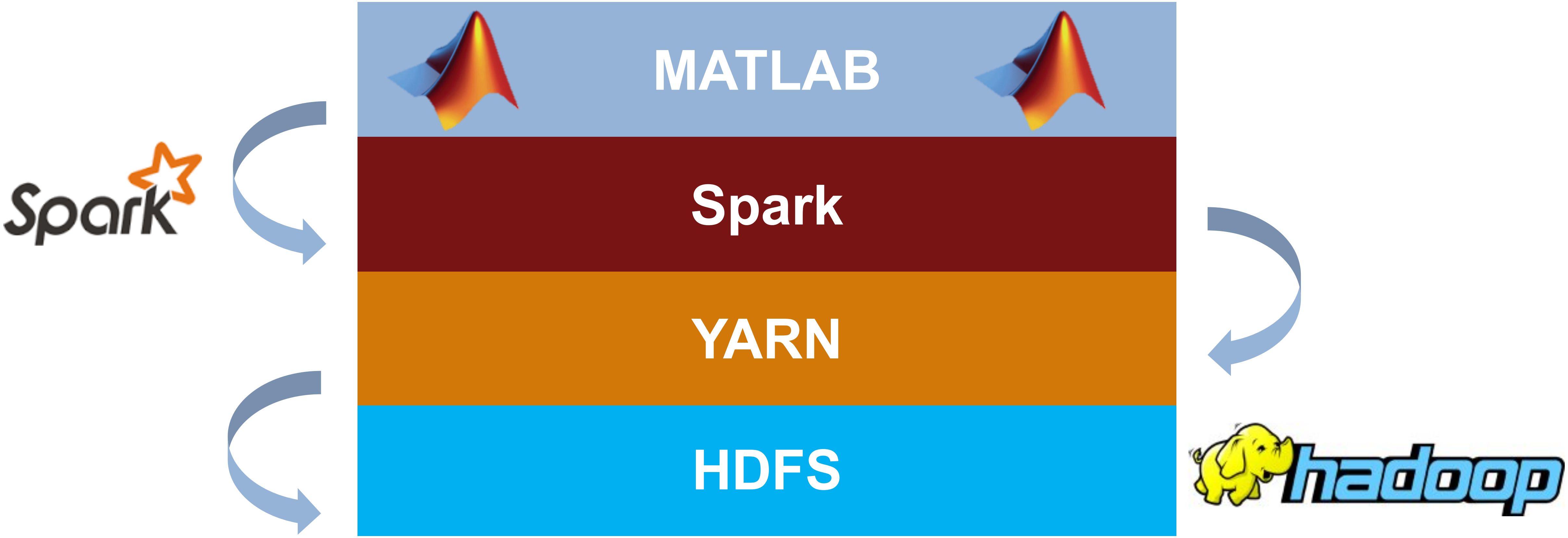
```
fileLoc = 'datasets/FoodImages';
```

```
ds = imageDatastore(fileLoc);
```



