தரவுச் சுரங்கம் – 1.0

"சேமமுற வேண்டுமெனில் தெருவெல்லாம்

தமிழ் முழக்கம் செழிக்கச் செய்வீர்!"

"பிறநாட்டு நல்லறிஞர் சாத்திரங்கள்

தமிழ் மொழியிற் பெயர்த்தல் வேண்டும்"

எனும் பாரதியின் கனவு மெய்ப்பட வேண்டும்.

முட்டையிலிருந்து கோழி வந்ததா அல்லது கோழியிலிருந்து முட்டை வந்ததா எனும் கேள்வியை நாம் கேள்விப்பட்டிருப்போம். இதே போன்ற ஒரு கேள்வி எனக்குள்ளும் எழுவதுண்டு! நாம் கற்றதால் பகிர்கின்றோமா அல்லது பகிர்வதால் கற்கின்றோமா? மேற்கண்ட கேள்வியின் பதிலைத் தேடும் போது, உண்மையில் இரண்டுமே நடக்கின்றது என்பதே நான் உணர்ந்தது. எனவே கற்றதைப் பகிரவும், பகிர்ந்து கொண்டு கற்கவும் முயற்சிக்கும் ஒரு தொடராக இதை எழுத முனைகின்றேன். அறிவார்ந்த சமூகத்தின் ஆதரவும், வழிகாட்டுதலும் எப்போதும் போல வழங்க வேண்டுமாய்க் கேட்டுக் கொள்கின்றேன்.

மனிதன் குகையில் வாழ்ந்த காலத்தே தகவல் தொழில்நுட்பம் பிறந்தாலும், மனிதன் கணினியைக் கண்டுபிடித்ததன் பின்னர் அசுர வளர்ச்சியுற்று இருக்கின்றது என்பதை அனைவரும் அனுபவபூர்வமாக உணர்கின்றோம். இணையம் செயல்பாட்டுக்கு வந்து இணையம் 1.0, 2.0, 3.0, 4.0 என்று புதிய பதிப்புகள் அரங்கேறும் போதெல்லாம் தகவல் தொழில்நுட்பமும் தாவித்தாவிச் சிகரத்தை எட்டியிருக்கின்றது. எனினும் மனிதனின் ஐம்புலன்களில் இரு அவயங்களை மட்டுமே பயன்படுத்த முடிந்துள்ளது! அதிலும் முழுமை பெற்றிருக்கின்றோமா என்றால் இல்லை என்றே பதில் வருகின்றது. இன்னும் தகவல் தொழில்நுட்பம் வளர வேண்டிய எல்லை வானத்தை முட்டி நிற்கின்றது. இருப்பினும் இதற்கே நமக்கு மூச்சு முட்டி நிற்கின்றது.

அதிலும் கடந்த பத்தாண்டுகளில் இத்துறை பெற்றிருக்கும் வளர்ச்சி பல்வேறு துறைகளையும் உலுப்பி எடுத்து வருகின்றது என்றால் அது மிகையாகாது. "எழுமின்! விழிமின்!" என்பது எதற்குப் பொருந்துமோ இல்லையோ இன்றைய தொழில் முனைவோர் அனைவருக்கும் இது பொருந்தும். இன்று எவரும் கணினித் தொழில்நுட்பம் நமது தொழிலை என்ன செய்து விடப் போகின்றது என்று வாளாயிருந்துவிட முடியாது. எத்தொழில் புரிவோரும் தமது தொழில் முன்னால் E என்னும் ஆங்கில எழுத்தினைச் சேர்த்துக் கொள்வதைக் கண்ணுறுகின்றோம். அந்த அளவிற்கு ஒவ்வொருவரின் தொழிலும் தகவல் தொழில்நுட்பம் கோலோச்சியிருக்கின்றது.

தரவுச்சுரங்கம் எனும் இத்தொடரின் மூலம் "தரவு" பற்றி அடியேன் கற்றுணர்ந்ததைப் பகிர்ந்து கொள்ள விழைகின்றேன். முதலில் தரவு என்றால் என்ன என்று பார்க்கலாம். தரவு பற்றிப் பார்க்க வேண்டுமாயின் நாம் தரவின் படிநிலைகளை உணர்ந்திருக்க வேண்டும்.

நடப்பு (Transaction) + பதித்தல் (Recording) -> உண்மை (Fact)

உண்மை (Fact) + தொகுத்தல் (Collection) -> தரவு (Data)

தரவு (Data) + பதப்படுத்தல் (Processing) -> தகவல் (Information)

தகவல் (Information) + ஆய்வு (Analyzing) -> அறிவு (Knowledge)

அறிவு (Knowledge) + அனுபவம் (Experience) -> ஞானம் (Wisdom)

செய்தி அல்லது உண்மை எனும் படிநிலையினின்று ஞானம் எனும் படிநிலைக்கு மனிதர்களை மட்டுமின்றி தகவல் தொழில்நுட்பத்தையும் கணினியையும் கொண்டு போய்ச் சேர்க்க வேண்டும் என்பதே இன்றைய தகவல் தொழில்நுட்ப உலகின் தேவையாய் இருக்கின்றது.

இன்றைய தொழில்முனைவோரின் தலைக்கு மேல் நான்கு கத்திகள் தொங்கிக் கொண்டு இருக்கின்றன. அதை ஆங்கிலத்தில் VUCA என்று அழைக்கின்றார்கள். நிலையின்மை, நிச்சயமின்மை, சிக்கல், குழப்பம் (Volatility, Uncertainty, Complexity, Ambiguity) ஆகியவை தான் அந்த நான்கு கத்திகள், இந்த நான்கு கத்திகளினின்று நம்மைப் பாதுகாத்துக் கொள்ள நாம் மூன்று விதமான உத்திகளையும், கவசங்களையும், ஆயுதங்களையும் கையாள வேண்டியிருக்கின்றது.

EEE, DDD, AAA என்று ஆங்கிலத்தில் இப்போது குறிப்பின்றார்கள்!

EEE- உத்தி – சிக்கனம், திறமை, பயன்பாடு *(Economy, Efficiency, Effectiveness)* – நமது தொழிலைச் சிக்கனமாகவும், திறமையாகவும், மற்றவர்களுக்குப் பயனுள்ளதாகவும் ஆக்கிக் கொள்ள வேண்டியது இன்றைய தொழில் முனைவோருக்கு இன்றியமையாததாகின்றது.

DDD - கவசம் – தரவு சார்ந்த முடிவெடுத்தல் *(Data Driven Decision making)*

AAA - ஆயுதங்கள் – பகுப்பாய்வு, செயற்கை நுண்ணறிவு, தானியக்கம் *(Analytics, Artificial Intelligence, Automation)* ஆகியவை நாம் அறிந்துணர்ந்து பயன்படுத்தி முன்னேறிச் சென்று வெல்ல ஆயுதங்களாகப் பயன்படுகின்றன.

தரவுச் சுரங்கம் – 2.0

தரவு (Data) என்ற சொல்லின் வேர்ச்சொல்லைத் தேடிச் சென்றால் தரப்படுத்தப்படுவது என்று பொருள் தருகின்றது. அனைத்து செய்திகளையும் சேர்த்தால் அது தரவு என்று கருதப்படமாட்டாது. நாம் எந்தச் சூழலில் பயன்படுத்தவிருக்கின்றோமோ அந்தச் சூழலுக்குத் (Context) தகுந்த அனைத்து விவரங்களையும் சேர்த்தால் மட்டுமே அது தரவு என்று அழைக்கப்படும். (Data – Facts collected with a context) செய்தித்தாளில் அனைத்து செய்திகளும் கலந்து காணப்படுகின்றது. நாம் ஏதேனும் ஒரு சூழலை எடுத்துக் கொண்டு அதற்குத் தொடர்புடைய செய்திகளை மட்டுமே தேர்ந்தெடுக்கும் போது அது தரவு என்றழைக்கப்படும்.

எடுத்துக்காட்டாக உலகக்கோப்பை கால்பந்து எங்கெல்லாம் நடந்தது. எந்த ஆண்டில் நடந்தது. எந்த நாட்டின் அணி வென்றது என்று தேடி எடுத்தால் அது தரவு ஆகும். ஒரு ஆண்டில் ஒரு நிறுவனத்தின் அனைத்து பண வரவு செலவுகளையும் தொகுத்தால் அது தரவு ஆகும்.

எனவே தரவைச் சேகரிக்கும் போது தரவு பயன்படுத்தப்போகும் சூழல் மற்றும் தரவுக்குத் தேவையான விவரங்கள் ஆகிய இரண்டு முக்கியமான கருத்துகளைக் கவனத்தில் கொள்ள வேண்டும்.

தரவை அதன் மூலம், அமைப்பு, வடிவம், பண்பு என்று பலவகைப்படுத்தலாம்.

மூலம் (Source)

***உட்தரவு**: (Internal Data) தரவு சேகரிப்பவராலேயே உருவாக்கப்படும் தரவு. எடுத்துக்காட்டு ஒரு நிறுவனத்தின் அனைத்து பண வரவு செலவுகளும் அந்த நிறுவனத்தைப் பொருத்தவரை உட்தரவு ஆகும்.*

***வெளித்தரவு**: (External Data) தரவை நாம் சேகரிக்காமல் வெளியிலிருந்து கிடைக்கும் தரவு வெளித்தரவு ஆகும். எடுத்துக்காட்டு: பங்குச் சந்தை நிலவரம், தங்கத்தின் விலை நிலவரம்.*

அமைப்பு (Structure)

***கட்டமை தரவு:** (Structured Data) நிரல்கள் – நெடுவரிசை (Columns) மற்றும் நிரைகள் – கிடைவரிசை (Rows) எனத் தகுந்த தலைப்புகளைக் (Headers) கொண்டு முறைப்படுத்தப்பட்ட பட்டியல்களாக (Tables) அடுக்கி வைக்கப்பட்ட தரவை கட்டமை தரவு என்று அழைக்கின்றோம். எடுத்துக்காட்டு: மாணவர்களின் மதிப்பெண் பட்டியல், விலைப்பட்டியல், வருகைப்பதிவேடு.*

***பகுதி கட்டமை தரவு:** (Semi structured Data) மொத்தத் தரவில் பகுதி கட்டமைவுடனும் பகுதி கட்டமைவில்லாமலும் கலந்து இருக்கும் தரவு வகை பகுதி கட்டமை தரவு என்றழைக்கப்படுகின்றது. எடுத்துக்காட்டு: மின்னஞ்சல், XML கோப்பு, JSON கோப்பு ஆகியவற்றைக் குறிப்பிடலாம். மின்னஞ்சல் அனுப்புநர், பெறுநர், தலைப்பு ஆகியவை கட்டமைவுடன் இருந்தாலும் உள்ளிருக்கும் செய்தி கட்டமைவில்லாமல் இருக்கின்றது.*

