المقالة الافتتاحية

البيانات الكبيرة: المنظور والتطبيقات العديدة

المستخلص

ثورة البيانات الكبيرة وتحليلاتها التي انتشر الحديث والبحوث المتعلقة بها في العقد الثاني من القرن الواحد والعشرين الحالي، ما هي إلا نتيجة عصر الرقمية الذي يشهده العالم الحديث في كل ابعاده وتوجهاته والمؤدية إلي التحول الرقمي المشاهد لكل أوجه الحياة المعاصرة.

وهذا العمل يستعرض في عجالة مفهوم البيانات الكبيرة وخصائصها المميزة إلي جانب التطرق لمشروعات البيانات الكبيرة وتطبيقاتها وأبعاد تحليلاتها والأدوات التكنولوجية المتقدمة المستخدمة فيها.

أ.د. محمد محمد الهادي

رئيس مجلس إدارة المجلة الشرفى

mohamed.m.elhadi@gmil.com

١. المقدمة .البيانات الكبيرة: المفهوم والخصائص:

١/١ المفهوم:

البيانات الكبيرة Big Data هو مصطلح مستخدم للإشارة إلي كميات معلومات ضخمة مرتبطة معا. فهي بيانات منظمة أو مهيكلة تتحرك سريعا وتشكل صعوبة لعملية استخدام قواعد البيانات العادية وبرمجة الإبداع. وفي معظم سيناريوهات المنظمة تكون البيانات بها غامضة أو مبهمة، تتدفق بواسطة سرعات متنوعة وغير متنبأ بها بسرعة، وتتجاوز قدرة المعالجة الحالية المتاحة للمنظمة.

٢/١ الخصائص:

البيانات الكبيرة توصف بواسطة الخصائص التالية:

(١) الحجم Volume: يمثل كمية الملومات التي تعتبر حيوية للمنشاأة في هذا السياق، حيث يكون مدى المعلومات التي تقرر جودتها وقدرتهما زال يفكر فيه، إما رؤيتها كبيانات كبيرة أم لا في الواقع. ويشتمل مسمي البيانات الكبيرة ذاته علي لفظ «كبير» الذي يعرف بواسطة الحجم هذه الخاصية بالتبعية.

(٢) التنوع Variety: يعتبر التنوع وجه البيانات الكبيرة التالي الذي يرتبط بتنوعها، كما يشير أيض إلي الوسيط الذي توضع في البيانات ويساعد الأشخاص الذين يبحثون عن المعلومات، كما يرتبط به توظيف المعلومات التي تشع وتثري الباحث بطريقة ملائمة، كما تعمل هذه الخاصية علي حفظ أهمية البيانات الكبيرة بالتبعية.

(٣) السرعة Velocity: يشير هذا اللفظ في سياق البيانات الكبيرة إلي سرعة توليد المعلومات، أو كيف

تكون البيانات منشأة ومعالجة لتلبية الطلبات والتحديات التي تقع في مسار النمو والتطوير إلي الأمام.

(٤) إمكانية التغيير م التقلب Variability: تمثل هذه الخاصية التغيير الذي يمكن أن يكون في حد ذاته قضية للأفراد يخترقون المعلومات. ويشير ذلك لعدم الانتظام الذي يظهر بواسطة المعلومات عن المناسبة أو امتلاك القدرة علي التداخل والتفاعل مع المعلومات بنجاح.

(٥) الصحة/الصدقVeracity : يمثل جودة البيانات الملتقطة التي يمكن أن تتنوع لحد كبير. وتعتمد دقة التحليل علي صدق مصدر البيانات ذاتها.

(٦) التعقيد Complexity: إدارة البيانات يمكن أن تصبح عملية معقدة جدا عندما تأتي أحجام البيانات الكبير من مصادر متعدد بصفة خاصة. هذه المعلومات يجب أن تترابط وتتوفق أو تتآلف لامتلاك القدرة للتداول علي البيانلا التي تعرض بواسطة هذه المعلومات.

