يمكن تحقيق بذلك باتباع بعض الطرق والاستراتيجيات *مثل* التنقيب عن المعرفة في المجال التعليمي وهو مجال بحثي جديد يُعرّف على أنه استخدام تقنيات التنقيب في البيانات الخام المستخرجة من الأنظمة والعمليات التعليمية بهدف مساعدة مصممي ومستخدمي هذه الأنظمة لتحسين أدائها واستخلاص معلومات مفيدة عن سلوك الطلاب في عملية التعلّم. كما أن استخدام تقنيات التنقيب في أنظمة التعلم الافتراضية يمكن أن يلعب دوراً هاماً وفعالاً في حل مشاكل التعلم وتقديم المعلومات التي تساعد أصحاب القرار من معلمين وواضعي الخطط الدراسية باتخاذ قرارات مدعومة بالمعلومات عن كيفية التفاعل مع المصادر التعليمية. وبذلك تكون عملية التنقيب السابقة تطبيق للأساليب الآلية بغية الكشف عن الأنماط في مجموعات كبيرة من البيانات التعليمية. قد يكون من الصعب أو المستحيل تحقيق ذلك نتيجة حجم البيانات الهائلة المستخلصة من الأنظمة التعليمية، إلا أن الجهود قد تكثفت في الآونة الأخيرة لإيجاد أدوات تسهّل وتمهّد الطريق أمام التنقيب بكمية هائلة من المعطيات وتعمل على تجهيز هذه المعطيات لمرحلة التنقيب قبل البدء بها.

### بعض الأدوات المستخدمة في عملية التحليل:

من أهم التطوّرات التي شهدها هذا المجال هو إيجاد طرق لإدارة الكميات الضخمة من البيانات مما يسهل إجراء عمليات التحليل عليها. فقط طورت غوغل MapReduce للحد من التحديات التي تظهر نتيجة إدارة البيانات على نطاق الإنترنت، بما في ذلك توزيع البيانات والتطبيقات التي تمتلك بيانات مرتبطة بشبكات الحواسيب. فقواعد البيانات السابقة كانت غير قادرة على إدارة البيانات على نطاق شبكة الإنترنت، مما أدى إلى تطوير Apache Hadoop الذي يستخدم الآن بشكل شائع في إدارة البيانات.

ظهرت مؤخراً أداة جديدة تدعى Apache Spark الذي يتمتع بميزات عديدة كمعالجة البيانات المتدفقة مما جعلها تتفوق على الأدوات الأخرى.

بالإضافة للأدوات المستخدمة في إدارة البيانات ظهرت العديد من الأدوات التي تدعم عمليات التحليل وتؤمن عمليات التنقيب عن المعرفة دون الحاجة لمعرفة عميقة في علوم الحاسب، معظم هذه الأدوات مطور بما يتوافق مع مجال ذكاء الأعمال والذي يتجلى بظهور SAS وأدوات IBM في المجال التعليمي، والتي استخدمت في التحليلات التنبؤية وتحسين عملية صنع القرار من خلال تحليل كميات كبيرة من البيانات وعرضها بشكل تفاعلي. فيما سبق كان استخدام الأدوات التحليلية عملية معقدة تتطلب فهم عميق في التحليل والبرمجة، أما الآن فحتى الأدوات المعقدة كـ SAS, RapidMiner, DBMiner, DB2 Intellegent Miner, SPSS تتطلب معرفة تقنية أقل إلا أن معظمها تجاري(غير مجاني). وهنا لا بد من الإشارة أن كثير من برامج الحاسوب مثل Microsoft Excel شهد تطور ملحوظ في التحليل والإظهار. يوجد أدوات أخرى تم تصميميها لدعم استخدام أدوات التحليل دون الحاجة إلى معرفة تقنية عالية، مما أدى إلى انتشار عمليات تحليل التنقيب في البيانات التعليمية وتحليلها بين عدد كبير من الباحثين في مجال علوم التعليم.