***கட்டமைவில்லாத் தரவு:** (Unstructured Data) நமக்குத் தேவையான செய்திகளைக் கொண்டிருந்தாலும், எந்த வித ஒழுங்கமைவும் இல்லாமல் கிடைக்கும் தரவு இவ்வாறு அழைக்கப்படுகின்றது. எடுத்துக்காட்டு: படங்கள், ஒலி, அசைபடங்கள் ஆகியவற்றைக் குறிப்பிடலாம்.*

வடிவம்: (Form)

***எழுத்து:** (Text) அனைத்து வகையான தரவையும் எழுத்து என்ற வடிவத்தில் சுருக்கி விடலாம்! அதிலும் கணினியில் எந்த வகை எழுத்தையும் இரும எண்களாக (Binary digits – bits/Bytes) மாற்றி 0 மற்றும் 1 என்றே குறிப்பிடமுடியும்.*

***எண்:** (Numeric) எழுத்து கொண்டு அனைத்து தரவையும் குறிப்பிடலாம் என்றாலும் எண்கள் எழுத்துக்கும் முன்பே தோன்றியவையாதலாலும், அதன் பயன் இன்றியமையாதிருப்பதாலும் எழுத்துக்குள் அடங்கினாலும் எண்களின் பெருமை அளவிடற்கரியது.*

எனவே வள்ளுவரும்,

“எண்ணென்ப ஏனை எழுத்தென்ப இவ்விரண்டும்

கண்ணென்ப வாழும் உயிர்க்கு”, என்கின்றார்,

ஔவையும் “எண்ணும் எழுத்தும் கண்ணெனத் தகும்” என்று குறிப்பிடுகின்றாள்.

எண்ணத்தில் விளைந்ததால் அதை எண் என்கின்றோம்! உலகத்திலிருக்கும் எந்தப் பொருளையும் நாம் எண்ணாலேயே அளக்க விழைகின்றோம். ஒரு பொருளை மற்றொரு பொருளோடு ஒப்பிட்டுப்

பார்க்கவும் நமக்கு எண்கள் தேவைப்படுகின்றன. "கண்ணளக்காததையா கையளக்கப் போகின்றது?" என்றொரு பழமொழி கூட உண்டு.

இந்த எண்களும் மூன்று வகைப்படும்.

பெயரெண்: (Nominal Numbers) : எந்த ஒரு கணக்கீடுக்கும் உட்படுத்தாமல் ஒரு பொருளையோ அல்லது மனிதரையோ மற்றவர்களிடமிருந்து வேறுபடுத்திப் பார்ப்பதற்கு எண்களைப் பயன்படுத்தினால் அவற்றைப் பெயரெண் எனலாம். எடுத்துக்காட்டு: அலைபேசி எண்கள், ஆதார் எண் ஆகியவை.

வரிசையெண்: (Ordinal Numbers – I, II, III,…) ஒன்றன்பின் ஒன்றாக வரிசைக்கிரமமாக அடுக்கி வைத்து அவற்றின் வரிசையைக் (Order) குறிப்பிடும் எண்கள் வரிசையெண்கள் என்று அழைக்கப்படுகின்றன. எடுத்துக்காட்டாக, முதல் மாணவன், இரண்டாம் மாணவன் என்று படிவரிசையில் அடுக்குவதைக் குறிப்பிடலாம்.

செவ்வெண்: (Cardinal Numbers – 1,2,3,…) ஒன்று, இரண்டு, மூன்று என்று எண்ணக் கூடிய அனைத்தையும் இவ்வாறு குறிப்பிடலாம். உன்னிடம் எத்தனை பலாச்சுளைகள் என்று கேட்டால் நம்மால் எண்ணிச் சொல்ல முடிகின்றது அல்லவா? இந்த எண்களே முதலில் தோன்றியமையால் இவற்றை முதலெண் என்றும் அழைப்பதுண்டு. இவ்வுலகில் நாம் காணும் அனைத்தையும் இவ்வாறே எண்ணப்பட்டு செய்திகள் சேகரிக்கப்படுகின்றன.

இடவெளித்தரவு: (Spatial Data) இடங்களைக் குறிப்பிடப் பயன்படுத்தப்படும் தரவை இவ்வாறு அழைக்கின்றார்கள். அவை படக்காட்சிகள், செயற்கைக்கோள் படங்கள் என்று எவ்வாறாக வேண்டுமானாலும் இருக்கலாம். இவற்றையும் திசைச்சாரி (Vector) மற்றும் படப்புள்ளிகளின் தொகுப்பான வரிக்கோலம் (Raster) என்று பிரிக்கலாம்.

அருவத்தரவு: (Abstract Data) எந்த உருவமும் இன்றி நமது கருத்தில் மட்டும் உலா வரும் தரவை அருவத்தரவு எனலாம்! பொருள் நோக்கு நிரலகத்தில் (Object Oriented Programming) இனக்குழு (Class) உருவாக்கத்தில் அருவப்படுத்தல் (Abstraction) என்பதுண்டு. இதைக் கணித்தமிழுக்கு அருந்தொண்டாற்றி கணிப்பொறியியல் கலைச்சொல் திரட்டு உருவாக்கிய மு. சிவலிங்கம் ஐயா (http://www.sivalingam.in) குறிப்பிடுவது போல் புறாவைக் காண முடியும். மயிலைக் காண முடியும். கழுகைக் காண முடியும். எங்காவது பறவையைக் காண முடியுமா? பறவை என்பது கருத்தியலில் மட்டுமே முடியும். எங்குமே எவருமே பறவை என்ற ஒன்றைக் காண முடியாது அல்லவா? ஆக, தேவையானவற்றைப் பொருத்திப் பார்க்கும் வகையில் நாம் உருவாக்கும் தரவமைப்பை அருவத்தரவு என்றழைப்பார்கள். அதையும் பட்டியல் (List), வரிசை (Queue) மற்றும் அடுக்கு (Stack) என்று வகைப்படுத்தலாம்.

உயர்தரவு: (Meta Data) தரவைப் பற்றிய தரவு ஒன்று இருக்குமானால் அதுதான் உயர்தரவு என்று அழைக்கப்படுகின்றது. ஆம். தரவின் அமைப்பு, வகை, உறுப்புகள், இடை உறவு (Relationships), முதன்மைத் திறவி (Primary Key), அயல்திறவி (Foreign Key), வரிசையாக்கத் திறவி (Sort Key), சுட்டிகைத் திறவி (Index Key) ஆகியவற்றை உயர்தரவு என்றழைக்கின்றார்கள். ஒரு புகைப்படம் என்பதை தரவு என்றால் அதை எடுத்தவர், என்று எடுக்கப்பட்டது, எந்தப் புகைப்படக்கருவி கொண்டு எடுக்கப்பட்டது ஆகியவை உயர்தரவு ஆகும்.

பண்பு:

தரவின் பண்பின் அடிப்படையில் தரவை இரண்டு வகைகளாகப் பிரிக்கலாம்.

தரவகைத் தரவு: (Qualitative Data) எண்களால் குறிப்பிடாமல், தோராயமாகவும், குத்துமதிப்பாகவும் குறிப்பிடப்படுவதாகவும், பண்புகளைக் குறிப்பிடுவதாகவும் இருக்கும் தரவு தரவகைத் தரவு எனப்படும். எண்களால் குறிப்பிடாமல் இருப்பதால் இவற்றைக் குழுவகைத் தரவு (Category/Group Data) என்றும் அழைப்பர். எடுத்துக்காட்டாக இலையின் நிறம் பச்சை (எந்த அளவு பச்சை?!)

எண்ணளவுத் தரவு: (Quantitative Data) எண்களால் குறிப்பிடக்கூடிய தரவு இவ்வாறு அழைக்கப்படுகின்றது. அவற்றைப் பற்றி ஏற்கனவே மேலே கண்டோம்.

தரவின் பல வகைகளைப் பற்றித் தெரிந்து கொண்டோம். இன்னும் தரவு எடுக்கும் பல்வேறு அவதாரங்களையும் அடுத்ததாகக் காண்போம்.

தரவுச் சுரங்கம் – 3.0

ஒன்றுக்கொன்று தொடர்புடைய செய்திகளின் அல்லது நிகழ்வுகளின் தொகுப்பு மட்டுமே தரவு என்று முடித்து விடமுடியாது. தரவின் பல்பரிமாணங்களைத் தொடர்ந்து காண்போம்.

வணிகம் தொடர்பான நடவடிக்கைகள் என்றாலும் சரி, ஒரு தனிமனிதரின் சமர்த்துக் கடிகாரத்தில் சேர்த்து வைக்கப்படும் அவரது உடற்பயிற்சிக்கான தரவு என்றாலும் சரி ஏதோ ஒரு பட்டியலில் பின்னாளில் தேவைப்படும் என்று குறித்து வைக்கப்படுகின்றது. அவ்வாறு குறித்து வைக்கப்படும் செய்திகள் மூன்று வகைப்படும்.

1. ***நடவடிக்கைகள்*** *(Transactional Facts)*

தினந்தோறும் நடந்து வரும் செயல்பாடுகளைக் கண்ணுற்று அவற்றில் தேவையானவை யாதென்று அறிந்து அவற்றைத் தொடர்ந்து குறித்து வைத்து வருவது நடவடிக்கைகள் ஆகும். பெரும்பாலும் அவை யார்? என்ன? எங்கே? எப்போது எவ்வளவு? (Who, What, Where, When, How much or How many) ஆகிய ஐந்து கேள்விகளுக்கான பதில்களாகத் தான் இருக்கும்.

வணிக நிறுவனங்களின் குறிப்பேடுகளை இந்த வகை தரவுக்கு எடுத்துக்காட்டாகக் குறிப்பிடலாம். இக்குறிப்பேடுகள் பின்னர் தேவைக்கேற்ப பேரேடுகளில் பதியப்பட்டு பின்னர் அவற்றின் மொத்தம் மற்றும் மீதம் ஆகியவை கணக்கிடப்பட்டு அறிக்கைகளாக உருப்பெற்றுப் பின்னர் நாம் வணிகம் சார்ந்து எதிர்காலத்தில் முடிவெடுக்க வசதியாகப் பின்பற்றப்படுகின்றன.

கணினி யுகம் வந்த பின்பு பெரும்பாலும் அனைத்து பெரும் மற்றும் நடுத்தர வணிக நிறுவனங்களின் மொத்த நடவடிக்கைக் குறிப்புகளும் கடந்த பதினைந்து ஆண்டுகளுக்கும் மேல் பலருக்கும் தெரியாமல் தேக்கி வைக்கப்பட்டு இருக்கின்றன. அவற்றில் எத்தனை மாணிக்கங்களும் வைரங்களும் பதுங்கி இருக்கின்றனவோ!?

