

# Infrastructure Education for Big Data and Analytics

John C. Hoag, PhD.

School of Information and Telecommunication Systems, Ohio University  
Electrical Engineering and Computer Science, Case Western Reserve  
[hoagjc@outlook.com](mailto:hoagjc@outlook.com)

IBM Insight EdCon 2014

## Proposed Presentation

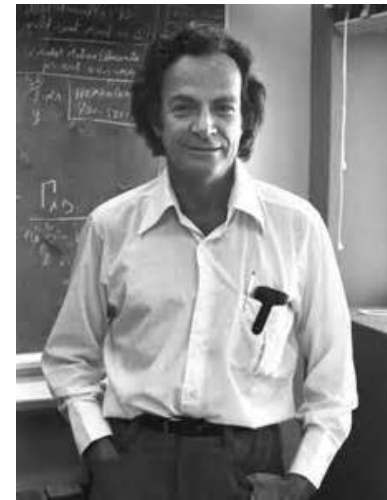
- Topic of infrastructure belongs in BDA curricula
- Compute + Network + Storage
  - apart from Algorithms
  - bottlenecks in unexpected places!
- Practical, scalable performance
- Locus: Datacenters, Cloud, BC/DR
  - Server virtualization, SDN, SDS

## Actual Presentation adds

- Programmatic lessons learned
- Science domains
  - Synergy, teaching and research
  - [Translational] Bioinformatics
  - Realtime data, ie Smart Grid
- OU nascent grad certificate
  - Thriving online MS in ISE
  - Coattails of med school branches
- Case BS in Datascience
- OSU/OSC HPC facility, network
  - NSF HPC facility

# Elements of Infrastructure Analyses

- Computational Complexity...closely tied to...
- [Processor] Cache Hits, not IMC
- [Channel] Utilization of Infiniband
- Development of probes, instrumentation
- Toward taxonomy, ontology of BDA performance measurement; bibliography
- Toward support of realtime transaction BDA
- Toward greater understanding of information entropy, information physics
- Collaborations welcome



IBM 1401 Operator



[datascience.case.edu](https://datascience.case.edu)

**Granted through the Case School of Engineering's Department of Electrical Engineering and Computer Science, the Data Science major will consist of a core curriculum that includes:**

- **mathematical modeling of data sources, including recitations that provide focus area applications**
- **informatics, including the acquisition, aggregation, assembly and security of data**
- **data analytics: examining raw data using analytics that focus on inference through the transformation of data to actionable information that improves decision-making**
- **visual analytics and user experiences**
- **capstone coursework that involves project-based applications of learning**

**The curriculum will further entail open source data management and analysis tools; a breadth of instruction in allied technical areas such as biology, physics or engineering; and a focus on a specific domain area: health, energy, and manufacturing and production. The new program will also feature a strong experiential learning component via partnerships with industry that provide co-op assignments and internships to students.**

Case Western Reserve President Barbara R. Snyder served as chair of BHEF in 2014, helping to focus its National Higher Education and Workforce Initiative on Data Science

Alumnus and former Microsoft COO Bob Herbold commits multi-million dollar gift to help launch data science program at Case Western Reserve University.

# Ohio University



- Graduate College nascent Analytics interdisciplinary certificate.
- Extant programs
  - Master of Information and Telecommunication Systems [online option]
  - MS/PhD in Industrial and Systems Engineering [online MS options]
  - MS/PhD in EECS with Bioinformatics Specialization/Emphasis
  - DO/PhD, Translational Biological Sciences - Main and two branch campuses
    - Medical School opened in Dublin, Ohio near the new IBM Analytics Solution Center
- Consensus BDA core curriculum plus recurring seminar, infrastructure elective
  - Graduate capstone project in home department with interdisciplinary support