البيانات الضخمة وأثرها في عملية اتخاذ القرار

خالد قاشي

أستاذ محاضر أ، كلية العلوم الاقتصادية والعلوم التجارية وعلوم التسيير، جامعة البليدة2

ساعد العوادي

أستاذ مؤقت، كلية العلوم الاقتصادية والعلوم التجارية وعلوم التسيير، جامعة البليدة2.

البريد الإلكتروني: khaledgachi2000@yahoo.fr

الملخص:

حوالي 90 % من البيانات العالمية ظهرت فقط خلال السنتين الأخيرتين، ومصدر هذه البيانات هي المؤسسات العالمية الرائدة مثل facebook, Twiter, Google، وهذه المؤسسات تولي أهمية كبيرة بالمعلومات التي تتضمنها هذه البيانات من أجل اتخاذ القرار المناسب، وعلى هذا الأساس، [1]فإن استخراج المعلومات اللازمة لا تتوفر لأي شخص، بل لا يمكن فهم هذه البيانات إلا صنف معين من المتخصصين في هذا الميدان الذي أصبح يسمى علم تحليل المعطيات، لكن لا يمكن فقط الاعتماد على مهارات التحليل بل يتعدى الأمر مهارات التحكم في وسائل التحليل، ووفقا لما سبق يبرز مفهوم مقرب للشخص المتخصص في تحليل المعطيات بشكل عام والبيانات الضخمة بشكل خاص، فهو يملك القدرة على استعمال الوسائل الإحصائية من جهة وأدوات الإعلام الآلي من جهة ثانية، حيث أن البيانات الضخمة بشكل خاص تتطلب مهارات عالية خاصة البرمجية منها، فنظرا لحجم هذه أصبح من الضروري الاعتماد على وسائل تحليل حديثة لعدم فاعلية الوسائل الكلاسيكية، ومن أهم هذه الوسائل هو برنامج Hadoop. التحليل وحساسية استخراج المعلومات رغم صعوبتيه يتطلب القدرة على التكيف بسرعة مع المعطيات أولا والأوضاع السائدة ثانيا، لأن هناك قرارات مهمة ستبنى حسب النتائج التي يفضي إليها التحليل وأي خطأ في ذلك يؤدي إلى نتائج سلبية قد تصعد من حدة الموقف وتدارك الموقف غير مضمون أو يكون مكلفا جدا. نظرا لحساسية الوضع ازدادت الحاجة بشكل كبير في السنوات الأخيرة إلى تحليل البيانات الضخمة التي تتخذ مسارا متعارفا ضمن المتخصصين، فالعملية تمر بعدة مراحل بدءا من جمع البيانات، تنظيمها وتبويبها، تحليلها ثم استخلاص النتائج التي تتطلب

الدقة وتعتمد على السرعة لأن هناك قرار سيبنى عليها، وبما أن القرار في حد ذاته يمس مختلف أنواع المؤسسات سواء الحكومية أو الخاصة، التجارية أو الخدماتية فإن تحليل المعطيات سيكون لها أهمية على أي قرار متخذ في أي نوع من هذه المؤسسات.

الكلمات المفتاحية: البيانات الضخمة، وسائل الإعلام الحديثة، Hadoop، القرار المبني على المعطيات

مقدمة:

يتعامل محللي الإحصاء مع أنواع عديدة من العينات الإحصائية مهما كان حجمها سواء صغيرة أو كبيرة، وبغض النظر عن الميدان أو المجال الذين ينشطون فيه، لكن العالم يعيش تطورا سريعا ما أدى إلى تنامي حجم المعطيات، فأصبحت العينات الإحصائية ضخمة جدا والتعامل مع هذا النوع أدى إلى تغير طريقة التفكير لدى الإحصائيين أو حتى في الطرق المستعملة في التحليل. سنركز في هذا البحث على دراسة هذا الموضوع بشكل خاص حيث أن كل شخص أو مؤسسة فهي مصدر لهذه البيانات بشكل مباشر أغير مباشر، فأصبح من الأحرى بالمؤسسات مهما كان طابعها القانوني أو الاقتصادي أن تولي اهتماما بهذا المجال. ومن أجل معالجة هذا الموضوع تم طرح التساؤل الآتي: **ما هو دور البيانات الضخمة في عملية اتخاذ القرار؟**

أهمية البحث:

تتجلى أهمية البحث في إبراز مكانة تحليل البيانات ضمن عملية اتخاذ القرار داخل المؤسسات التي تعتمد على هذه البيانات من أجل اتخاذ القرار، مثلما يهدف البحث إلى إظهار قيمة البيانات الضخمة باعتبارها موضوع العصر سواء من ناحية كيفية التعامل معها، أو من ناحية وسائل التحليل أو حتى من ناحية مكانتها في بناء النماذج التنبؤية ومنه بناء عملية بناء القرار.

دواعي البحث:

ربما يمكن من الأسباب الرئيسة في معالجة الموضوع هو مرونة الموضوع وأهميته، كما أنه يجمع بين العديد من الفروع أي inter-disciplinary ، فهو يجمع بين العلوم التكنولوجية وما تحتويه من التعامل مع البرامج ولغات البرمجة، كما يتناول الأدوات الإحصائية، الإحصاء التطبيقي، الاقتصاد التطبيقي وإدارة الأعمال، ولكن قد تكون أهمية هذا الموضوع وما توليه المؤسسات على مستوى العالم المتقدم من أهمية لهذا الموضوع، غير أن الاعتماد على البيانات في البلدان العربية والجزائر بصفة خاصة

لا يزال ضعيفا، فكان الهدف تسليط الضوء على الموضوع ووضعه على الواجهة ليحظى بالاهتمام الذي يجب أن يحظى به.

