DATA MINING AND WAREHOUSE

PRACTICAL 2

Parth Parekh(60004200006)

Prateek Ranka (60004200007)

Sarthak Bhan (60004200005)

Hetansh Shah (60004200024)

Aarav Sharma (60004200003)

Aim

Build Data Warehouse/Data Mart for a given problem state

Theory

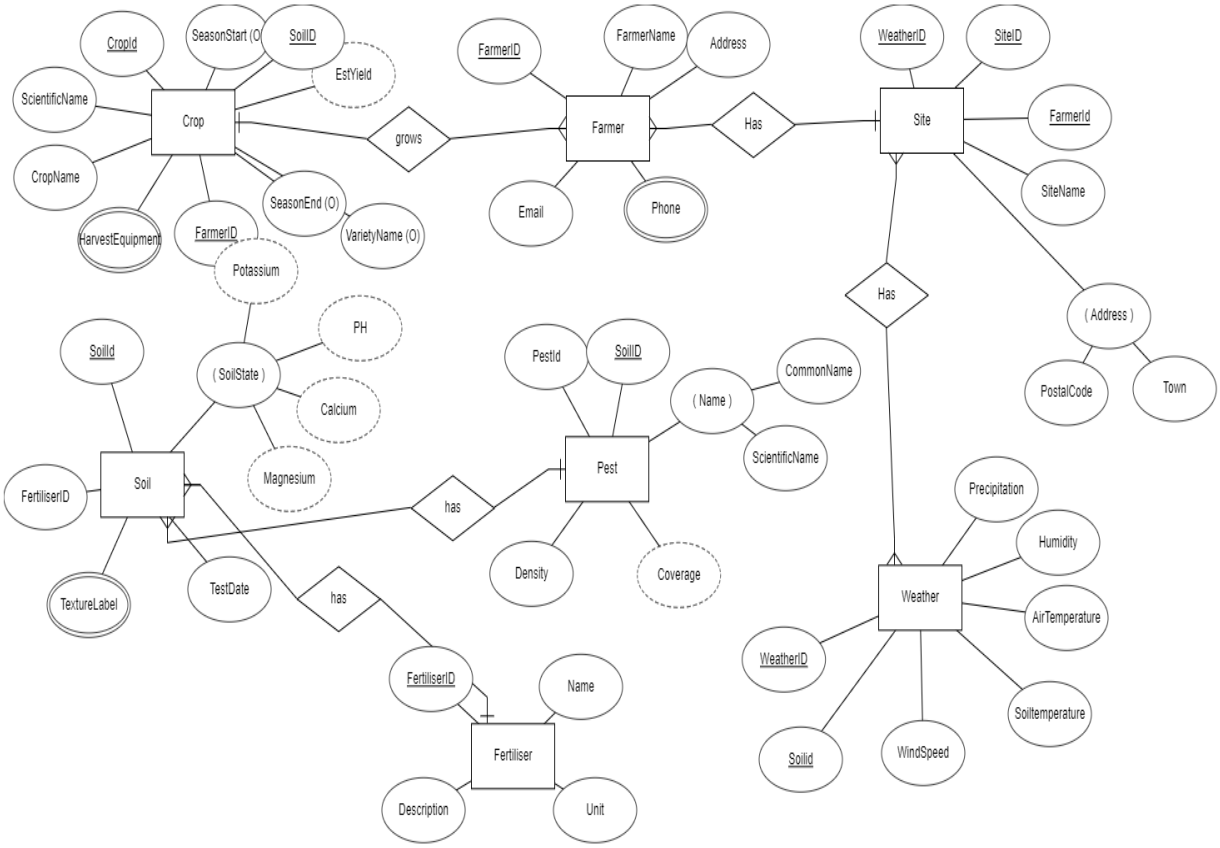
Data Warehouse: A data warehouse is a large collection of business data used to help an organization make decisions. The concept of the data warehouse has existed since the 1980s, when it was developed to help transition data from merely powering operations to fueling decision support systems that reveal business intelligence.A Data Warehousing (DW) is a process for collecting and managing data from varied sources to provide meaningful business insights. A data warehouse is typically used to connect and analyze business data from heterogeneous sources. The data warehouse is the core of the BI system which is built for data analysis and reporting.It is a blend of technologies and components which aids the strategic use of data. It is electronic storage of a large amount of information by a business which is designed for query and analysis instead of transaction processing. It is a process of transforming data into information and making it available to users in a timely manner to make a difference.

Star Schema: Each dimension is represented with only one-dimensional table. This table contains a set of attributes. There is a fact table at the center, It contains keys to other dimensions.

