Booz | Allen | Hamilton
Commercial Solutions

Big Data Imagery Analysis and Application to Life Sciences

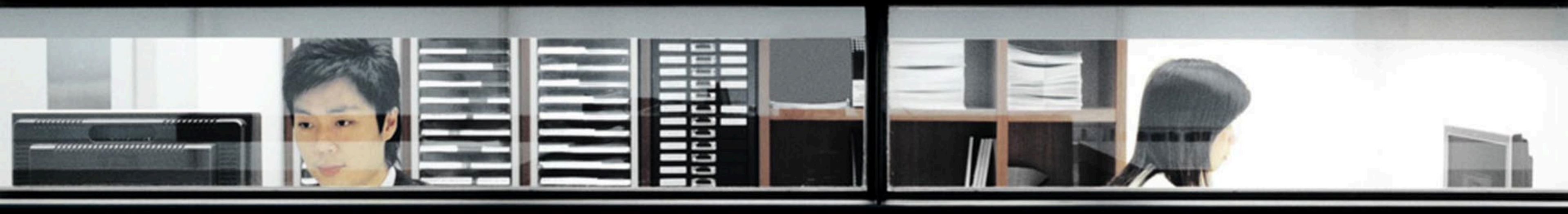
PRISME Forum

Peter Guerra

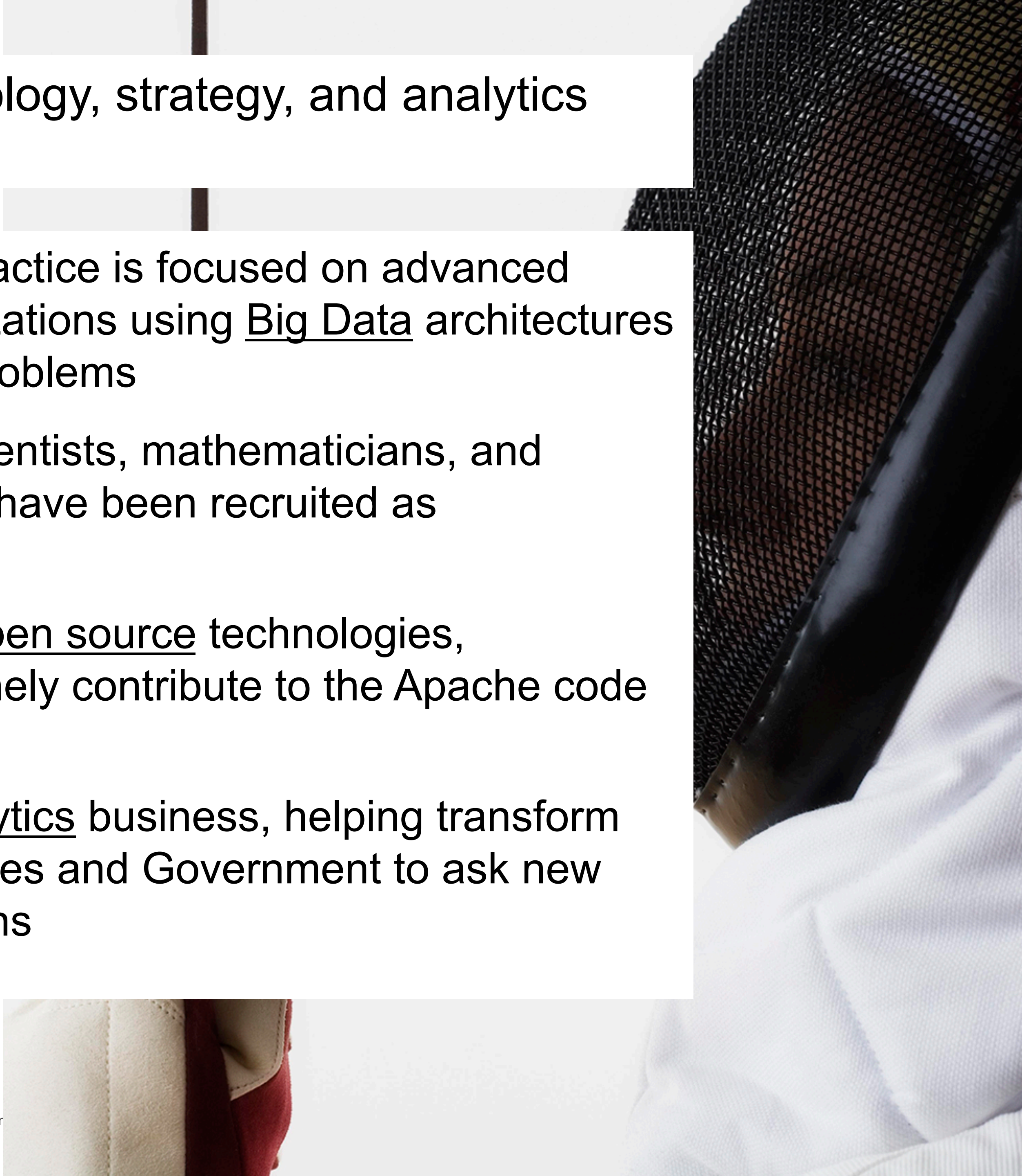
Rick Whitford

“The problem is not that we don’t have enough data – it’s that we have too much data and we need to make sense of it.”