هيكل البحث:

من أجل عرض محتوى هذا البحث تم تقسيمه إلى محاور خمسة أساسية، حيث تم تعريف البيانات الضخمة وإبراز أهم خصائصها وطبيعتها، إضافة مدخل مختصر حول علم تحليل البيانات في المحور الأول، أما المحور الثاني فقد خصص للمشاكل التي تواجه البيانات الضخمة، ليأتي دور منهجية التحليل التي لها ارتباط وثيق بأدوات التحليل، حيث تستحيل عملية التحليل اليدوي، فلقد كرس البحث المحور الأخير لأدوات التحليل الأكثر استعمالا في هذا الميدان وهو برنامج Hadoop الذي خضع لعملية التشريح كتلميح صريح لأهميته ضمن تحليل البيانات الضخمة، وفي الأخير خصص المحور الخامس كمثال أو كتطبيق لما سبق على مستوى قطاع الأعمال.

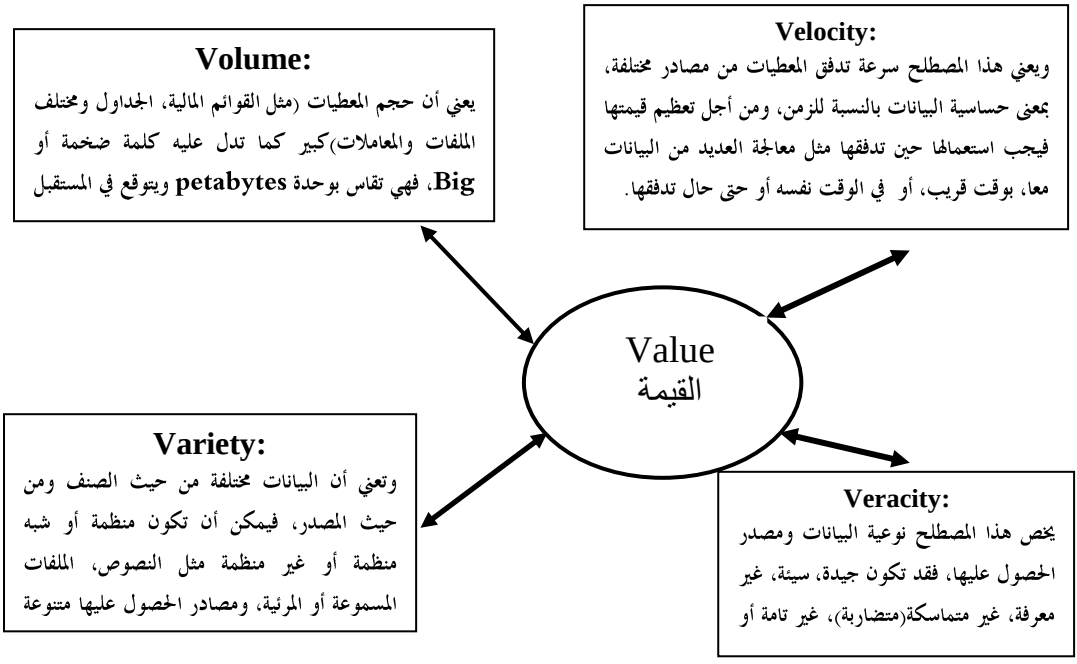
1- تعريف البيانات الضخمة: البيانات الضخمة هي بيانات تنمو بسرعة كبيرة، لذلك تجد أن تعريفها يتغير أيضا فلا يمكن أن تجد تعريفا واحدا لها. حسب خصائص هذه البيانات الثلاث الأكثر شيوعا والتي تعرف ب3V وهي Volume، Variety ، Velocity (سنحافظ على المصطلحات باللغة الأصلية مع إعطاء تعريف لكل مصطلح)، فقد عرفها المحلل Edd Dumbell الذي يعمل في شركة O'Reilly كما يلي:

"البيانات الضخمة هي تلك البيانات التي تفوق قدرة قواعد البيانات العادية على معالجتها. البيانات هي كبيرة جدا، تتحرك بسرعة عالية، ولا تتلاءم مع متطلبات بنية قاعدة البيانات الخاصة بك، وحتى يمكنك الاستفادة من هذه البيانات فإنه من الأجدر اختيار البديل الأنسب لمعالجتها".[1]

التعريف السابق يوافق التعريف الأول الذي أتى به Meta Group (حاليا هو جزء من Gartner)، وحسب نوعية البيانات فقد أضافت شركة IBM الV الرابع وهو الدال على Veracity، بينما شركة Oracle فقد أضافت أيضا الV الرابع وهو الدال على Value، وبالاعتماد على التعريفات السابقة يمكن إعطاء تعريف شامل للبيانات الضخمة، وأساس هذا التعريف هو خصائص هذه البيانات كما هو موضح في الشكل البياني رقم 01[2]، حيث تشكل **قيمة البيانات أو Value** المحور الأساسي ولب هذه البيانات وهي تظهر أهمية القيمة "Value" من أهمية النتائج التي يتوصل إليها الباحث من خلال عملية تحليل البيانات المخزنة، على أي يقدم تقارير تساهم وتساعد

متخذي القرار باختيار التوجه الأنسب للمؤسسة أخذا بعين الاعتبار الاستراتيجيات المسطرة. فهدف رجال الأعمال تعظيم

الفائدة أما قطاع الإعلام والتكنولوجيا IT فهدفه الأداء بأنجع الطرق، فالتكامل بين القطاعين يبين قيمة البيانات. [3]



