More Than A Buzzword: Big Data in the Environmental Arena


Brooke Roecker
Senior Environmental Data Analyst

Mark Packard, PG, CPG
President/CEO

www.ddmsinc.com
Presentation Outline

- Big Data Defined
- Environmental Data Past & Present
- Today’s Tools and Approaches
- Example Projects
- Future Considerations
Big Data Defined: (yet again)

"Big data is high volume, high velocity, and/or high variety information assets that require new forms of processing to enable enhanced decision making, insight discovery and process optimization."

Big Data Defined: (yet again)

"Big data is high volume, high velocity, and/or high variety information assets that require new forms of processing to enable enhanced decision making, insight discovery and process optimization."

• **Velocity**: high frequency data
• **Variety**: mixed data/attributes

• **Volume**: very large datasets
• **VERACITY**: Accuracy of data

Diya Soubra, “The 3Vs that define Big Data”. http://www/datasciencecentral.com/forum/topics/the-3vs-that-define-big-data
Where have we come from?

- Hand-written field logs
- Text files
- Spreadsheets
- Simple Reports
Where are we today?

- Older technologies remain
- Database storage
- Out of the box storage/analysis tools
  - EQuIS™
  - ENFOS
  - Locus
  - Project Portal™
Where are we today?

• Older technologies remain
• Database storage
• Out of the box storage/analysis tools
  – EQuIS™
  – ENFOS
  – Locus
  – Project Portal™

• Limitations?
• What data are you not managing/analyzing?
“Bigger” data tools available today

• High frequency advancements:
  – EQuIS Live
  – Project Portal Analytics Module

• Analysis and modeling tools
  – Spatial: ArcGIS, EVS
  – Visualization: Tableau

• Custom scripting (R, Python, T-SQL...)
  – SSAS, Weka
  – MatPlotLib (Python), ggplot2 (R)
Project Examples
Project: **Surface Water Monitoring**

- High velocity data, 5-minute intervals
- Teledyne ISCO samplers
- Historical data archived in raw MS Excel files

**Challenge:**

- Dataset too large for “Big picture” trends
- Storing/archiving data long-term
- Centralized access for project team
Project: **Surface Water Monitoring**

**Solution:**
- Streamlined data resampling & import routine
  - Resample from 5-minute to 12-hour averages (or totals)

---

**Raw Data**

**Resampled Data**
Project: **Surface Water Monitoring**

**Solution:**

- Data available to project team via Project Portal
  - Environmental Database module for resampled data
  - Documents module for raw data
Project: RAD Site Monitoring

Challenge:

- High volume, high velocity sensor data w/ telemetry
- Automated database storage
- Visual analysis of high volume weather data