يوجد برمجيات أخرى مفتوحة المصدر تؤمن عمليات التنقيب مثل Weka, Keel. كما يوجد بعض البرمجيات المختصة بالتنقيب عن المعطيات التعليمية مثل: MultiStar الذي يعمل على التصنيف ودراسة علاقات الارتباط، EPRules لدراسة علاقات الارتباط، KAON للعنقدة والتنقيب في النصوص، Synergo/ColAT للحصول على الإحصائيات وعمليات عرض النتائج، GISMO للعرض، Listen tool للعرض، TADA-Ed للتنقيب والعرض، Sequential Mining tool لتنقيب الأنماط، MINEL لتنقيب طرق التعلم، CIECoF لتنقيب علاقات الارتباط، Simulog لكشف أنماط السلوك غير المتوقع.

أما عن كمية البيانات التي عادة ما يتم معالجتها فهي ضخمة جداً، فمع تزايد استخدام أجهزة المحمول والتقنيات المتزامنة في العمليات التعليمية، أُتيحت كميات هائلة من البيانات تِبعاً لإمكانية الحصول على نتيجة كامل العمليات وأشكال تفاعل الطالب مع المحتوى التعليمي وتسجيلها “logging” وجعلها متاحة للأغراض التحليلية. كما أنه أصبح من المتوفر الآن كميات كبيرة من البيانات للباحثين من خلال السجلات العامة المتاحة من خلال المركز “Pittsburgh Science of Learning Center Datashop”.

## النموذج المفاهيمي للنظام المقترح:

يتمتع النظام المقترح بمرونة عالية نتيجة اعتماده على مكونات تؤدي خدمات متباينة، يسهل إضافتها وحذفها. بالإضافة لاعتماده على معيار موحد بهيكلة البيانات تسهيلاً لخاصية **البينية** **interoperability** وسهولة دعم أنظمة متنوعة. وبالتالي فالنظام المقترح يصبو لتحقيق Open Learning Analytics، أي دعم الـ 4 R المتمثلة بـ Reuse, Redistribute, Revise, Remix. بالإضافة لاستفادة كل من يشارك في العملية التعلمية من الخدمات التي يقدمها النظام، فعلى سبيل المثال وبنقيض الكثير من الأنظمة الأخرى يشكل الطالب مستفيد مركزي وأساسي من الخدمات المقدمة ولا يشكل مجرد مولد للبيانات للاستفادة منها في التحليل المقدّم للمعلمين. يشمل الانفتاح Openness الجوانب الأربعة المذكورة سابقاً في النموذج المرجعي في تحليل التعلم. فيكون الإجابة عن الإجابات الأربعة الواردة في عملية التحليل كما يلي:

* “What”: تكون مصادر البيانات المحلة متنوعة وتشمل البيانات الناتجة من أنظمة إدارة التعلم ونظم الاختبارات ونظم معلومات الطلاب وغيرها.
* “Who”: يخدم النظام مستخدمين متنوعين وذوي احتياجات متنوعة ومختلفة.
* “Why”: يقدم النظام خدمات مختلفة باختلاف المستفيد النهائي من الخدمة وبالتالي فتكون الأهداف متنوعة ولكن جميعها يصب في مصلحة الطالب وتحسين وضعه وبالتالي تحسين أوضاع المؤسسات التعليمية.
* “How”: يتم تحقيق الأهداف بأشكال متنوعة لتحقيق الأهداف المرجو تحقيقها بأفضل طريقة ممكنة.

نورد فيما يلي أمثلة عن حالات استخدام مقترحة في هذا النموذج تبعاً لأنواع المستخدمين النهائيين:

* المعلم: يمكن لمدرس يستخدم نظام إدارة تعلم على سبيل المثال في العملية التعليمية استخدام النموذج المطروح الذي من الممكن أن يعطيها لمحة وتغذية راجعة عن المقررات المسؤولة عن تدريسها بالاعتماد على مؤشرات مختلفة لتحسين العملية التدريسية. فيمكن للمدرس على سبيل المثال لا الحصر الحصول على معدل مشاركة الطلاب في المحاضرة والحوارات، الوثائق الأكثر مشاهدة وتحميل، التقدم الذي يحرزه الطلاب في الاختبارات. في حال عدم وجود مؤشر ما قد يفيد المعلم في الحصول على معرفة معينة تفيده، يؤمن النظام مرونة كبيرة بإمكانية دعم واضافة المؤشرات حسب حاجة وطلب المدرس.
* الطالب: يمكن لطالب يستخدم MOOC على سبيل المثال وبرغب بالحصول بمعرفة وإحصائيات حول وضعه وربما بعض النصائح التي تحسن من وضعه. يمكن إضافة مؤشر إضافية هنا أيضاً حسب رغبة الطالب.
* المطور: يمكن لمطور أنظمة التعلم الاستفادة ودمج نظام تحليل التعلم مع النظام التعليمي باعتماد على API يتم فيه تبادل البيانات بين النظام التعليمي ونظام التحليل بعد الاتفاق على صيغة معينة لهيكلة البيانات وفقها.