4 Scale Up

Use MATLAB on a Spark-enabled Hadoop cluster for machine learning at scale

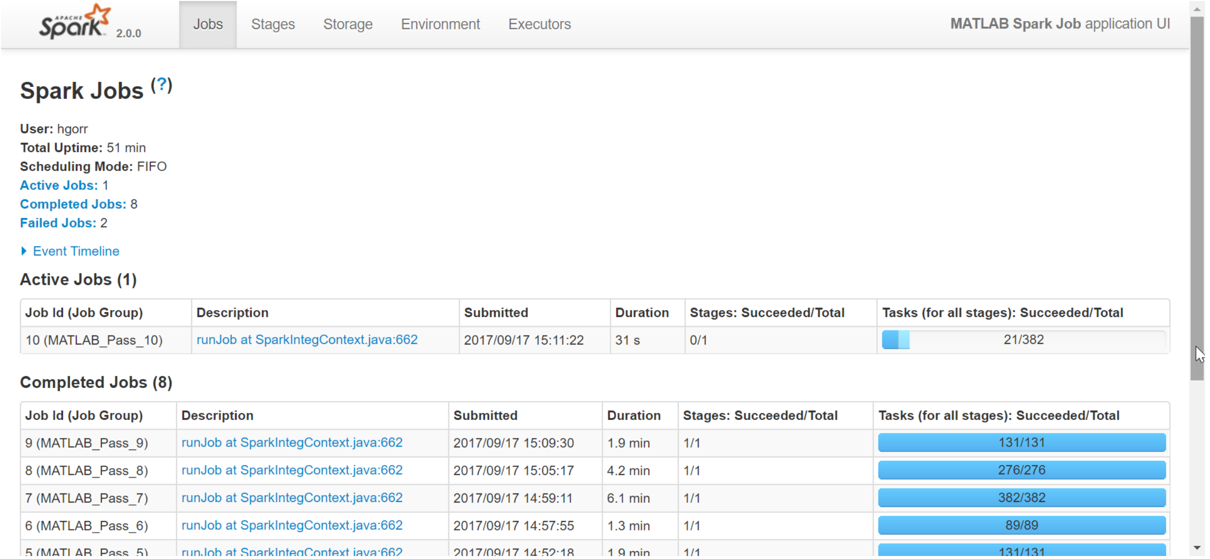


4

Scale Up

Monitor large jobs in MATLAB

```
Evaluating tall expression using the Spark Cluster:
- Pass 1 of 13: Completed in 4.0333 min
- Pass 2 of 13: Completed in 2.3 min
- Pass 3 of 13: Completed in 1.8667 min
- Pass 4 of 13: Completed in 4.2167 min
- Pass 5 of 13: Completed in 4.2167 min
- Pass 6 of 13: Completed in 4.3 min
- Pass 7 of 13: Completed in 1.2 min
- Pass 8 of 13: Completed in 3.75 min
- Pass 9 of 13: Completed in 2.5167 min
- Pass 10 of 13: Completed in 38.7 min
- Pass 11 of 13: Completed in 51 sec
- Pass 12 of 13: Completed in 26.833 min
- Pass 13 of 13: 72% complete
Evaluation 98% complete
```



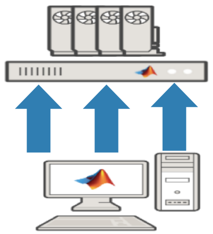
4 Scale Up

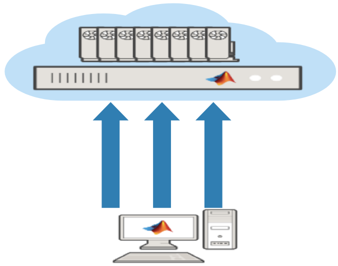
Parallel and Distributed Computing

Single CPU

Single CPU
Single GPU

Single CPU, Multiple GPUs





Parallel Computing Toolbox

- Speed up parallel applications
- Take advantage of GPUs
- Prototype code for clusters

