

Original Article

Establishing a Framework for a Successful Center of Excellence in Advanced Analytics

Ankit Bansal

Leander, USA.

Abstract: *The increasing complexity and volume of data in today's business environment highlight the critical need for organizations to adopt advanced analytics to enhance decision-making and drive strategic initiatives. This paper outlines the framework for establishing a Center of Excellence (CoE) for Advanced Analytics, which serves as a centralized hub for promoting best practices, fostering collaboration, and developing analytical expertise. The CoE addresses prevalent challenges, such as skill shortages, data governance issues, and the integration of analytics into business processes. By leveraging cutting-edge analytical techniques—such as machine learning, predictive analytics, and data visualization—the CoE enables organizations to unlock actionable insights from their data. This article discusses the essential components for building a successful CoE, including defining its mission, engaging stakeholders, creating a robust training program, and implementing technological infrastructure. Furthermore, it examines the significant impact of advanced analytics on organizational performance and innovation, providing a roadmap for organizations seeking to cultivate a data-driven culture that sustains competitive advantage.*

Keywords: *Center of Excellence, Advanced Analytics, Data-Driven Decision Making, Organizational Culture, Business Intelligence, Predictive Analytics, Data Governance, Stakeholder Engagement, Skill Development, Cross-Functional Collaboration, Technology Infrastructure, Business Intelligence (BI), Operational Efficiency.*

I. INTRODUCTION

In an era where data is often referred to as the new oil, organizations face immense pressure to harness its power to drive business outcomes effectively. Advanced analytics, which encompasses techniques such as machine learning, predictive modeling, and big data analysis, offers unprecedented opportunities for organizations to derive insights that inform strategic decisions and operational improvements. However, many companies grapple with the challenges of fragmented analytics efforts, a shortage of skilled data professionals, and the absence of a cohesive strategy for data utilization. To navigate these complexities, establishing a Center of Excellence (CoE) for Advanced Analytics becomes crucial. A CoE acts as a strategic framework that not only centralizes analytical capabilities but also cultivates a culture of data-driven decision-making across the organization. By fostering collaboration among cross-functional teams and ensuring alignment with business objectives, a CoE can effectively streamline the adoption of advanced analytics. This initiative not only empowers organizations to leverage their data assets more effectively but also positions them to respond swiftly to market changes, driving innovation and enhancing overall competitiveness.

II. PROBLEM STATEMENT

The establishment of a Center of Excellence (CoE) for Advanced Analytics presents a multifaceted challenge that requires a comprehensive problem statement to guide its development. Organizations often struggle with the integration of advanced analytical capabilities due to fragmented data sources, inadequate skill sets, and a lack of strategic alignment between analytics initiatives and business objectives. The CoE must address these issues by fostering a culture of data-driven decision-making, enhancing collaboration across departments, and standardizing best practices. Key problems include identifying the right talent and training programs to build a proficient analytics team, ensuring data quality and accessibility, and creating a framework for effectively translating analytical insights into actionable business strategies. Additionally, there is a need to define clear governance structures and performance metrics to evaluate the impact of analytics initiatives. By addressing these challenges, the CoE can empower the organization to leverage advanced analytics for improved operational efficiency, innovation, and competitive advantage in an increasingly data-centric landscape. Case study described here talks about improvement in four main areas data, people, process, and technology through CoE , and then understanding current state , assess gaps and design future state of the company



