



# Lowering the barrier to analysing big data in the UK Public Sector

Alex Purkiss and Dr. Tim Hunter

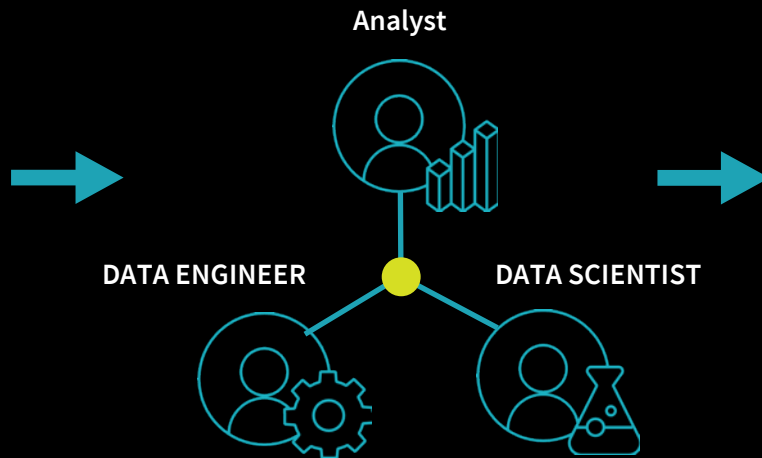
Data4good 2018



# AI is a Game Changing Opportunity

## LOTS OF NEW DATA

Citizen data  
Video feeds  
Social Media  
IoT & sensors  
Third party data.....



## OPPORTUNITY

Economic growth  
Efficiency  
Recommendation Engine  
New revenue streams  
...

Machine Learning Requires  
**Collaborative Experimentation on Big Data**

# New applications or better applications?

# ATKINS

Member of the SNC-Lavalin Group



Andrew Meyer

@AndrewMeyer3

Follow



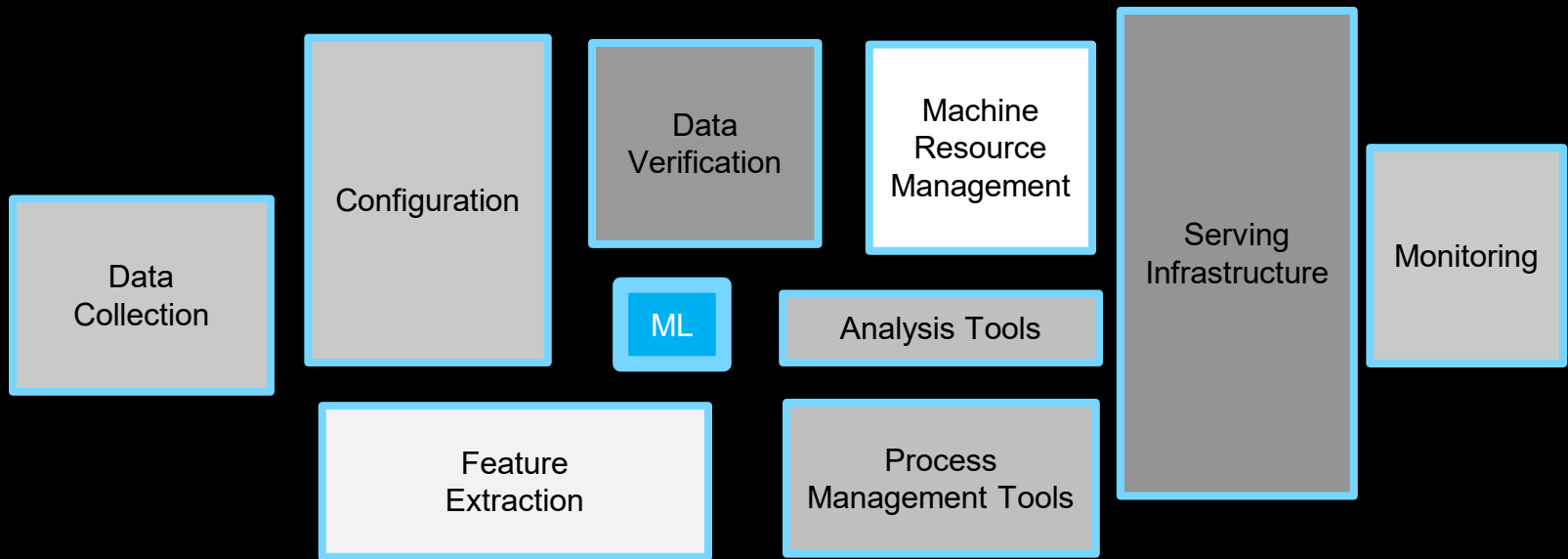
#ApacheSpark is used throughout @NHSDigital core data products and we will be exploring how @databricks can be used to enhance our services @SparkAISummit today #bigdata #unifiedanalytics