2. ***நேரப் படப்பதிகை****: (Periodic Snapshots)*

நடவடிக்கைகள் அனைத்தையும் குறிப்பெடுத்து வைத்து விட்டோமே? இன்னும் அப்படி என்ன சேர்த்து வைக்க இருக்கின்றது என்று நினைத்து விடவேண்டாம். சிறந்த முறையில் ஒரு வணிக நிறுவனம் நடத்துவதற்கு இவை மட்டுமே அனைத்து தகவல்களையும் தந்து விட முடியாது. அந்த நிகழ்வுகளால் ஏற்பட்ட விளைவுகளை அவ்வப்போது படம் பிடிப்பது போல் குறித்து வைத்திருப்பதும் தேவையாகின்றது.

எடுத்துக்காட்டாக ஒரு ஐந்து நட்சத்திர தங்கும் விடுதியில் தினந் தோறும் எத்தனை அறைகள் நிரம்பி இருக்கின்றன எத்தனை அறைகள் காலியாக இருந்தன என்ற விபரம் தொடர்ந்து பதியப்பட்டு வருவதைக் குறிப்பிடலாம். தினந்தோறும் உங்கள் பையில் பணம் எவ்வளவு இருந்தது என்று தொடர்ந்து கவனித்து வந்தாலே உங்கள் பணப்புழக்கம் பற்றி அறிந்து கொள்ள முடியும்.

கூகுள் வரைபடத்தில் 2000ம் ஆண்டு ஒரு நிலப்பரப்பு எவ்வாறு இருந்தது என்றும் 2020ம் ஆண்டு அதே நிலப்பரப்பு எவ்வாறு இருந்தது என்றும் குறித்து வைத்திருந்தால் அதனால் அந்தப் பகுதியில் என்னென்ன மாற்றங்கள் நிகழ்ந்திருக்கின்றது என்று அனுமானிக்க முடியும். அதுமட்டுமின்றி இன்னும் இருபது ஆண்டுகள் கழித்து எவ்வாறு இருக்கும் என்றும் கண்டறிய முடியும் அல்லவா?

3. ***கூட்டுப் படப்பதிகை*** *(Cumulative/Accumulating Snapshots)*

ஒரே நிகழ்வு அல்லது நடவடிக்கையின் சங்கிலித் தொடரைக் காலக்கிரமத்துடன் அவற்றின் முந்தைய குறிப்புகளுடனே சேர்த்து வைத்து வருவதை கூட்டுப்பதிவு எனலாம். இவை நடவடிக்கைப் பதிவுகளாக ஒருபுறம் இருந்தாலும் நமது வசதிக்காகவும், மேற்கொண்டு நடவடிக்கை மேற்கொள்ள ஏதுவாகவும் ஒரே இடத்தில் சேர்த்துப் பதிந்து கொள்ளலாம்.

எடுத்துக்காட்டாக, அமேசான் நிறுவனத்தின் தகவல் முறையைக் கண்ணுறலாம். ஒருவர் இந்தப் பொருள் பிடிக்கின்றது என்று குறித்து வைத்துக் கொண்டால், அவருக்கு அதை அவ்வப்போது ஞாபகப்படுத்தி அதை வாங்கும் வரை தொடர்ந்து பின்னர் அவர் வாங்க முடிவெடுத்து அவரிடம் பணத்தை வாங்கி சரக்கைக் கொண்டு போய் சேர்ப்பது என்று அனைத்தும் நடவடிக்கைகள் தான் என்றாலும் ஏதோ ஒரு முடிச்சில் அனைத்தும் ஒன்றாகச் சேர்த்து வைத்துப் பார்த்தால் மட்டுமே சிறந்த முடிவுகள் எடுக்க முடியும்.

அவ்வளவு பெரிய நிறுவனத்தில் தினம் நடைபெறும் கோடிக்கணக்கான நடவடிக்கைகளை இவ்வாறு தொகுப்பது என்று மனிதர்களால் ஆகாது என்றாலும் கணினிகளால் அது முடிகின்றது.

அடுத்ததாக மேகக் கணிமையில் தரவு எவ்வாறு சேமிக்கப்படுகின்றது என்பதையும் பெருந்தரவு குறித்தும் காணலாம்.

கணினி உலகின் வாயிலாக விஸ்வரூபம் எடுத்து நிற்கும் தரவின் பல பரிமாண அவதாரங்களைக் கண்ட நாம் இப்போது தரவின் பரிணாம வளர்ச்சியையும் காண்போம்.

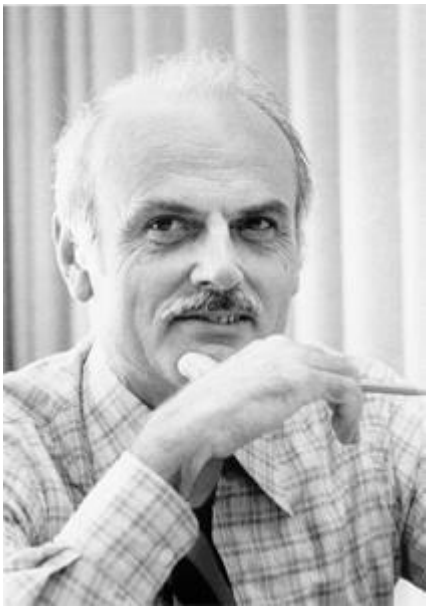
கல்வெட்டிலும், பட்டயங்களிலும், ஓலைச்சுவடிகளிலும் இருந்த தகவல்கள் இயேசு கிறிஸ்து பிறப்புக்கு 200 ஆண்டுகளுக்கு முன்பு காகிதம் கண்டுபிடிக்கப்பட்டவுடன் காகிதத்திற்கு மாறியது. அப்போதிருந்து கி.பி. 1725ம் ஆண்டில் தறி இயந்திரங்களில் அழகான ஆடை வடிவமைப்பைச் சேமித்து வைக்கும் துளை அட்டைகள் வரும் வரை காகிதத்தின் கைகளே ஓங்கியிருந்தன. அதன் பின் 1837ல் கணினி உலகின் தந்தை சார்லஸ் பாப்பேஜ் 1837ல் முதல் கணினியைக் கண்டறியும் போது துளை அட்டைகளே அதிகம் பயன்படுத்தப்பட்டன. 1980 வரையிலும் துளை அட்டைகளின் பயன்பாடு இருந்தது என்பது ஆச்சரியமான தகவல். 1960களில் காந்தத் தட்டுகள் பயன்பாடு ஆரம்பித்தது. காந்தத் தட்டுகளில் வன் தட்டு மென்தட்டு என்று பலவகைகள் பல அளவுகளில் வர ஆரம்பித்தன. 20MB அளவிலான வன்தட்டுகளில் அனைத்து மென்பொருட்களையும் உள்ளடக்கி, பத்தாண்டுகளுக்கான தகவல்களை ஏற்றிய பின்னும் பாதி இடம் மீதி இருந்தது நினைவில் இருக்கின்றது. பின்னர் லேசர் ஒளி அலைகளின் மூலம் எழுதி வைக்கப்படும் முறையில் சிடி, டிவிடி, ப்ளூரே என்று பலவித தொழில்நுட்பங்கள் தகவலை எழுதி வைப்பதற்குப் பயன்பட்டன. பிளாஷ் எனப்படும் அதிவிரைவு நினைவகம் வந்த பின்னர் தற்போது SSD, NAND ஆகிய தொழில்நுட்பங்களைக் கொண்டு அதில் பதியப்பட்டு வருகின்றது.

மெய்நிகர் மற்றும் மேகக் கணிமை (Virtual and Cloud Computing) வந்த பின் இப்போது நமது தகவல் எங்கே பதியப்படுகின்றது எந்த வன்பொருளில் பதியப்படுகின்றது என்பதே தெரியாவிட்டாலும் கூட நம்மால் தகவலைப் பதிந்து கொண்டு எப்போது வேண்டுமானாலும் எடுத்துக் கொள்ள முடிகின்றது.

எத்தகைய தரவாக இருந்தாலும் அது (Binary Digit-Bit) இரும இலக்க எண்ணாக மாற்றப்பட்டு ஏதேனும் ஒரு கோப்பின் பெயரிலேயே கணினியில் சேமிக்கப்படுகின்றது என்பதை நாம் அறிவோம்.

இருப்பினும் மென்பொருளில் தரவுகளை மேலாண்மை செய்யப்படும் வரலாறையும் சற்று சுருக்கமாகக் காண்போம். கணினி கண்டறியப்பட்டு தொழில்முறை பயன்பாட்டுக்கு வந்த புதிதில் தரவு எடுப்பார் கைப்பிள்ளையென கணினி மொழி மற்றும் இயக்கக மென்பொருளின் அடிமையாகவே இருந்தது எனலாம்.

எடுத்துக்காட்டாக கோபால் எனும் மொழி பற்றி கேள்விப்பட்டிருப்பீர்கள். கோபால் மொழியில் எழுதப்பட்ட மென்பொருளைக் கொண்டு உருவாக்கப்பட்ட தரவுகளை கோபால் மொழி கொண்டு மட்டுமே வாசிக்க முடியும். மற்ற மொழிகளைக் கொண்டு அதை வாசிக்க முடியாது. அதே போல் டாஸ் இயங்குதளத்தில் ஒரு மென்பொருள் இருக்குமானால் அதில் இருக்கும் தரவை ஒரு யுனிக்ஸ் இயங்குதளத்தில் இருக்கும் மற்றொரு மென்பொருள் கொண்டு காண முடியாது. ஒரு கருவியிலிருந்து மற்றொரு கருவிக்கு தகவல்களை மாற்றம் செய்வதும் வலைப்பின்னல் இல்லாத காலங்களில் சாத்தியமில்லாமல் இருந்தது.



எனவே முதலில் தனிக் கோப்புகளில் மட்டுமே தரவுகள் சேமிக்கப்பட்டு வந்தன. அவற்றைத் தட்டை கோப்பு முறை (Flat File System) என்றே அழைத்து வந்தனர். அதன்பின்னர் தகவல்களுக்கிடையே இருக்கும் உறவுமுறைகளைக் குறித்து வைக்க ஏதுவாக அடுக்கு முறை தரவு மேலாண்மை (Hierarchial Database Management) சிலகாலம் இருந்தது. அதன்பின்னர் இன்னும் சிலகாலம் வலைமுறை தரவு மேலாண்மையும் பயன்பாட்டில் இருந்தது. இந்தச் சூழ்நிலையில் தான் 1970களில் தரவின் அடிமைச் சங்கிலியைத் தகர்த்தெறிந்து அதன் சுதந்திரத்திற்கு வழிவகுத்த எட்கர் காட் (Edgar F. Codd) உறவுநிலைத் தரவுத் தளமேலாண்மை முறையைக் கண்டறிந்து அறிமுகப்படுத்தினார்.