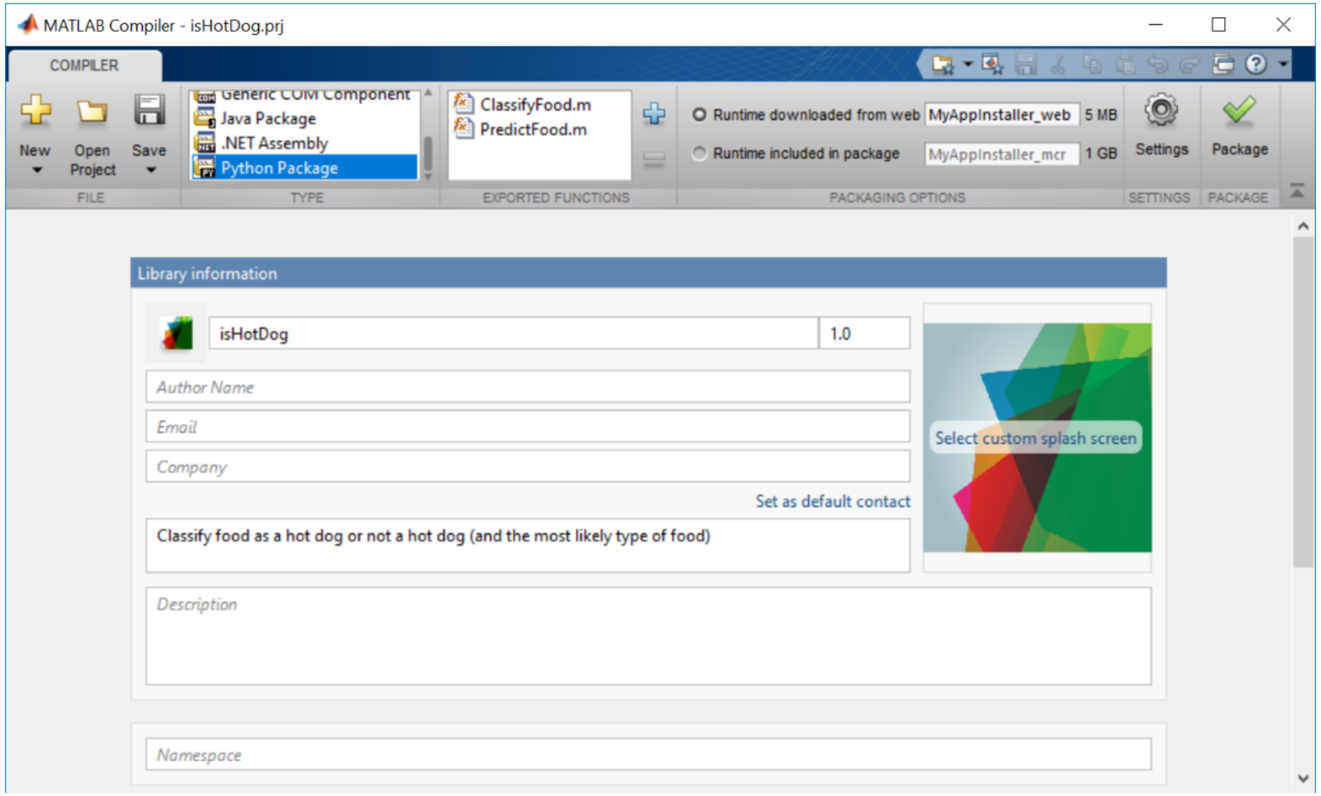
MATLAB Distributed Computing Server

- Scale up computation

5

Integrate with Production Systems

Package and deploy model to run anywhere



Enterprise Applications

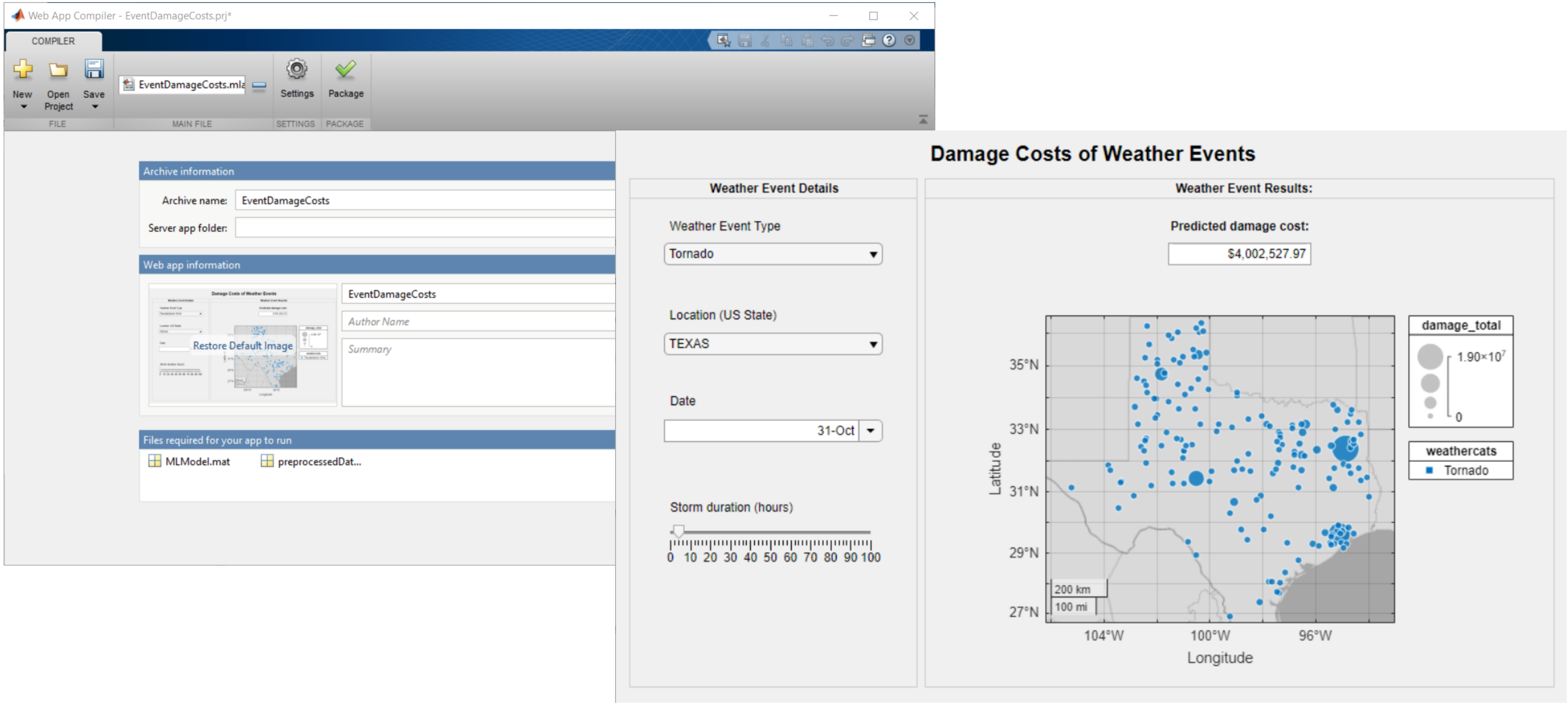
Dashboards

Devices

Cloud, IOT

5 Integrate with Production Systems

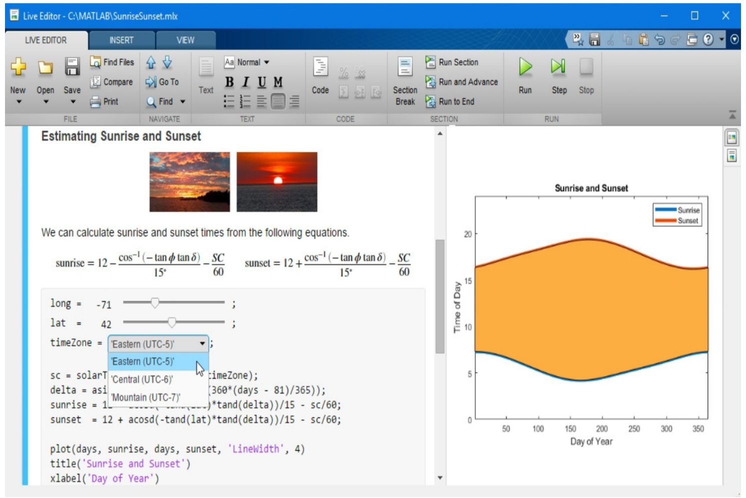
Create web application



Share your discoveries