12:52 am - 3 Oct 2018

## How to exploit data at low cost, low risk and maximum benefit!

# Hardest Part of AI isn't AI, it's the data

*"Hidden Technical Debt in Machine Learning Systems," Google NIPS 2015*



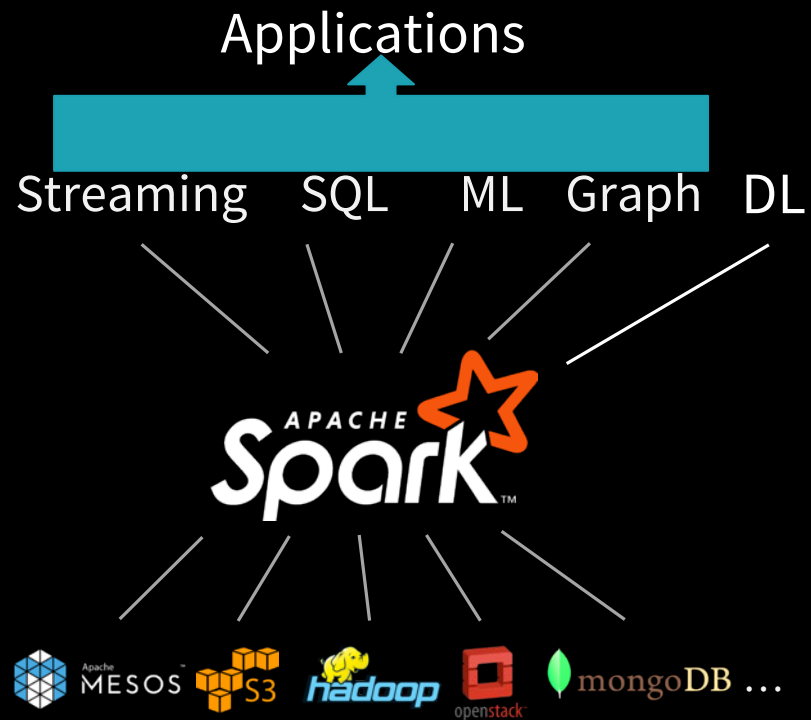
*Only a small fraction of real-world ML systems is composed of the ML code, as shown by the small blue box in the middle. The required surrounding infrastructure is vast and complex.*