Advanced Analytics CoE Services & Scope

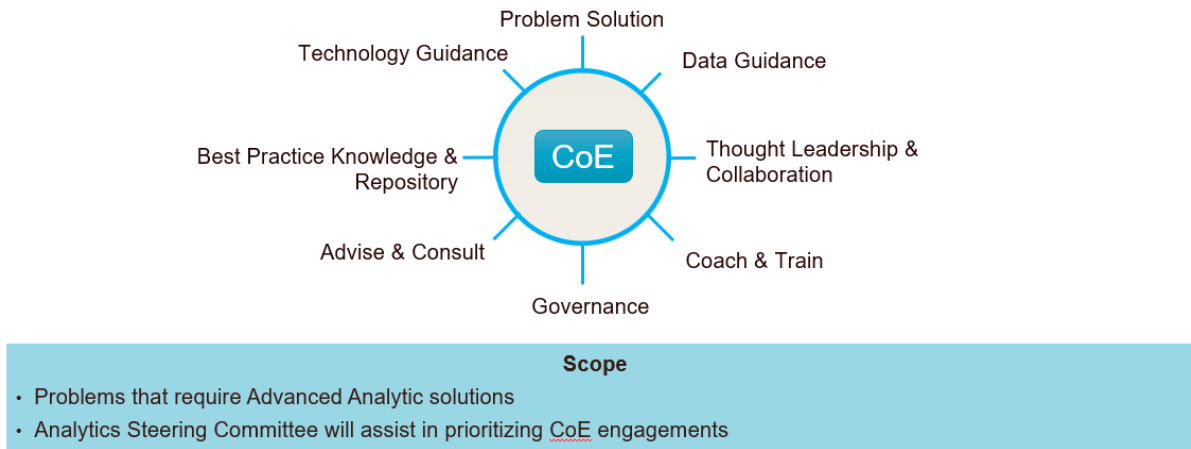


Figure 1: CoE Areas of Impact

III. SOLUTION

Building a CoE for Advanced Analytics involves several key steps:

- **Defining the Vision and Objectives:** Establish a clear vision for the CoE that aligns with organizational goals. This includes identifying key performance indicators (KPIs) to measure success.
- **Stakeholder Engagement:** Engage stakeholders from various departments to ensure buy-in and align the CoE's objectives with their needs.
- **Skill Development:** Invest in training and development programs to build a team with diverse skills in data science, business intelligence, and domain expertise.
- **Technology Infrastructure:** Implement the necessary technology stack, including data storage solutions, analytical tools, and visualization platforms.
- **Data Governance Framework:** Establish a robust data governance framework to ensure data quality, security, and compliance.
- **Pilot Projects:** Launch pilot projects to demonstrate the value of advanced analytics and refine processes based on feedback.
- **Continuous Improvement:** Foster a culture of continuous learning and improvement to adapt to evolving analytical needs and technological advancements.

IV. USES

The CoE for Advanced Analytics can facilitate a wide range of applications across various sectors, including:

- **Healthcare:** Enhancing patient outcomes through predictive analytics, personalized medicine, and operational efficiency.
- **Finance and Banking:** Utilizing advanced analytics for risk management, fraud detection, and improving customer experience.
- **Retail:** Analyzing consumer behavior and inventory management to optimize sales strategies and supply chain operations.
- **Manufacturing:** Implementing predictive maintenance and quality control analytics to minimize downtime and improve production efficiency.
- **Telecommunications:** Analyzing customer data to reduce churn, optimize network performance, and enhance service offerings.
- **Insurance:** Utilizing data analytics for underwriting processes, claims management, and customer segmentation.
- **Energy and Utilities:** Implementing analytics for demand forecasting, resource allocation, and operational optimization.
- **Transportation and Logistics:** Enhancing route optimization, fleet management, and supply chain analytics for improved efficiency.
- **Education:** Leveraging analytics for student performance tracking, curriculum development, and operational efficiency.
- **Real Estate:** Utilizing market analytics for property valuation, investment analysis, and risk assessment.

- Pharmaceuticals: Analyzing clinical trial data, drug efficacy, and market trends to inform research and development.
- Hospitality: Using customer analytics to enhance guest experiences, optimize pricing strategies, and improve service delivery.
- Marketing and Advertising: Implementing analytics to measure campaign effectiveness, customer segmentation, and market trends.
- Government: Utilizing data analytics for policy-making, resource allocation, and public safety improvements.
- Agriculture: Analyzing crop data and market trends to optimize yield, supply chain management, and resource utilization.
- Construction: Using project analytics for risk assessment, cost estimation, and resource management.
- Media and Entertainment: Leveraging audience analytics to inform content creation, distribution strategies, and advertising effectiveness.
- Automotive: Implementing analytics for supply chain management, customer behavior analysis, and product development.
- Non-Profit Organizations: Using data analytics to measure impact, optimize fundraising efforts, and allocate resources effectively.
- E-commerce: Analyzing user behavior and transaction data to optimize website performance, marketing strategies, and customer experience.