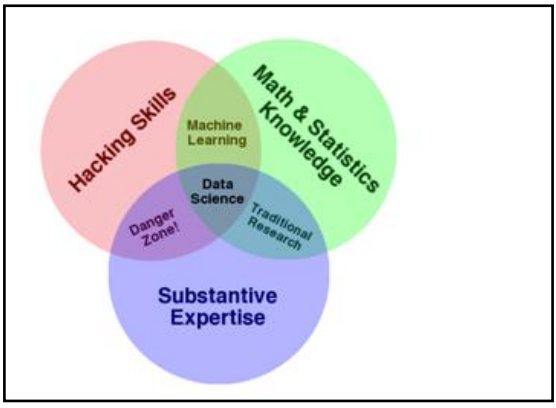
شكل(01): ال5V للبيانات الضخمة

يبين الشكل 01 خصائص البيانات الضخمة، حيث يمكن للشركات والمؤسسات أن تستفيد من فرصة تجميعها من مصادر متنوعة مع تقييمها ومن ثم استعمالها في التحليل، لكن سوء التعامل أو سوء تقييم هذه البيانات يبعث العديد من المشاكل، من بين هذه المشاكل هو تخزين هذا الكم الكبير من البيانات من جهة، أم من جهة ثانية فهو استعمال أحسن الطرق في عملية التحليل، ومن أجل توضيح ذلك يمكن أخذ المثال التالي: إذا سألك صاحب شركة من أنت؟ يمكن أن تجيب أنا محلل المعطيات (Data analyist or data scientist) فإني اقتصادي من جهة كما أني إحصائي متقن

لأدوات الإعلام الآلي من ناحية أخرى. إذا كان بهذا المفهوم فإنه يهدف إلى الاستفادة من المعلومات التي تتضمنها البيانات بفعالية من جهة أنه اقتصادي، كما يهدف إلى استعمال أحسن طرق التحليل من جهة أنه مهندس إعلام آلي وإحصائي، وهذا ما يجعله مدرك للوضع ومرن للأوضاع، وهذا ما جعل هذا النوع من المحللين الأكثر طلبا في قطاع الأعمال، وهو التخصص الذي أصبح يطلق عليه "**العمل الأكثر جاذبية في القرن 21" (The sexiest job of 21st century)** [4] . على هذا فإنه يجب تعريف هذا التخصص، حيث يتنبأ بأنه في العشر سنوات القادمة سيحتاج العالم إلى عديد كبير من المتخصصين في هذا الميدان مما هو عليه في الوقت الحالي.

1-1-مفهوم علم تحليل المعطيات Data science : حسب Venn-Diagram المبين في الشكل رقم 02 فإن هذا العلم يقع في منطقة التقاطع بين ثلاث مناطق وهي:

شكل 02: مخطط Venn



*** Hacking Skills**: وتعني إمكانية التعامل مع الحواسيب وشبكات الانترنيت بمهارة عالية، خاصة جانب أنظمة التشغيل

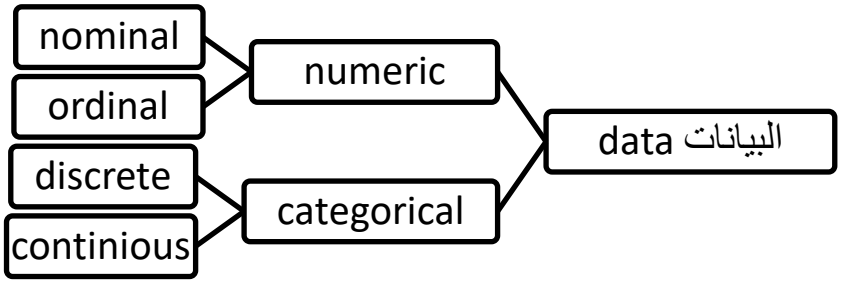
والبرامج ولغات البرمجة (جانب ال software).

***Math & statistics knowledge** : وتعني اكتساب خلفية قوية تمكن التحكم في الجانب الرياضي والإحصائي.

***Substantive expertise** : وتعني الاعتماد على الخبرة الميدانية، أي التدريب على ما هو في الواقع فعلا وهذا ما اعتمد عليه العديد من المؤلفين في هذا المجال، مثلهم مثل القائمين على الدورات التدريبية حيث أغلبهم يعتمدون على معطيات حقيقية.

على ضوء التعريف السابق فإنها تتضح العديد من الأمور، وهي متطلبات الخوض في عملية التحليل لكن الأمر لا يخص تحليل المعطيات بشكل عام بل يختص بالمطيات الضخمة، وعلى الرغم من أن محللي المعطيات يتميزون بمهارات خاصة إلا أن هذا لا يمنع من أنهم يجدون أنفسهم أحيانا أمام مواقف أقل ما يقال عنها أنه صعب التعامل معها، وفيما يلي الخطوات المتبعة في عملية التحليل.