அவரது வழிகாட்டுதல்களான மிகவும் பெயர்பெற்ற காட் 12 (Codd's 12 rules) விதிகளுக்குட்பட்டு ரேய்மண்ட் பாய்ஸ் (Raymond F. Boyce) உருவாக்கிய SQL எனும் தரவுத்தள வினவல் மொழி தினந்தோறும் மாறிக் கொண்டிருக்கும் கணினி உலகில் ஐம்பதாண்டுகளுக்குப் பின்னர் இன்றளவும் நடைமுறையில் இருக்கின்றது என்றால் அதன் முக்கியத்துவத்தை நாம் உணர்ந்து கொள்ளலாம்.

வியாபாரக் கணக்குகளைக் குறித்து வைப்பதற்கும் முறைசார் தரவுகளைக் குறித்து வைப்பதற்கும் இந்த உறவுநிலைத் தரவுத்தள மேலாண்மை முறை மிகவும் கனகச்சிதமாகப் பொருந்துவதால் என்ன தான் பொருள் நோக்கு தரவுத்தள மேலாண்மை *(Object Oriented Database)* மற்றும் மட்டற்ற மட்டுமல்லா *SQL (NoSQL)* என்று பலவித தொழில்நுட்பங்கள் வந்தாலும் இன்னும் இந்த உறவுநிலைத்தரவுத் தள மேலாண்மையின் இடத்தைப் பிடிக்க யாரும் வரவில்லை என்றே சொல்லலாம்!

இன்றைய பெருந்தரவு காலத்திலும் மேகக்கணிமைத் தொழில்நுட்பத்திலும் கூட இந்தத் தரவு மேலாண்மையின் தாக்கத்தை நம்மால் காணமுடியும். தரவுத் தொழில்நுட்பத்தின் வரலாற்றைச் சுருக்கமாகக் கண்டோம். இனி தரவுத்தள மேலாண்மையின் அடிப்படையை அடுத்த பகுதியில் சற்று விளக்கமாகக் காண்போம்.

தரவுச் சுரங்கம் – 5

*தேவைக்குத் தகுந்தவாறு ஒன்றுக்கொன்று தொடர்புடைய உண்மைச் செய்திகளின் தரப்படுத்தப்பட்ட தொகுப்பே தரவு என்று முதல் பகுதியிலேயே கண்டோம். **ஒன்றுக்கொன்று** என்று போகின்ற போக்கில் கூறிவிட்டாலும், எந்த ஒன்றுடன் மற்றொன்று தொடர்புடையது? அது எது குறித்தது? எனும் கேள்வி எழாமலில்லை!*

தரவுத் தொழில்நுட்பத்தில் இன்றளவும் கோலோச்சிக் கொண்டிருக்கும் உறவுநிலைத் தரவுத் தள மேலாண்மை (RDBMS) மட்டுமின்றி அதற்கு முந்தைய காலந் தொட்டே செய்திகளை எதன் அடிப்படையில் ஒன்றிணைத்து அல்லது கட்டமைக்க வேண்டும் என்று அறிந்திருந்தனர் கணினி வல்லுநர்கள். அந்த அடிப்படையான அமைப்பின் பெயர் தான் உருபொருளும் உறவுநிலையும்.

***உருபொருள்:** **(Entity)** தனித்துவமானதாவும் மற்றவைகளிடமிருந்து வேறுபடுத்திக் காட்டுமளவுக்கு பண்புகள்/தன்மைகள் கொண்டதாகவும் தனது இருப்பை உணர்த்துவதாகவும் இருக்கும் பொருளையே உருபொருள் என்று அழைக்கின்றோம். எடுத்துக்காட்டாக ஒரு பேனாவிலிருந்து இந்தப் பேரண்டம் வரை எதையெல்லாம் தனித்துவமாகக் கண்டு உணர முடிகின்றதோ அதையெல்லாம் உருபொருள் என்று அழைக்கின்றோம். பேனாவுடன் சேர்ந்திருக்கும் போது பேனா என்று அறியப்பட்டாலும், அதிலிருந்து கழற்றி விட்டால் பேனாவின் மூடி கூட ஒரு தனிப்பட்ட உருபொருள் என்று ஆகிவிடக் கூடும். தரவுகளைப் பொருத்தவரை உயிருள்ள, உயிரற்ற என்ற பாகுபாடுகள் கிடையாது. எனவே உயிருள்ள நீங்களும், நானும் கூட தரவு மேலாண்மையில் உருபொருள் என்றே அழைக்கப்படுவோம்!*

***பண்புகள்:** **(Attributes)** மேற்கண்ட உருபொருட்களுக்கே உரித்தான பண்புண்மைகளை உருபொருளின் பண்புகள் என்று அழைக்கின்றோம். இப்பண்புகள் தான் நாம் உருபொருளுடன் கோந்து போல் இணைத்து சேர்த்துக் கட்டி வைக்கும் உருபொருளைப் பற்றிய செய்திகள் என்று உணர்ந்து கொண்டிருப்பீர்கள் என்று நினைக்கின்றேன். ஆங்கிலத்தில் சொல்வதானால், (Everything uniquely identifiable is called entity. Every entity has its own attributes. These attributes are nothing but the facts about the entity itself.)*

ஆக, ஒவ்வொரு உருபொருளும் அதனதன் பண்புகளைச் சுற்றிலும் வைத்துக் கொண்டு தாமே நடுநாயகமாக விளங்குகின்றது.

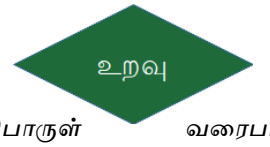
இவ்வாறு உருபொருட்களையும் அவற்றின் பண்புகளையும் கொண்டு தரவு மென்பொருள் நிர்வாகிகள் மற்றும் படைப்பாளிகள் யாவருக்கும் புரியும் வண்ணம் ஒரு வரைபடத்தைத் தயாரிக்கின்றார்கள். அதற்கு உருபொருள் உறவுநிலை வரைபடம் (உ.உ வரைபடம்) (Entity Relationship Diagram – Shortly ER Diagram) என்று பெயர்.

சில அடிப்படை வடிவங்களை மட்டுமே பயன்படுத்தி இந்த உ.உ வரைபடங்கள் தயாரிக்கப்படுகின்றன.

செவ்வகம் – உருபொருள்

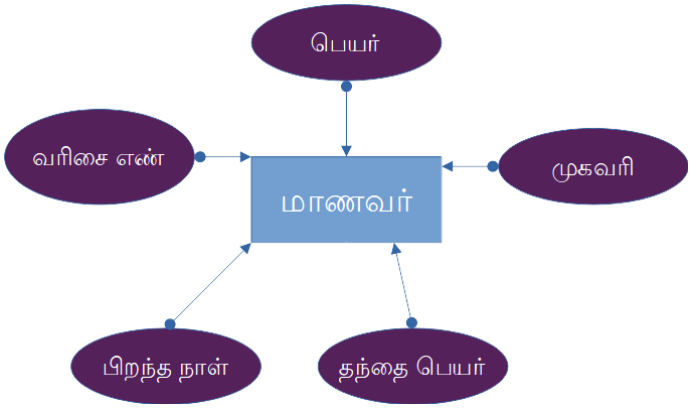
நீள்வட்டம் – பண்புகள்

வைரம் – உறவு



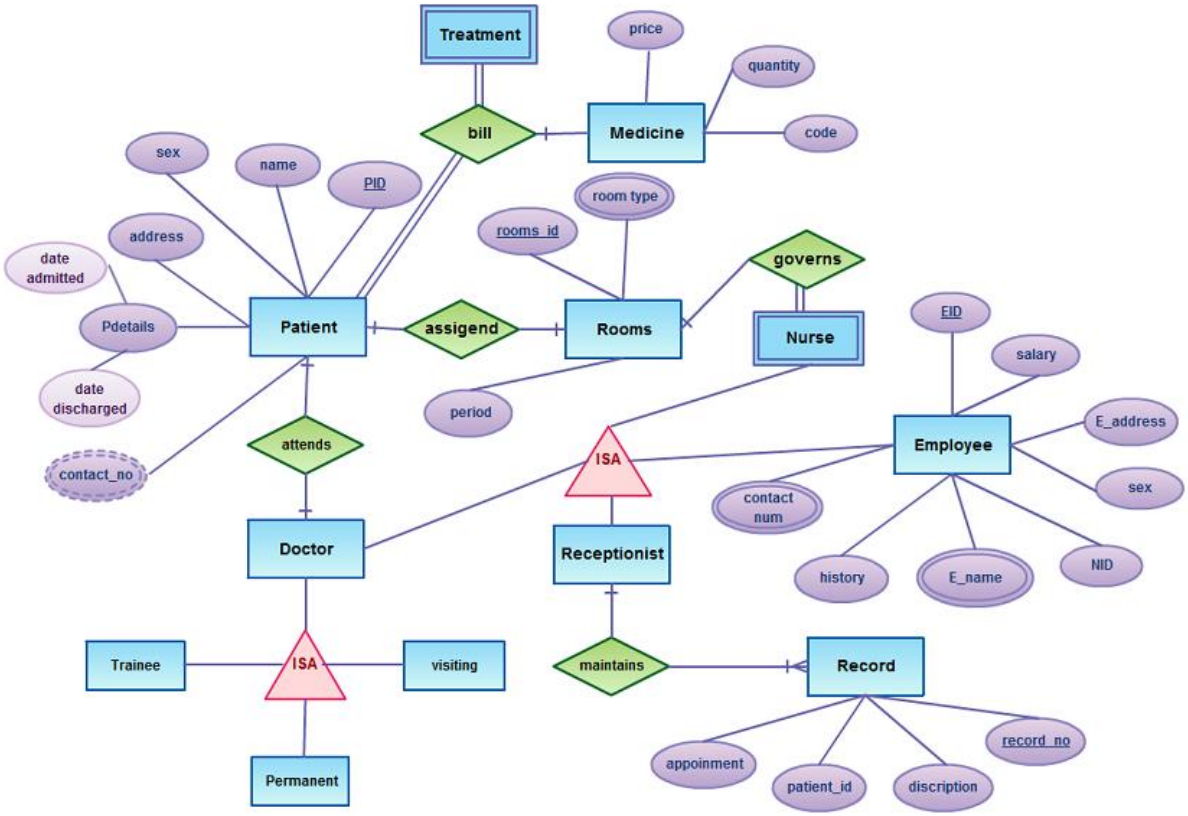
கீழ்க்காணும் உருபொருள் வரைபடத்தைப் பார்த்தவுடன் இது ஒரு மாணவரின் தரவுப் பட்டியலில் இருக்கும் அவரது தனித்தன்மையான பண்புகள் என்று புரிகின்றது அல்லவா?