V. IMPACT

The establishment of a CoE for Advanced Analytics can significantly impact an organization by:

- Enhanced Decision-Making: A CoE provides data-driven insights that empower leaders to make informed strategic decisions, reducing reliance on intuition and guesswork.
- Increased Operational Efficiency: By standardizing processes and implementing best practices, a CoE streamlines analytics workflows, leading to faster turnaround times and improved productivity.
- Improved Data Quality: The CoE establishes governance frameworks that ensure data accuracy, consistency, and reliability, enhancing the overall quality of analytical outputs.
- Greater Innovation: A dedicated CoE fosters a culture of experimentation and innovation, encouraging teams to explore new analytical techniques and technologies.
- Skill Development and Talent Retention: By providing training and development opportunities, a CoE helps build a skilled analytics workforce, improving employee satisfaction and retention.
- Cross-Functional Collaboration: The CoE promotes collaboration across departments, breaking down silos and ensuring that analytics initiatives align with organizational goals.
- Scalability of Analytics Initiatives: A centralized approach allows for the scalable implementation of analytics solutions across the organization, facilitating broader adoption of data-driven practices.
- Enhanced Customer Insights: Advanced analytics enables organizations to gain deeper insights into customer behaviors and preferences, allowing for more personalized marketing and improved customer experiences.
- Increased Competitive Advantage: Organizations with a CoE for Advanced Analytics can leverage insights to stay ahead of market trends, outperform competitors, and adapt quickly to changing conditions.
- Measurable Business Impact: By developing KPIs and metrics to assess the effectiveness of analytics initiatives, a CoE can demonstrate tangible business outcomes, such as revenue growth, cost savings, and improved customer satisfaction.

VI. SCOPE

Building a Center of Excellence (CoE) for Advanced Analytics provides a comprehensive framework for organizations to elevate their analytical capabilities. One of the primary focuses of the CoE is the standardization of best practices, which helps in developing consistent methodologies that enhance the quality of analytics projects across the board. Additionally, the CoE serves as a hub for skill development and training, offering structured programs that promote data literacy and empower employees to harness analytics effectively. By fostering cross-functional collaboration, the CoE enables experts from different departments to share insights and innovative strategies, leading to more effective problem-solving. Furthermore, the CoE plays a crucial role in evaluating and adopting cutting-edge technologies, ensuring that the organization utilizes the best tools available to maintain a competitive edge. Finally, by aligning analytics initiatives with the broader business strategy, the CoE facilitates data-driven decision-making that directly supports organizational goals, ultimately driving improved performance and growth.

VII. CASE STUDY

A real time case study was performed in building CoE advanced analytics for a top Freight & Logistics company, client was having issues in four areas data , people, technology and process/ also they need to design how their council structure should look like and how should be the full stack of tools and technologies .

High level assessment was done into four major areas along with their challenges as shown below:

Table 1: Current state assessment and Challenges

Area	Current State Assessment	Challenges / Opportunities
Process / Use Cases	<ul style="list-style-type: none"> Existing use cases: Customer churn, customer segmentation & propensity to grow Hadoop environment is used for limited activities Some groups (OR) do not need new system 	<ul style="list-style-type: none"> Access more detailed customer data: behavioral, experience, site visit data. Combine with offline data Real time prediction using streaming data needed Different groups want to leverage predictive fraud management analytics capabilities, pricing transformation, minimizing counter transmission, lane optimization & optimizing package center workflow
People	<ul style="list-style-type: none"> Unavailability of data scientist in some groups Analytics resources gravitated to dev. side Fast attrition rate data scientists 	<ul style="list-style-type: none"> Align career path with skill-sets, interests and education to provide more challenging opportunities in data science Focus on retention, promotion, HR job models Have policies training with R, Python, Spark, Hadoop
Technology	<ul style="list-style-type: none"> Data is housed on Mainframes, SQL Server and Oracle databases currently Data is kept on SQL server and using R for modelling / visualization Most of the Analytics and visualization is done using Datameer Hadoop clusters are not secured Hadoop cluster resource allocation mechanism is not defined 	<ul style="list-style-type: none"> Need new system which is scalable to handle growing amount of data Currently looking for best alternative to use with Spark. Spark has API's with R, Python and its own machine learning library called Spark ML LIB. Need for distributed computing, Cloud based environment Hadoop clusters are not secured, it is advised to Kerborize the cluster and use Apache Sentry for user level security Use Dynamic resources allocation options available with new versions of Cloudera and use appropriate Hadoop schedulers Datameer has many disadvantages, explore better tools for Data Analytics For certain Data processing explore better batch processing tools in Hadoop like Hive, Impala Considering moving to No-sql databases for some of the real-time data analysis
Data	<ul style="list-style-type: none"> Most of data the is currently housed on SQL servers and Oracle Data sources are not integrated Provisioning data for new users a big challenge 	<ul style="list-style-type: none"> Data is needed at granular level (PLD & transaction level) Access web browsing data Data Persistence: Handle different data foormats from various sources Centralized location for real-time data Use better compression techniques in HDFS to improve Query performance and better storage

Based on challenges, below are the recommended solutions for each area.