وبالتالي يتمتع النظام المقترح بالخصائص التالية:

البينية، المعيرة، قابلية إعادة الاستخدام، المرونة وقابلية التوسع، والاعتماد على مبدأ Modularity وبالتالي إمكانية إضافة وحدات وخصائص جديدة بسهولة وفعالية، وبالطبع الاهتمام بموضوع الأداء وتوفير نتائج التحليل في الوقت المطلوب دون تأخير. كما يميز النظام سهولة الاستخدام، فعلى الرغم من استخدام بعض التقنيات المعقدة كالتنقيب إلا أن النتائج يتم تقديمها بشكل سهل ومبسط للمستخدم دون الحاجة لأي معرفة تقنية. ويجب الاهتمام بموضوع الخصوصية والشفافية في أي نظام يقوم بعلمية تحليل لبيانات الآخرين.

في الشكل التالي توضيح للمكونات الأساسية للنموذج المطروح وكيفية التفاعل فيما بينها

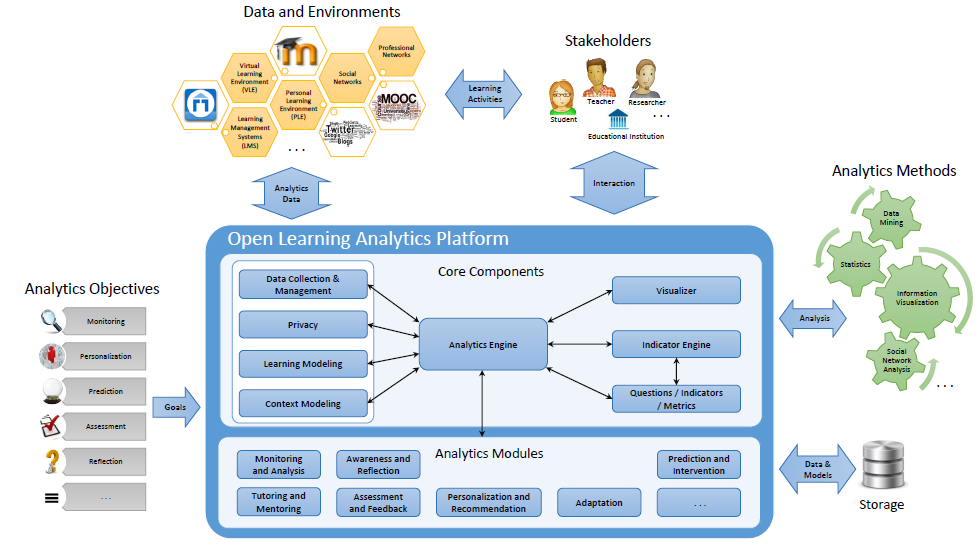


Figure :1 النموذج المجرد لمعمارية النظام

## عملية ما قبل المعالجة:

كلما كانت عملية تحضير البيانات الخام قبل البدء بمعالجتها أكثر، كلما أعطت عملية المعالجة نتائج أفضل. ولكن عملية التحضير هذه تتطلب جهد يدوي كبير، إذ أن هذه العملية تستهلك حوالي 60-90% من الوقت والجهد والموارد. فيما يتعلق بالبيانات الناتجة عن الأنظمة التعليمية نجد أن كمية البيانات ضخمة جداً وربما لا تكون بالشكل الأنسب وبالتالي تكون عملية التحليل غير مجدية. لذلك يجب العمل على تهيئة هذه البيانات للحصول على النتائج الأمثل من عمليات التحليل. هناك أمور أخرى يجب الاهتمام بها كعملية المكاملة بين البيانات الناتجة عن أنظمة مختلفة.