Snowflake Schema: Some dimension tables normalized. The normalization splits up the data into additional tables. Due to normalization, the snowflake schema, the redundancy is reduced and therefore, it becomes easy to maintain and save storage space.

Information Package: An Information Package is a conceptual container of two types of information called Content Information and Preservation Description Information (PDI). The Content Information and PDI are viewed as being encapsulated and identifiable by the Packaging Information. The resulting package is viewed as being discoverable by virtue of the Descriptive Information

E-R DIAGRAM:


CropId
SeasonStart (O
SoilID
EstYield
ScientificName
Crop
grows
CropName
HarvestEquipment
FarmerID
SeasonEnd (O)
VarietyName (O)
Potassium
FarmerID
FarmerName
Address
Farmer
Email
Phone
Has
WeatherID
SiteID
Site
FarmerId
SiteName
Has
(Address)
PostalCode
Town
SoilId
(SoilState)
PH
Calcium
Magnesium
Soil
FertiliserID
TextureLabel
TestDate
has
PestId
SoilID
(Name)
CommonName
ScientificName
Pest
Density
Coverage
has
FertiliserID
Name
Fertiliser
Description
Unit
Precipitation
Humidity
AirTemperature
Soiltemperature
WindSpeed
Soilid
WeatherID
Weather

DESCRIPTION OF SOME DIMENSION TABLES

No.	Dim. tables	Particular attributes
1	Business	BusinessID, Name,Address, Phone, Mobile, Email
2	Crop	CropID, CropName, VarietyID, VarietyName, EstYield, SeasontSart, SeasonEnd, BbchScale, ScientificName, HarvestEquipment, EquipmentWeight
3	CropState	CropStateID, CropID, StageScale, Height, MajorStage, MinStage, MaxStage, Diameter, MinHeight, MaxHeight, CropCoveragePercent
4	Farmer	FarmerID, FarmerName, Address, Phone, Mobile, Email
5	Fertiliser	FertiliserID, Name, Unit, Status, Description, GroupName
6	Field	FieldID, FieldName, SiteID, Reference, Block, Area, AreaUnit, WorkingArea, WorkingAreaUnit, FieldGPS, Notes
7	Inspection	InspectionID, CropID, Description, ProblemType, Severity, ProblemNotes, AreaValue, AreaUnit, Order, Date, Notes, GrowthStage
8	Nutrient	NutrientID, NutrientName, Date, Quantity
9	OperationTime	OperationTimeID, StartDate, EndDate, Season
10	Pest	PestID, CommonName, ScientificName, PestType, Description, Density, MinStage, MaxStage, Coverage, CoverageUnit
11	Plan	PlanID, PlanName, PlanNumber, RegistrationNo, ProductName, ProductRate, Date, WaterVolume
12	Product	ProductID, ProductName, GroupName
13	Site	SiteID, FarmerID, SiteName, Reference, Country, AddressName, AddressTown, PostalCode, GPS, Created, CreatedBy
14	Spray	SprayID, SprayProductName, ProductRate, AppliedArea, AppliedDate, WaterVolume, VolumeUnit, ConfirmDuration, ConfirmWindSPeed, ConfirmDirection, ConfirmTemperature, ConfirmHumidity, ActivityType
15	Soil	SoilID, PH, Phosphorus, Potassium, Magnesium, Calcium, CEC, Silt, Clay, Sand, TextureLabel, TestDate
16	Supplier	SupplierID, SupplierName, SupplierContactName, Address, ContactPhone, ContactMobile, ContactEmail
17	Task	TaskID, TaskDesc, TaskStatus, TaskDate, TaskInterval, CompletedDate, AppCode
18	Treatment	TreatmentID, TreatmentName, FormType, LotCode, Rate, ApplCode, LevlNo, Type, Description, ApplDesc, TreatmentComment
19	WeatherStation	WeatherStationID, StationName, MeasureDate, AirTemperature, SoilTemperature, StationReadingBatch

Dimensions

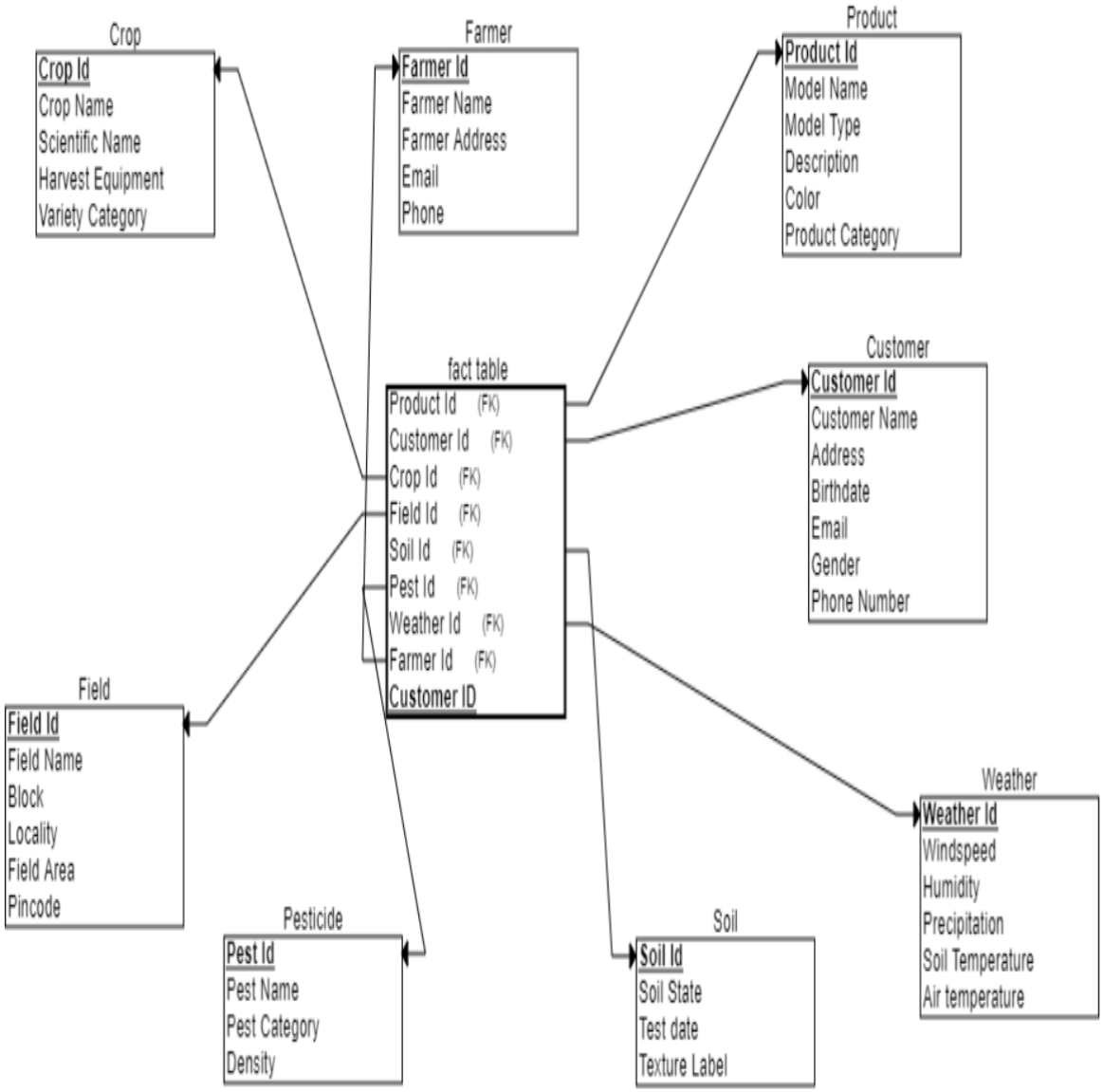
INFORMATION PACKAGE

Categories	Time	Product	Customer	Crop	Field	Pesticide	Soil	Weather	Farmer
	Year	Model Name	Name	Crop Name	Field Name	Pesticide Name	Soil State	Windspeed	Name
	Month	Model Type	Address	Scientific Name	Block	Pesticide Category	Test date	Humidity	Address
	Quarter	Description	Birthdate	Harvest Equipment	Area	Density	Texture label	Precipitation	Email
	Day of Week	Color	Email	Variety Category	Pincode				Phone
	Day of Month	Product category	Gender						
	Product Launch date		Phone Number						
Facts: Actual Sale Price,MSRP,EstYield,EquipmentWeight,Crop height, Crop diameter,Crop Coverage Percent,Field Working Area									

Questions:

Q.1) Which crop is best suited for hot and humid weather
Q.2) Which type of soil is needed for that specific crop?
Q.3) Based on soil and crop data , what is the best pesticide to use and how much?
Q.4) Which farmer is the most efficient and has he highest turnover?
Q.5) Which product and sold by which agent has the most negative reviews/complaints?
Q.6) What equipment is widely used?
Q.7) How is the weather affecting different soils?
Q 8) Which is the best-selling product in Q3?
Q9) Which product has incurred the most amount of losses in Q2 and Q3?

STAR SCHEMA



Conclusion:

In this experiment we explored and built a data mart,schema and information package for Agritech companies