A. Data Related Recommendations:

Data - Concerns, Gaps, and Recommendations

Concerns/Observations	Gaps	Recommendations
CENTRALIZED DATA		
<ul style="list-style-type: none"> Data is difficult to access and there are multiple copies of some data sources Many constraints in getting data and running on desktops 	<ul style="list-style-type: none"> Data architecture in a disparate state because of <u>siloed</u> data sources No tool to load data directly to Hadoop environment 	<p>DR1. Establish a Data Lake:</p> <ul style="list-style-type: none"> Source data directly from EDW Create centralized data environment Store current and historical system of records data Include raw, transformed and derived data for granular analysis Establish processes for data acquisition, marshalling, analysis, and action
UNSTRUCTURED DATA		
<ul style="list-style-type: none"> Systems need to accommodate new varieties of data (unstructured, multi-structured) 	<ul style="list-style-type: none"> Platform cannot accommodate unstructured and multi-structured data 	
GRANULAR DATA		
<ul style="list-style-type: none"> For a <u>number</u> of high value use cases there is a need for more granular data within the Hadoop system (package level, customer level) Need access to event level data to <u>open up</u> some new cases and deeper understanding of our customer engagement (ex: Customer data, damages, weather data, scan, DIR, DRIVE, liability etc.) Need all <u>infolib</u> data, aggregated data, outside data, large data to measure yearly trends in Hadoop environment 	<ul style="list-style-type: none"> Limited availability of granular/detailed data for analytical activities Need the data availability plan sync with customer demand and needs 	

Figure 2: Recommendations for data

B. People Related Recommendations

People - Concerns, Gaps, and Recommendations

Concerns/Observations	Gaps	Recommendations
STAFFING		
<ul style="list-style-type: none"> Lack of satisfactory job models for analytics team members and no defined job focused on data science Recruitment of new analytics employees tough due to scarce availability. The <u>high level</u> attrition of analytic employees is of great concern at UPS. The collaboration among advanced analytics groups is low 	<ul style="list-style-type: none"> Analytics job models do not exist and need a defined career progression path Advanced Analytics exists as a solitary team model with each team on their own with divergent journeys. Advanced Analytics roles for the <u>CoE</u> are defined and properly staffed 	<p>PL1. Develop AA job models to better define positions and career paths focused on Education, Skills/Technology & Experience</p> <p>PL2. Establish a <u>Advanced Analytics CoE</u> with a federated model, designed to direct and empower each advanced analytics groups with improved collaboration and sharing.</p> <p>PL3. Formalize AA <u>CoE</u> team with appropriate staffing plans.</p>
TRAINING		
<ul style="list-style-type: none"> Training is required for Advanced analytics (AA) resources who are unfamiliar with Spark and other open source/cloud technology 	<ul style="list-style-type: none"> Lack of base training material and collaboration among tools & techniques 	<p>PL4. Establish Self-serve training curriculum</p>
PLATFORM / PRACTITIONER SUPPORT		
<ul style="list-style-type: none"> Area that manages and maintains the Hadoop environment (AA Platform) is understaffed 	<ul style="list-style-type: none"> Existing EBI team is not sufficiently staffed to maintain/support AA platform AA Platform SMEs are needed to support AA practitioners in execution of projected demand of analytics work 	<p>PL5. Staff the existing EBI team sufficiently to support AA Practitioners</p> <p>PL6. Assign a SME (AA platform administrators) for all AA CoE projects</p>

