Data Science Center of Excellence: What’s the value?

#StrataData

Simon Lidberg
Benjamin Wright-Jones
Microsoft Services CTO Office for Data & AI
Microsoft WW Services CTO Office for Data & AI
Global team of business, solution architects and lead data scientists
Working with organisations around the world

Simon Lidberg
@simonlid

Ben Wright-Jones
@superlatch
Data Is Not the Biggest Obstacle

Despite popular opinion, getting the data right is not a top challenge that organizations face when adopting analytics. Only about one out of five respondents cited concern with data quality or ineffective data governance as a primary obstacle.

The adoption barriers that organizations face most are managerial and cultural rather than related to data and technology. The leading obstacle to widespread analytics adoption is lack of understanding of how to use analytics to improve the business, according to almost four of 10 respondents. More than one in three cite lack of management bandwidth due to competing priorities. (See “The Impediments to Becoming More Data Driven.”)

The BICC is dead...

by Alan D. Duncan | March 11, 2016 | 1 Comment

There has been some chatter on social media following our Sydney and London Summits about Gartner declaring the Business Intelligence Competency Centre, or “BICC”, to be dead. Ahead of our next Business Intelligence and Analytics Summit in Dallas next week, it’s probably a good time to blog about this little nugget...
NATURAL EVOLUTION

Emphasis on historical, descriptive analysis

Shared Service Center (SSC)

Transaction-Oriented SSC (focus on presence)

Traditional SSC for BI (BICC or BI CoE)

Knowledge-Oriented SSC (focus on future)

Analytics COE / Data Science COE

Emphasis on data mining and predictive scenarios

ORGANISATIONAL MATURITY

Foundational
- Questioning what AI is and how to apply it
- Wrong expectations or disappointment
- Low digitization
- Basic analytical capabilities

Approaching
- Hopeful on AI and its promise
- Digitization under way
- Looking to increase or optimize business processes
- Cautious about disruption

Aspirational
- Experimented and applied AI
- High digitization
- Desires new business models
- Achieved an open data culture

Mature
- Data science and operational capability
- Understands model lifecycle and management
- Building a foundational data architecture

Cultural Shift

Ethical Shift
DATA SCIENCE CENTER OF EXCELLENCE

COE model allows for business enabling governance model instead of typical technology centric

Centrally manage technology complexity instead of various teams wasting time & energy separately

Reduce delay in setup so that teams can quickly start to collaborate

Drives end user adoption by making it easy to engage

A Center of Excellence (COE) provides knowledge, assets, and guidance to business units and enables exploration and implementation to accelerate adoption and value realization.

### Capabilities

<table>
<thead>
<tr>
<th>Services</th>
<th>Readiness</th>
<th>Governance</th>
<th>Development</th>
<th>Operations</th>
<th>Management</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Knowledge base of best practices</td>
<td>Program management</td>
<td>Reference solutions</td>
<td>Source data configurations</td>
<td>Supplier and vendor management</td>
</tr>
<tr>
<td></td>
<td>Readiness roadmap and training catalog</td>
<td>Solutions design and architecture</td>
<td>Reusable data publication</td>
<td>Security</td>
<td>Change management</td>
</tr>
<tr>
<td></td>
<td>Solution catalog/portal</td>
<td>Solutions reviews</td>
<td>Common data and asset sharing</td>
<td>External sharing</td>
<td>Client configuration</td>
</tr>
<tr>
<td></td>
<td>Advice, guidance and training to business users</td>
<td>Migration from legacy solutions, cloud adoption etc.</td>
<td>Data integration</td>
<td>CMDB catalog</td>
<td>License management</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Problem resolution</td>
<td>Consumption management</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Technical Quality Assurance</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td><strong>Benefits</strong></td>
</tr>
</tbody>
</table>

- Realize Value Sooner
- Minimized IT involvement
- Optimized process efficiency and economies of scale.
- Increased quality
- Aligned with guidelines and best practices
- Improved business adoption
- Agility enabler “business works faster”
**CENTER OF EXCELLENCE FRAMEWORK EXAMPLES**

**Decentralized**
- **CXO**
- **Bizunit**
- **Function**

**Benefits**
- Subject matter expertise in each business and functions helping them align analytics needs
- Analytics function and team close to business, issues and customers

**Challenges**
- Redundancy in physical resources and people
- Inconsistent SLA, cost and organization
- Inconsistent results and their interpretation
- Focus is on local issues
- No standardization, not leveraging scale

**Semi-centralized**
- **CXO**
- **Bizunit**
- **Function**

**Benefits**
- Shared service, process and methodologies developed at one place. Better coordination among departments.
- Functional expertise in existing shared service
- On demand provisioning of environment, cost control
- Continuous improvement by feedback, best practices and maturity assessment