எந்த ஒரு நிறுவனத்தின் தரவுகளையும் நாம் பகுப்பாய்வு செய்ய முற்படும் போது அந்த நிறுவனத்தின் ஒட்டு மொத்த உருபொருள் உறவுநிலை வரைபடத்தைக் கண்டுணர்தல் அவசியமாகும். அவ்வாறு ஒரு வரைபடம் இதுவரை இல்லையாயின் நாமே உருவாக்குதல் நலம் பயக்கும்.

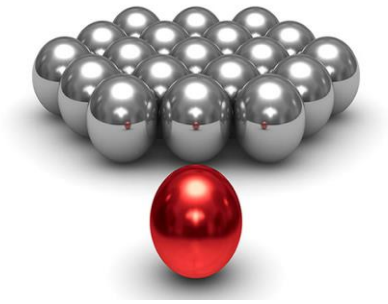


நாம் அடுத்துக் காண்பது ஒரு மருத்துவமனையின் மாதிரி உ.உ. வரைபடம் ஆகும். *(நன்றி: https://practice.geeksforgeeks.org/problems/er-diagram-of-hospital-management-system)* இதன் மூலம் எவ்வாறு மருத்துவர், நோயாளி, செவிலியர், அறைகள், மருந்து, மருத்துவம் பார்த்த சீட்டுகள், பட்டிகள், வேலையாட்கள், ஆவணங்கள் என்று மருத்துவமனை தொடர்பான அனைத்து உருபொருட்களும், அவற்றின் பண்புகளும், எவ்வாறு ஒரு உருபொருள் மற்றொரு உருபொருளுடன் உறவு கொள்கின்றது என்றும் அதன் மூலம் எவ்வாறு அனைத்து உருபொருட்களும் ஒட்டு மொத்த நிறுவனத்தில் இணைந்து செயலாற்றுகின்றன என்றும் உணர்ந்து கொள்ள முடியும்.

E-R Diagram for Hospital Management System



உறவுமுறை தகவல் தரவுத்தள மேலாண்மையில் மட்டுமின்றிப் பொதுவாகவே தரவு மேலாண்மையில் தரவுகளின் இடையே இருக்கும் உறவைப் பற்றியே முந்தைய பகுதியிலிருந்து தொடர்ந்து பார்த்துக் கொண்டிருக்கின்றோம். ஒவ்வொரு உருபொருளும் அதனதன் பண்புகளைத் தன்னகத்தே கொண்டு மற்றொரு உருபொருளுடன் இருக்கும் தொடர்பை எவ்வாறு ஏற்படுத்திக் கொள்கின்றது என்பதில் மொத்தம் நான்கு வகையான உறவுமுறைகள் இருக்கின்றன. அவற்றுள் இரண்டு அடிப்படையான உறவுமுறைகள் ஆகும். மற்ற இரண்டு உறவுமுறைகள் மேற்கண்ட அடிப்படை உறவுமுறைகளைக் கொண்டே அமைகின்றன.

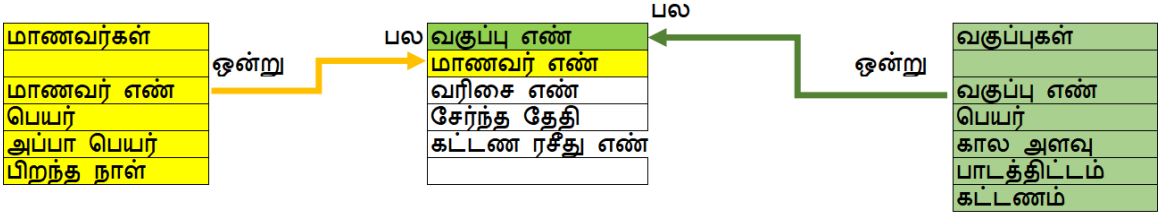


உருபொருட்களுக்கிடையேயான உறவு அவற்றின் எண்ணிக்கையைப் பொருத்தே அமைகின்றது. ஒரு உருபொருளுடன் பல உருபொருட்கள் ஒரே நேரத்தில் இணைந்திருக்க வேண்டியிருந்தால் அதனை **ஒன்றுடன் – பல உறவுமுறை** *(One to Many Relationship)* என்று அழைக்கின்றோம். எடுத்துக்காட்டாக, ஒரே நுழைவுச்சீட்டில் பலர் உள்ளே நுழைவதையும், ஒரு குறிப்பிட்ட எண் கொண்ட ரயிலில் (எடுத்துகாட்டாக 12661 – பொதிகை விரைவு வண்டி) பல பெட்டிகள் இருப்பதையும் குறிப்பிடலாம்.

ஒரு உருபொருளுடன் ஒரே ஒரு உருபொருள் மட்டுமே இணைந்திருக்க முடியுமாயின் அதை **ஒன்றுடன் – ஒன்று உறவுமுறை** *(One to One)* என்று அழைக்கின்றோம். எடுத்துக்காட்டாக, ஒரு நாற்காலி உருபொருளில் ஒரு மனிதர் மட்டுமே அமரலாம் என்பதையும், ஒரு பேருந்தை ஒரு ஓட்டுநர் மட்டுமே ஓட்டலாம் என்பதையும் குறிப்பிடலாம்.

ஆக, ஒரு பேருந்தில் ஒரு நடத்துநரும், ஒரு ஓட்டுநரும் இருப்பது ஒன்றுடன் ஒன்று எனவும், அதில் பல பயணிகள் ஒரே நேரத்தில் பயணிப்பது ஒன்றுடன் பல எனவும் கணினித் தரவில் குறிப்பிடப்படும். அதே போல் பல பேருந்துகள் ஒரு வழித்தடத்தில் செல்லுமானால் ஒன்றுடன் பல என்று குறிப்பிடப்படும். ஆக ஒட்டுமொத்த தரவுகளையும் இது போன்று உருபொருட்களாகவும், அவற்றின் பண்புகளாகவும் பிரித்துப் பின்னர் உருபொருட்களுக்கிடையேயான உறவுமுறைகளைக் கொண்டு சேர்த்து ஒரு ஒட்டுமொத்தமான தரவுத்தளமாகக் காண்பது தான் தரவுத்தளமேலாண்மையின் அடிப்படையாகும்.

மேற்கண்ட இரு உறவுமுறைகள் தவிர பலவற்றுடன் – ஒன்று *(Many to one)* , பலவற்றுடன் – பல *(Many to Many)* ஆகிய இரண்டு உறவுமுறைகளும் உண்டு என்றாலும், அவை மேற்கண்ட அடிப்படை உறவுமுறைகளின் சேர்க்கை அல்லது மாறுபட்ட பார்வையே ஆகும். எடுத்துக்காட்டாக, பல மாணவர்கள், பல வகுப்புகளில் சேர்ந்திருக்கின்றார்கள் என்று கூறினாலும், ஒவ்வொரு மாணவரும் பல வகுப்புகளில் சேர்ந்திருக்கின்றார்கள் என்றும், ஒரு வகுப்பில் பல மாணவர்கள் இருக்கின்றார்கள் என்றும் கொள்ளப்பட்டு, இரண்டு ஒன்றுடன் – பல உறவுமுறைகளாகப் பிரிக்கலாம்.



மேற்கண்ட படத்தில் காட்டியிருப்பதைப் போல் இரண்டு ஒன்றுடன் - பல உறவுமுறைகள் இணையும் போது பலவற்றுடன் – பல எனும் உறவுமுறையாக உண்டாகின்றது.

இதே போல் ஒன்றுடன் – பல உறவுமுறையை இடமிருந்து வலமாகக் காணாமல் வலமிருந்து இடமாகப் பார்த்தால் அது பலவற்றுடன் – ஒன்று ஆகிய உறவுமுறை ஆகின்றது. எடுத்துக்காட்டாக நம்முடைய கடையில் இன்று வாங்கிய வாடிக்கையாளரின் பட்டியலைத் தயார் செய்ய முனைந்தால், ஒவ்வொரு வாடிக்கையாளரின் பெயரும், அவர் பல முறை நம் கடைக்கு இன்று வந்தாலும், ஒருமுறை மட்டுமே பட்டியலில் இடம் பெறும் அல்லவா? இது பலவற்றுடன் – ஒன்று எனும் உறவுமுறையாகும். உறவுமுறை ஒன்று தான் என்றாலும் கேட்கும் கேள்வியின் காரணமாக உறவுமுறையானது மாறுபட்ட கோணத்திலிருந்து காணப்படுகின்றது.

இவ்வாறு உருபொருட்களிடையே காணப்படும் நான்கு வகையான உறவுமுறைகளையும் வகைப்படுத்திய நாம் அடுத்ததாக அவற்றின் குணங்களின் அடிப்படையில் எவ்வாறு பிரிக்கப்படுகின்றது என்பதைக் காணலாம்.

உருபொருட்களின் பலவித பண்புகளே சீராக அடுக்கி வைக்கப்பட்டவுடன் தரவாக மாறுகின்றது என்று கண்டோம். அவ்வுருபொருட்களினிடையே ஏற்படும் உறவுமுறைகள் அவற்றின் எண்ணிக்கையின் அடிப்படையில் ஒன்றுடன் ஒன்று, ஒன்றுடன் பல என்று அமைந்ததையும் கண்டோம். ஏன் அவ்வாறு அமைய வேண்டும் அல்லது அமைக்கப்பட வேண்டும் என்று ஆராய்ந்தால் மேலும் சில உறவுமுறைகள் புலப்படுகின்றன. அவை அவ்வுருபொருட்களின் குணங்கள், எது முன்னால் தோன்றியது, எது மீண்டும் பயன்படுத்தப்படுவது, எது முக்கியமானது என்ற அடிப்படையில் உருவாகின்றன.

அவை பெரும்பாலும் கீழ்க்கண்ட உறவுமுறைகளைக் கொண்டுள்ளன:

1. முதன்மை – நடவடிக்கை தரவு *(Master – Transaction Data)*
2. பெற்றோர் – குழந்தை உறவு *(Parent – Child Relationship)*

இவ்விரண்டும் கிட்டத்தட்ட ஒரே பொருளைக் குறிப்பதாக இருந்தாலும் தனித்தனியான குணத்தினைக் கொண்டே விளங்குகின்றன.