Figure 3: Recommendations for people

C. Technology related recommendations

Technology - Concerns, Gaps, and Recommendations

Concerns/Observations	Gaps	Recommendations
SECURITY		
<ul style="list-style-type: none"> Existing FA security implementation is restrictive and cumbersome 	<ul style="list-style-type: none"> Technologies for implementing data security are not adequate 	<p>TR1. Transition from <u>Datameer</u> to <u>Cloudera EDH</u> for robust Hadoop distribution support</p> <p>TR2. Setup/Enhance Active Directory and Kerberos Authentication</p> <p>TR3. Implement security on Hadoop using built-in <u>Cloudera</u> features and <u>Vormetric</u></p> <p>TR4. Adopt more industry standard role-based access controls using <u>Cloudera Sentry</u></p>
SANDBOX		
<ul style="list-style-type: none"> The majority of advanced analytics work is currently done on desktops and quasi-servers, using a variety of software. Need self service ability to load data to Hadoop environment 	<ul style="list-style-type: none"> Developer-friendly infrastructure for creating models with appropriate toolsets, procedures and data is not available 	<p>TR5. Adopt “Sandbox” environment</p> <ul style="list-style-type: none"> for analytics model development for self-service data ingestion for data exploration
STREAMING DATA		
<ul style="list-style-type: none"> Organization is asking for real-time streaming capabilities 	<ul style="list-style-type: none"> Real-time streaming technologies are not present in current platform 	<p>TR6. Leverage Spark technologies with the <u>Cloudera Enterprise Data Hub</u> for real-time streaming</p>

Figure 4: Recommendations for Technology

D. Process related recommendation

Process - Concerns, Gaps, and Recommendations

Concerns/Observations	Gaps	Recommendations
STANDARD MODEL DEVELOPMENT LIFE CYCLE		
<ul style="list-style-type: none"> Lack of standard process and methodology is a concern for timely model deployment 	<ul style="list-style-type: none"> Standard model development life cycle methodology does not exist 	<p>PR1. Adopt a standard model development life cycle</p> <ul style="list-style-type: none"> Acquire INFORMS methodology to align with O.R. team Evaluate and adopt model management technology to model proposal, design, validation-simulation and approval processes Determine an organizational owner of the model management technology
MODEL DOCUMENTATION		
<ul style="list-style-type: none"> Lack of reusable templates for model documentation 	<ul style="list-style-type: none"> Recommended model documentation requirements do not exist 	<p>PR2. Model documentation standards acquired</p>
MISSION APPROVAL AND DATA ACQUISITION		
<ul style="list-style-type: none"> Long wait times to get missions approved and acquire data. Process can take ~6 months. In many cases they have to start over due to rigorous/rigid process 	<ul style="list-style-type: none"> An SLA driven data request and mission approval process does not exist. 	<p>PR3. Need a streamlined mission approval and data acquisition process with a defined SLA</p>

Figure 5: Recommendations for process

E. Overall Roadmap:

Data, Technology, Process Roadmap (4 phased Maturity Model)

	Initial State	Defined State	Managed State	Optimized State
Data	<ul style="list-style-type: none"> Provide access to Data Warehouses (Edge node) Create a Data Lake plan for migration of Data Warehouses based on current project priorities 	<ul style="list-style-type: none"> Build Data Lake Migrate structured data from Data Warehouses to Data Lake Move other structured data to Data Lake as prioritized by the stakeholders 	<ul style="list-style-type: none"> Provide accessibility to unstructured data Move unstructured data to Data Lake as prioritized by the stakeholders 	<ul style="list-style-type: none"> Continue to add new data sources to the Data Lake
Technology	<ul style="list-style-type: none"> Upgrade from Cloudera Basic to EDH Assess Datameer value proposition Create a Sandbox development environment Complete RFI cloud assessment Begin exploration of Microsoft Advanced Analytics solutions through Pricing Transformation 	<ul style="list-style-type: none"> Utilize Hadoop/Spark ecosystem components for the use cases for maximum impact Explore Cloud capabilities Begin exploration of IBM offerings in knowledge-based AI based on use case 	<ul style="list-style-type: none"> Perform POV for Cloud Utilize Hadoop/Spark real-time/streaming capabilities 	<ul style="list-style-type: none"> Adopt Cloud First PaaS/SaaS Strategy
Process	<ul style="list-style-type: none"> Setup SLAs for mission, data, and user account setup process Initiate Analytics councils Clear contractual and security roadblocks to Cloud 	<ul style="list-style-type: none"> Implement SLAs Establish a "Front Door" Data Intake Process Define and Automate Data Governance 	<ul style="list-style-type: none"> Define self-service framework where appropriate Implement environment with proper controls 	<ul style="list-style-type: none"> Continue to optimize data access interfaces