<table>
<thead>
<tr>
<th>Date</th>
<th>Time</th>
<th>Baro</th>
<th>EPM1</th>
<th>EPM2</th>
<th>EPM3</th>
<th>EPM4</th>
<th>Prcp</th>
<th>RH</th>
<th>Temp</th>
<th>WDir</th>
<th>WSpd</th>
<th>Pvel</th>
</tr>
</thead>
<tbody>
<tr>
<td>3/4/2015</td>
<td>16:07:03</td>
<td>29.818</td>
<td>8</td>
<td>11</td>
<td>2</td>
<td>13</td>
<td>0.23</td>
<td>45.807</td>
<td>39.635</td>
<td>291.894</td>
<td>9.322</td>
<td>1</td>
</tr>
<tr>
<td>3/4/2015</td>
<td>16:09:20</td>
<td>29.818</td>
<td>6</td>
<td>9</td>
<td>2</td>
<td>10</td>
<td>0.23</td>
<td>45.994</td>
<td>40.424</td>
<td>253.396</td>
<td>6.126</td>
<td>1</td>
</tr>
<tr>
<td>3/4/2015</td>
<td>16:10:37</td>
<td>29.819</td>
<td>7</td>
<td>7</td>
<td>2</td>
<td>10</td>
<td>0.23</td>
<td>45.415</td>
<td>37.941</td>
<td>262.800</td>
<td>3.628</td>
<td>1</td>
</tr>
<tr>
<td>3/4/2015</td>
<td>16:13:55</td>
<td>29.821</td>
<td>7</td>
<td>8</td>
<td>2</td>
<td>10</td>
<td>0.23</td>
<td>44.88</td>
<td>38.819</td>
<td>222.352</td>
<td>4.42</td>
<td>1</td>
</tr>
<tr>
<td>3/4/2015</td>
<td>16:15:13</td>
<td>29.819</td>
<td>7</td>
<td>7</td>
<td>2</td>
<td>10</td>
<td>0.23</td>
<td>44.371</td>
<td>39.629</td>
<td>285.836</td>
<td>4.844</td>
<td>1</td>
</tr>
<tr>
<td>3/4/2015</td>
<td>16:16:29</td>
<td>29.818</td>
<td>6</td>
<td>7</td>
<td>2</td>
<td>10</td>
<td>0.23</td>
<td>44.28</td>
<td>38.819</td>
<td>289.678</td>
<td>3.869</td>
<td>1</td>
</tr>
<tr>
<td>3/4/2015</td>
<td>16:19:47</td>
<td>29.814</td>
<td>6</td>
<td>5</td>
<td>3</td>
<td>7</td>
<td>0.23</td>
<td>43.91</td>
<td>40.218</td>
<td>269.342</td>
<td>3.879</td>
<td>1</td>
</tr>
<tr>
<td>3/4/2015</td>
<td>16:21:04</td>
<td>29.817</td>
<td>5</td>
<td>10</td>
<td>4</td>
<td>8</td>
<td>0.23</td>
<td>43.901</td>
<td>37.768</td>
<td>255.619</td>
<td>6.041</td>
<td>1</td>
</tr>
<tr>
<td>3/4/2015</td>
<td>16:22:22</td>
<td>29.816</td>
<td>4</td>
<td>13</td>
<td>2</td>
<td>8</td>
<td>0.23</td>
<td>45.294</td>
<td>36.176</td>
<td>291.813</td>
<td>4.188</td>
<td>1</td>
</tr>
<tr>
<td>3/4/2015</td>
<td>16:24:39</td>
<td>29.82</td>
<td>3</td>
<td>12</td>
<td>7</td>
<td>0.23</td>
<td>47.733</td>
<td>38.544</td>
<td>293.003</td>
<td>8.017</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>3/4/2015</td>
<td>16:26:57</td>
<td>29.821</td>
<td>4</td>
<td>12</td>
<td>2</td>
<td>6</td>
<td>0.23</td>
<td>47.715</td>
<td>39.113</td>
<td>265.432</td>
<td>7.315</td>
<td>1</td>
</tr>
<tr>
<td>3/4/2015</td>
<td>16:27:15</td>
<td>29.819</td>
<td>3</td>
<td>11</td>
<td>2</td>
<td>7</td>
<td>0.23</td>
<td>48.261</td>
<td>36.859</td>
<td>279.921</td>
<td>7.99</td>
<td>1</td>
</tr>
<tr>
<td>3/4/2015</td>
<td>16:28:32</td>
<td>29.819</td>
<td>2</td>
<td>10</td>
<td>2</td>
<td>8</td>
<td>0.23</td>
<td>47.978</td>
<td>37.012</td>
<td>255.555</td>
<td>4.883</td>
<td>1</td>
</tr>
<tr>
<td>3/4/2015</td>
<td>16:29:49</td>
<td>29.814</td>
<td>3</td>
<td>10</td>
<td>2</td>
<td>8</td>
<td>0.23</td>
<td>47.732</td>
<td>39.042</td>
<td>266.987</td>
<td>5.636</td>
<td>1</td>
</tr>
<tr>
<td>3/4/2015</td>
<td>16:32:06</td>
<td>29.819</td>
<td>2</td>
<td>9</td>
<td>2</td>
<td>9</td>
<td>0.23</td>
<td>47.142</td>
<td>37.127</td>
<td>246.404</td>
<td>4.384</td>
<td>1</td>
</tr>
<tr>
<td>3/4/2015</td>
<td>16:33:23</td>
<td>29.819</td>
<td>4</td>
<td>8</td>
<td>2</td>
<td>7</td>
<td>0.23</td>
<td>47.142</td>
<td>37.127</td>
<td>258.235</td>
<td>5.192</td>
<td>1</td>
</tr>
<tr>
<td>3/4/2015</td>
<td>16:34:41</td>
<td>29.819</td>
<td>5</td>
<td>7</td>
<td>2</td>
<td>8</td>
<td>0.23</td>
<td>46.491</td>
<td>38.128</td>
<td>234.716</td>
<td>4.825</td>
<td>1</td>
</tr>
<tr>
<td>3/4/2015</td>
<td>16:35:38</td>
<td>29.823</td>
<td>8</td>
<td>6</td>
<td>2</td>
<td>8</td>
<td>0.23</td>
<td>46.294</td>
<td>35.301</td>
<td>251.514</td>
<td>6.705</td>
<td>1</td>
</tr>
<tr>
<td>3/4/2015</td>
<td>16:40:16</td>
<td>29.824</td>
<td>8</td>
<td>6</td>
<td>2</td>
<td>8</td>
<td>0.23</td>
<td>46.573</td>
<td>37.297</td>
<td>246.073</td>
<td>6.427</td>
<td>1</td>
</tr>
<tr>
<td>3/4/2015</td>
<td>16:41:33</td>
<td>29.825</td>
<td>5</td>
<td>5</td>
<td>2</td>
<td>7</td>
<td>0.23</td>
<td>46.314</td>
<td>38.09</td>
<td>254.202</td>
<td>8.087</td>
<td>1</td>
</tr>
<tr>
<td>3/4/2015</td>
<td>16:44:51</td>
<td>29.827</td>
<td>5</td>
<td>4</td>
<td>2</td>
<td>8</td>
<td>0.23</td>
<td>46.959</td>
<td>37.012</td>
<td>245.586</td>
<td>5.926</td>
<td>1</td>
</tr>
<tr>
<td>3/4/2015</td>
<td>16:46:08</td>
<td>29.826</td>
<td>6</td>
<td>3</td>
<td>2</td>
<td>7</td>
<td>0.23</td>
<td>46.172</td>
<td>35.853</td>
<td>260.303</td>
<td>4.684</td>
<td>1</td>
</tr>
<tr>
<td>3/4/2015</td>
<td>16:47:26</td>
<td>29.828</td>
<td>6</td>
<td>3</td>
<td>6</td>
<td>9</td>
<td>0.23</td>
<td>47.528</td>
<td>36.646</td>
<td>255.07</td>
<td>3.493</td>
<td>1</td>
</tr>
<tr>
<td>3/4/2015</td>
<td>16:48:43</td>
<td>29.829</td>
<td>5</td>
<td>3</td>
<td>11</td>
<td>13</td>
<td>0.23</td>
<td>47.715</td>
<td>37.344</td>
<td>264.796</td>
<td>5.037</td>
<td>1</td>
</tr>
</tbody>
</table>
Project: **RAD Site Monitoring**