முதன்மைத் தரவு எனும் போது அது ஒரு உருபொருளின் நிரந்தரமான தன்மையைக் குறிக்கின்றது. எத்தனை ஆண்டுகள் ஆனாலும் அவற்றின் பயன்பாடு மாறுவதில்லை *(Relatively permanent)*. எடுத்துக்காட்டாக ஒரு வாடிக்கையாளரின் பெயர்/அலைபேசி எண்/முகவரி ஆகிய தரவுகள் அடிக்கடி மாற்றப்படுவதில்லை.

நடவடிக்கைத் தரவோ அந்த நடவடிக்கை நடப்பில் இருக்கும் வரை மட்டுமே முக்கியத்துவம் பெறுகின்றது. நடவடிக்கை முடிந்ததும் அது கண்டுகொள்ளப்படுவதில்லை *(temporary usage)*. எடுத்துக்காட்டாக, காலாண்டுத் தேர்வு முடிந்து மாணவர்கள் எடுக்கும் மதிப்பெண்கள் தரவு அவர்கள் அரையாண்டுத் தேர்வு எழுதியவுடன் முக்கியத்தன்மை இழந்து விடுகின்றது. முழு ஆண்டுத் தேர்வு முடிந்து அடுத்த வகுப்புக்குச் சென்று விட்டால் முந்தைய ஆண்டின் அத்தனை நடவடிக்கைத் தரவுகளும் சில சமயங்களில் தேவையில்லாததாக் கூட மாறி குப்பைத் தொட்டிக்கும் செல்வதுண்டு!

ஆக, ஒரு உருபொருளின் மாறாத தரவை முதன்மைத் தரவு என்றும், அன்றாட செயல்பாடுகளைச் சேமித்து வைத்தலை நடவடிக்கைத் தரவு என்றும் குறிப்பிடுகின்றோம். சொல்லப்போனால் இரண்டு தரவுகளுமே நிரந்தரமாகத் தரவுக் கிட்டங்கியில் *(Data Warehouse)* சேமிக்கப்படுகின்றன. நடவடிக்கைகளின் கூட்டுத் தொகையானது குறிப்பிட்ட கால இடைவெளியில் முதன்மைத் தரவில் புதுப்பிக்கப்பட்டு சமீபத்திய ஆரம்ப இருப்பாக மாறி விடுகின்றது.

இனி, பெற்றோர் - குழந்தை உறவுமுறை தரவுகளுக்கிடையே ஏன் வருகின்றது என்பதையும் காண்போம். இங்கே எது அடிப்படை என்பதைக் காட்டுவதற்காகவும், எதன் பின்னே எது வருகின்றது என்று காட்டுவதற்காகவும் இந்த உறவு முறை குறிப்பிடப்படுகின்றது. குழந்தை இல்லாமல் பெற்றோர் இருக்கலாம் ஆனால் பெற்றோர் இல்லாமல் குழந்தை இருக்க முடியாது அல்லவா? அடிப்படையான உருபொருளை பெற்றோர் என்றும், அதன் இருப்பின் மூலமே உறுதி செய்யப்பட்டு அடுத்து வரும் உருபொரு(ட்க)ளை குழந்தை(கள்) என்றும் குறிப்பிடலாம். எடுத்துகாட்டாக, புத்தகம் என்று ஒன்று இருந்தால் அதன் ஆசிரியர் என்று ஒருவர் இருக்க வேண்டுமல்லவா? ஒரே ஆசிரியர் பல புத்தகங்கள் எழுதக் கூடும் அல்லவா? எனவே தான் ஒன்றுடன் – பல *(One to Many)* என்ற உறவுமுறை உருவாகின்றது. இந்த பெற்றோர் – குழந்தை உறவின் மூலம் ஒட்டு மொத்தத் தரவொழுங்கு *(Data Integrity)* பராமரிக்கப்படுகின்றது.

நமது அன்றாட வாழ்க்கை முறையில் அநாதைகள் இருக்கலாம். ஆனால் தரவுத்தள மேலாண்மைகள் அநாதைகளே இல்லாதவாறு தரவுத்தள மேலாண்மை மென்பொருள் கவனித்துக் கொள்வதையே தரவொழுங்கு என்று அழைக்கின்றோம்.

தரவுத்தள மேலாண்மை குழூஉக்குறியில் சொல்வதாக இருந்தால் முதன்மைத் திறவி இன்றி அந்நியத் திறவி இருக்க இயலாது. *(For maintaining the referential integrity in the database, foreign key without primary key is not allowed.)* உருபொருளைச் சுட்டும் எண் அல்லது தனிக்குறி பெற்றோர் அட்டவணையில் இருக்குமானால் அதை முதன்மைத் திறவி என்றும், அதே எண் அல்லது குறி குழந்தை அட்டவணை அல்லது வேறோர் அட்டவணையில் பெற்றோரைச் சுட்டுவதற்காக இருந்தால் அதை அந்நியத் திறவி என்றும் அழைக்கின்றோம்.

தரவுத்தள மேலாண்மையின் இந்த அடிப்படையைப் புரிந்து கொண்டால் எவ்வாறு குறைந்த நினைவகத்தில் அதிகத் தகவல்களை முறையாகவும், ஒழுங்காகவும் பதிந்து கொள்ளலாம் என்பதிலிருந்து, எவ்வாறு அவற்றை விரைவாகச் சேமித்து வைக்கவும், மீண்டும் கையாளவும் சாத்தியமாகின்றது என்பதையும் புரிந்து கொள்ளலாம்.

இவ்வாறு உருபொருட்களை முதன்மை/பெற்றோர் மற்றும் நடவடிக்கை/குழந்தை என்று தரம்பிரிப்பதை தரவுத்தள இயல்பாக்கம் என்று அழைக்கலாம்.

தரவுத்தள இயல்பாக்கம் பற்றிய விளக்கத்தினை சென்ற பகுதியில் கண்டோம். தரவுத்தள இயல்பாக்கம் மூலம் தரவொழுங்கு உறுதி செய்யப்படுகின்றது. மேலும் தரவுகளை அறிவியல்பூர்வமாக மிகக் குறைந்த நினைவகச் சேமிப்பை மட்டுமே பயன்படுத்தி சேமிக்கவும், மீண்டும் தரவு தேவைப்படும் போது விரைவாக மீட்டெடுக்கவும் இத்தரவுத்தள இயல்பாக்கம் உதவுகின்றது. ஏற்கனவே குறிப்பிட்டது போல் முதல் மூன்று வகை தரவுத்தள இயல்பாக்கம் குறித்தும் இப்போது எடுத்துக்காட்டுடன் காண்போம்.

1. முதல் இயல்புப் படிவம் – First Normal Form

இப்படிவத்தின் விதிப்படி, தரவுத்தளப்பட்டியலின் ஒரு அறையில் ஒன்றுக்கும் மேற்பட்ட நிரல்களுக்குண்டான தகவல்கள் இருக்குமானால் அவற்றை மற்றொரு தரவுப் பட்டியலில் சேமிக்க முற்படவேண்டும். அதாவது ஒரே மாதிரியான தகவல் ஒன்றுக்கு மேற்பட்ட நிரல்களில் சேமிக்க வேண்டியிருந்தாலோ அல்லது ஒரே அறையில் ஒன்றுக்கு மேற்பட்ட தகவல்களிருந்தாலோ அவற்றைத் தனியாகப் பிரித்து இன்னொரு பட்டியலில் சேமிக்கவேண்டும். இவ்விரு பட்டியல்களையும் இணைப்பது முதன்மை மற்றும் அந்நியத் திறவியாக இருக்கும்.

இப்போது ஒரு எடுத்துக்காட்டைக் காண்போம்:

வரிசை எண்	பெயர்	அலைபேசிஎண்
1	குமார்	*9123456789, 9234567890*
2	ரவி	*9421541251*
3	வேலவன்	

என்றொரு பட்டியல் இருப்பதாய்க் கொள்வோம். மேற்கண்ட பட்டியலில் குமாருக்கு மட்டும் இரண்டு அலைபேசி எண்கள் இருக்கின்றன. வேலவனுக்கு அலைபேசி எண் இல்லை. தரவுத்தள மேலாண்மையில் மிக முக்கியமானதாக் கருதப்படும் ககாட் அவர்களின் 13 விதிகளில் ஒன்று வெற்றிடத்தை அறிவியல்பூர்வமாக நிர்வகிக்க வேண்டும் என்பதாகும். கூடுமானவரை வெற்றறையைத் தவிர்க்க வேண்டும்.

முதல் இயல்புப் படிவத்தின் விதிப்படி ஒரு அறையில் ஒரு தகவல் தான் இருக்க வேண்டும். இரண்டு அலைபேசி எண்களை வைப்பது சரியல்ல. எனவே மேற்கண்ட பட்டியலை இரண்டாகப் பிரித்து கீழ்க்கண்டவாறு அமைப்பது முதல் படிவ இயல்பாக்கம் என்றழைக்கப்படுகின்றது.

வரிசை எண்	பெயர்
1	குமார்
2	ரவி
3	வேலவன்

வரிசை எண்	அலைபேசி எண்
1	*9123456789*
1	*9234567890*
2	*9421541251*

இப்போது வெற்றிடமே இல்லை என்பதையும், ஒரு அறையில் ஒரு தகவல் மட்டும் இருப்பதையும் காணலாம். வரிசை எண் முதல் அட்டவணையில் முதன்மைத் திறவியாகவும், இரண்டாம் அட்டவணையில் அந்நியத் திறவியாகவும் செயல்படுகின்றது.

கீழ்க்கண்ட பட்டியலிலும் முதல் படிவ இயல்பாக்கம் செய்ய இயலும்.

எண்	நாள்	பெயர்	விபரம்1	எண்ணம் 1	விலை 1	விபரம்2	எண்ணம்2	விலை2	விபரம்3	எண்ணம்3	விலை3
1	10-07-2022	குமார்	அரிசி	1	60	கோதுமை	1	55	சர்க்கரை	1	90
2	10-07-2022	ரகு	கோதுமை	2	55						
3	12-07-2022	குமார்	அரிசி	2	60	கோதுமை	3	55			
4	12-07-2022	ரவி	சர்க்கரை	1	90						

எடுத்துக்காட்டாக நமது கடையில் மொத்தம் மூன்று பொருட்கள் இருப்பதாகவும் அதற்கான விற்பனைப் பட்டியலில் மேற்கண்ட தகவல்களிருப்பதாகவும் கொண்டால் அதை ஒரே பட்டியலில் சேமிக்கும் போது எத்தனை

வெற்றிடங்கள் காணக்கிடைக்கின்றன என்று பாருங்கள்? அனைவரும் மூன்று பொருட்களும் வாங்க முடியாது தானே? இதுவே நூற்றுக்கு மேற்பட்ட பொருட்கள் கொண்டிருக்கும் ஒரு கடையென்றால் அதன் நிலை என்ன? அதுவே முதல் படிவ இயல்பாக்கம் செய்யும் போது எளிதாகின்றது.

கீழ்க்கண்ட இரு பட்டியல்களைப் பாருங்கள். எங்கும் வெற்றிடம் இல்லை. எத்தனை பொருட்கள் இருந்தாலும் கவலையில்லை என்பது புரியும்.

எண்	நாள்	பெயர்
1	10-07-2022	குமார்
2	10-07-2022	ரகு
3	12-07-2022	குமார்
4	12-07-2022	ரவி

எண்	விபரம்	எண்ணம்	விலை
1	அரிசி	1	60
1	கோதுமை	1	55
1	சர்க்கரை	1	90
2	கோதுமை	2	55
3	அரிசி	2	60
3	கோதுமை	3	55
4	சர்க்கரை	1	90

மேற்கண்ட பட்டியல்களிலும் எண் முதன்மை மற்றும் அந்நியத் திறவியாகப் பயன்படுகின்றது.

2. இரண்டாம் இயல்புப் படிவம் *(Second Normal Form)*

இப்படிவத்தின் படி நிரைகளில் திரும்ப வரும் வாய்ப்பிருக்கும் தகவல்களைத் தனியாகப் பிரித்து அவைகளை தனிப்பட்டியலில் சேமிக்கச் செய்ய வேண்டும். முதல் படிவத்தில் நிரல்களில் திரும்ப வரும் தகவல்களைப் பிரித்தோம். இதில் நிரைகள்/வரிகளில் மீண்டும் வரும் தகவல்களைத் தனியே பிரிக்கின்றோம். இதன் மூலம் ஒரு உருபொருளின் எந்த ஒரு தகவலும் ஒட்டு மொத்தத் தரவுத் தளத்தில் ஒரே ஒரு இடத்தில் மட்டுமே சேமிக்கப்படும். திறவிகள் மட்டுமே மீண்டும் வரும். எடுத்துக்காட்டாக மேலே குறிப்பிட்ட வாடிக்கையாளர்களில் குமார் திரும்பவும் வருகின்றார். அரிசி, கோதுமை, சர்க்கரை என்ற பொருட்களும் திரும்பவும் வருகின்றன. மேலும் வாடிக்கையாளரின் தகவல் வாடிக்கையாளர் உருபொருளுக்கு உடையது. நாள், எண் ஆகியவை பட்டியின் உருபொருளுக்கு உடையது. அரிசி, கோதுமை போன்றவை மற்றுமொரு உருபொருளாகும். எனவே அவற்றைக் கீழ்க்கண்டவாறு பிரிக்க வேண்டும். இதன் மூலம், எந்த ஒரு உருபொருளின் தகவலும் மீண்டும் வரவில்லை என்பதைக் காணவும்.

எண்	நாள்	வா. எண்
1	10-07-2022	1
2	10-07-2022	2
3	12-07-2022	1
4	12-07-2022	3

எண்	பொ.எண்	எண்ணம்	விலை
1	1	1	60
1	2	1	55
1	3	1	90
2	2	2	55
3	1	2	60
3	2	3	55
4	3	1	90

வா. எண்.	பெயர்
1	குமார்
2	ரகு
3	ரவி

பொ. எண்	விபரம்
1	அரிசி
2	கோதுமை
3	சர்க்கரை

எடுத்துக்காட்டில் மஞ்சள் நிறத்திலிருப்பவை முதன்மைத் திறவிகள். பச்சை நிறத்திலிருப்பவை அந்நியத் திறவிகள்.

இதனால் கிடைக்கும் மற்றுமொரு நன்மை, பொருளே வாங்காமல் வாடிக்கையாளர் தகவலை மட்டும் சேமிக்க முடியும், பொருளை விற்பனைப் பட்டியலில் சேர்க்காமலே நான்காவதாக ஒரு பொருளைச் சேமித்து வைக்க முடியும்.

3. மூன்றாம் இயல்புப் படிவம் *(Third Normal Form)*

மேற்கண்ட இரண்டு இயல்பாக்கங்களிலேயே தரவுகள் ஓரளவுக்கு நேர்த்தியாகச் சேமிக்கப்படும் என்றாலும், நமது அறிக்கைகளின் நேர்த்திக்காகவும் மேலும் தகவல்களை அறிவியல்பூர்வமாகச் சேமித்து வைக்கவும் இந்தப் படிவம் உதவிகரமாக இருக்கின்றது. இப்படிவத்தின் படி, முதன்மைத்திறவி குறிக்கும் உருபொருளை முதன்மையாகவும் முழுமையாகவும் கொண்டிராத தகவல்களை மற்றுமொரு பட்டியலுக்குக் கொண்டு செல்ல வேண்டும்.

எடுத்துக்காட்டாக கீழ்க்கண்ட வாடிக்கையாளர் பட்டியலைப் பாருங்கள்.

வா. எண்.	பெயர்	ஊர்
1	குமார்	சிவகாசி
2	ரகு	சிவகாசி
3	ரவி	மதுரை

இதில் வாடிக்கையாளரின் ஊர் அவருக்கு மட்டுமே உரியதில்லை தானே? அவரது பெயருக்கு அவர் உரிமை கொண்டாடலாம். ஆனால் ஊருக்கே அவர் உரிமை கொண்டாட முடியாது அல்லவா? ஒரே ஊரில் பலரும் இருக்கலாம் தானே? எனவே அதையும் கீழ்க்கண்டவாறு மற்றுமொரு பட்டியலாகப் பிரிப்பதையே மூன்றாம் இயல்புப் படிவம் என்கின்றோம்.

வா. எண்.	பெயர்	ஊர் எண்
1	குமார்	1
2	ரகு	1
3	ரவி	2

ஊர் எண்	ஊர்
1	சிவகாசி
2	மதுரை

இவ்வாறு மூன்று இயல்புப் படிவங்கள் கொண்டு தரவுகளைத் தரம் பிரித்து வைத்தால் மிக எளிதாகச் சேமிக்கவும், மீட்டெடுக்கவும் வசதியாக இருக்கும்.

ஏற்கனவே குறிப்பிட்டது போல் இன்னும் சில மேம்பட்ட அபூர்வமான இயல்பாக்கங்கள் இருக்கின்றன. நமது தொடரின் ஓட்டத்தைக் கருத்தில் கொண்டு அவற்றை விளக்குவதை விடுத்து அடுத்ததாக இப்போது வலையுலக தரவுத்தளங்கள் சிலவற்றையும் அவற்றின் தகவல் தொழில்நுட்பத்தையும் காணலாம்.

தரவுச்சுரங்கம் – 9

இதுவரை நாம் கண்டது அனைத்தும் தரவுத்தளம் குறித்த அடிப்படை ஆகும். இவையெல்லாம் நாற்பது ஆண்டுகளுக்கும் மேலாக நடைமுறையில் இருக்கின்றன! இந்தப் பகுதியில் தற்போதைய தரவுத்தளப் புதுமைகளைக் கண்டு பின் நமது தரவுச்சுரங்கப் பயணத்தைத் தொடரலாம்.

- *பொருள் சார் தரவுத்தளம்: (Object Oriented Database)*

1980 களில் பொருள் சார்ந்த நிரலாக்கம் (Object Oriented Programming -OOP) பிரபலமடைந்தது. இதனால் தரவுத்தளங்களில் சேமித்திருக்கும் தரவுகளைக் கையாளும் முறையும் சற்று மாறியது. மென்பொறியாளர்களும் வடிவமைப்பாளர்களும் தரவுத்தளங்களில் உள்ள தரவுகளை ஒரு பொருளாக அணுக முற்பட்டனர். உதாரணமாக ஒரு நபரை பொருளாகவும், அவரைப்பற்றிய தகவல்களின் அந்த பொருளின் பண்புகளாகவும் பார்க்கப்படுகிறது. தரவுகளை வெறும் அட்டவணையாகவும் அதிலுள்ள வரிசையாகவும் பாவிக்காமல், ஒவ்வொரு வரிசையையும் ஒரு பொருளாகவும், அதிலுள்ள அனைத்துத் தகவல்களும் அதன் பண்புகளாகவும் பாவிக்கப்படுகிறது. இதனால் தரவுகள் அட்டவணைக்கும் அதிலுள்ள நெடுவரிசைக்கும் சொந்தமானது என்று பார்க்காமல், அவை அணைத்தும் அதனதன் பொருளிற்கு சொந்தமாகப் பார்க்கப்படுகிறது.

இதனால் சில சிக்கல்கள் ஏற்படுகிறது. இந்த சிக்கல் பொருள் - தொடர்பு பொருத்தமின்மை என்று அழைக்கப்படுகிறது. பொருள் தரவுத்தளங்கள் மற்றும் பொருள்-தொடர்புசால் தரவுத்தளங்கள் ஆகியவை இந்த சிக்கலை தீர்க்க அதன் வினவு மொழிகளை மாற்றி அமைத்தனர். இந்த சிக்கலைத் தீர்பதற்கு இன்னொரு முயற்சியே பொருள்-தொடர்பு மேப்பிங் ஆகும். (நன்றி விக்கிபீடியா)

- *கட்டமைப்பு மட்டுமல்லா வினவு மொழி: (NoSQL)*

ஏற்கனவே நாம் இது வரை கண்ட உறவுமுறை தரவுத்தள மேலாண்மையில் எளிமைக்கு மிகப் பெரிய பங்கு வகிப்பது எளிய கட்டமைப்புள்ள வினவு மொழி ஆகும். (Structured Query Language – SQL). நம்மிடையே இருக்கும் பல தரவுகள் கட்டமைப்புள்ள தரவுகள் தான் என்றாலும், கட்டமைப்பில்லா மற்றும் பகுதி கட்டமைவுள்ள தரவுகளையும் (எடுத்துக்காட்டு: மின்னஞ்சல்) நிர்வகிப்பதற்கு 2000களில் உருவான தொழில்நுட்பம் இதுவாகும். மையத் தரவுத்தளங்களிலிருந்து விடுபட்டு இணையத்தில் தரவுகளைச் சேமித்து வைக்க வேண்டிய கட்டாயத்துக்கு நாம் தள்ளப்பட்டுள்ளோம். இதற்குப் பெருந்தரவு ஒரு முக்கியமான காரணியாகும். மேகக்கணிமையும், பகிர்ந்தமை தரவுத்தளங்களும் கோலோச்சும் இந்தக் காலகட்டத்தில் கட்டமைப்பு உள்ள தரவுமொழி சரியானது அல்ல. எனவே பழையன கழிதலும் புதியன புகுதலுமாக கட்டமைப்பு மட்டுமல்லா வினவு மொழி தற்போது நடைமுறையில் உள்ளது. (Not only SQL).