يمكن تلخيص أهم المهام التي يتم إنجازها في عملية ما قبل المعالجة بما يلي:

data cleaning, user identification, session identification, path completion, transaction identification, data transformation and enrichment, data integration, data reduction.

نتيجة لاعتماد معيار موحد في هيكلة البيانات تتقلص كمية العمل الواجب القيام به على البيانات بغية تجهيزها قبل البدء بمعالجتها.

* **Moodle**: Modular Object Oriented Developmental Learning Environment

نظام إدارة تعلم مفتوح المصدر، يساعد المدرسين على خلق بيئة تعليمية تفاعلية فعالة على شبكة الإنترنت. يدعم حوالي 75 لغة فيما يزيد عن 160 دولة. ويعتمد في تصميمه على نمط تعلّم يدعى social constructionist pedagogy الذي يقوم على أن أفضل العمليات التعليمية تكون عند التفاعل مع: المحتوى التعليمي والأقران حول المحتوى التعليمي.5

يحوي مودل على:

-خمس أصناف للأدوات التعليمية: الصفحات النصية، صفحات الويب، رابط لشيء آخر على الويب، view into one of the course’s directories وعرض لنص أو صورة.

-ست طرق للتفاعل مع المقرر: الوظائف، الخيارات، الصحف، الدروس، الاختبارات والدراسات الاستقصائية.

-خمس طرق للتفاعل مع الآخرين: الدردشة، المنتدى، glossary، الويكي وورشات العمل.

يخزن مودل كل تحركات الطلاب وكل النقرات التي يقومون بها في النظام، ويمكن فلترة هذه المعلومات على حسب المقرر، النشاط، الطالب، واليوم. كما يمكن للمعلمين استخدام هذه البيانات لمعرفة الطلاب الأكثر نشاط في المقرر. لا يتم تخرين البيانات السابقة على شكل نصوص وإنما يتم تخزينها في قاعدة معطيات علائقية. أفضل النظم إدارة قواعد المعطيات الداعمة لمودل MySQL, PostgreSQL إلا أنه يمكن استخدام نظم أخرى مثل Oracle, Access, Interbase. يحوي مودل حوالي 145 جدول مترابط، إلا أن عملية التنقيب لا تحتاج إلى كل هذه البيانات وهنا يبرز دور مرحلة ما قبل التنقيب. نتيجة لبنية مودل تكون كمية الجهد المبذول في مرحلة ما قبل المعالجة أقل من الأنظمة التعليمية الأخرى. إلا أن بعض العمليات لا تزال مطلوبة مثل اختيار البيانات التي سيتم عملية التنقيب عليها وتحويل البيانات الرقمية إلى وصفية ونقل وتحويل البيانات للصيغة الأمثل لعميات التنقيب كصيغة ARFF(Attribute-Relation File Format) مثلاً.

تطبيق تقنيات التنقيب على بيانات Moodle:

* الإحصائيات: لا يؤمن مودل إحصائيات تزود المدرسين بتقارير محددة عن تفاصيل إحصائية لأداء كل طالب. وإنما يزود بإحصائيات لها علاقة فقط بأداء الطلاب (كالعلامات والاختبارات)، فيمكن للمدرسة معرفة أداء الطلاب في كل سؤال في الاختبار، مما يساعد المعلمين على تحديد الأسئلة الواجب تعديلها أو حذفها نتيجة لصعوبتها أو سهولتها الفائقة.
* العرض: يتعلق برسومات الحاسوب وواجهات المستخدم ويهتم بطرق عرض البيانات بشكل رسومي أو متحرك ليسهل على المستخدم فهمها. تبرز فعالية هذه الطريقة عند التعامل مع كمية كبيرة من البيانات ولتمثيل البيانات متعددة الأبعاد. لا يؤمن مودل طرق مرئية لعرض هذه البيانات غير النصوص، إلا أنه يمكن استخدام بعض الأدوات مثل GISMO الذي يمثل البيانات الناتجة عن تتبع الطلاب في مودل بطرق متعددة مرئية ومفهومة.
* العنقدة: تتصف بأنها [*unsupervised learning*](http://en.wikipedia.org/wiki/Unsupervised_learning)، تعمل العنقدة على كشف العلاقات بين العناصر بهدف تجميع العناصر المتشابهة. المبدأ المتبع في العنقدة هو زيادة التشابه في العنقود الواحد وتقليص التشابه بين العناقيد المختلفة. في التعلم الافتراضي يمكن الاستفادة من العنقدة في إيجاد الطلاب ذوي الخصائص التعليمية المشتركة وتشجيع التعلم التعاوني. يمكن بعد تجميع الطلاب في مجموعات تحديد مرشد خاص بهم بناء على الخصائص التي يتمتع بها طلاب المجموعة ومهاراتهم. كما يمكن توجيه اختبارات خاصة بهم. تؤمن الأداة Weka عدة خوارزميات عنقدة. يمكن على سبيل المثال تجميع الطلاب بناء على نشاطاتهم في مودل ونتائج الاختبارات النهائية.
* التصنيف: يتصف بأنه[supervised learning](http://en.wikipedia.org/wiki/Supervised_learning)، يعتمد على معرفة انتماء عناصر جديدة لفئات موجودة ومحددة مسبقاً. يمكن الاستفادة من التصنيف في التعلم الافتراضي في كشف الطلاب ذوي الخصائص المتشابهة وردّات الفعل المتشابهة على الاستراتيجية التربوية المتبعة، وللتنبؤ بأداء الطلاب ونتيجتهم النهائية ولتحديد المتعلمين الذين لا يملكون الدوافع والمحفزات اتخاذ الإجراءات لتخفيض معدلات التسرب. تؤمن الأداة Keel عدة خوارزميات تصنيف. يمكن للمعلم استخدام المعرفة الناتجة عن التصنيف باتخاذ قرارات لها علاقة بالنشاطات المتعلقة بالمقرر ضمن مودل.
* علاقات الارتباط: تعتمد علاقات الارتباط على مدى ارتباط حدوث العنصر x بعنصر آخر y. يمكن الاستفادة من علاقات الارتباط بتحديد المواد التعليمية الأنسب لاقتراحها على المستخدم، ولاكتشاف العلاقات المثيرة للاهتمام من استخدام الطالب للمعلومات بهدف توفير التغذية الراجعة لمدرس المقرر، ولكشف العلاقات بين سلوكيات التعلم. كما يمكن استخدامها لكشف الأخطاء التي يشترك الطلاب بارتكابها. تؤمن الأداة Weka عدة خوارزميات لكشف علاقات الارتباط.
* تنقيب الأنماط التتابعية
* تنقيب النصوص

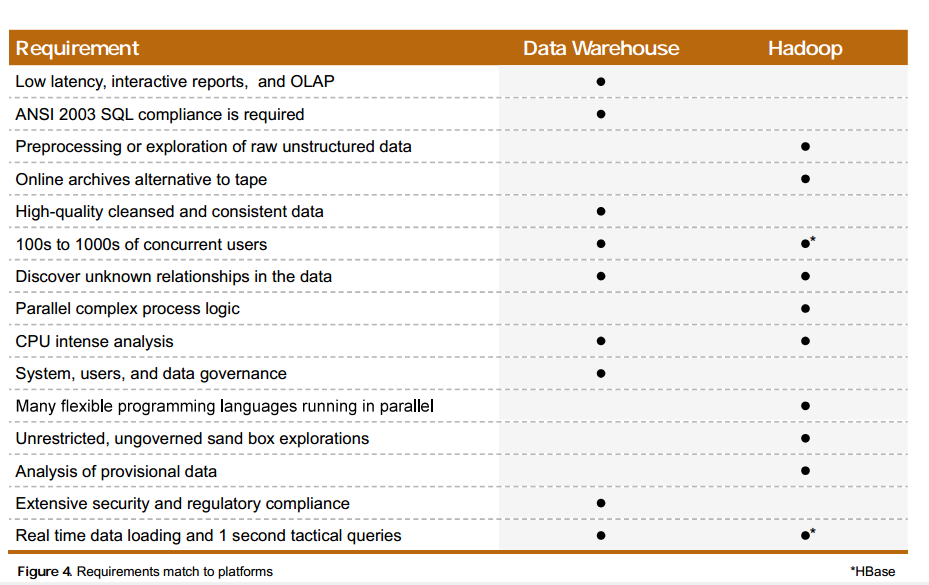
## تقنيات الـ Bid Data و Data Warehouse:

نظراً لتنوع البيانات وغرارتها نتيجة لمراقبة وتخزين كافة تحركات وتفاعلات الطلاب مع البيئات التعليمية الافتراضية، كانت الحاجة لاستخدام بعض تقنيات ذكاء الأعمال التي من شأنها أن تسهل عملية التحليل وتزيد فعاليتها وسرعتها. وهنا يكون السؤال، ما هي الطرق والتقنيات الأكثر فعالية في نظام التحليل المقترح؟ هل يكفي استخدام مستودعات البيانات Data Warehouses أم أن استخدام منصات المعالجة التفرّعية مثل Apache Hadoop قد يكون أكثر فعالية؟

إن جواب هذا السؤال يكون بعد دراسة للجوانب التي تغطيها هذ التقنيات وارتباطها بالنظام المقترح والوظائف التي يعمل على تلبيتها. وهنا يجب الإشارة إلى عدم تفوق إحدى الاستراتيجيات على الأخرى وإنما يكون القرار المتخذ بناءً على هيكلية النظام المقترح والخدمات التي يقدمها. نورد فيما يلي دراسة مبسّطة لبعض حالات الاستخدام لكل من مستودعات البيانات Data Warehouses ومنصة التحليل والتخزين التفرعية Apache Hadoop:

* Apache Hadoop: منصة تحليل وتخزين تفرعية تعتمد مبدأ Map/Reduce في عملها، إذ يتم القيام بالعمل اعتماداً على عشرات إلى مئات العقد من السيرفرات. تتم عمليات تجميع المعطيات وتنظيفها وإعادة هيكلتها ومعالجتها في الـ run time.
  + تستقبل Hadoop البيانات غير المهيكلة ولا تؤمن عملية فحص لهيكلية البيانات ونوعها بعكس قواعد المعطيات العلاقاتية التي لا تسمح بتخزين البيانات إن لم تكن تتصف بصفات وهيكلية معينة.
  + يعتبر Hadoop مثالي في التطبيقات التي تستخدم بيانات قصيرة الأمد والتي سيتم استخدامها لشهر أو اثنين ثم التخلص منها.
  + يشكل Hadoop الحل الأمثل للتطبيقات التي تكون إمكانية التوسع فيها عالية نظراً لقلة الكلفة المحتاجة في حال التوسع.
* Data Warehouses: مستودعات للبيانات تتصف بأنها Subject Oriented إذ يتم تصميمها بناء على بنية النظام واحتياجاته، وتكون مستودعات فيزيائية أي أنها ليست مجرد مستودعات منطقية. كما أن البيانات فيها يتم تجميعها من مصادر مختلفة وتتصف بأنها بيانات غير متطايرة.
  + تستقبل مستودعات البيانات معطيات مهيكلة ذات حجوم كبيرة.
  + تخزن مستودعات البيانات معطيات تاريخية طويلة الأمد.
  + تعتبر الحل الأمثل والأكثر فعالية للقيام بإعداد التقارير بعكس Apache Hadoop الذي يتمتع بإمكانيات محدودة في إعداد التقارير.
  + تكون كلفة التوسع فيها كبير.

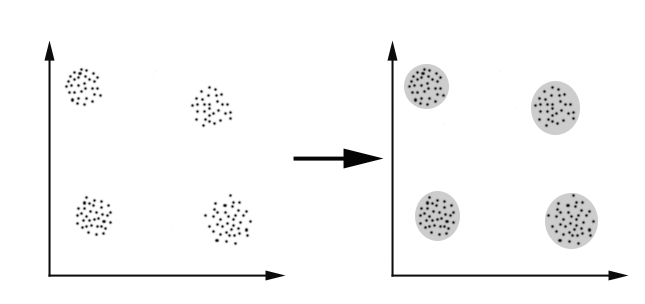
يبين الشكل التالي حالات استخدام كل من Hadoop و Data warehouses



الشكل 9:Requirements match to Hadoop or Data Warehouse

## العنقدة:

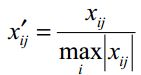
تعتبر العنقدة من أشهر التقنيات المستخدمة في الـ Unsupervised learning إذ تهدف إلى تجميع مجموعة من الأغراض المتشابهة بنواحي محددة في مجموعة واحدة تسمى عنقود. ومنه يمكن تعريف العنقود على أنه مجموعة الأغراض المتشابهة فيما بينها والمختلفة عن الأغراض الموجودة في العناقيد الأخرى. تركز العنقدة على تجميع العناصر المتشابهة في البينية والسلوك. الشكل التالي يوضح مثال عن تجميع الأغراض المتشابهة في عناقيد.