1-2-طبيعة البيانات: يبين الشكل التالي نوعية البيانات التي يمكن تقسيمها إلى قسمين رئيسين وهما: معطيات رقمية (عددية Numeric **)** ومعطيات على شكل فئات Categorical data



شكل -03- طبيعة البيانات

2- التخزين Storage: عملية تخزين حجم كبير من المعطيات هي ليس بالشيء الجديد، فقد اعتادت المؤسسات على تخزين معلومات عن المتعاملين معها مثل الموردون والزبائن، كما قواعد المعلومات قديما تختلف عن ما هي عليه حاليا، وهي ما تدعى **بمخازن البيانات data warehouse(DW)**، وهدفها هو تخزين العديد من البيانات من أنماط مختلفة، وهذا يعني أنه يمكن تخزين البيانات بأشكال مختلفة، أساسا أشكال متعددة الأبعاد التي اعتبرت أنجع الطرق من أجل

استعادة البيانات بأسرع طريقة. إن عملية إنشاء واسترجاع هذه البيانات بواسطة لغة (Structured Query Language)

SQL أو كما تدعى لغة البرمجة لقواعد البيانات ذات العلاقة، أما بالنسبة للبيانات التي لا تربطها علاقة وأخذا بعين الاعتبار الزمن الحقيقي الذي يعتبر أساسي في حلقة تحليل البيانات فإنه يتم استعمال NoSQL، وتتميز بالقدرة على التعامل مع البيانات غير المنظمة والتي ليس علاقة فيما بينها، ومن بين أهم استعمالاتها هي صفحات الواب.[5]

اختلاف وسائل التعامل مع البيانات وتنوع **DW** سهل كثيرا من حياة محللي البيانات، خاصة وأن العالم يتوجه إلى الرقمنة أو ما تسمى ب cloud computing والتي تسمح بالقيام بجميع العمليات اللازمة عبر شبكة الأنترنيت دون تحمل عناء التخزين محليا، وهذا ما دفع بالعديد من قطاع الأعمال بالتوجه إلى هذا الاقتصاد الرقمي.[6]

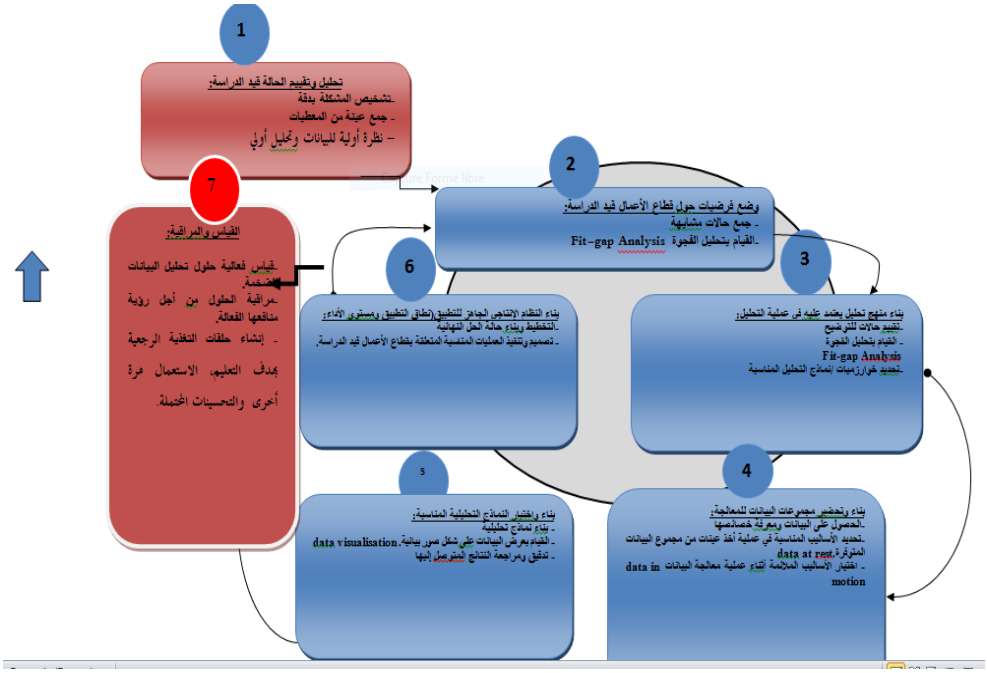
3- منهجية تحليل البيانات الضخمة:

يتبع محللي المعطيات منهجية معينة من أجل القيام ببناء نموذج تنبئي، الشكل (04) يبين الخطوات المتبعة وأول مرحلة تعتبر من مراحل التحليل وهي تحديد طبيعة المشكلة بدقة، وعدم العناية بالبداية سيؤدي إلى عدة مغالطات من جهة وإلى تضييع وقت معتبر لأنه في الغالب يؤدي إلى إعادة تشخيص المشكلة من جديد ثم المرور على المراحل المبينة في الشكل (04) في كل مرة بهدف الوصول إلى أحسن نموذج.