**Challenge**
- Less transparent allocation of analytics resources among different units. Bias towards parent business unit
- Difficulty in cross functional alignment and consensus

**Centralized**
- **CXO**
- **Bizunit**
- **Function**

**Benefits**
- Shared service organization, standard processes, standard methodologies, continuous improvement by feedback, best practices and maturity assessment
- On demand provisioning of environment, Cost control, SLA commitment
- CXO level commitment, faster and better organizational adoption.
- Self service, user empowerment

**Challenge**
- Perception that the group lacks functional expertise
- Pushback from business units on possible relocation of resources
## ROLES AND RESPONSIBILITIES

<table>
<thead>
<tr>
<th>Role</th>
<th>Key responsibilities</th>
</tr>
</thead>
<tbody>
<tr>
<td>Executive sponsor</td>
<td>Serves as the executive level champion for the solution. The primary responsibility is strategic – position the solution as a critical mechanism for achieving business value and help communicate the value of the solution to the management levels of the organization.</td>
</tr>
<tr>
<td>Governance board/steering committee</td>
<td>Serves as a governance body, where the final responsibility for meeting the goals of the solution rests. It typically comprises representatives of each of the major participants in the solution, including corporate communications, HR, and IT.</td>
</tr>
<tr>
<td>Business owner</td>
<td>Manages the overall design, functionality, and integrity of the solution from a business perspective. The business owner does not have to be an IT expert, although they should know the solutions at a high level. This role is responsible for internal communications.</td>
</tr>
<tr>
<td>Solution administrator (Technology)</td>
<td>Manages the overall design and functional integrity of the solution from a technological perspective. Works in partnership with the business owner.</td>
</tr>
<tr>
<td>Technology (Shared Service) support team</td>
<td>Ensures the technical integrity of the solution. Makes regular backups of the solution and its content. Also, typically sets up and maintains the security model, at least the components in Active Directory. Develops new components and provides support to sponsors/owners seeking enhancements to their solution or new uses of the solution.</td>
</tr>
<tr>
<td>Information/Data Architecture</td>
<td>Individual or team responsible for maintaining the information/data architecture throughout the life of the solution.</td>
</tr>
<tr>
<td>Center of Excellence</td>
<td>Provides coaching and design consulting to new users with full control design privileges to ensure that best practices are followed and appropriate features are applied. In many organizations, a particular feature becomes the de facto solution for any business problem – a hammer in search of a nail. One successful organization implemented drop-in office hours where new solution owners could spend an hour or two with an experienced solution architect to ensure that they got appropriate guidance (in addition to formal training). Several others have established in-house consulting services to help business units get started. In many cases, the first hour or two of consulting is “free.”</td>
</tr>
</tbody>
</table>

A manufacturing company wanted to set-up an Data Science Shared Service Center to enable delivery of new digital services.

- Semi-centralized function was chosen.
- The main drivers:
  - Provide Data Science capabilities across the company.
  - Increase adoption of advanced analytics in LoBs.
  - Being able to accelerate the go to market of new digital services.
- Additional benefits:
  - New reporting capabilities.
  - Better collaboration between Business and IT.
  - New cloud based platform with scalability and cost savings as a result.

- Use 1 business unit (truly involved in data science activities) as a first pilot in this new model.
GETTING STARTED IS EASIER SAID THAN DONE

We hear these common pain points....

Cultural change management
Lack of access to the right talent
Budget constraints for building a data science practice
Breaking down silos between the business and IT
PILLARS TO A SUCCESSFUL DATA SCIENCE COE

Technology

Process

People
PEOPLE

Technology

Process

People
**DATA SCIENCE COE ROLES**

"I have a PhD. And I use it 45 minutes a day"

**Data Sleuthing.** "What data is where in this company, what does it mean, who do I bribe to get an extract, and can I trust it?"

**Data Management.** "I need to develop a data sample I trust, move that to the SAS server, ensure I get updates, and figure out how to integrate it with other data."

**Use of tools for analysis and visualization. Core modeling.**

**Responding to customers.** "I do analysis for a variety of departments and they ping me all the time with simple new requests and changes in scope."

**Production Support.** "One of the experiments I worked on one will bring a ton of value and needs to be moved to production. But IT is very busy and insist on a re-write to their standards. That means I spend my time doing production support instead of innovating."

**Provisioning.** "I need to get a terabyte or two of space to stand this experiment up, run it for a couple of days, then tear it down. I'm tired of waiting until the weekend because that's the only time there is compute headroom. I'm tired of coordinating with others that want to run an experiment at the same time as me."