- Erik Brynjolfsson, MIT

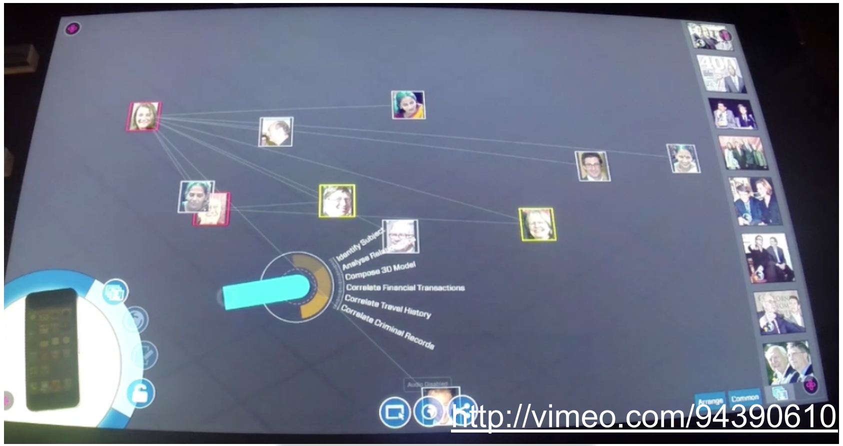


Booz Allen is a technology, strategy, and analytics firm



- Our Data Science practice is focused on advanced analytics and visualizations using Big Data architectures to solve real-world problems
- We are computer scientists, mathematicians, and domain experts who have been recruited as Data Scientists
- We are big fans of open source technologies, so much so we routinely contribute to the Apache code base
- We run a broad Analytics business, helping transform commercial businesses and Government to ask new and different questions

Our R&D demonstrates that scalable imagery and video analysis can be achieved using open source tools



http://vimeo.com/94390610

CHALLENGES

- Traditional large scale image analysis and are performed in stove-piped systems that rely on manual processes to discover connections between data
- Current system architectures are typically tied to a single vendor, don't scale linearly, and don't allow for combining other data sources with images/video

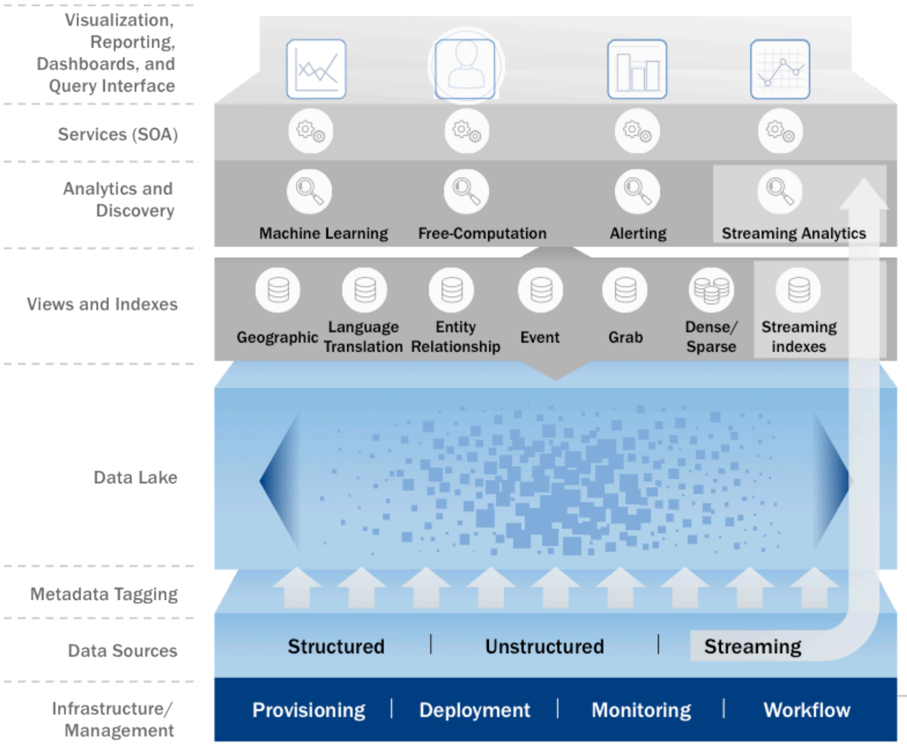
CAPABILITY

Our architecture leverages the power of Big Data architectures to:

- Perform scalable, extensible data ingest into the data analytics platform
  - Structured bulk image data or bulk video data
  - Metadata and unstructured data
- Recognize image detection using feature extraction and nearest neighbor algorithms
- Integrate multiple open source imagery algorithms into a scalable platform:
  - Apache Solr, Accumulo, Hadoop, and HTML5
  - jpgimage, OpenCV, Apache Mahout
  - Microsoft Touch Table for visualization
- Allows for mixed data type analysis of images or video for advanced processing in little time

Part of a broad Analytics business is an imagery analysis and advanced visualizations capabilities

A layered architecture to address volume and variety of data



Human Insights and Actions

Enabled by customizable interfaces and visualizations of the data

Analytics and Services

Your tools for analysis, modeling, testing, and simulations

Data Management

The single, secure repository for all of your valuable data

Infrastructure

The technology platform for storing and managing your data

Rapid Deployment

Applications and analytics can be deployed in weeks, rather than months or years

Better Scalability

New analytics, data sets, and more compute power can be rapidly added as business needs change

Reduced Costs

Open source technologies and use of cloud (public or virtual private) reduces infrastructure and licensing costs

Questions?

Check out the Field Guide to Data Science

Thank you!

- Peter Guerra
- (@petrguerra, guerra_peter@bah.com)

- Rick Whitford
- (@whitfordrick, whitford_richard@bah.com)