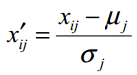
الشكل : مثال عن عملية العنقدة

عادة ما يتم تمثيل العينات المراد عنقدتها في فضاء متعدد الأبعاد وفق قيم الواصفات الموافقة لها. كما يتم في بعض الحالات إجراء بعض عمليات التسوية على الواصفات لتقريب مجالات القيم فيما بينها، مما يجعل جميع قيم الواصفات تنحصر في المجال [-1,+1]. نورد فيما يلي بعض طرق التسوية: (فيما يلي *i* يمثل المشاهدة*، j* يمثل الواصفة)

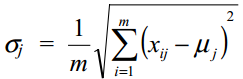
* تقسيم كل قيم الواصفة على أكبر قيمة مطلقة للواصفة بين العينات.



* طرح المتوسط µj من كل قيم الواصفة ثم تقسيمها على الانحراف المعياري σj لهذه الواصفة

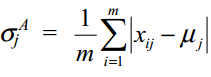


حيث:

* طرح المتوسط µj من كل قيم الواصفة ثم تقسيمها على الانحراف المطلق σj2 لهذه الواصفة

حيث:



### نقاط القوة في خوارزميات العنقدة:

* قابلية التوسع: تحافظ خوارزميات العنقدة على فعاليتها حتى مع تزايد حجم البيانات.
* القدرة على التعامل مع أنواع الواصفات المختلفة: مثل الواصفات الرقمية أو الوصفية أو حتى البوليانية.
* القدرة على التعامل مع عدد كبير من الأبعاد: فخوارزميات العنقدة تتميز بقدرتها على معالجة عدد كبير من الأبعاد، مما يسمح بأخذ معرفة كافية عن البيانات التي تتم عنقدتها بعين الاعتبار.
* القدرة على التعامل مع البيانات غير النظيفة: كالبيانات التي تحوي قيم مفقودة. إذ أن بعض الخوارزميات الأخرى غير خوارزميات العنقدة تكون حساسة في هذه الحالة وتؤدي إلى نتائج غير جيدة.
* البينية: فلا تتأثر خوارزميات العنقدة باختلاف مصدر البيانات.

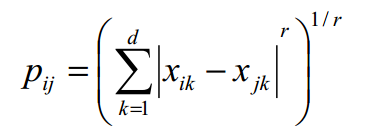
### أنواع العنقدة:

* العنقدة الصلبة: كل عينة تنتمي إلى عنقود واحد فقط فالعلاقة مع العنقود تكون انتماء أو عدم انتماء.
* العنقدة المرنة: بدلاً من وضع كل عينة في عنقود واحد فقط يتم وضع احتمالات انتماء لعناقيد مختلفة.

### طرق حساب التشابه:

يتم حساب التشابه بين قيم العينات المدروسة لتحديد العناقيد التي تنتمي كل منها إليها وفق قيمها بعدة طرق منها:

* Distance Measures:



* Ordinal Measures:
* Cosine Measure:



### تقنيات العنقدة:

* :Center-Based Partitional Clustering تركز الفكرة الأساسية فيها على إمكانية تمثيل العنقود بمركزه. بعض الأمثلة عليها خوارزمية K-means التي تعتمد على centroid الذي يمثل متوسط مجموعة النقاط، وخوارزمية K-medoid التي يتم الاعتماد فيها على medoid الذي يمثل أكثر 16 نقطة مركزية لمجموعة النقاط.
* Hierarchical Clustering: تهدف إلى الوصول لسلسة هرمية من العناقيد. يدعى الكل الناتج عن التمثيل dendogram
* Density Based Clustering:
* Graph-Based Clustering: