1.Tại sao khai thác dữ liệu lại có nhu cầu cao và nó là một phần của sự phát triển tự nhiên của công nghệ thông tin

Khai thác dữ liệu biến dữ liệu thành kiến thức giúp con người tiếp cận với kiến thức nhanh hơn cũng như chính xác hơn ,tiết kiệm thời gian .

Cơ sở dữ liệu và ngành cơ sở dữ liệu đã phát triển trong quá trình phát triển của 1 số chức năng như thu thập và phát triển dữ liệu .Thực chất của sự phát triển công nghệ thông tin là sự tinh chỉnh ,phân tích dữ liệu để đưa ra kết quả . Công nghệ thông tin gắn liền mật thiết với khai thác dữ liệu .

2. thực chất của viêc khai thác dữ liệu là biến dữ liệu thành tri thức

Quá trình biến đổi dữ liệu

-Loại bỏ dữ liệu dư thừa

-kết hợp nhiều dữ liệu từ nhiều nguồn khác để có được dữ liệu chính xác hơn

-lựa chọn dữ liệu cần biến đổi

-chuyển đổi dữ liệu để thuận tiện hơn trong quá trình khai thác

-trích xuất dữ liệu và đánh giá chất lượng dữ liệu

-khai thác dữ liệu

Khai thác dữ liệu là quá trình phát hiện ra các mẫu thú vị và biết từ nhiều lượng dữ liệu. Các nguồn dữ liệu có thể bao gồm cơ sở dữ liệu, kho dữ liệu, Web, các kho thông tin khác hoặc dữ liệu được truyền trực tiếp vào hệ thống một cách động.

3.khai thác dữ liệu có thể áp dụng được cho nhiều loại dữ liệu sao cho dữ liệu đáp ứng được các yêu cầu chúng ta đặt ra .

Các dạng dữ liệu cơ bản thường gặp nhử cơ sở dữ liệu , dữ liệu kho ,dữ liệu giao dịch

Cơ sở dữ liệu là hệ thống các dữ liệu có liên quan với nhau . Mô hình erd là mô hình dữ liệu đại diện cho mối liên quan của các dữ liệu với nhau

Vd : học sinh và lớp học có mối liên hệ là mã lớp

Dữ liệu giao dich giúp xem thông tin giao dịch dự đoán rui ro khi giao dịch

Vd:giao dịch của tháng 1 trong thẻ ngân hàng ,chi tiết tùng giao dịch

Kho dữ liệu được xây dựng thông qua quá trình xử lý làm sạch dữ liệu, tích hợp dữ liệu, chuyển đổi dữ liệu, tải dữ liệu và làm mới dữ liệu chu kỳ

Có một số chức năng khai thác dữ liệu. Chúng bao gồm phân biệt đặc điểm và phân biệt đối xử ,việc khai thác các mẫu, liên kết và tương quan thường xuyên ,phân loại và hồi quy , phân cụm anal-ysis ,và phân tích ngoại lệ. Các chức năng khai thác dữ liệu được sử dụng để chỉ định các loại mẫu được tìm thấy trong các tác vụ khai thác dữ liệu. Nói chung, các nhiệm vụ như vậy có thể được phân thành hai loại: mang tính mô tả và gây nghi ngờ. Các tác vụ min-ing mô tả đặc trưng cho các thuộc tính của dữ liệu trong tập dữ liệu đích.

Hệ thống cơ sở dữ liệu tập trung vào việc tạo, duy trì và sử dụng cơ sở dữ liệu cho các tổ chức và người dùng cuối. Đặc biệt, các nhà nghiên cứu hệ thống cơ sở dữ liệu đã thiết lập các nguyên tắc được công nhận cao trong mô hình dữ liệu, ngôn ngữ truy vấn, xử lý truy vấn và phương pháp tối ưu hóa, lưu trữ dữ liệu cũng như các phương pháp lập chỉ mục và truy cập. Hệ thống cơ sở dữ liệu thường được biết đến với khả năng mở rộng cao trong việc xử lý các tập dữ liệu rất lớn, có cấu trúc tương đối. Do đó, khai thác dữ liệu có thể tận dụng tốt các công nghệ cơ sở dữ liệu có thể mở rộng để đạt được hiệu quả cao và khả năng mở rộng trên các tập dữ liệu lớn.