**Solution:**
- Automated data import via web upload
- Database available to field staff via Project Portal
- Wind rose graphics to visualize data via EnvirolInsite
- QA of erroneous data points (particulates)
Project: Mine Tunnel Monitoring

- High volume, high velocity data
  - On-site sensor data from PLC system
  - Public big data streams

Challenge:
- Centralized database storage
- Real-time data access
- Real-time notifications/alarms
Project: Mine Tunnel Monitoring

Solution:
• Generic database design optimized for high frequency data
• Predictive trend modeling calculations
• Data available via Project Portal
• Email alerts when incoming data parameters out of spec.
Project: O&M Site Monitoring

High velocity, automated SVE and GW treatment systems

Challenge:

• Centralized storage
• Centralized monitoring
• System troubleshooting...
Project: O&M Site Monitoring

System Solution:
- Multivariate analysis to review system variables
- Secondary analysis to identify fluctuations
What’s Next?

• **Use of emerging technologies:**
  – Distributed data sourcing
    • Hadoop HDFS
    • NoSQL
  – Distributed processing
    • Batch processing (MapReduce, Apache Hive)
    • Real-time processing/streaming (Cloudera Impala, Apache 54)
    • PolyBase (cross-querying HDFS and SQL Server)
Summary / Key Takeaways

• Free big data – go out and use it!
• Big data to investigate the unknown
• Greater project intelligence & decision making
Thank you!

Contributors:
Myles Hook, ddms
Jon Turner, ddms
Heidi Gaedy, ddms
Ed Larson, ddms
Angela Remer, ddms
Emily Mulford, Earthsoft Inc.

Brooke Roecker
BRoecker@ddmsinc.com

Mark Packard, PG, CPG
MPackard@ddmsinc.com

ddms
www.ddmsinc.com