Let's hire a bunch of PhD's and build a data science COE
# DATA SCIENCE COE ROLES

<table>
<thead>
<tr>
<th>Focus &amp; Measures</th>
<th>Executives (C-level)</th>
<th>Senior Leaders</th>
<th>Business Architects</th>
<th>Solution Architects</th>
<th>Data Scientist</th>
<th>Data Engineer</th>
<th>Developers</th>
</tr>
</thead>
<tbody>
<tr>
<td>TCO</td>
<td><img src="#" alt="High Importance" /></td>
<td><img src="#" alt="High Importance" /></td>
<td><img src="#" alt="High Importance" /></td>
<td><img src="#" alt="Low Importance" /></td>
<td><img src="#" alt="Low Importance" /></td>
<td><img src="#" alt="Low Importance" /></td>
<td><img src="#" alt="Low Importance" /></td>
</tr>
<tr>
<td>Speed to value</td>
<td><img src="#" alt="High Importance" /></td>
<td><img src="#" alt="High Importance" /></td>
<td><img src="#" alt="High Importance" /></td>
<td><img src="#" alt="Low Importance" /></td>
<td><img src="#" alt="Low Importance" /></td>
<td><img src="#" alt="Low Importance" /></td>
<td><img src="#" alt="Low Importance" /></td>
</tr>
<tr>
<td>Organizational Change Mgmt.</td>
<td><img src="#" alt="High Importance" /></td>
<td><img src="#" alt="High Importance" /></td>
<td><img src="#" alt="High Importance" /></td>
<td><img src="#" alt="Low Importance" /></td>
<td><img src="#" alt="Low Importance" /></td>
<td><img src="#" alt="Low Importance" /></td>
<td><img src="#" alt="Low Importance" /></td>
</tr>
<tr>
<td>Barriers to adoption &amp; change</td>
<td><img src="#" alt="High Importance" /></td>
<td><img src="#" alt="High Importance" /></td>
<td><img src="#" alt="High Importance" /></td>
<td><img src="#" alt="Low Importance" /></td>
<td><img src="#" alt="Low Importance" /></td>
<td><img src="#" alt="Low Importance" /></td>
<td><img src="#" alt="Low Importance" /></td>
</tr>
<tr>
<td>Synchronization</td>
<td><img src="#" alt="High Importance" /></td>
<td><img src="#" alt="High Importance" /></td>
<td><img src="#" alt="High Importance" /></td>
<td><img src="#" alt="High Importance" /></td>
<td><img src="#" alt="High Importance" /></td>
<td><img src="#" alt="High Importance" /></td>
<td><img src="#" alt="High Importance" /></td>
</tr>
<tr>
<td>Data Movement</td>
<td><img src="#" alt="Low Importance" /></td>
<td><img src="#" alt="Low Importance" /></td>
<td><img src="#" alt="Low Importance" /></td>
<td><img src="#" alt="Low Importance" /></td>
<td><img src="#" alt="Low Importance" /></td>
<td><img src="#" alt="Low Importance" /></td>
<td><img src="#" alt="Low Importance" /></td>
</tr>
<tr>
<td>Security</td>
<td><img src="#" alt="High Importance" /></td>
<td><img src="#" alt="High Importance" /></td>
<td><img src="#" alt="High Importance" /></td>
<td><img src="#" alt="High Importance" /></td>
<td><img src="#" alt="High Importance" /></td>
<td><img src="#" alt="High Importance" /></td>
<td><img src="#" alt="High Importance" /></td>
</tr>
<tr>
<td>Professional Value</td>
<td><img src="#" alt="Low Importance" /></td>
<td><img src="#" alt="Low Importance" /></td>
<td><img src="#" alt="Low Importance" /></td>
<td><img src="#" alt="Low Importance" /></td>
<td><img src="#" alt="Low Importance" /></td>
<td><img src="#" alt="Low Importance" /></td>
<td><img src="#" alt="Low Importance" /></td>
</tr>
<tr>
<td>Internal Experience</td>
<td><img src="#" alt="Low Importance" /></td>
<td><img src="#" alt="Low Importance" /></td>
<td><img src="#" alt="Low Importance" /></td>
<td><img src="#" alt="Low Importance" /></td>
<td><img src="#" alt="Low Importance" /></td>
<td><img src="#" alt="Low Importance" /></td>
<td><img src="#" alt="Low Importance" /></td>
</tr>
<tr>
<td>WHERE TO FIND TALENT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----------------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Where to find people with skills/interest in Data Science:

