كيف يمكن تجنب أوجه فشل مبادرات ومشرعات

تحليلات البيانات الكبيرة

أ.د. محمد محمد الهادي

مبادرات البيانات الكبيرة والتحليلات احتوت علي أوجه فشل كثيرة في الماضي التي أدت إلي إهدار الجهد والوقت بدون طائل منها، إلي جانب عدم توافر المهنيين المهرة مما أدي لأحباط إدارة المنظمات والمؤسسات التي اضطلعت بها.

فكيف يمكن حاليا تجنب عوامل الفشل هذه في مبادرات ومشروعات تحليلات البيانات الكبيرة؟ ومن خلال بعض الممارسات الأحسن التي اضطلعت بها حديثا كثيرا من منظمات ومنشآت الأعمال المتاحة في هذا الشأن يتضح أننه من خلال المواقف الناجحة لإدارات الأعمال الأساسية التي تتمثل في تأكيد التمويل الملائم لاستثمارات التكنولوجيا المحتاج لها، وتوفيرها للخبراء المحتاج إليهم، إلي جانب التدريب الجيد في تلك المنظمات يمكن تجنب كثيرا من أوجه الفشل المعوقة للنجاح، حيث لوحظ أنه عندما لا يراعي مخاطبة تلك الأساسيات الرئيسة بشكل سليم ومرضي يؤدي ذلك إلي قصور وفشل مبادرات تحليلات البيانات الكبيرة، فهي تفصل بين النجاح والفشل للتعامل مع القضايا والتحديات الفنية في تحليلات البيانات الكبيرة.

والعرض التالي يلخص الأبعاد والعوامل المطلوب مراعاتها لتجنب أوجه فشل مبادرات تحليلات البيانات الكبيرة وخاصة الفنية منها مما يسهم في نجاح المعادلة المستهدفة:

أولا: اختيار أدوات تحليلات البيانات المحتاج لها بعناية فائقة:

تنبع عوامل الفشل التكنولوجية الكثيرة من حقيقة أن المنظمات المختلفة تشتري منتجات تكنولوجيا البيانات الكبيرة وتحليلاتها التي قد تبرهن فيما بعد عدم ملاءمتها لما تحاول تحقيقه واستهدافه، وقد يرجع ذلك إلي أن بائعي وموردي منتجات وأدوات تكنولوجيا االمعلومات يضعون شعارت «البيانات الكبيرة» و «التحليلات» في مواصفات تلك المنتجات والأدوات بعكس الواقع محاولة للاستفادة من ميزتها العالية عند مشاهة تلك الشعارات، علي الرغم من اختلافها الجوهري لا في الجودة والفعالية المتطلبة فحسب بل أيضا في تركيزها علي غايات تلك المنظمات. كما أنه عند اختيار منتج قوي فنبا فقد لا يكون جيدا وملائما لما تحتاج لعمله المنظمة في الواقع.

وفي هذا السيق يمكن تحديد بعض القدرات الأساسية لكل تحليلات البيانات الكبيرة، مثل قدرات تحويل البيانات ومعمارية التخزين، مما يؤدي إلي التفكير في الاستفادة من قدرات كلا من برمجيات Hadoop and Apache Spark. إلي جانب ذلك توجد أيضا بعض الثغرات وأوجه الفشل في تحليلات البيانات الكبيرة مما يستوجب الحصول علي المنتجات والأدوات التي تعالج استراتيجيا تلك الثغرات التكنولوجية التي تتعلق بعملية التنقيب، التحليلات التنبؤية، حلول الوقت الحقيقي، الذكاء الاصطناعي، ولوحات عرض الرسومات التوضيحية المرئية Virtualization بذكاء الأعمال.

وقبل شراء أي منتج أو أداة تحليلات البيانات الكبيرة أو منصة التخزين، سوف يحتاج لمعرفة ما تحتاجه الأعمال حقيقة وما هي تلك الثغرات المثارة، والعمل علي اختيار المنتجات المصممة للتغلب علي قضايا المنظمة بفعالية. علي سبيل المثال، إن منتجات البيانات الكبيرة المعرفية Cognitive كالتحليلات التي تستخدم الذكاء الاصطناعي في تحليل البيانات غير المهيكلة Unstructured بسبب تعقيد مجموعات البيانات المجمعة الهائلة، لا يمكن استخدام منتجات التحليلات المعرفية للبيانات المهيكلة Structured والمقننة معماريا حيث يصعب استعراض أحد منتجات تحليلات البيانات الكثيرة المتاحة في إنشاء وخلق بصائر جودة في الوقت الحقيقي بسعر معقول، وبذلك يصبح من الحكمة والحصافة تشغيل دليل مفاهيم مستخدما منتجين علي الآقل قبل اختيار احداهما لبيئة الإنتاج الحالية. كما يجب أيضا علي المنتج المعين أن يقدر علي التفاعل مع منصات المنشأة المتوافقة معه. وكل أداة تحليلات كبيرة تتطلب تطوير «نموذج بيانات Data Model» في خلفية النظام الذي يمثل جزء المشروع الأكثر أهمية. ويحتاج ذلك إلي تأكيد مكونات عمليات الأعمال وعمل خبراء الموضوع المرتبط بالأعمال معا في هذا الجهد.

وفي هذا الصدد، يجب أخذ الوقت الكافي وأداء الفعل والإجراء الصحيح المتطلب في المقام الأول، كما أنه من المهم

أيضا تذكر أن البيانات الصحيحة يجب أن تكون متوافرة دائما ومترجمة بلغة الأعمال حتى يمكن فهم مخرجاتها من قبل مستخدمي الأعمال مما يمكنهم في استيعابها واكتشاف الفرص أو تحسينات العملية.

ثانيا: التأكد من سهولة استخدام الأدوات:

البيانات الكبيرة والتحليلات المتقدمة تعتبران ظاهرتين حديثيتين معقدتين لحد ما، إلا ان المنتجات والأدوات التي يعتمد عليها مستخدموا الأعمال يجب أن تكون مفيدة ومفهومة ولا تتضمن أي تعقيد. مما يستوجب تقديم أدوات سهلة وفعالة لفرق تحليلات الأعمال فيما يتعلق باستخدام البيانات الكبيرة من أجل الاكتشاف إلي جانب التحليلات ولوحات الرسوم التوضيحية المرئية Virtualization.

وعلي الرغم من أن تجميع الأدوات الصحيحة واستخدامها في مشروعات البيانات الكبيرة والتحليلات كن صعبا في الماضي، إلا ان أدوات ذكاء الأعمال Business Intelligence صارت حاليا متوافرة ومباحة لبيانات المنشأة التي تتضمن لوحات الرسومات التوضيحية المرئية والقدرة بدرجة كافية لكافة التحليلات المتعمقة بما يسهل في ايجاد المنتجات التي تساعد مستخدمي الأعمال في اكتشاف البيانات الملائمة بسهولة مما يمكن إعداد لوحات الرسومات التوضيحية المرئية التي تسهم في تحرير فرق التحليلات من أداء التحليلات المتطورة أكثر مما كان عليه الوضع في الماضي.

وفي هذا السياق، لا ينصح بتقديم أدوات مستوي المبرمج لمستخدمي الأعمال غير الفنيين، حتى لا يحبطوا ويرجعوا لاستخدام أدواتهم التقليدية السابقة التي لا تلبي مهام وظائفهم في الواقع الفعلي.

رابعا: ربط المشروع والبانات مع حاجة الأعممال الفعلية:

من الأسباب الأخري لفشل جهود تحليلات البيانات الكبيرة ما يتمثل في أنها قد ترتبط بمشكلات لا تتواجد بالفعل، بدلا من تحديد إطار التحديات واحتياجات الأعمال الفعلية فيما تواجهه المنظمة للتحليلات الصحيحة. ويحتم ذلك أهمية تضمين خبراء الموضوع وتحديد خلفيات تحليلية قوية منذ البداية في منهجية المشروع للعمل مع علماء البيانات Data Scientists أو مهنيي البيانات لتفسير المشكلات الكامنة المتطلب حلها. ومن امثلة مبادرات تحليلات البيانات الكبيرة، ما يرتبط بالحلول المتعلقة بمشكلات الخداع والغش Fraud، حيث يتمثل التحدي في التغلب علي ذلك ما قد يتم تجميعه من معلومات عن هوية التعريف الشخصية كالأسماء، العناوين، وأرقام تحقيق الشخصية القومية التي تعتبر كلها معلومات منطقية وشرعية، أو قد يكون التحدي في تقدير ما إن كان الشخص يطلب قرضا من أحد البنوك مثلا مستخدما مجموعة مؤشرات تعريفية شرعية، أو كلا التحدين معا.

حيث أن التحدي الأول المرتبط بتحديد الهوية سوف يحتاج لنموذج تحليلات يحدد ابعاد المخاطرة الناجمة من عدم توضيح مستويات تعريف هوية المستخدم وما هو متطلب من مستوي معين؛ أما التحدي الثاني فيرتبط بمشكلة تطبيق الخداع أو الغش وتحديد درجة تقدير المخاطرة المعينة المحتاج تطويرها وتطبيقها عند مستوي التطبيق المطلوب اتخاذه. وفي هذا السياق، فإن الألجوريثمات الصحيحة يجب أن تطبق لاستخلاص الذكاء وعمل التنبؤات الملائمة من حيث جمع وتصميم مجموعات البيانات المتوافقة في عملية النمذجة التي تكون دائما أكثر أهمية من ألجوريثمات تعلم الآلة حيث أن هذا الجهد المرتبط بالبيانات هو الذي يعالج كأولية قصوي.

رابعا: بناء بحيرة بيانات (مستودع بيانات) وعدم التغاضي عن سعة النطاق:

كما يشير له مصطلح «البيانات الكبيرة» من توافر كم ضخم من البيانات المتفرقة من أشكال عديدة لم يكن في مقدرة

كثير من المنظمات جمعها وتخزينها ثم تحليلها في النهاية علي الرغم من أن القليل من المنظمات كانت تقدر علي جمع وتنظيم وتحليل ما يتاح لها من البيانات في الماضي. إلا أنه بفضل توافر وانتشار تكنولوجيات تخزين كميات بيانات الأداء العالي والمعالجة الموازية علي نطاق واسع في إطار تكنولوجيا الحوسبة السحابية Cloud Computing عبر نظم متواجدة في أماكن العمل الفعلية صار في مقدرة معظم منظمات ومنشآت المجتمع التعامل مع تحليلات البيانات الكبيرة في الوقت الحالي.

وقد لا يكون التخزين كافيا لدي المنظمة المعنية ويحتاج لتداول أنوع بيانات كبيرة متفرقة التي تغذي تحليلات البيانات الكبيرة عندئذ توجد حاجة لملحة لبرمجيات Apache Hadoop التي تسمح بتخزين مجموعات بيانات ضخمة وتحديد إطارها وخريطتها. وقد صار يطلق علي مستودعات البيانات تلك «بحيرات البيانات Data Lakes» وبحيرة البيانات العادية تتعدي لحد كبير تدفقات بيانات متعددة كما تشتمل علي علي كثير من المكونات وتتغذي فعليا بواسطة مصادر بيانات معددة تتضمن أنواعا كثيرة من البيانات، إلآ أنها يجب أن تغرق للبيانات، حيث توجد حاجة للتفكير في كيفية جمع البيانات واستخلاص الخواص بطريقة ذات مغذي ومفهوم واضح كما يمكن أن تكون البيانات المتوافرة في البحيرة متفرقة أيضا. وكل ذلك يطرح تساؤل عن كيفية تحويل كل هذه البيانات والمكونات المتوافرهة للتحليلات الذي ينجز حاليا باستخدام أدوات كثيرة منها كلا من أدوات Map Reduce and Apache Spark التي تعمل بمعمارية ثابة في مكانها وتنشئ بحيرة بيانات تتشكل من عمليات الاستيعاب والتكشيف تطبيع مكونات مخططة لاستراتيجية البيانات الكبيرة. وبدون فهم واضح لكل هذه الأبعاد تتعرض كثير من مبادرات ومشروعات تحليلات البيانات الكبيرة الكثيفة للفشل.

إلي جانب ما تقدم، يعتبر توافر سعة نطاق Bandwidth كافية بعدا حيويا لنجاح مبادرة تحليلات البيانات الكبيرة الكثيفة أيضا، حيث لا تتحرك البيانات من مصادرها المختلفة لكي تصب في بحيرة البيانات وتغذيها، لكي تتاح فيما بعد لمستخدمي الأعمال بسرعة كافية وتصبح مفيدة لهم في أفعالهم واتخاذ قراراتهم المبنية عليها. ويتطلب إمداد موارد البيانات الكثيفة في موقع العمل توافر أقراص تخزين قادرة علي تداول ملايين المدخلات والمخرجات (IPOs)، إلي جانب توافر محاور متداخلة ومترابطة معا ومحركات معالجة يمكنها الوصول لمصدادر البيانات عند انتاجها, وفي هذا السياق تكون السرعة مهمة بصفة معينة لتحليلات البيانات الكبيرة في الوقت الحقيقي مما يستدعي بناء بحيرة البيانات علي الترابط الأسرع المتوافر وبدون ذلك يؤدي لفشل الجدهد والوقت المبذول في ذلك.

خامسا: تصميم أبعاد الأمن في كل أوجه البيانات الكبيرة:

درجة عدم التجانس العالية في مكونات البنية الأساسية الكمبيوترية تؤثر بالتبعية علي قدرة المنظمات في سرعة كتشاف بصائر ذات مغذى ومعنى من البيانات المتاحة، إلي جانب ذلك يوجد جانب سلبي يتمثل في تعقيد النظم لإدارتتها وتأمينها، ومع الكميات الضخمة من البيانات المتاحة والتحليلات النقددية الحرجة قد تفشل كثير من المنظمات في أخذ الاحتياطات الملائمة لحماية النظم والبيانات مما يسهم في أخطار عظيمة تعاني منها المنظمة في عملها وقد تؤثر علي تواجدها.

وفي هذا الصدد يلاحظ أن معظم البيانات التي تجمعها المنظمات والمنشآت المختلفة وتخزنها وتحللها ويشارك فيها هي معلومات تتعلق بالمستخدمين أو عملائها والبعض منها يتضمن معلومات شخصية تعريفية عنهم، وعند وصول تلك البيانات إلي أيدي خطأ غير مصرح ومسموح لها بالاطلاع عليها قد تكون نائج ذلك المتنبأ بها كارثية وخطيرة علي أصحابها من المستخدمين والعملاء بالمنظمة. ويدعو ذلك إلي أهمية القيام بقياسات أمن المعلومات التي يجب أن تتضمن نشر الأدوات المتعلقة بذلك في المنظمة المتضمنة تشفير البيانات الذي عندما يكون عمليا ويتعامل مع إدارة الهوية والوصول إلي البيانات وأمن الشبكة النقالة. إلا إن قياسات الأمن يجب أيضا أن تشتمل علي تقوية سياسة الأمن وتأمين المعلومات والتدريب علي الوصول الآمن والمناسب للبيانات واستخدامها.

سادسا: جعل إدارة البيانات والجودة أولوية قصوي:

تأكيد إدارة وجودة بيانات جديدة يجب أن يكون جوهر ولب مشروعات تحليلات البيانات الكبيرة، وفيما عدا ذلك فإن فرص الفشل والإحباط تكون أعظم جدا.ولذلك توجد حاجة ملحة لوضع الأساليب الرقابية التي تؤكد حداثة البيانات ودقتها واتاحتها في نمط فوري. وكجزء أساسي من أي مبادرة بيانات كبيرة يجب تطبيق نظم الإنذار الفوري عند حدوث أي عطل أو انتهاك أو حدوث فشل في تحديث البيانات أو أن ذلك سوف ينجز فيما بعد.

ولذلك يجب تطبيق فحوصات البيانات لكي تتوافق مع أبعاد الجودة المستهدفة في إطار مقاييس رئيسية ترسل الإنذارات عن عدم توافق هذه القياسات مع التوقعات. ويرتبط تأكيد وجودة حوكمة البيانات الكبيرة بأهمية تعيين كوادر مهنية كفء في إدارة البيانات تشتمل علي مدير بيانات وأغراد مهنيين مهرة في تكنولوجيا المعلومات لإدارة مشروعات تحليلات البيانات الكبيرة، مع إعطاء الأهمية الاستراتيجية لتلك المبادرات والمشروعات التي ترتبط بالبيانات الكبيرة وتملكها والإشراف عليها وإدارتها وحوكمتها وسياسة أمنها.

تلك هي القضايا الرئيسية التي تواجه منظمات اليوم في مبادرات ومشروعات تحليلات البيانات الكبيرة التي يجب التعامل معها بحرفية لتجنب فشلها وإهدار الجهد والوقت والمال