# Apache Spark Philosophy

**Unified engine** for complete data applications

**High-level** user-friendly APIs

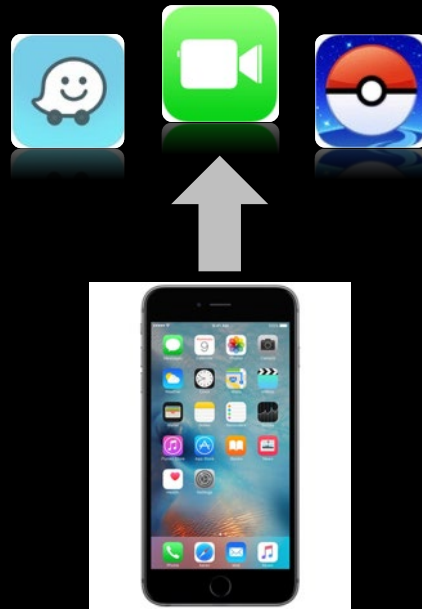
**Modular** solutions



# An Analogy ....



New applications



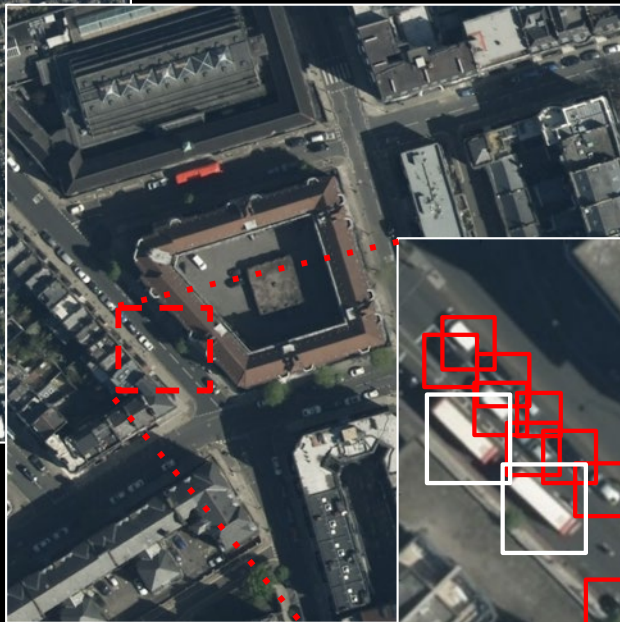
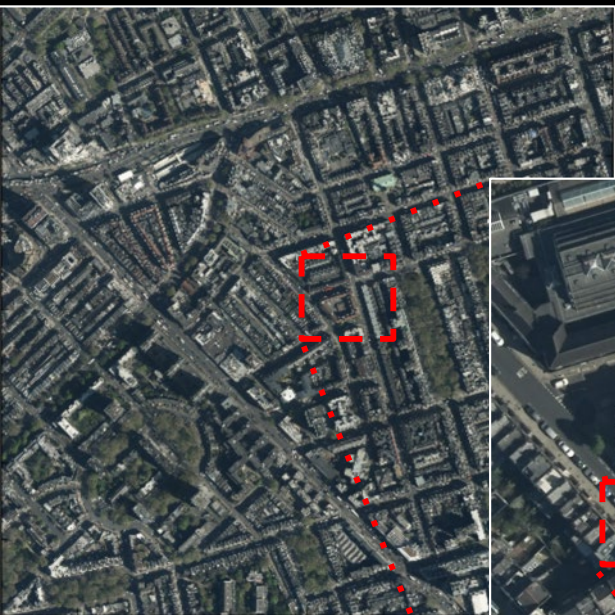
# An example: Identifying Vehicles in Aerial Imagery







```
vehicle_classes = {  
  18:('car', 'red'),  
  23:('truck', 'orange'),  
  19:('bus', 'white', 0.0)}
```



```
vehicle_classes = {  
  18:('car', 'red'),  
  23:('truck', 'orange'),  
  19:('bus', 'white', 0.0)  
}
```

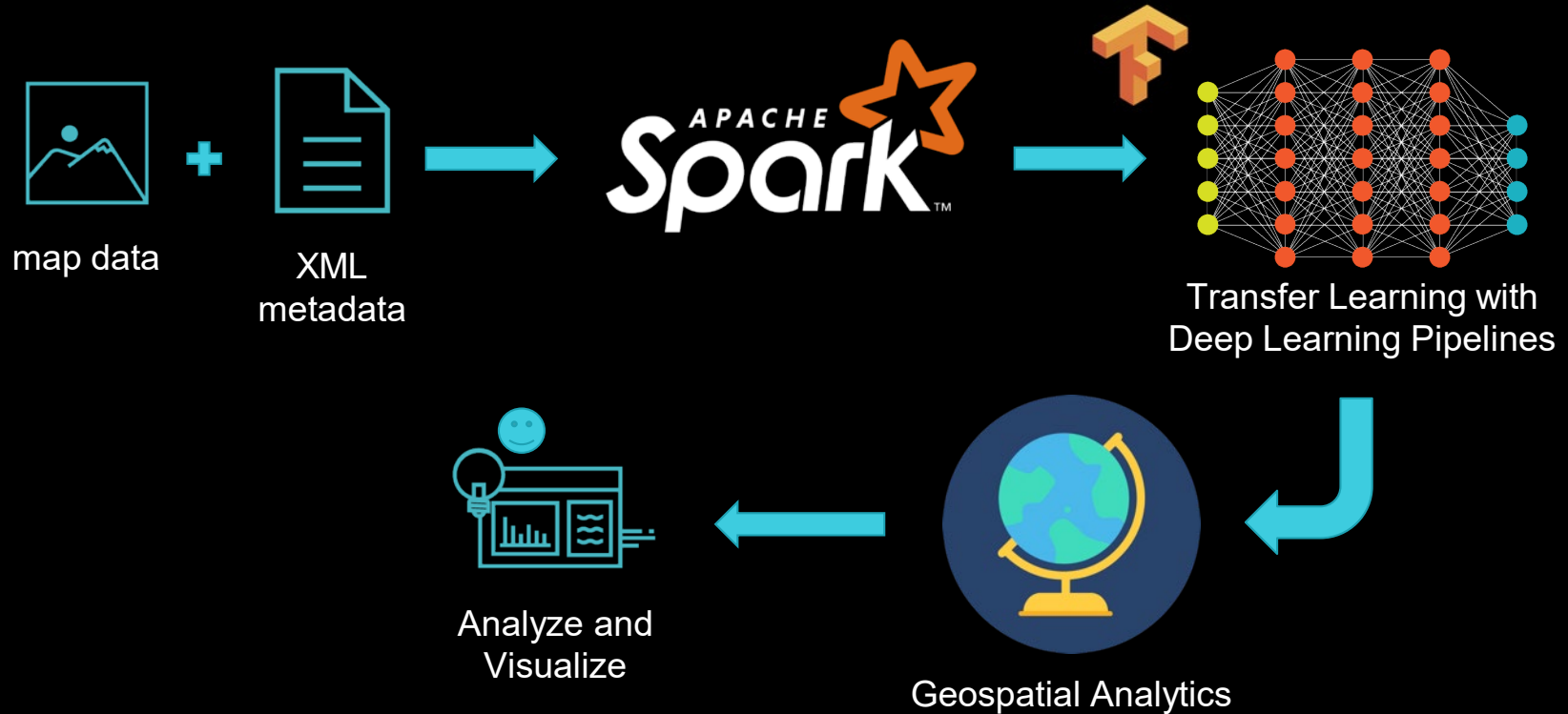
# Apache Spark - the glue of big data

One framework to bind all these libraries together

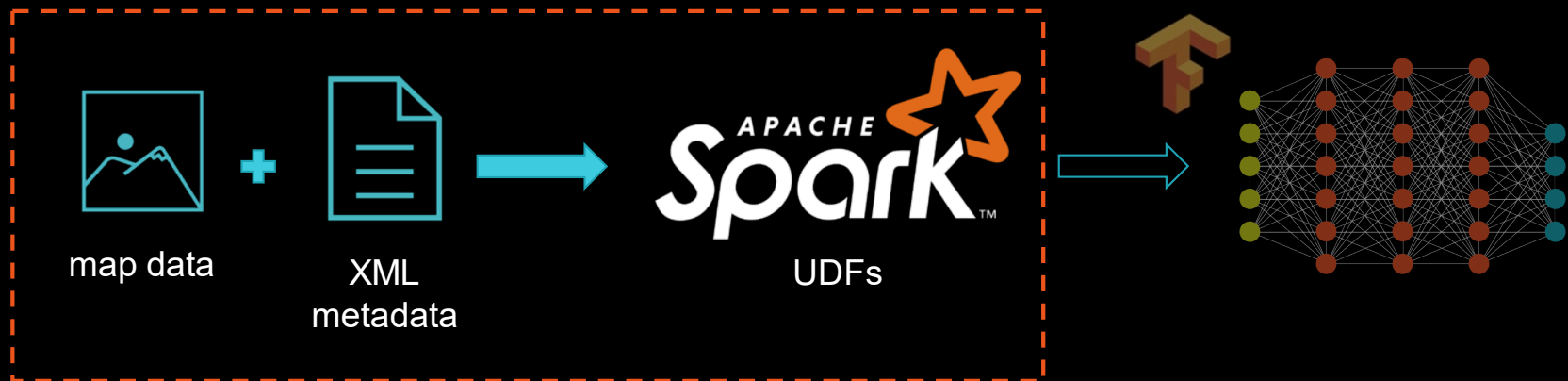
- At scale
- Allows pipelining
- Easily move data