٢.أهداف البيانات الكبيرة:

يحتاج للبيانات الكبيرة من أجل دعم ومساندة عمليات اتخاذ قرار المنظمة. وتنوع الأهداف يتحقق بواسطة المنظمات المختلفة مع استخدام تكنولوجيا المعلومات المتقدمة. وتتمثل بعض من هذه الأهداف في التالي:

(١) تقليل التكلفة من تكنولوجيا المعلومات المتقدمة الكبيرة: المنظمات التي تسعي للبيانات الكبيرة تعتقد أن ملايين المعلومات في الثانية الواحدة MIPS وتخزين تيرابايت Terabyte بيانات من البيانات غير المهيكلة صارت تقدم حاليا من خلال تكنولوجيا البيانات الكبيرة علي نطاق واسع ورخيص حاليا باستخدام تكنولوجيا إطار عمل Hadoop. والمنظمات التي كانت مرتكزة علي تقليل التكلفة أخذت قرار تطبيق أدوات تكنولوجيا البيانات الكبيرة علي أساس معايير فنية واقتصادية.

(٢) تقليل الوقت من البيانات الكبيرة: الهدف الثاني الشائع لتكنولوجيا وحلول البيانات الكبيرة يتمثل في تقليل الوقت.

(٣) تطوير مخرجات ونتائج مبنية علي البيانات الكبيرة: أحد الأبعاد الأكثر طموحا لما يمكن أن تعمله البيانات الكبيرة للمنظمة هو تطبيقها في تطوير نتائج منتج وخدمة جديدة مبنية علي البيانات. وكثير من المنظمات التي تطبق هذا المدخل تكون منظمات علي الخط التي لها حاجة واضحة لتطبيق المنتجات والخدمات المبنية علي البيانات. ومن الأمثلة لذلك، موقع LinkIn الذي استخدم البيانات الكبيرة وعلماء البيانات لتطوير نتائج وأوجه معلومات عن الأفراد والمجموعات التي يعرفها الفرد ويمكنه بحثها للتعرف علي الوظائف المهتم بها حيث تعرض سمة الفرد وغيره. وقد جذبت هذه النتائج ملايين الأشخاص لهذا لموقع.

(٤) مساندة مسارات الأعمال الداخلية في المنظمة: الغرض الأولي والأساسي لتحليلات البيانات الصغيرة التقليدية كان لمساندة القرارات الداخلية في المنظمة وذلك للتساؤل عن أسئلة مثل: ما النتئج الي يجب أن تعرض لعميل ما؟، أي العملاء أكثر احتمالا للتوقف عن أن يكنوا عملاء سريعا؟، كم كمية المخزون التي يجب حفظها في مستودع/ مخزن الشركة؟، كيف يمكن تسعير المنتجات والخدمات؟ أنواع القرارات التي تتخذ حيال حلول تلك الأسئلة تطبقها البيانات الكبيرة حاليا عندما توجد مصادر بيانات جديدة جديدة التي يمكن يمكن أن تطبق للقرار المتخذ. علي سبيل المثال، أي بيانات يمكن أن تلقي ضوءا علي رضا العميل تكون مفيدة. وفي هذه الحالة تكون بيانات كثيرة نابعة من تفاعلات العميل هي بيانات غير مهيكلة. وعلي ذلك فإن بيانات الأعمال مع البيانات الكبيرة يمكن أن يتضمن أيضا مجالات تقليدية أخري للتحليلات مثل سلاسل الإمداد، إدارة المخاطرة، أو التسعير. والعامل الذي يجعل مشاكل هذه البيانات الكبيرة بدلا من صغره يتكمثل في استخدام البيانات الخارجية لتحسين التحليل. علي سبيل المثال، الشركات التي تستخدم البيانات الخارجية لقياس ومراقبة مخاطر سلسلة الإمداد، فإن مصادر بيانات المورد الخارجي يمكن أن تشكل معلومات عن الموردين فيما يتعلق بقدراتهم الفنية، وأوضاعهم المالية الصحيحة، إدارة الجودة الخاصة بهم، موثوقية الإمداد لهم، وماراساتهم التجارية حتي يتخذ القرار المرتبط باختيار المورد (Snijdes et al, 2012

. كما أن الشركات الأكثر تقدما التي تعتمد في قرارتها علي البيانات هي التي لا تراقب فقط موردیهابل موردي موردیها أيضا.

٣. مشروعات البيانات الكبيرة:

تتوافر حاليا بعض أمثلة من مشروعات البيانات الكبيرة القائمة حاليا من قبل كثير من المنظمات والشركات، التي منها التالي:

(١) شركات منتجات المستهلك ومنظمات التجزئة: صارت تراقب الوسائط الاجتماعية مثل الفيسبوك Facebook وتويتر Twitter بغية الحصول علي رؤي جديدة لم يسبق التطرق لها تتصل بسلوك المستهلك وأفضلياته وإعداد المنتج أو الخدمة الجديدة.

(٢) صارت المصانع الحديثة تراقب ذبذبات أو اهتزازات اجهزتها التي تتغير عند التقادم وذلك للتنبؤ بالوقت الأمثل للاستبدال أو الصيانة. حيث أن استبدالها السريع قد يهدر الأموال، كما أن استبدالها المتأخر يؤدي لتوقف العمل بما يعود بالخسارة الناجمة من ذلك.

(٣) تراقب المصانع أيضا اشبكات الاجتماعية ولكن بغاية مختلفة عن المسوقين في الشركات ومنظمات التجزئة، حيث أنها تستخدم تلك الشبكات الاجتماعية لاكتشاف قضايا دعم المنتج بعد البيع قبل ما يصبح فشل الضمان الخاص بها بصفة عامة.

(٤) المصالح والأجهزة الحكومية صارت تجعل بياناتها متاحة بصفة عامة للمستخدمين علي كافة المستويات القومية والمحلية لتطوير منتجات جديدة تنشئ توعية جماهيرية وتقدم خدمات للمواطنين من خلال مشروعات الحكومات الإلكترونية.

(٥) منظمات الخدمات المالية تستخدم البيانات التي نقبت من تفاعاعلات العميل لكي تصنف عملائها في شرائح أو فئات مختلفة وغفقا لنتائج التنقيب، مما يساعد في تقديم عروض أو عطاءات مخصصة لكل فئة علي حدة.

(٦) وكالات الإعلان والتسويق تتبع الوسائط االاجتماعية لفهم مدى الاستجابة للحملات والترويجات ووسائط الإعلانات الأخرى.

(٧) شركات التأمين تستخدم تحليلات البيانات الكبرية لرؤية أي تطبيقات التأمين القائمة يمكن أن تعالج بصفة فورية، وأي منها تحتاج إلي التصحيح في زيارة مندوي التأمين للعميل.

(٨) منظمات التجزئة تعتبر أكبر داعي للسلع الجديدة وتعمل علي تغيير الإدراك المضاد عن المنتج الجديد، كما تقوم بإذكاء حماس العملاء نحو المنتج من خلال احتضانها الوسائط الاجتماعية.

(٩) المستشفيات المختلفة صارت عن طريق تحلي البيانات الطبية وسجلات المرضي التنبؤ يالمرضي المحتمل إعادة ترددهم وقبولهم علي مدي شهور قليلة من خروجهم، وبذلك عن طريق البيانات الكبيرة يمكن للمستشفي من الحد في بقاء المرضي الطويل المكلف بها.

(١٠) العمال المبنية علي الويب تطور منتجات معلومات مبنية علي الأعمال التي تجمع البيانات المجمعة من العملاء لتقديم توصيات أكثر قبولا وبرامج أكثر نجاحا.

