Earth Science Data Analytics: Preparing for Extracting Knowledge from Information
Steve Kemple1, Lindsay Barbieri2
1 NASA Goddard Space Flight Center, 2 The University of Vermont

Earth Science Data Analytics/Science Skills Needed: Overall Experiences /Operational Needs

Data Analytics / Data Science

- Data scientist studies methods of analyzing data, ways of storing it, and ways of presenting it.
- Data analytics is performed by the practitioners who applies tools and techniques to co-analyze data.
- Both, data science and data analytics require very similar skill sets.
- Once acquired, it becomes up to the individual to determine how best to use these skills, based on their interest and aptitude.

General Experiences

- Need skills in: mathematics, numerical modeling, statistics, software engineering and the ability to integrate data across multiple domains.
- Need expertise in tools and techniques: rule learning, classification, cluster analysis, data fusion, machine learning, neural networks, anomaly detection, modeling, time series analysis, visualization.
- Need knowledge in particular science domains where data analytics can advance our understanding of science.
- The role is a hybrid one... skills to support domain scientists with data and computational needs to communicate across domains.

Operational Needs

- Need to facilitate making data more useful.
- Should be interdisciplinary from the start.
- Learn your math and statistics.
- Know the importance of the data lifecycle.
- Understand what the data says and how to understand the data.
- Know the territory... What information is available. Where to get it. How it is generated, how to use it. How it can be used.
- Understand data, metadata, and data integration.
- Know how to apply the techniques to the discipline.
- Learn through internships.

References

http://101.datascience.community/2012/04/09/
http://datascience.community/colleges

Percent of Degree Programs Pertaining to Data Analytics/Science:

<table>
<thead>
<tr>
<th>Program Focus Areas</th>
<th>B on line</th>
<th>B ca impusa</th>
<th>M on line</th>
<th>M ca impusa</th>
<th>PhD on line</th>
<th>PhD ca impusa</th>
</tr>
</thead>
<tbody>
<tr>
<td>Business analytics</td>
<td>15</td>
<td>22</td>
<td>41</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Data/Information science</td>
<td>15</td>
<td>10</td>
<td>21</td>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Data analytics/Analytics</td>
<td>1</td>
<td>4</td>
<td>12</td>
<td>34</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Quantitative Analysis/Applied Statistics</td>
<td>2</td>
<td>17</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Information technology/systems management</td>
<td>4</td>
<td>19</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Computer science</td>
<td>1</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Health/Bio analytics</td>
<td>15</td>
<td>15</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Business management</td>
<td>3</td>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Environmental Informatics</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

What the Universities Offer (July, 2016 study and comparison with 2013 Study)

What Else Universities Should Consider Offering

Programs Pertaining to Data Science/Data Analytics: Course Topics Most Offered

Overall:
- Statistics, Data Mining, Database Management/Analysis
- Data Science, Data Analytics, and Computer Science
- Data Mining, Mathematics, Statistics, Machine Learning, Data Visualization
- Data Science, Data Analytics, Information Systems:
  - Database Management/Analysis
  - Quantitative Analysis:
    - Data Mining, Mathematics, Statistics
- Relevant Courses Offered:
  - Programming, Neural Networks, Data Analysis, Artificial Intelligence, Clustering, Time Series, Database Warehousing, Pattern Recognition, GIS, Remote Sensing, Text Mining, Information/Knowledge Management

Data Analytics/Data Science Techniques Practiced:
- Anomaly Detection
- Artificial Intelligence
- Classification
- Cluster Analysis
- Data Compression
- Data Engineering
- Data Fusion
- Data Mining
- Database Warehousing
- Database Management
- Machine Learning
- Mathematics
- Modeling
- Neural networks
- Pattern Recognition
- Rule Learning
- Signal Processing
- Statistics
- Time series
- Visualization
- Ability to integrate data across multiple domains
- Support domain scientists with data and computational needs to communicate across domains (be interdisciplinary)
- Knowledge of data life cycle
- Software engineering - Programming

Every Earth science program should contain training in Data analytics/science and Programming (see above).