

Solving Big Data Problems in Computer Vision with MATLAB

Loren Shure

Why Are We Talking About Big Data?

- 100 hours of video uploaded to YouTube per minute¹
- Explosive increase in number of imaging devices
 - Webcams
 - Smartphone Cameras
 - IP Cameras
 - Industrial Cameras



1- KPCB 2013 Internet Trends <http://www.kpcb.com/blog/2013-internet-trends>

Challenges Posed by Large Sets of Images

1. How do I import several thousand images into MATLAB?
2. Can I find patterns or models to represent my image data?
3. How do I test and visualize my algorithm on many images?
4. What if my desktop or laptop doesn't have enough computing power?
5. What if my data is too big to fit into memory?

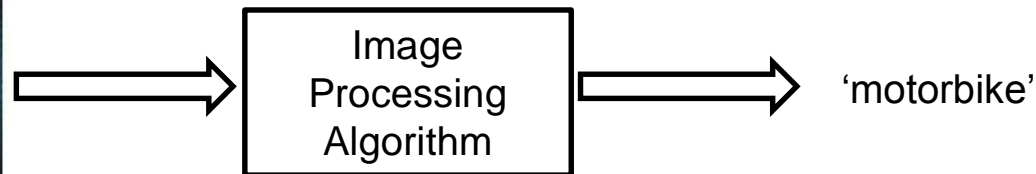
Goal: Show you **new functionality** in MATLAB to **augment existing workflows** to solve these challenges

Problem: Image Category Classification

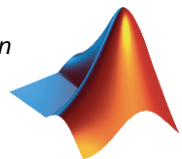
Given a large set of images of cars, planes and motorbikes



How to model the data to recognize which category an image belongs to ?



Data from Caltech 101: L. Fei-Fei, R. Fergus and P. Perona. *Learning generative visual models from few training examples: an incremental Bayesian approach tested on 101 object categories.* IEEE. CVPR 2004, Workshop on Generative-Model Based Vision. 2004
http://www.vision.caltech.edu/Image_Datasets/Caltech101/



Challenges Posed by Large Sets of Images

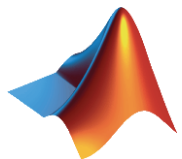
1. How do I import several thousand images into MATLAB?

- `imageSet`

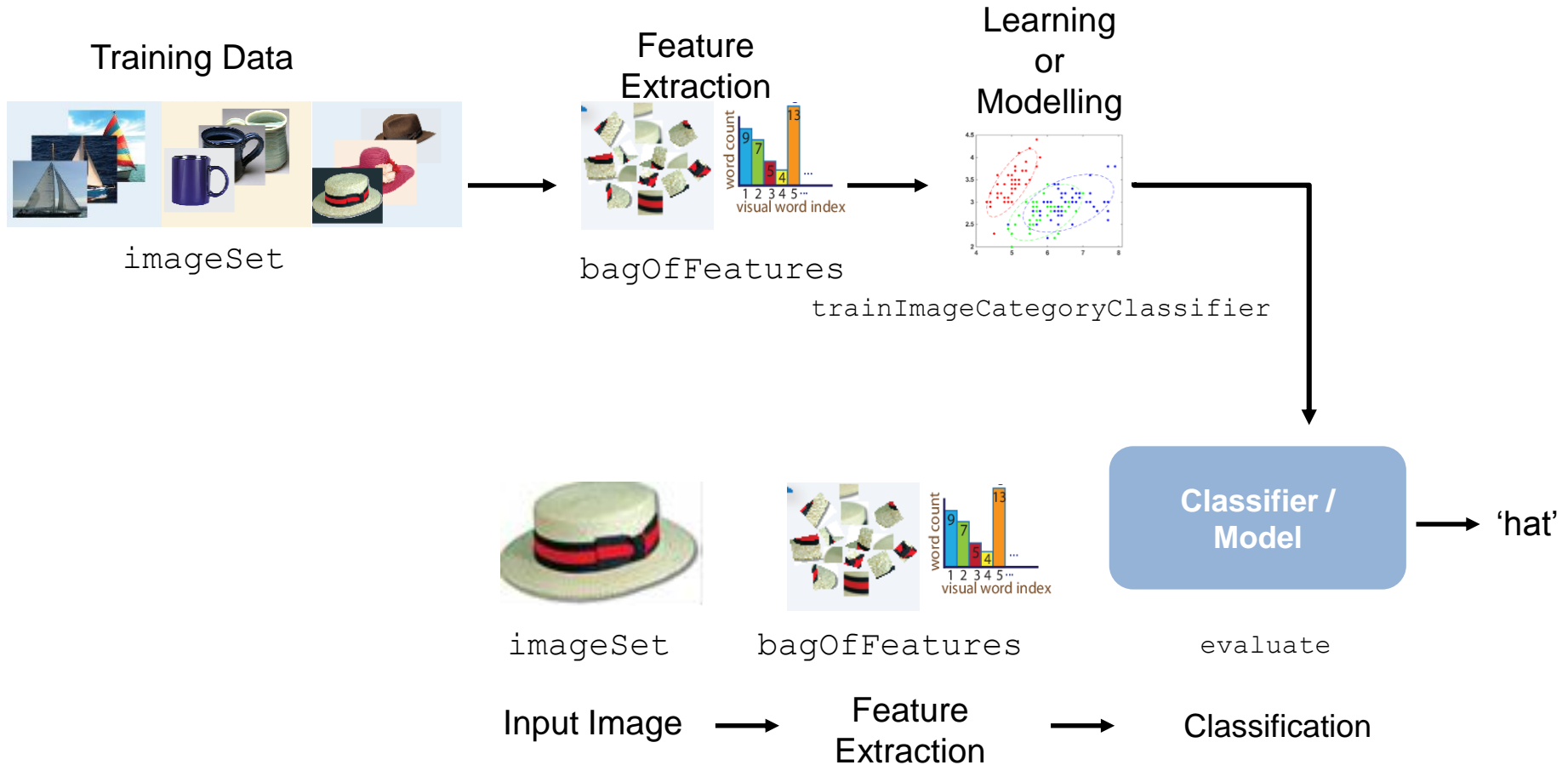
2. Can I find patterns or models to represent my image data?
3. How do I test and visualize my algorithm on many images?
4. What if my desktop or laptop doesn't have enough computing power?
5. What if my data is too big to fit into memory?

Challenges Posed by Large Sets of Images

1. How do I import several thousand images into MATLAB?
 - `imageSet`
- 2. Can I find patterns or models to represent my image data?**
3. How do I test and visualize my algorithm on many images?
4. What if my desktop or laptop doesn't have enough computing power?
5. What if my data is too big to fit into memory?



Machine Learning Workflow Using Images

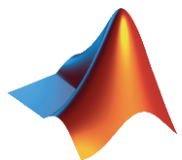


Challenges Posed by Large Sets of Images

1. How do I import several thousand images into MATLAB?
 - `imageSet` to manage large collections of images
- 2. Can I find patterns or models to represent my image data?**
 - Computer Vision System Toolbox, Statistics and Machine Learning Toolbox
3. How do I test and visualize my algorithm on many images?
4. What if my desktop or laptop doesn't have enough computing power?
5. What if my data is too big to fit into memory?

Challenges Posed by Large Sets of Images

1. How do I import several thousand images into MATLAB?
 - `imageSet` to manage large collections of images
2. Can I find patterns or models to represent my image data?
 - Computer Vision System Toolbox, Statistics and Machine Learning Toolbox
- 3. How do I test and visualize my algorithm on many images?**
4. What if my desktop or laptop doesn't have enough computing power?
5. What if my data is too big to fit into memory?



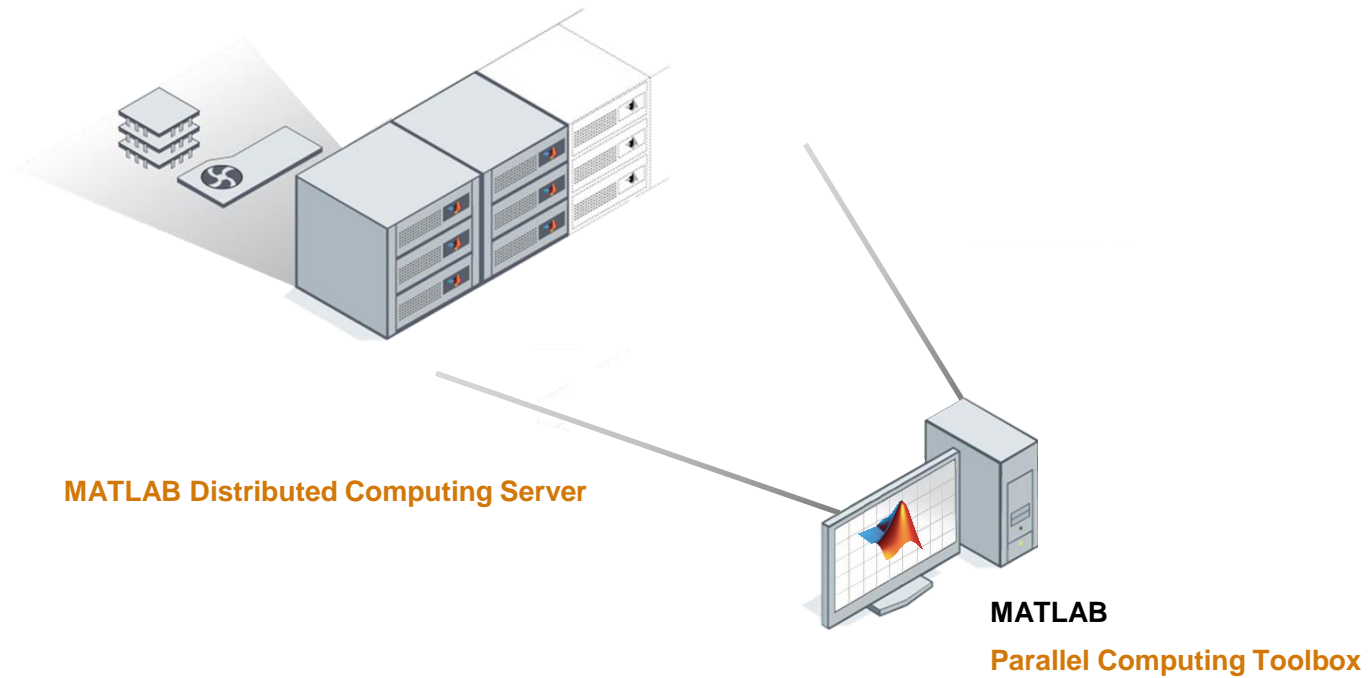
Challenges Posed by Large Sets of Images

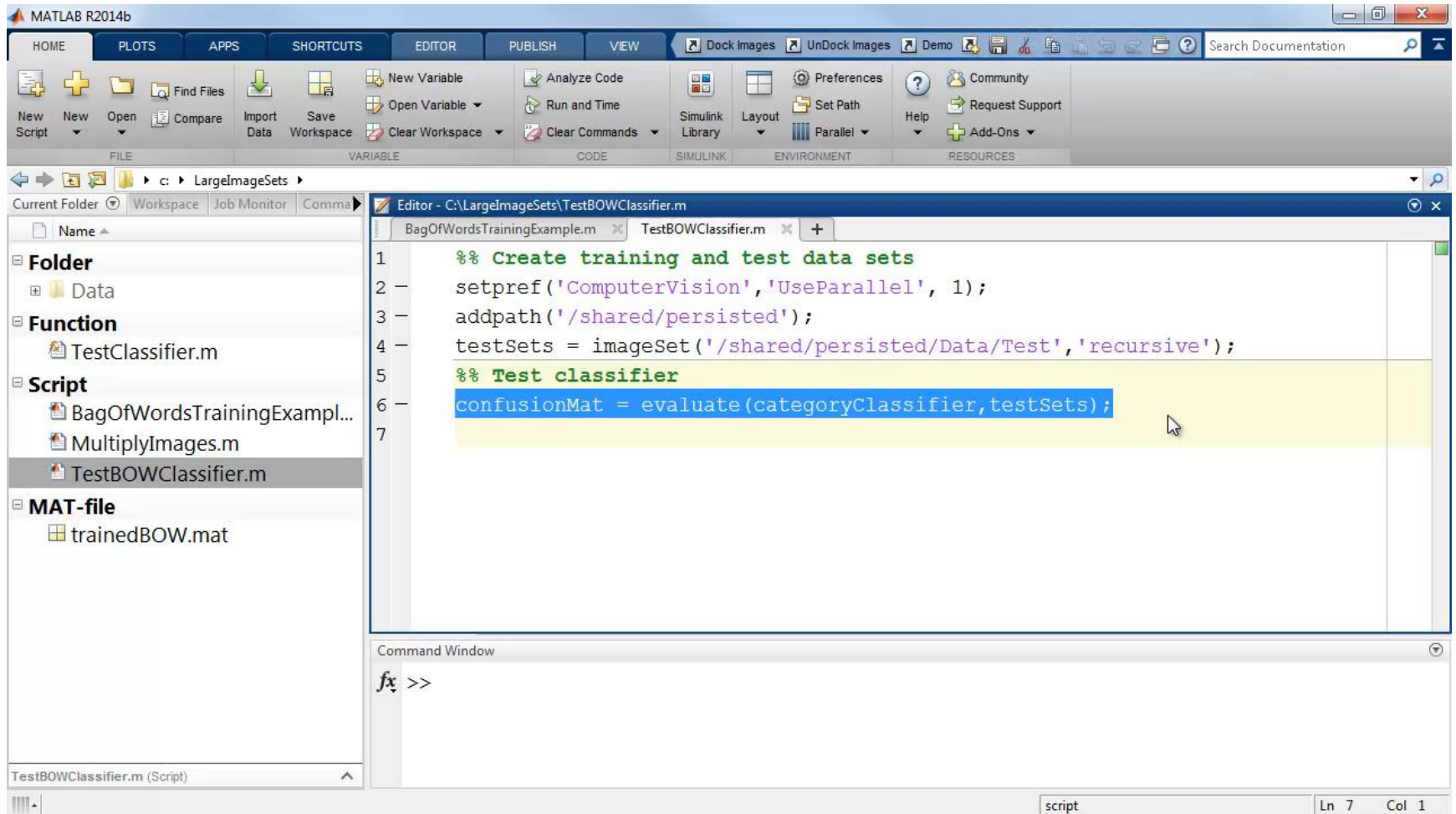
1. How do I import several thousand images into MATLAB?
 - `imageSet` to manage large collections of images
2. Can I find patterns or models to represent my image data?
 - Computer Vision System Toolbox, Statistics and Machine Learning Toolbox
- 3. How do I test and visualize my algorithm on many images?**
 - Image Batch Processor App
4. What if my desktop or laptop doesn't have enough computing power?
5. Can I acquire large sets of images using MATLAB ?

Challenges Posed by Large Sets of Images

1. How do I import several thousand images into MATLAB?
 - `imageSet` to manage large collections of images
2. Can I find patterns or models to represent my image data?
 - Computer Vision System Toolbox, Statistics and Machine Learning Toolbox
3. How do I test and visualize my algorithm on many images?
 - Image Batch Processor App
4. **What if my desktop or laptop doesn't have enough computing power?**
5. What if my data is too big to fit into memory?

Parallel Computing with MATLAB





MATLAB R2014b

HOME PLOTS APPS SHORTCUTS EDITOR PUBLISH VIEW Dock Images UnDock Images Demo Search Documentation

FILE VARIABLE CODE SIMULINK ENVIRONMENT RESOURCES

Current Folder: c:\LargeImageSets

Editor - C:\LargeImageSets\TestBOWClassifier.m

```

1  %% Create training and test data sets
2  setpref('ComputerVision','UseParallel', 1);
3  addpath('/shared/persisted');
4  testSets = imageSet('/shared/persisted/Data/Test','recursive');
5  %% Test classifier
6  confusionMat = evaluate(categoryClassifier,testSets);
7

```

Command Window

fx >>

TestBOWClassifier.m (Script) script Ln 7 Col 1

Challenges Posed by Large Sets of Images

1. How do I import several thousand images into MATLAB?
 - `imageSet` to manage large collections of images
2. Can I find patterns or models to represent my image data?
 - Computer Vision System Toolbox, Statistics and Machine Learning Toolbox
3. How do I test and visualize my algorithm on many images?
 - Image Batch Processor App
4. **What if my desktop or laptop doesn't have enough computing power?**
 - Parallel Computing Toolbox , MATLAB Distributed Computing Server
5. What if my data is too big to fit into memory?

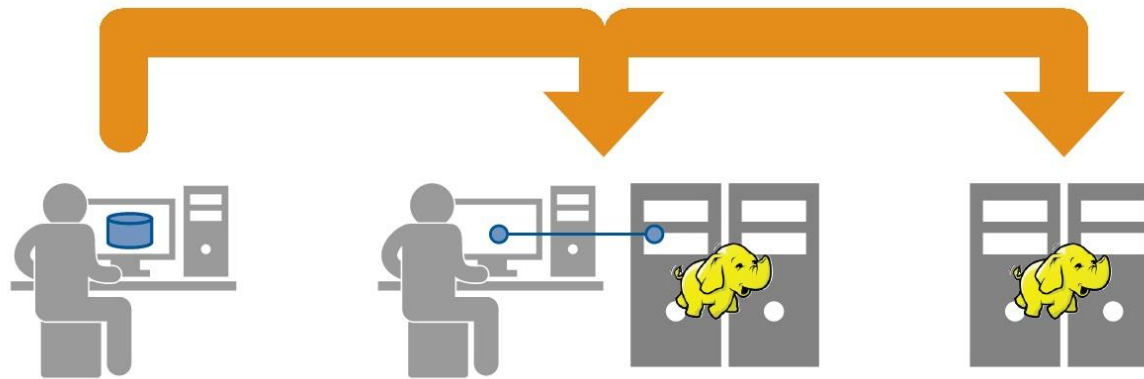
Challenges Posed by Large Sets of Images

1. How do I import several thousand images into MATLAB?
 - `imageSet` to manage large collections of images
2. Can I find patterns or models to represent my image data?
 - Computer Vision System Toolbox, Statistics and Machine Learning Toolbox
3. How do I test and visualize my algorithm on many images?
 - Image Batch Processor App
4. What if my desktop or laptop doesn't have enough computing power?
 - Parallel Computing Toolbox , MATLAB Distributed Computing Server
5. **What if my data is too big to fit into memory?**

Big Data Stored on Hadoop

R2014b

Easily migrate programs from workstation to Hadoop



On Your Workstation

- Access, explore, and analyze out-of-memory data
- Prototype MapReduce based algorithms
- Access portions of data sets on HDFS

Connected to Hadoop

- Run algorithms on Hadoop against data in HDFS

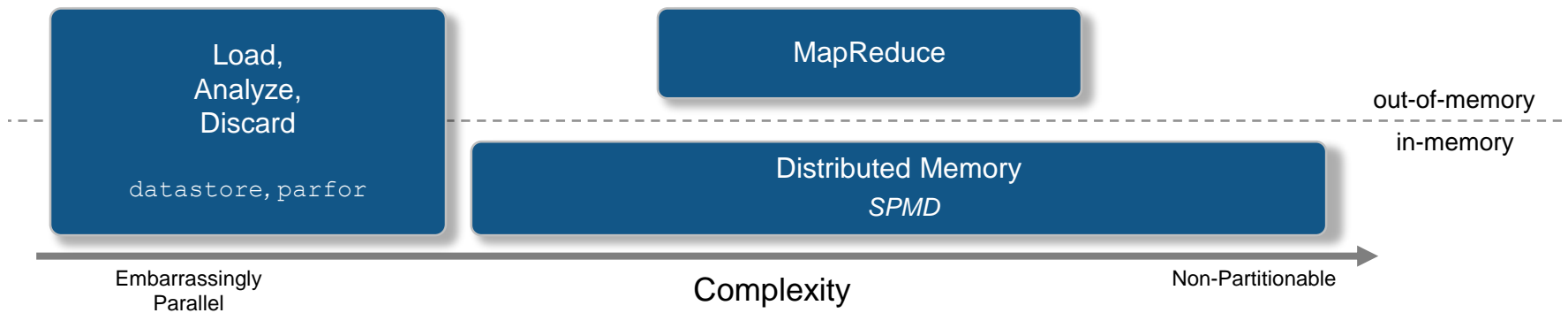
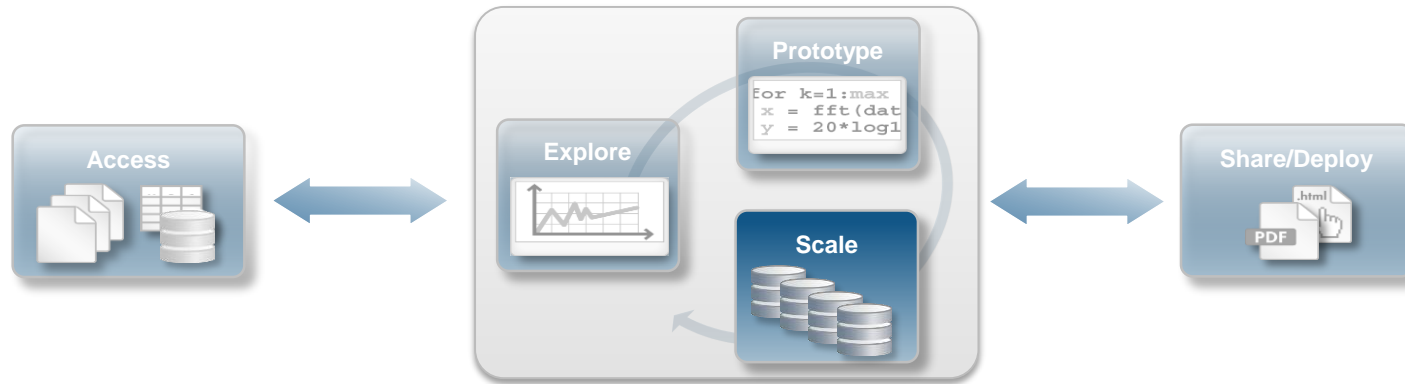
MATLAB Distributed Computing Server

Production Hadoop

- Create applications or components that execute on Hadoop

MATLAB Compiler

Scaled Analysis



Challenges Posed by Large Sets of Images

1. How do I import several thousand images into MATLAB?
 - `imageSet` to manage large collections of images
2. Can I find patterns or models to represent my image data?
 - Computer Vision System Toolbox, Statistics and Machine Learning Toolbox
3. How do I test and visualize my algorithm on many images?
 - Image Batch Processor App
4. What if my desktop or laptop doesn't have enough computing power?
 - Parallel Computing Toolbox , MATLAB Distributed Computing Server
5. **What if my data is too big to fit into memory?**
 - MapReduce, distributed arrays, SPMD, Hadoop

Thank You !