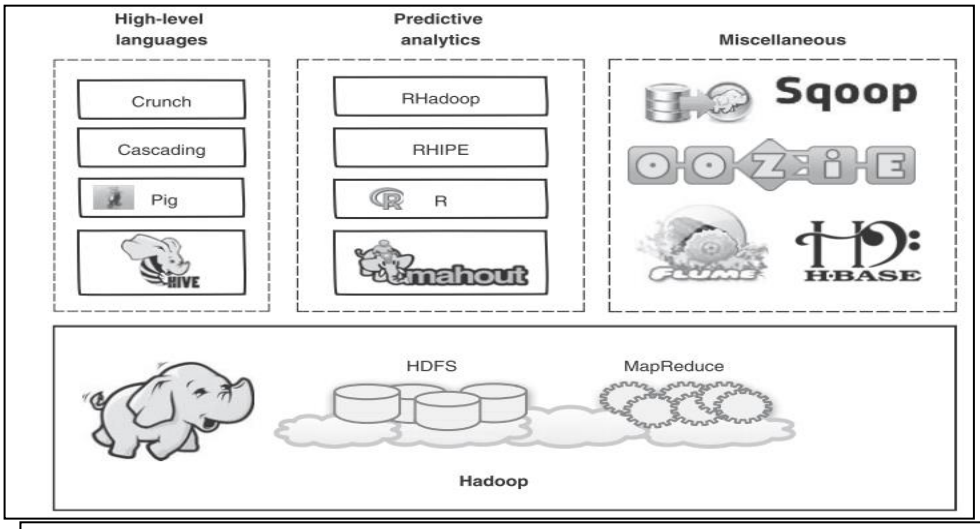
شكل-04- خطوات تحليل البيانات

الاعتماد على النماذج التنبؤية في العصر الحالي أصبح واسع الاستعمال على مستوى العديد من الميادين، حتى على مستوى العلوم الطبية، الزراعية والكيميائية، لكن تطور هذه النماذج بشكل كبير خاصة على مستوى المؤسسات المالية أين هناك كم هائل من المعطيات، فكان من الأحرى أن يركز قطاع الأعمال على هذه النماذج بشكل مكثف. منهجية بناء هذه النماذج لا تختلف لكن وجود اختلاف بين بعض المراحل فقط، وعلى هذا الأساس فقط تم بناء عدة برامج قادرة على معالجة النماذج الترددية مثل برنامج SAP الذي يتميز بخاصية **الذكاء الاصطناعي في قطاع الأعمال Business Intelligence (BI)**، وهذا في حال الاعتماد على طرق التحليل التنبؤية.[7] يمثل الشكل 04 وجود ثمان مستويات من أجل بناء نموذج تنبؤي، لكن أشرنا إمكانية إقصاء بعض المستويات أو اعتبار مستويين كمستوى واحد، فيمكن خلال مرحلة جمع البيانات عملية التبويب والتي يمكن حتى تحديد نوعية المتغيرات أيضا ربحا لعامل الزمن، ومن أجل تفسير منهجية التحليل بشكل أكثر نعتمد على التمثيل البياني التالي[8]:

شكل-05- منهجية تحليل البيانات الضخمة



4-أدوات تحليل البيانات الضخمة: تعتمد عملية تحليل البيانات الضخمة على تكنولوجيا حديثة متطورة، ومن البرامج الأكثر استعمالا هي البرنامج مفتوح المصدر القائم على Appache Hadoop، يسمح هذا البرنامج بدرجة كبيرة بمعالجة مجموعة كبيرة من البيانات من خلال تقنية الجمع بين العديد من الحواسيب المتصلة فيما بينها باستعمال نماذج برمجية بسيطة. صممت هذه التقنية لتعدي خدمة الحواسيب القاعدية أو servers إلى آلاف الحواسيب الأخرى التي تضمن عملية حساب وتخزين محلي، ومن أجل تقديم هذه التقنية سنقوم أولا بإعطاء مخطط أولي وبعدها سنقوم بشرح ميكانيزم عملها، الشكل التالي يبين هيكل Hadoop ولواحقها.[9]



الشكل-06- ***Hadoop and related technologies***

يوضح الشكل -06- أداة التحليل الأكثر استعمالا في ميدان تحليل البيانات الضخمة وهي ما تدعى Hadoop، وفيما يلي توضيح ميكانيزم عملها:

4-1- Hadoop: هي عبارة عن برنامج قابل للتشغيل فقط عبر أنظمة التشغيل القائمة على linux (أي linux based-system) بما في ذلك Mac os، والميزة التي يقدمها هذا البرنامج هي إمكانية توزيع التخزين والعمليات الحسابية عبر أكثر من جهاز، لأن المشكل الذي الأساسي هو مشكل تخزين الحجم الكبير، لكن من المعلوم أن وسائل التخزين تتطور بسرعة، لكن قراءة وتخزين البيانات على القرص نفسه مثلا بقية بطيئة جدا، غير أن قراءة البيانات نفسها من أقراص عدة سيؤدي إلى تقليص ذلك الزمن لأن الأمر يتعلق بكم كبير من البيانات ممكن بوحدة Peta-

Bytes وفي هذه الحالة يصبح عامل الزمن لا نقاش فيه، وهذا هو معنى إمكانية التوزيع التي يقدمها البرنامج ولا تخص التخزين فحسب بل عملية المعالجة أيضا. إذا تم اعتبار جانب البنية الهيكلية فإنه أساسا يتكون برنامج Hadoop من جانبين وهما:

الجانب الأول: جانب النظام Hadoop file System

الجانب الأول: جانب برمجي Map reduce

فيما خلا هذين الجانبين فإن جميع الأعمال الأخرى هي عبارة فروع أو مكملات للنظام الأساسي كل فرع تم بناؤه من أجل حل مشكل أو التعامل مع أوضاع معقدة، وكل ذلك ما يدعم فعالية هذا البرنامج.

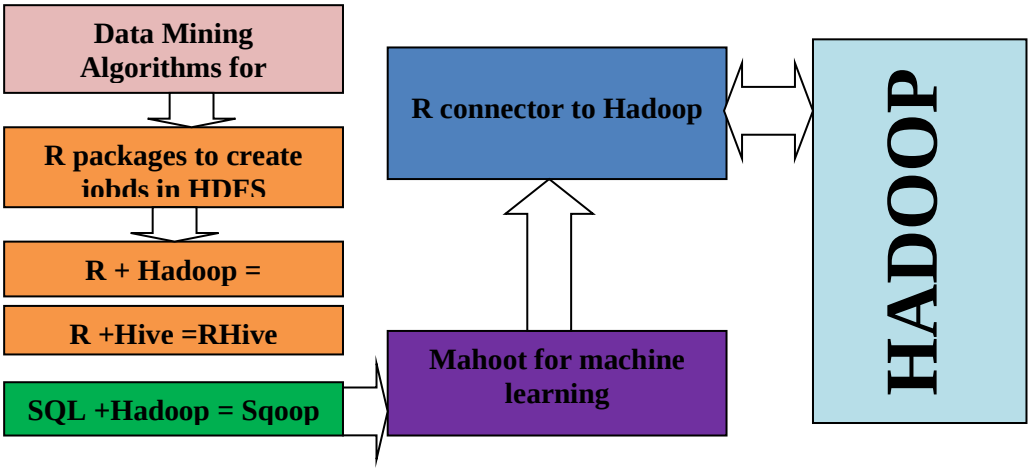
4-2-البنية الهيكلية لبرنامج Hadoop: هدف البرنامج هو التعامل مع التخزين والتحليل أو المعالجة، فالجانب الأول يتعامل مع التخزين والجانب الثاني يتعامل مع التحليل:

***Hadoop Distributed File System(HDFS)** : ويعني أن هذا النظام قد صمم للتعامل مع أحجام كبيرة جدا من البيانات مهما كان حجمها(mega, giga, peta) ضمن أنماط معينة للوصول إلى هذه البيانات، وعادة مت تعتم النمط التالي: **التخزين مرة- القراءة مرات عديدة** (write-once, read-many-times) ، وضمن نظام متصل فيما بينه من أجهزة التخزين (ضمن حواسيب)، فهي لا تحتاج إلى أجهزة مكلفة جدا بل تعتمد على نظام clustering.

***MapReduce**: والجانب الذي يتعلق بلغات البرمجة في عملية تحليل البيانات، حيث يمكن القيام بتحليل البيانات بالعديد من لغات البرمجة مثل Ruby, Python , Java, C++. كما أن هذا النموذج MapReduceيقوم بتقسيم العملية إلى مرحلتين: Reduce-task، Map-task وهذا المبدأ يرجع الفضل فيه لغة البرمجة Lisp ، حيث أن كل مرحلة تحتوي على العديد من الثنائيات Key-value كمدخلات ومخرجات، حيث يقوم المبرمج بالقيام بالعديد من العمليات وهو المسئول عن تحديد كل مرحلة، أي مرحلة map أين تتوافق العديد من الثنائيات كمدخلات لتأتي مرحلة التقليص Reduce من أجل الحصول على مخرجات نهائية، كما أن المرحلة الأولى map التي تتعامل مع HDFS أي مصدر الثنائيات الداخلة، أما المرحلة الثانية Reduce فهي تعتمد أساسا البرنامج أو الأوامر المكتوبة من طرف المبرمج في حد ذاته، كمثال للانتقال من المرحلة الأولى

إلى الثانية فيمكن اعتبار أنك تتعامل بلغة البرمجة python مع نوعية المتغيرات التالية: tuples, lists and dictionaries، فانطلاقا من هذه المتغيرات يمكن القيام بالعديد من الثنائيات عن طريق التفكيك والتركيب. هذه المرحلة تعتمد على البرمجة وكيفية تعامل المبرمج مع الوضع ونوعية البيانات، وفعالية العملية أيضا تعتمد أيضا على مرونة هذه المرحلة.[10]

***لواحق Hadoop**: بالرجوع إلى الشكل -06- نلاحظ أن القاعدة هي عبارة برنامج Hadoop بينما القسم العلوي يتمثل في عملية التحليل، ومن المعلوم أن الشركات اتجهت إلى استعمال لغات البرمجة في عملية تحليل بياناتها لعدة أسباب سواء المالية أول للكفاءة وهذه الميزة يوفرها برنامج Hadoop الذي أقصى مشكل التخزين، فمرحلة تبويب وتنظيم البيانات تتم باستعمال العديد من لغات البرمجة التي تتبع خوارزميات عديدة، ومن بين اللغات المحبوبة لدى الإحصائيين هي لغة البرمجة SQL , R وانضمامها إلى Hadoop أدى إلى تسهيل عملية التحليل أكبر، والمخطط التالي يبين ذلك:

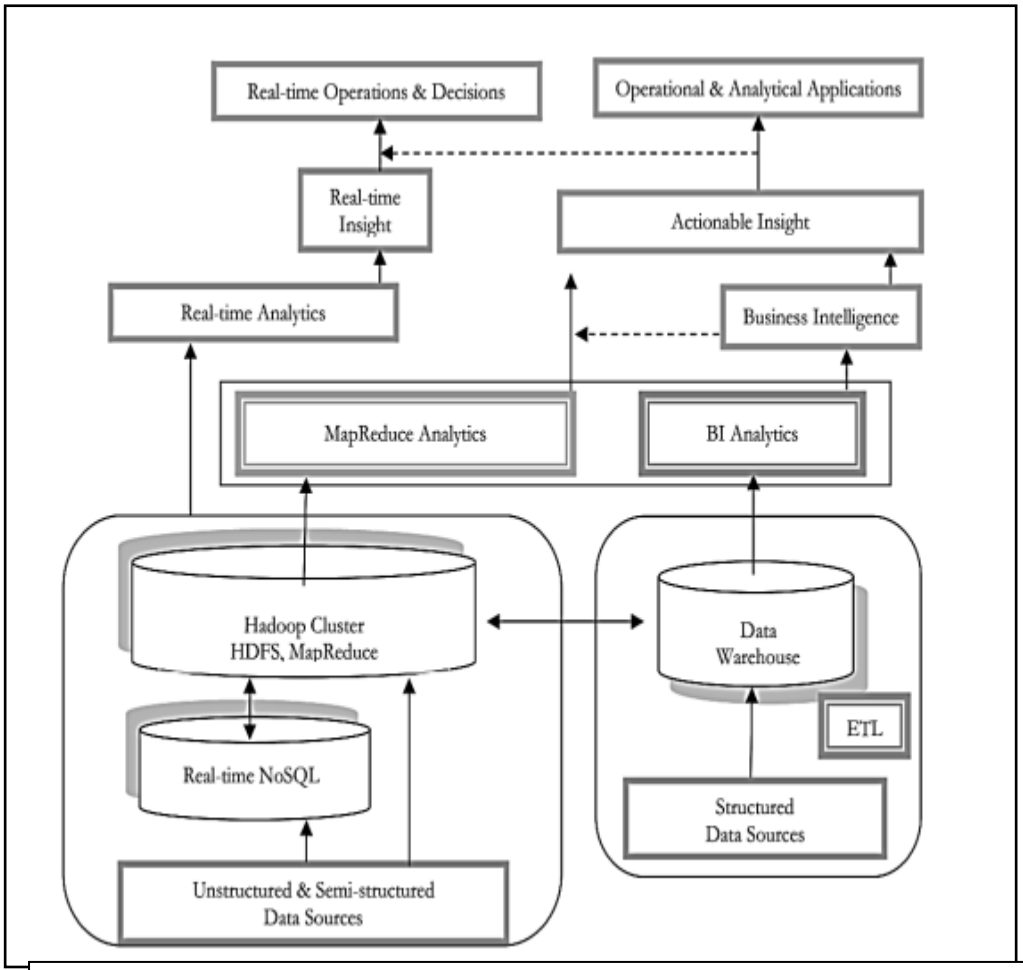


الشكل-07- لواحق برنامج Hadoop

5- الأوضاع الحرجة في قطاع الأعمال كأساس للتنقيب عن المعلومات: يمكن أن يبعث العنوان نوعا من الغموض حول من يحدث أولا، وهو عملية اتخاذ القرار يعتمد على عملية تحليل المعطيات، ولا تعارض لحد الآن لكن ما السبب الأساسي الباعث للتحليل بمعنى أن عملية التحليل تعتمد على اتخاذ قرار مسبق، وهذا القرار المسبق مبني على أساس ماذا؟ فلا شك أنه على أساس تحليل مسبق أيضا وهو يعتمد على طرق كلاسيكية لا تجديد فيها، وهذا ليس هدف هذا البحث، فلا نريد معرفة من يسبق من بل أن التطور يؤدي باستمرار بقطاع الأعمال بالوقوع في مواقف تبدو صعبة الحل في بدايتها، وهذا هو أساس عملية التنقيب عن المعطيات من أجل إيجاد الحل، حيث أن أي قطاع أعمال أساس عملية اتخاذ القرار فيه تعتمد على المعطيات يواجه مشكل وحيد أخذا بعين الاعتبار الثنائيات الخاصة به من أهداف، رغبات، قيود أو حتى الشخصيات، لكن هناك العديد من الوظائف التي تشكل أساس مشاكل قطاع الأعمال، لهذا يقوم المتخصصون في عملية تحليل المعطيات بالتعاون مع أصحاب المصالح بتجزئة المشكل إلى مشاكل صغيرة، وحل هذه الأخيرة يؤدي إلى التوصل إلى حل كلي، هذه المشاكل الجزئية قد تختص بقطاع أعمال معين، بينما أخرى هي وظائف عادية في عملية التنقيب عن المعطيات، على سبيل المثال فإذا كانت المؤسسة X تنشط في قطاع معين، فإن المشكل الجزئي هو خاص فقط بالمؤسسة X أي هناك خصائص تجعله مختلفا عن المؤسسات الأخرى في القطاع نفسه، وحل هذا المشكل الجزئي يصبح حلا لجميع المؤسسات في القطاع نفسه بعد فترة معينة، عن طريق جمع المعطيات والقيام بهذه العملية بشكل طبيعي، أي أن المشكل يكون فريد ثم يصبح كوظيفة عادية لعملية تحليل المعطيات، وعملية إيجاد الحلول نظريا وعلميا وتطبيقيا معلومة لدى أصحاب الاختصاص والواقع يثبت نجاعة هذه العملية. [11]

لقد تم التطرق إلى نوعية البيانات ووسائل التحليل المستعملة، ومن أجل التوصل إلى اتخاذ القرار على مستوى المؤسسة بدءا من مرحلة تجميع البيانات إلى مرحلة اتخاذ القرار، ومن أجل توضيح هذه العلاقة بين تحليل البيانات الضخمة وقطاع الأعمال يبين الشكل التالي[12] رقم -08- مختلف مراحل التحليل، حيث وحسب نوعية البيانات تستعمل الطريقة المناسبة والوسيلة المناسبة أيضا، فإذا كانت البيانات منظمة (structured) وحسب أنظمة متطورة فإنه من مصدر البيانات تمر البيانات إلى المخازن عبر مرحلة يمكن تسميتها مصفاة Extract transform and load(ETL)، تتعرض البيانات في المخازن DW إلى التحليل عن طريق عدة أنظمة والتي تدخل تحت Business Intelligence(BI)، هذا النوع الأول هو أقل صعوبة لأن النماذج مبنية مسبقا فقط تتم عملية التحليل ثم التوصل اتخاذ قرار في الوقت المطلوب. أما بالنسبة للبيانات غير المنظمة أو شبة المنظمة

Unstructured and semi-structured خاصة التي يتم الحصول عليها من وسائل التواصل الاجتماعي، الهواتف النقالة، الأقمار الاصطناعية وغيرها، في هذه الحالة تستعمل لغة NoSQL عادة ليتم تحميل البيانات إلى برنامج Hadoop(HDFS)، ثم تخضع البيانات إلى عملية المعالجة وفق ميكانيزم MapReduce الذي تم التطرق إليه، وحسب التنسيق بين فرعي برنامج Hadoop تظهر فعالية كبيرة في عملية التحليل واستخراج المعلومات اللازمة التي تعتمد في عملية اتخاذ القرار، أخذا بقاعدة "**نستعمل جميع البيانات من أجل عدم فقد المعلومات، ونستعمل أحسن الوسائل من أجل الوصول إلى أفضل النتائج**".