٤. تحليلات البيانات الكبيرة:

تحليلات البيانات الكبيرة هي عملية فحص واختبار البيانات الكبيرة وتحليلها لاكتشاف الأنماط الخفية والارتباطان غير المعروفة وما يتخللها من معرفة تستخدم في اتخاذ قرارت أحسن. ومع تحليلات البيانات الكبيرة يمكن للمحللين المختلفين تحليل أحجام البيانات الكبيرة التي لا تستوعبها التحليلات التقليدية ومجال ذكاء الأعمال. حيث يسمح تحليل البيانات الكبرة للمحللين والباحثين ومدراء الأعمال اتخاذ قرارات رشيدة وأسرع باستخدام البيانات التي كانت غير متوافرة ولم يصلوا لها من قبل. واستخدام أبعاد التحليلات المتقدمة مثل تحليلات النص، تعلم الآلة، التحليلات التنبؤية، تنقيب البيانات، الاحصائيات، ومعالجة اللغة الطبيعية يمكن الأعمال من تحليل مصادر البيانات غير المستخدمة في السابق بطريقة مستقلة أو بطرق متكاملة معا ترتبط بإطار بيانات المنظمة لاكتساب بصائر جديدة ناجة من

قرارات أحسن وأسرع بصفة جوهرية (Snijders, Motzat and Reips, 2012).

ويمكن أن تساعد تحليلات البيانات الكبيرة المنظمات والمؤسسات المختلفة والمتنوعة لفهم المعلمات المتضمنة في البيانات بطريقة أحسن كما تساعد أيضا في تعريف البيانات الأكثر أهمية للأعمال المختلفة بالمنظمة المعنية، ولمستقبل قرارا الأعمال. وفي هذا الصدد يحتاج محللو البيانات المعرفة التي تنبع من تحليل البيانات بصفة أساسية.

ويعتبر مجال البيانات الكبيرة لمعظم المنظمات مجالا متسما بالتحدي لحد كبير. ويرتبط أحد هذه التحديات في تجزيئ محكم للبيانات وصولا لكل أبعاد مجموعات البيانات المتاحة في اماكن (فروع، إدارات) مختلفة ومتاحة في أطر نظم معلومات عديدة. والتحدي الثاني للبيانات الكبيرة يتمثل في إنشاء منصات المعالجة التي تجتذب البيانات غير المهيكلة بسهولة كالبيانات المهيكلة. ومن الملاحظ أن الحجم المتنامي والمتشعب من البيانات المتعاظم جدا يصعب معالجته باستخدام قواعد البيانات وطرق البرمجيات التقليدية المعمول بها في مظم منظمات الأعمال وخاصة في البيئات المحلية الحالية.

٥. تطبيقات البيانات الكبيرة:

تحليل البيانات الكبيرة مفيد في اتخاذ القرارات نابعة من البيانات المتاحة للمنظمات. وفي هذا التوجه ظهرت مشروعات عديدة لتطبيقات البيانات الكبيرة التي منها علي سبيل المثال لا الحصر المشروعات التالية:

١. تحليل البيانات المهيكلة: لا يزال تحليل البيانات مجالا من مجالات البحث النشطة الذي يتطلب أي تطبيق جديد تطوير طرق تحليلية جديدة. علي سبيل المثال، تطبق أسلوب تعلم الآلة المبني علي النماذج الرياضية والألجوريثمات القوية لاكتشاف الأبعاد الشاذة في البيانات ولتحديد الأنماط المقبولة التي تسهم في كشف المعرفة المحتاج لها.

٢. تحليل بيانات النص: يعتبر مجال تحليل بيانات النص عملية لاستخلاص معلومات ومعرفة مفيدة من النص غير المهيكل. وفي هذا الإطار تستخدم أساليب عديدة لتحليل البيانات التي منها تعلم الآلة، استرجاع المعلومات، الاحصائيات، الحوسبة اللغوية، وتنقيب البيانات (Chen, Mao and Liu, 2014). ومعظم تطبيقات تنقيب النص مبنية غلي معالجة اللغة الطبيعية للنص المستخدم.

٣. تحليل بيانات الويب: بزغ مجال تحليل بيانات الويب كمجال بحث نشط يهدف الاسترجاع، والاستخلاص والتقييم الآلي للوثائق والخدمات المختلفة المتاحة علي الويب لاكتشاف معرفة مفيدة تسهم في اتخاذ قرارات رشيدة نابعة منها. ويتضمن هذا المجال استخدام أساليب مثل تنقيب النص، وتنقيب الوصل المتشعب Hyperlink، وتنقيب المحتوي Content Minig.

٤. تحليل البيانات البنكية: عند التعامل مع مجموعات بيانات كبيرة، تواجه البنوك المختلفة صعوبات عديدة ترتبط بالقدرة علي تكامل وتداول ومعالجة تلك البيانات المتاحة بكفاة وفعالية. وعند زيادة كمية البيانات المنشأة عن العميل، فإن النسبة المئوية من البيانات التي يمكن أن يعالجها البنك تقل سريعا وذلك بسبب أن نظمها التقليدية الحالية تعتبر غير ملائمة لتخزين ومعالجة وتحليل كميات البيانات المهيكلة وغير المهيكلة المتدفقة بكثافةلأن مخططاتها ومدي التخزين العالي الكثيف غير متوافر لها. وبذلك صار من المهم توافر أبعاد معمارية تكنولوجيا متقدمة تتعامل مع البيانات الكبيرة. ويتضمن ذلك الحاجة لتوظيف طرق متقدمة لاستخلاص القيمة المرجوة من مجموعات البيانات الكبيرة المتاحة. وتقود طرق التحليل الجديدة إلي اتخاذ قرارات أكثر اعتمادا علي البيانات تتسم بالصدق والموثوقية. مما يسهم في اتخاذ قرارات أحسن ترضي العميل. وعلي ذلك صارت كثر من البنوك العاملة تستخدم تطبيقات البيانات الكبيرة لاستخلاص الفوائد التي تعود بالنفع والقيمة علي العملاء. وبالتالي تستخدم تحليلات البيانات الكبيرة في البنوك لتحقق متتطلبات ترتبط برضا العميل، توفير الأمن لحسبات العملاء من خلال اكتشاف التحايل

والغداع والغش، تحليل التغذية اراجعة، إلي جانب تحليل آراء العملاء والتعرف علي أفضلياتهم.

٥. تنقيب الشبكة الاجتماعية: تحليل البيانات الكبيرة المحملة علي الشبكات الاجتماعية يستخدم التحليلية الحوسبية Computational Analytics لفهم العلاقات الكامنة في المجتمع البشري بواسطة النظريات والطرق التي تقدمها مجالات الرياضيات والمعلوماتية وعلم الاجتماع وعلم الإدارة المتضمنة في تحليل الرأي العام، تحديد أبعاد تحليل ذكاء الشبكة، ومسارات اتخاذ القرارات والتعليم علي الخط، الخ.

٦. تطبيقات البيانات الكبيرة الطبية: ترتبط بيانات العناية الصحية والطبية بالبيانات المعقدة النامية والمتطورة بسرعة مستمرة في الأساس. وللبيانات الكبيرة الطبية أهمية غير محدودة للتخزين والمعالجة زالتساؤل والتحليل في كل الأبعاد اكلنيكية والصحية والإدارية والمالية المرتبطة بها. المعلومات الصحية يمكن إدخالها وتحميلها حاليا مع الأجهزة النقالة Mobile Devices الذكية وقد ترد من السجلات الطبية المتاحة للمرضي في المستشفيات أو من البحوث الطبية التي تحدد طرق العلاج والعقاقير الفعالة وغير ذلك من مؤثرات تؤدي للمرض (Big Data)

٦.أدوات وتكنولوجيا البيانات الكبيرة وتحليلاتها:

تتوافر حاليا أدوات وتكنولوجيا بيانات كبيرة متقدمة تستخدم في تحليلات البيانات الكبيرة ومن أهم هذه الأدوات أتوات هادوب Hadoop المطورة من قبل مؤسسة آباشي Abache.

والعرض التالي يحدد أبعاد هذه الأدوات التكنولوجية المتطورة:

- **أدوات هادوب Hadoop:**

تمثل أدوات هادوب إطار عمل برمجية مبنية علي لغة الجافا Java التي تساند معالجة مجموعات بيانات ضخمة في بيئة حوسبة موزعة Distributed Computation، وقد قام فريق عمل وبدعم مباشر من شركة جوجل Google من إعداد مشروع مفتوح المصدر أطلق عليه اسم إحدي الألعاب السائدة التي تتمثل في الفيل. وقد تطور ذلك لشركة Apache Hadoop المنبثة من مجموعة شركات مؤسسة برمجيات آباشي Apache Software Foundation.

وتحدد أدوات عادوب أدار عمل ممكنه تشغيل التطبيقات علي نظم المعلومات المتاحة بآلاف المحاور التي تتضمن آلاف التيرابايتات Terabytes ، وسهم في تسهيل نظام الملف الموزع في نقل بينات سريع بين المحاور، كما يسمح للنظام باستمرار التشغيل غير المنقطع في حالة فشل محور من المحاور المتاحة، بالإضافة أن هذا المدخل يقلل المخاطرة الناجمة من فشل النظام حتى ولو لم تعمل بعض مجاوره الأساسة.

ويستخدم إطار عمل Hadoop في كثير من الشركات العملاقة القائمة مثل شركات Google، ياهو Yahoo، آي.بي.إم. IBM، آمازون Amazon، الخ لتطبيقاتها التي تتضمن محركات بحث وإعلانات ونظم التشغيل المعتمدة علي برمجيات النوافذ Windows ، والبرمجيات المفتوحة المصدر Linux، إلي جانب ذلك تعمل أدوات هادوب مع كل نمن نظم BSD & OSX .

وتعتبر منهجية إطار عمل هادوب خطة مشروعات فرعية للحوسبة الموزعة، كما تمثل المنهجية الأحسن المستخدكمة لبيئة وقت التشغيل براميات MapReduce ، ونظام الملف الموزع HDDS.

وقد تشكل مشروع هادوب بواسطة تطوير الأدوات والبرمجيات الأساسية التالية:

- مركز هادوب يطلف عليه المحور Core الذي يتضمن الأجزاء والتفاعلات التي تمد إلي جانب محولا لأطر عمل وثائق متداولة تربط بالمدخل/ المخرج I/O (الذي يتمثل في التسلسل، لغة الجافا، هياكل المعلومات المتطابفة).ويتجزأ هذا المحور إلي تدفقات تعتبر جزمة خوادم جغرافية، تقلل حركة النظام بين الخوادم المتاحة في المجموعات الجغرافية العديدة.
- برمجيات MapReduce تمثل نموذج برمجة وإطار عمل برمجيات لكتابة التطبيقات التي تعالج بسرعة كميات بيانات كبيرة بالتوازي علي عناقيد

محاور حاسبات آلية كبيرة. كما تستخدم تلك البرامج نظام الملف الموزع HDPS وصولا لأجزاء الملف ولتخزين النتائج المختصرة.

- نظام الملف الموزع HDFS هو نظام تخزين مستخدم بواسطة تطبيقات هادوب في الأساس. ويقترح إطار عمل وثائق مبثة وموزعة يعطي وصولا متداخلا لتطبيق المعلومات. كما يقوم بعمل إعادة إنتاج مكعبات معلومات عديدة علي عملية تتمحور كليا خلال فرع يعمل علي تقوية الحسابات السريعة. وتقدم مشروعات هادوب الأخرى خدمات وتجريبات مستويات أعلي متممة.

أما المشروعات الفرعية لإطار عمل هادوب فتتمثل في التالي:

- HBASE تعتبر جزء من قاعدة بيانات فعلية قائمة، وتستخدم المخزون إصدار فواتير المخزون Stock Billing، كما تحدد خريطة نظام الملف الموزع HDFS في قاعدة البيانات المتعلقة بالهيكلية، وتقدم تفاعل برمجة تطبيق جافا Java API وصولا لقاعدة البيانات المستخدمة، إلي جانب ذلك فإن هذه الأداة تدعم وتساند مجموعة نمط الحسابات التي تستخدم برمجيات MapReduce وتشير لتساؤلات الاستراتيجيات المسبقة غير العادية. إضافة لكل تلك الاستخدمات، فإن هذه الأداة تستخدم أيضا كجزء من منهجية هادوب عند الوصول للقراءة/ الكتابة الشاذة متطلبا. وتستهدف الأداة تسهيل الجداول المسهبة المشكلة علي قمة فرع جهاز الفرع أو الوحدة المعينة.
- أداة Pig تمثل تدفق بيانات لغة معينة يطلق عليها Apache Pig تشكل مرحلة تجزيء المعلومات الكبيرة تمازليا، وتتضمن حالة برامج غير ناضجة لتوصيل معلومات فحص واختبار البرامج. ويعتبر الهيكل السليم لهذه الأداة توازي هيكلة تقويتها بعدد كبير لمشروعاتها التي تتداول فائدة مجموعات المعلومات جوهريا.، وتعطي حرية لتحسنات الهادوب، وتعتبر أقل طلبا مقارنة ببرامج MapReduce.
- أدة ZooKeeper تمثل أداة ومكون عنقود Cluster تفيد إدارة تسلسل عناقيد هادوب كبيرة. وتعمل هذه الأداة علي تركيز خدمات حفظ وصيانة معلومات مكون العنقود المعين وتوضح أبعاد تسميته، وتقديم التزامن الموزع ومجموعات الخدمات المتنوعة.
- أداة HIVE تمثل بيئة مستودع بيانات أساسية مبنية علي علي قمة إطار عمل هادوب، وتقدم أدوات لمساعدة تلخيص بيانات سهل، كما توفر تساؤلات عشوائية، وتعمل علي تحليل مجموعات بيانات مخزنة في ملفات هادوب. وفي نفس الوقت تقدم آليات لوضع هيكلية هذه البيانات مع لغة تساؤل بسيطة يطلق عليها HIVE QL مبنية علي لغلة التساؤل المهيكلة SQL تساعد المستخدمين الملمين بلغة SQL في تساؤلاتهم عن البيانات.
- أداة Chukwa تستخدم لمراقبة عناقيد خوادم موزعة كبيرة. وتمثل نظام جمع بيانات لمراقبة النظم الموزعة الكبيرة المتاحة. وتتضمن هذه الأداة مجموعة أدوات مرنة وقوية لعرض ومراقبة وتحليل نتائج الاستخدام الأحسن للبيانات.
- أداة HCatalogتعتبر أداة طبقة إدارة تخزين الهادوب التي تساعد المستخدمين فيما يتعلق بأدوات معالجة بيانات مختلفة. ويعرض تحريد جدول هذه الأداة رؤي بيانات علاقية في نظام الملف الموزع HDFS، كما تؤكد عدم حاجة المستخدمين للقلق عن أين وفي أي شكل تخزن بياناتهم.

المراجع:

1. Big Data [http://www.sas.com/en_th/insigths/big-data/what-is-big-data.html]

2. Chen, M., Mao, S. and Liu, Y. (April 2014). «Big Data: A Survey,» Mobile Networks of Applications, Vol. 19, pp. 171 - 209.

3. Snijders, S., Matzat, U. and Reips, U. D. (July 2012). «Big Gaps of Knowledge in the Field of Internet Science,» International Journal of Internet