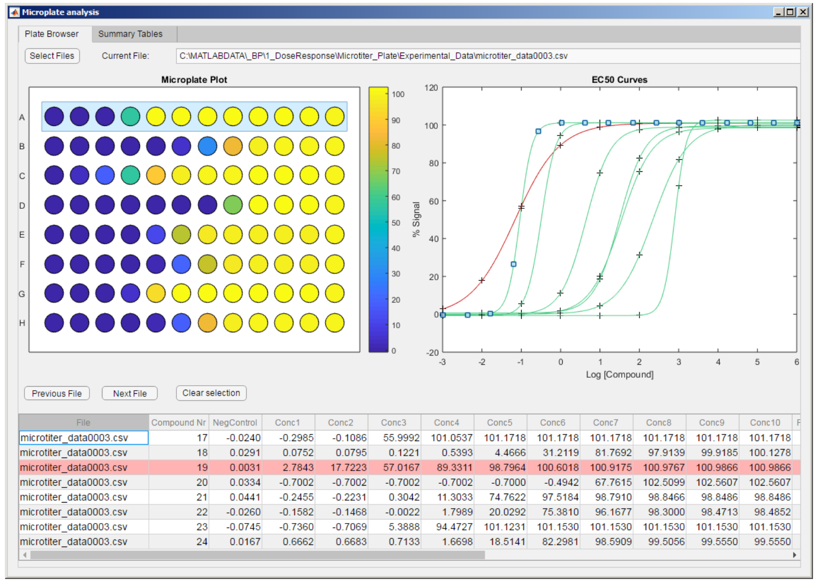
5 Integrate with Production Systems

Document and publish results

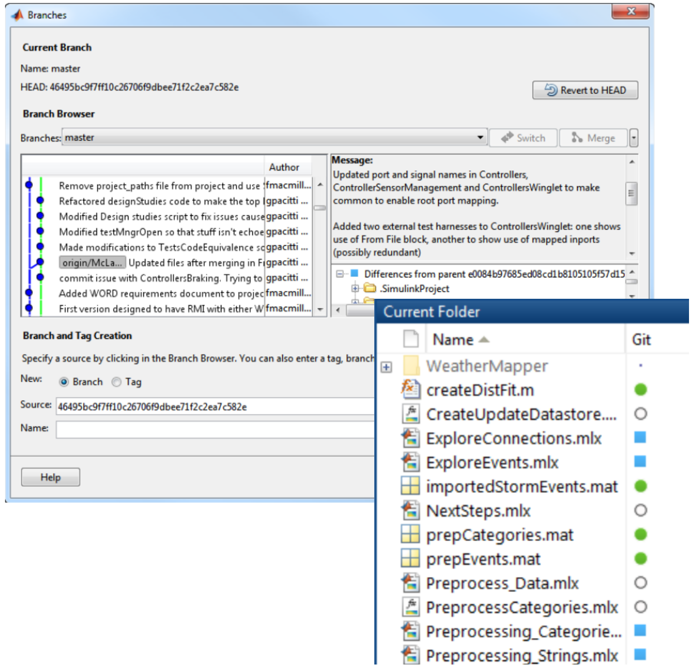


.pdf, html, LaTeX

Create apps



Use source control (GitHub, SVN)



Running MATLAB on Cedar/Graham/Niagara

- https://docs.computecanada.ca/wiki/MATLAB

- https://docs.computecanada.ca/wiki/Cedar
- https://docs.computecanada.ca/wiki/Graham
- https://docs.computecanada.ca/wiki/Niagara

Q & A