شكل-08- هيكل تحليل البيانات الضخمة على مستوى قطاع الأعمال

الخاتمة:

تناولت هذه الورقة البحثية أهمية البيانات بصورة عامة والبيانات الضخمة بصورة خاصة في ميدان اتخاذ القرار على مستوى المؤسسات أو الشركات باختلاف طابعها القانوني أو الاقتصادي، حيث تم تقديم مفهوم البيانات الضخمة بمختلف خصائصها، كما تم إظهار أهمية علم تحليل المعطيات والذي أخذ بعدا هاما بفضل البيانات الضخمة، ولم تظهر تلك الأهمية فقط على مستوى قطاع الأعمال، بل أصبح هناك تداخل بين ميدان التكنولوجيا والبحث والتطوير والميدان الاقتصادي، فأسفر ذلك استعمال وسائل وبرامج متطورة من أجل معالجة المشاكل التي تواجه محللي البيانات، والتي تستعمل بدورها في استخراج معلومات ونتائج تستعمل في عملية اتخاذ القرار، حيث يظهر الواقع ضرورة وجود هذا الصنف من الناس الذين نصفهم اقتصاديون ونصفهم الآخر ينتمي إلى تكنولوجيا المعلومات.

نتائج البحث: أضفى البحث إلى عدة نتائج نلخصها فيما يلي:

-علم تحليل البيانات أصبح علما قائما بذاته.

- أصبح هناك ترابط وتداخل بين العديد من التخصصات العلمية، وهو ما أدى إلى بالعالم إلى التقدم أكثر، حيث أصبح المتخصصون قادرون على القيام بالعديد من الأعمال بدلا من الاقتصار على وظيفة وحيدة.

- تطور وسائل التحليل وأساليب التحليل بشكل ملفت للانتباه، حيث أدى بالمهندسين بالتوجه للقطاع الاقتصادي من أجل عملية التطبيق للوسائل النظرية.

- زيادة اهتمام الشركات وقطاعات الأعمال بالمتخصصين في تحليل البيانات والبيانات الضخمة مستعملة نتائج التحليل ونماذج التنبؤ في أغلب القرارات المتخذة، حيث لا تكاد تخلو أي مؤسسة من هذه المصلحة مهما كان نوعها أو نوع نشاطها.

التوصيات والاقتراحات: من خلال النتائج المتوصل إليها خلال البحث يمكن أن نعطي جملة من الاقتراحات التالية:

- التنسيق والتعاون بين قطاع المعلومات والتكنولوجيا I&T وقطاع الإحصاء التطبيقي وإدارة الأعمال.

– تبني تخصص علم تحليل المعطيات ضمن الجامعات أو المعاهد (باعتبار عدم وجود هذا التخصص على مستوى الجزائر).

– القيام بدورات تكوينية بصورة دورية للعمال على مستوى المؤسسات، بهدف تحديث المعلومات والتعريف بطرق التحليل المبتكرة.

– نرى على مستوى الولايات المتحدة الأمريكية مثلا وجود مخابر بحث متخصصة في هذا المجال، تصدر مجلات دورية ونشرات، وحسب خبراء النمو الاقتصادي فإنه لا حرج في التقليد بهدف تحقيق معدلات نمو مقبولة، وتبنيا لهذا المبدأ:

– إنشاء مخابر متخصصة في علم تحليل المعطيات التطبيقي والنظري، والسعي إلى التعريف بهذه المخابر وأهميتها من خلال وسائل الإعلام المتعددة: دوريات وجرائد وحتى قنوات تلفزيونية ومواقع إلكترونية.

BIBLOGRAPHIE

[1] Feinleib David, *big data bootcamp, Springer Science+Business* Media New York, 2014.

[2] Baaziz Abdelkaader, Quoniam Luc, *"How to use big data technologies to optimize operations in upstream petrolium industry"*, International Jornal of innovation, sep 2013.

[3]Katal A, Wazid M, Goudar R H, "*Big Data: Issues, Challenges, Tools and Good Practices"*, IEEE, 2013.

[4] Grus Joel, *Data science from scratch: first principles with python*, O'Reilly media, 2015.

[5] lian Duan & Ye Xiong, *"big data analytics and business analytics*", journal of management analytics, may 2015.

[6] Vezzoso Simonetta, *" competition policy in a world of big data*", Research handbook on degital transformations, 2016.

[7] Sastry, Hanumanth Sistla, and M. S. Prasad Babu, *"Big data and predictive analytics in ERP systems for automating decision making process*", IEEE, 2014.

[8] Mohandy Soumendra, Jagadeesh madhu, Srivatsa harsha, *Big data imperatives*, Springer Science +Business Media New York, 2013.

[9]Holmes alex, *Hadoop in practice*, Manning publications, 2012.
[10] Tom white, *Hadoop: The definitive guide*, 3rd ed. O'Reilly media, 2012.
[11] Kenny petter, *Business Problems and Data Science Solutions*, Springer Science +Business Media New York, 2014.
[12] Biswas Sanjib, Sen Jadip, "*A proposal architecture for big data driven supply chain analytics"*.