Figure 6: Roadmap to Achieve Final State

F. Recommended To-be State diagram

"To-be State" Data Platform – Tools & Technologies

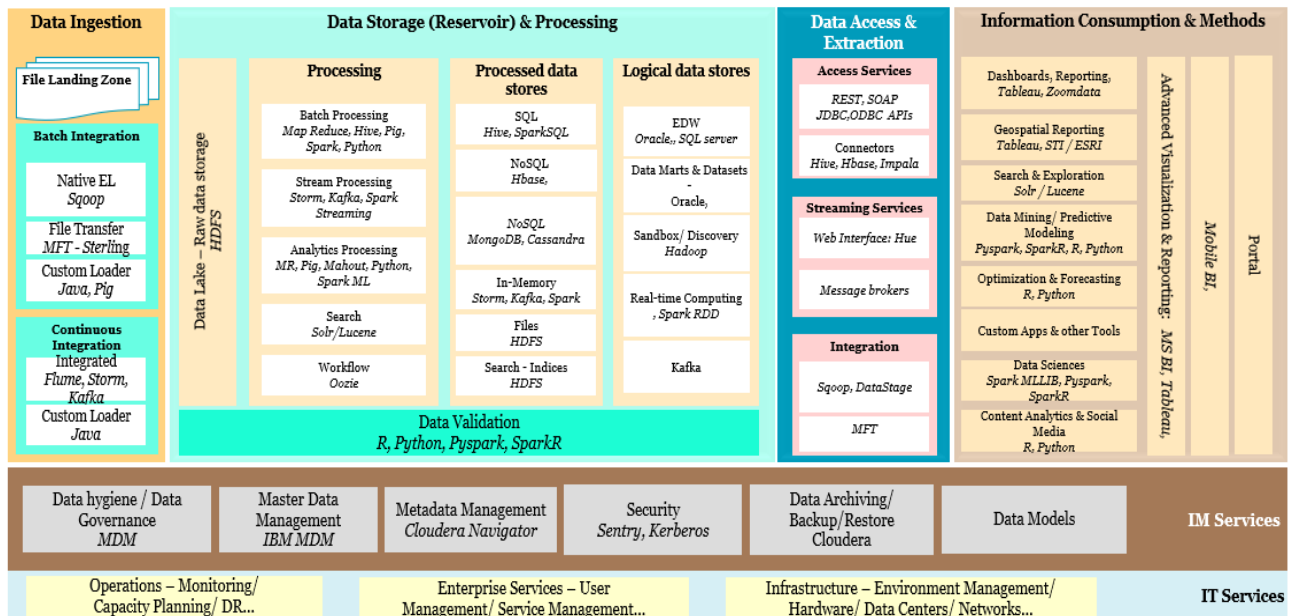


Figure 7: Final state tools stack

G. Recommended Council structure

Advanced Analytics Council (AAC) Structure

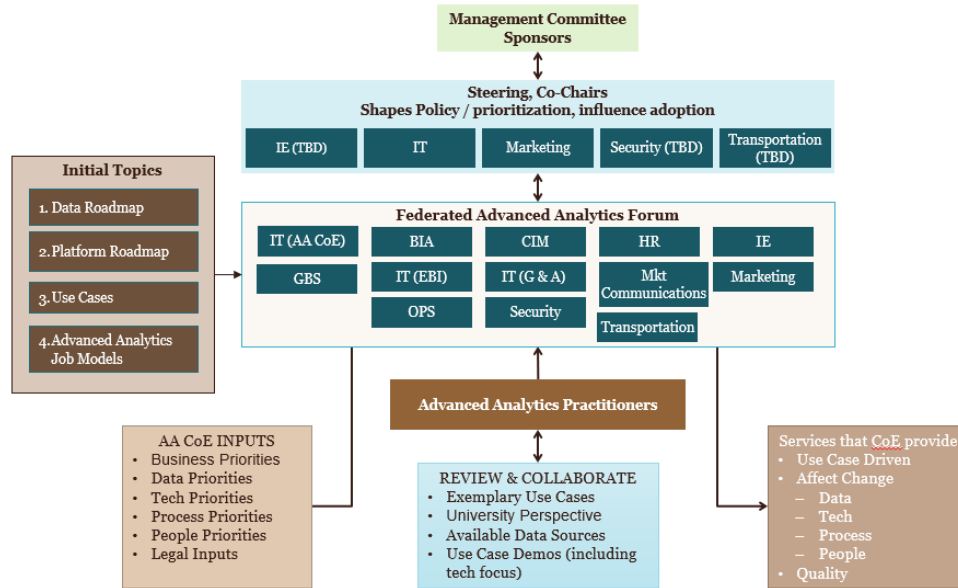


Figure 8: Recommended Council structure

VIII. CONCLUSION

Establishing a Center of Excellence for Advanced Analytics is a strategic initiative that can transform an organization's approach to data. By fostering a culture of analytics, investing in skills development, and ensuring robust data governance, organizations can leverage their data assets to drive innovation and improve decision-making. As the landscape of data continues to evolve, the CoE will play a crucial role in ensuring organizations remain competitive and responsive to change.

IX. REFERENCES

[1] Davenport, T. H., & Harris, J. G. (2007). *Competing on Analytics: The New Science of Winning*. Harvard Business Review Press.

[2] Laursen, G. H. & Thorlund, J. (2010). *Business Analytics for Managers: Take Charge of Your Company's Data*. Wiley.

[3] Waller, M. A., & Fawcett, S. E. (2013). Data Science, Predictive Analytics, and Big Data: A Revolution that Will Transform Supply Chain Design and Management. *Journal of Business Logistics*, 34(2), 77-84.

[4] Kumar, V., & Reinartz, W. (2016). *Creating Enduring Customer Value*. Journal of Marketing, 80(6), 36-68.

[5] Kwon, S., & Shin, B. (2020). The Role of Data-Driven Decision-Making in Enhancing Operational Performance. *Information & Management*, 57(8), 103-113.

[6] Aijaz, A., & Gupta, R. (2019). Framework for Building a Data Analytics Center of Excellence. *International Journal of Information Management*, 45, 164-173.

[7] Hossain, M. U., & Poon, S. K. (2014). Data Analytics in Healthcare: A Review. *International Journal of Health Information Systems and Informatics*, 9(2), 1-15.

[8] Côte-Real, N., Ruivo, P., & Oliveira, T. (2019). The Role of Data Analytics Capabilities in Building Customer Relationships. *Journal of Business Research*, 105, 272-280.

[9] Shanks, G., & Bekmamedova, N. (2018). Business Analytics Research: A Review and Future Directions. *European Journal of Information Systems*, 27(1), 1-22.

[10] Ranjan, J. (2016). Business Intelligence and Analytics: From Big Data to Big Impact. *Journal of Business Research*, 70, 350-356.

[11] Ransbotham, S., & Mitra, S. (2018). Data Science in Business: A Review and Research Agenda. *MIS Quarterly*, 42(3), 911-922.

[12] Ghasemaghaei, M., & Calhoun, A. (2019). The Role of Analytics in Business Transformation. *Journal of Business Analytics*, 2(4), 172-185.

[13] Waller, M. A., & Fawcett, S. E. (2013). Data Science, Predictive Analytics, and Big Data: A Revolution that Will Transform Supply Chain Design and Management. *Journal of Business Logistics*, 34(2), 77-84.

[14] Chae, B. (2019). The Role of Big Data Analytics in Supply Chain Management: A Review and Future Directions. *International Journal of Production Research*, 57(3), 886-910.

[15] Chen, H., Chiang, R. H. L., & Storey, V. C. (2012). Business Intelligence and Analytics: From Big Data to Big Impact. *MIS Quarterly*, 36(4), 1165-1188.

[16] Franks, B. (2012). *Business Analytics: Models, Methods, and Decision-Making*. Wiley.

- [17] Zhai, Y., & Zhang, A. (2018). The Impact of Big Data Analytics on Firm Performance: A Literature Review. *Journal of Business Research*, 94, 130-145.
- [18] Ransbotham, S., & Mitra, S. (2018). Data Science in Business: A Review and Research Agenda. *MIS Quarterly*, 42(3), 911-922.
- [19] McCarthy, I. P., & Anagnostopoulos, D. (2019). The Role of Business Analytics in the Evolution of Business Models: A Case Study. *Journal of Business Research*, 101, 90-101.
- [20] Goh, J. M., & Sweeney, E. (2018). The Role of Data-Driven Decision-Making in Enhancing Operational Performance. *Information & Management*, 55(4), 1-12.
- [21] Hsieh, C.-C. (2020). Building a Data-Driven Organization: The Role of Data Analytics. *Journal of Business Research*, 110, 1-10.