<table>
<thead>
<tr>
<th>Meetups</th>
<th>PASS SQL Groups</th>
<th>Universities</th>
<th>Within your Organisation</th>
</tr>
</thead>
<tbody>
<tr>
<td>(DS, SA, DE, BA)</td>
<td>(SA, DE)</td>
<td>(Potential for all roles)</td>
<td>Analysts</td>
</tr>
<tr>
<td>Find your local Big Data, R, Python, Hadoop meetup group, Business Architecture Association sponsored event and you can find top talent presenting</td>
<td>Various chapters are beginning to work on Analytics (especially R and Python)</td>
<td>Math Department Computer Science Engineering Psychology Economics College of Business</td>
<td>Database Administrators Developers Business Architects Solution Architects Other Business Units e.g. Product Research</td>
</tr>
</tbody>
</table>
HOW TO DEVELOP TALENT

Retooling people who are interested in supporting a Data Science function:

<table>
<thead>
<tr>
<th>Training</th>
<th>Community</th>
<th>Meet ups</th>
<th>Mentoring</th>
</tr>
</thead>
<tbody>
<tr>
<td>Start where people are. If you have current SQL skills, learn how to do analytic projects on SQL.</td>
<td>Start reading relevant blogs/papers/books, listen to talks, participate in MOOCs.</td>
<td>Attend local R User/Python/Machine Learning/Data Science groups. Ask questions, don't be afraid to present what you are learning.</td>
<td>If you have senior BA, SA, DS talent, make sure they are helping bring up others.</td>
</tr>
</tbody>
</table>
## HOW TO RETAIN TALENT

Tips for retaining highly sought after Data Science resources:

<table>
<thead>
<tr>
<th>Challenge</th>
<th>Development</th>
<th>Career Path</th>
<th>Recognition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constantly doing the same sort of project will lead to attrition.</td>
<td>One difficulty with analytics consulting is that long term projects often don’t allow for personal technical development. Much like being challenged, a data scientist and architects must constantly learn</td>
<td>Distinct paths for each role: BA, SA, DS</td>
<td>Recognize accomplishments.</td>
</tr>
<tr>
<td>Data science is about exploration, and data scientists need challenging problems to solve.</td>
<td></td>
<td>Understand that some people in these roles don’t want to move to management roles. There is a need for a matrix of skills</td>
<td>It is easy to neglect it takes a village of talent to delivery intelligent business solutions</td>
</tr>
</tbody>
</table>
Process is important
Several exists:
CRISP-DM, KDD, SEMMA, TDSP etc.
Team Data Science Process (TDSP) provides a recommended lifecycle that you can use to structure your data science projects.

https://github.com/Azure/Microsoft-TDSP
CUSTOMER EXAMPLE

- A manufacturing company wanted a process for data science
- An adopted version of TDSP was chosen
- The main drivers:
  - All data science projects follow the same methodology
  - A process that is supported by tooling
  - Shorter time to market with new predictive models

- Learnings:
  - It takes time to adopt new processes
  - Sometimes the data scientists felt that the process was hindering them
  - Process change over time to adopt to the organisation
<table>
<thead>
<tr>
<th></th>
<th>Leadership</th>
<th>Business Architects</th>
<th>Business Analyst</th>
<th>Solution Architects</th>
<th>Data Scientist</th>
<th>Data Engineer</th>
<th>Developers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Capability Modeling</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
</tr>
<tr>
<td>Process Modeling</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
</tr>
<tr>
<td>Machine Learning</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
</tr>
<tr>
<td>Modeling</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
</tr>
<tr>
<td>R &amp; Python</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
</tr>
<tr>
<td>SQL</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
</tr>
<tr>
<td>Hadoop/Spark</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
</tr>
<tr>
<td>GitHub</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
</tr>
<tr>
<td>Cloud Platform</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
</tr>
<tr>
<td>Visualisations</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
</tr>
<tr>
<td>LAMDA (batch &amp;</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
</tr>
<tr>
<td>stream processing)</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
</tr>
</tbody>
</table>
SUCCESS FACTORS

- Executive sponsorship & support – typically a combination of roles
- Collaboration (avoid “ivory tower” syndrome)
- Measurable - should be accountable to business outcomes
- Establish and adopt a data science process
- Expectations – data science is exploratory
Acknowledgements:
Mark Tabladillo, Neera Talbert, Derek Norton, Manoj Kumar, Ken Collins, Eric Charran, Sander Timmer, Benjamin Rest, Nadeem Ishqair, plus un-named organisational contributions

Simon Lidberg
Architect, Microsoft Services CTO Office for Data & AI
E: simonlid@microsoft.com
M: +4687522928

Benjamin Wright-Jones
Architect, Microsoft Services CTO Office for Data & AI
E: benjones@microsoft.com
M: +447764359926