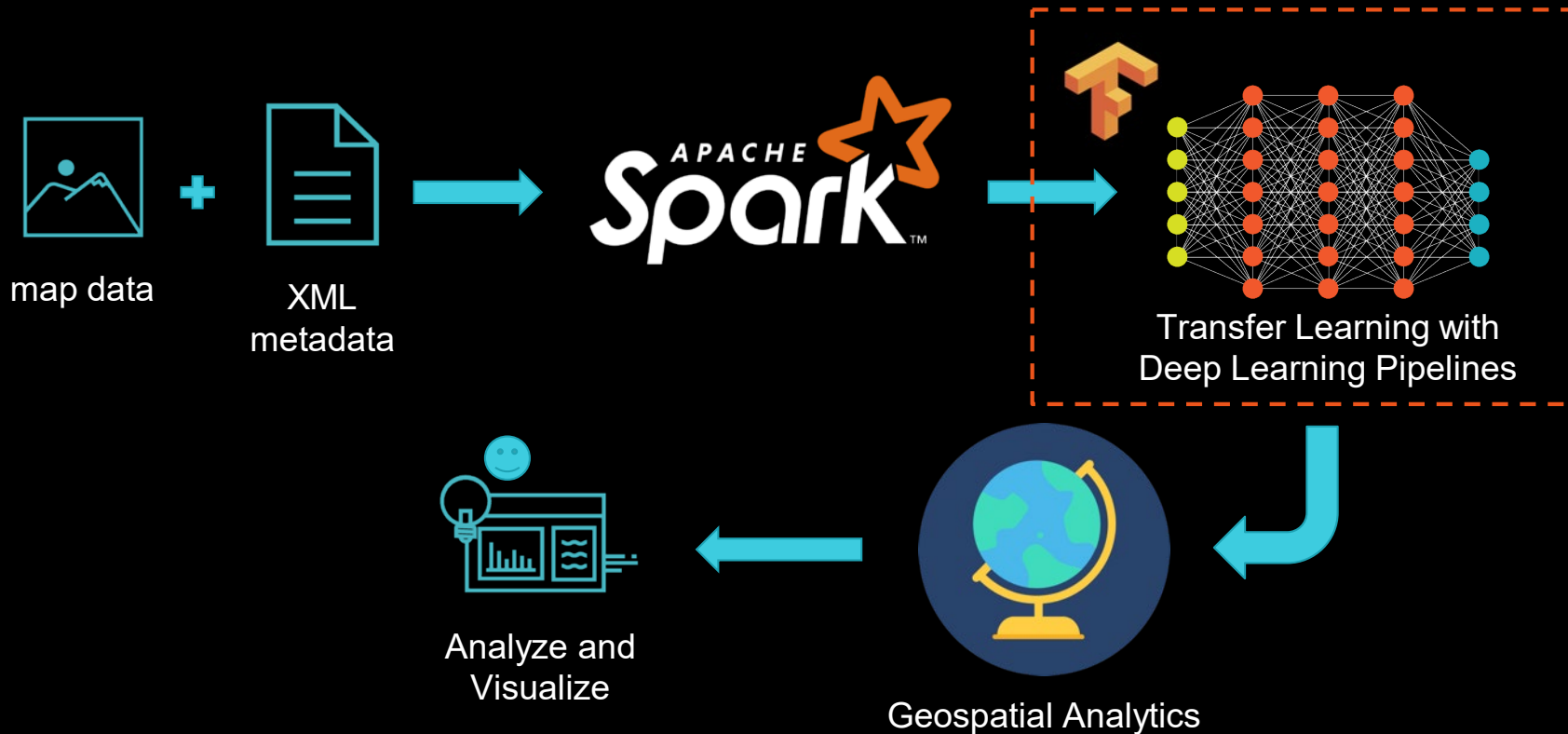
# High-level View of the Pipeline



# Parsing Image Data

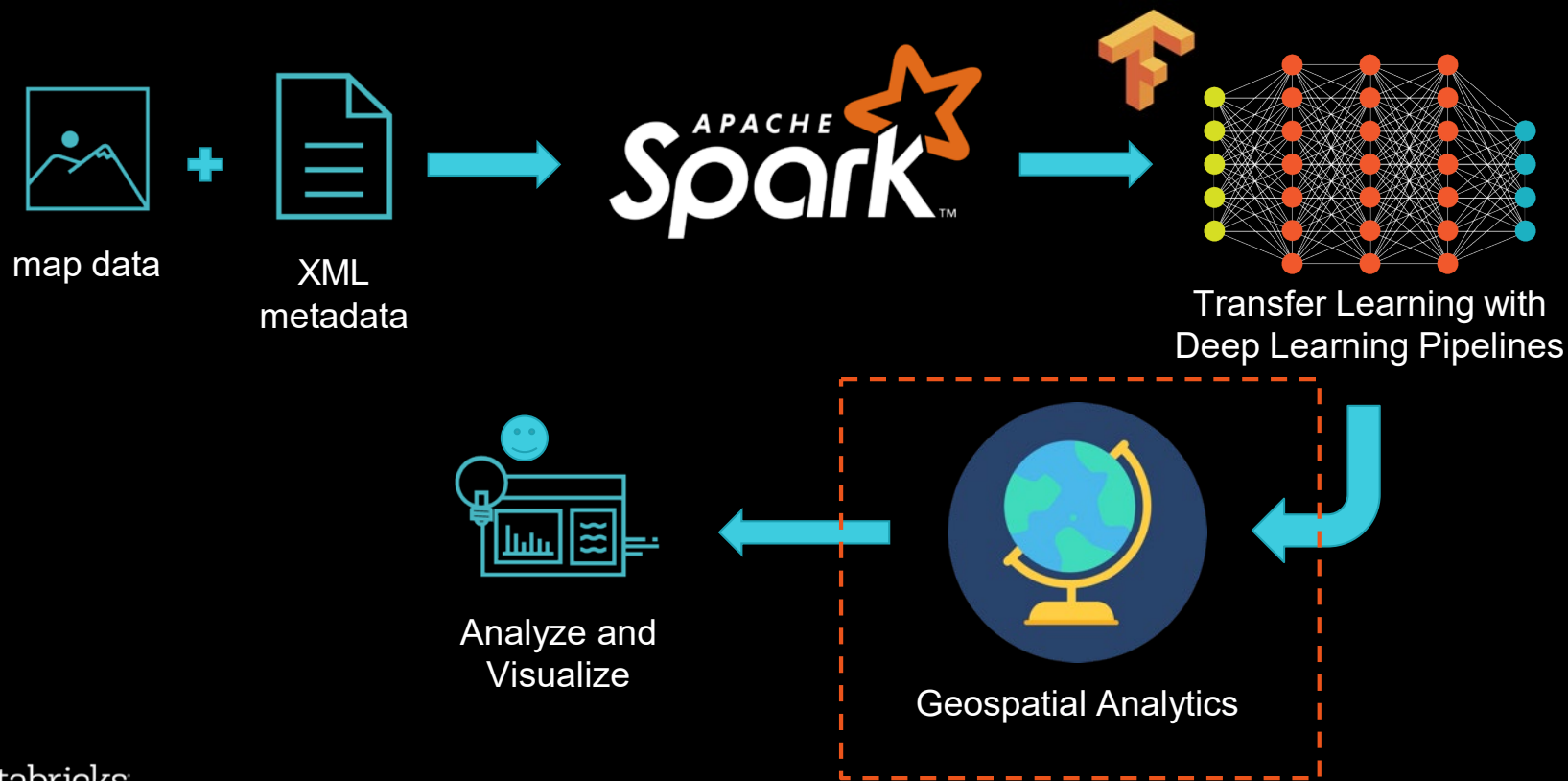


# Deep Learning





# Geospatial Analytics with Magellan



The background features a dark, textured surface with vibrant splashes of orange, red, and teal in the upper left corner. A faint, light-colored grid pattern is visible in the upper left quadrant, partially overlapping the colorful splashes.

# Demo



# Childsplay with

- AI is easy with the right tools
- Big data emerges from combining smaller datasets
- Easy integration between existing tools and Apache Spark



The background features a dark, textured surface with vibrant splashes of orange, red, and teal in the upper left corner. A series of light-colored, semi-transparent diamond shapes are arranged in a path that recedes into the distance.

Thank you