எடுத்துக்காட்டாக நீங்கள் ஒரு வலைப்பூ உருவாக்கி நிர்வகிக்கின்றீர்கள் என்று வைத்துக் கொள்வோம். அதில் நீங்கள் இட்ட ஒரு இடுகைக்குப் பலரும் பதிலளித்துள்ளனர் என்று வைத்துக் கொண்டால், இவையனைத்தையும் உறவுமுறைத் தரவுத்தளத்தின் அடிப்படையில் பல தரவுப்பட்டியல்களில் பிரித்துச் சேமிப்பது என்பதும், அதைத் தேவையான போது திரும்பப் பெறும் போது பல பட்டியல்களில் இருந்து மீண்டும் எடுப்பது என்பதும் வேலைமெனக்கெட்ட வேலையாகும். ஏனெனில் ஒவ்வொரு தரவுப்பட்டியலும் வெவ்வேறு வழங்கிகளில் இருக்கக் கூடும்! இதனால் நமது இணையதளம் மிகவும் மெதுவாக இயங்க ஆரம்பித்து விடும்.

ஆகவே, ஒரு இணையப்பக்கத்துக்குத் தேவையான அனைத்துத் தரவுகளும் ஒரே ஆவணமாகச் சேமிக்கப்படுவது சிறப்பாக இருக்கும். இதில் பலவகைகள் இருந்தாலும், சாவி-மதிப்பு கிட்டங்கிகள் (Key-Value Stores) மற்றும் கோப்பு சார் தரவுத்தளம் ஆகியவை மிகவும் குறிப்பிடத் தக்கவை ஆகும். இவற்றில் முக்கியமாக XML மற்றும் JSON வடிவத்தில் தரவுகள் சேமிக்கப்படுகின்றன. இதனாலேயே வருமான வரித் தாக்கல் செய்வதும், சரக்கு மற்றும் விற்பனை வரி தாக்கல் செய்வதும் தற்போது இந்த வடிவங்களில் தாக்கல் செய்யப்படுகின்றது.

- *நெடுவரிசைத் தரவுத்தளம்: (Columnar Database)*

நடவடிக்கைகளைப் பதிந்து வைக்கச் சிறந்தது உறவுமுறை தரவுத்தளம் என்றால் அதே நடவடிக்கைகளைப் பகுத்தாய்வு செய்யச் சிறந்தது இந்த நெடுவரிசைத் தரவுத்தளம் ஆகும். நமது இத்தொடரின் அடிப்படையே இவ்வகை தரவுத்தளம் எனலாம். ஏனெனில் பெருந்தரவைப் பகுத்தாய்வு செய்வதற்கு உறவுமுறை தரவுத்தளம் அவ்வளவு உகந்ததல்ல. காரணம், உறவுமுறை தரவுத்தளம், தரவுகளை நிரையாக படுக்கை வசமாக வன்தட்டில் சேமித்து வைக்கின்றது. நாம் பகுத்தாய்வு செய்யும் போது படுக்கை வசமான தகவல்கள் தேவைப்படுவது இல்லை. மாறாக, நெடுவரிசயில் தான் தேவைப்படுகின்றது. எடுத்துக்காட்டாக 25000 ரூபாய்க்கு மேல் சம்பளம் வாங்குபவர்களின் எண்ணிக்கை அல்லது கூட்டுத்தொகை என்ற கேள்வி எழும் போது, சம்பளம் என்னும் நெடுவரிசை மட்டுமே போதுமானது அல்லவா? ஒரு லட்சம் வரிகளிலும் இந்தச் சம்பளம் என்னும் நெடுவரிசை மட்டுமே காண்பதற்கு நிரைவரிசைத் தரவுத்தளம் மிகவும் தடுமாறும். ஆனால் நெடுவரிசை தரவுத்தளம் மிக வேகமாக இயங்கும்.

எனவே தரவைப் பகுப்பாய்வு செய்ய இந்த நெடுவரிசை தரவுத்தளம் மிகவும் உதவிகரமாக இருக்கின்றது. தரவுச்சுரங்கத்தில் தோண்டி விலைமதிப்பில்லா ஞானத்தைப் பெறுவதற்கு நாம் இன்னும் இது பற்றி அடுத்த பகுதியில் அதிகம் அறிந்து கொள்வோம்.

நெடுவரிசைத் தரவுத்தளம்: (Columnar Database)

நெடுவரிசைகளில் (Columns) வேண்டிய தரவைச் சேமிப்பது என்பது ஏதோ புதிதாய் கண்டுபிடிக்கப்பட்டது அல்ல. நாம் நிரைகளிலேயே (Rows) தரவுகளைப் பார்ப்பது வழக்கமாக இருப்பதால் இது புதிதாகத் தோன்றுகின்றது. நமது தேவைக்குத் தகுந்தவாறு நிரைகளிலோ நிரல்களிலோ தரவுகளைச் சேமித்து வைப்பது நல்லது. அதிலும் தரவுக் கிட்டங்கிகளில் (Data Warehouses) சேமித்து வைத்துப் பின்னர் தரவுச்சுரங்கத்தில் (Data Mining) பயன்படுத்துவதற்கு இந்த வகை தரவுத்தளம் பொருத்தமானதாக இருப்பதால் இதைப் பயன்படுத்துகின்றனர்.

இனி, ஒரு சிறு எடுத்துக்காட்டின் உதவியோடு இந்த நெடுவரிசைத்தளத்தினைப் பற்றி அறிந்து கொள்வோம். ஒரு விற்பனைத் தரவு அட்டவணையில் ஒரு லட்சம் வரிகளும், பதினைந்து நெடுவரிசைகளும் இருக்கின்றது என்று வைத்துக் கொள்வோம். இதில் பதிநான்காவது நெடுவரிசையில் விற்பனைத் தொகை இருப்பதாக வைத்துக் கொள்வோம். இப்போது ரூ. 10000/-க்கு அதிகமான விற்பனைத் தொகைகளைப் பட்டியலிட வேண்டியிருந்தால் நிரை தரவுத்தளத்தில் ஒரு லட்சம் வரிகளையும் பார்வையிட வேண்டும். ஒவ்வொரு வரியைப் பார்வையிடும் போதும் தேவையில்லாத மற்ற பதிநான்கு நெடுவரிசைகளையும் தாண்டித் தாண்டி வன்தட்டிலிருந்து தகவல்கள் எடுக்கப்படுவதால் மிகவும் நேரமெடுக்கும். அதுவே 1 லட்சம் விற்பனைத் தொகைகளும் ஒரே நெடுவரிசையாகச் சேமிக்கப்பட்டிருந்தால் மிகக் குறைந்த நேரத்திலேயே தேவையான பதிலை எடுத்துவிட முடியும்.

வ.எண்	தேதி	பெயர்	சம்பளம்
1	01-09-2022	முருகன்	100000
2	01-09-2022	குமரன்	150000
3	01-09-2022	தமிழ்ச்செல்வன்	120000
4	01-09-2022	கண்ணன்	125000
5	01-09-2022	தென்றல்	120000
6	01-09-2022	ரவி	100000
7	01-09-2022	சந்திரன்	110000

மேற்கண்ட அட்டவணை நிரைவரிசைத் தரவாகும். இது நடவடிக்கைகளைப் பதிவிட வாடிக்கையாக அனைவரும் பயன்படுத்தும் முறை. இதை அப்படியே வன்தட்டில் நிரைவரிசையாகவே பதிவிடுவதாகக் கொண்டால் கீழ்க்கண்டவாறு வன்தட்டில் எழுதப்படும்.

1	01-09-2022	முருகன்	100000	2	01-09-2022	குமரன்	150000	3

இதையே நெடுவரிசைத் தரவுத்தளத்தில் கீழ்க்கண்டவாறு பதிவிடப்படும்.

வ.எண்	தேதி
1	01-09-2022
2	01-09-2022
3	01-09-2022
4	01-09-2022
5	01-09-2022
6	01-09-2022
7	01-09-2022

வ.எண்	பெயர்
1	முருகன்
2	குமரன்
3	தமிழ்ச்செல்வன்
4	கண்ணன்
5	தென்றல்
6	ரவி
7	சந்திரன்

வ.எண்	சம்பளம்
1	100000
2	150000
3	120000
4	125000
5	120000
6	100000
7	110000

இதை வன்தட்டில் பதியும் போது கீழ்க்கண்டவாறு பதிந்தால் போதும்.

01-09-2022:1,2,3,4,5,6,7
முருகன்:1;குமரன்:2;தமிழ்ச்செல்வன்:3;கண்ணன்:4;தென்றல்:5;ரவி:6; சந்திரன்:7
100000:1,6;15000:2;120000:3,5;125000:4;110000:7

01-09-2022 என்ற தேதி ஒரே ஒருமுறை மட்டும் பதியப்படுவதையும், வரிசை எண் மட்டுமே திரும்ப வருவதையும் கவனியுங்கள். இதில் ஒவ்வொரு நெடுவரிசைக்கும் சுட்டுகையும் (Index) உருவாக்கிவிட்டால் இன்னும் வேகமாகத் தரவுகளை மீட்டெடுக்க முடியும். இந்த முறையில் ஒரு தகவல் ஒருமுறை மட்டுமே சேமிக்கப்பட்டு வன்தட்டின் பயன்பாட்டு அளவும் கணிசமாகக் குறைகின்றது.

லட்சம் வரிகளில் சேமிப்பதை 15 நெடுவரிசைகளில் சேமிப்பது உகந்தது தானே? மேலும், தரவுகளைப் பகுப்பாய்வு செய்யும் போது நாம் எப்போதும் முழு வரியையும் தேர்ந்தெடுப்பதே இல்லை. (எ.கா. 120000க்கு மேல் சம்பளம் வாங்குபவர்களின் பெயர் மட்டுமே தேவைப்படுமே ஒழிய முழு முகவரி, அவர் வேலைக்கு வந்த நாட்கள், விடுமுறை எடுத்த நாட்கள் போன்ற தகவல்கள் தேவையில்லை தானே?)

ஆனால் அதுவே மாதாந்திரச் சம்பள அறிக்கை எடுக்கும் போது முழு முகவரியும் மேலே குறிப்பிட்ட அனைத்து தகவல்களுமே தேவைப்படக்கூடும். எனவே நடவடிக்கைகளைச் சேமித்து வைத்து முழு நடவடிக்கைகளின் தகவல்களும் தேவைப்படும் நிகழ்நிலை நடவடிக்கைச் செயலாக்கத்திற்கு (Online Transaction Processing – OLTP) நிரைவரிசை முறை சிறப்பானதாகும்.

நெடுவரிசைத் தரவுத்தளத்தில் மீட்டெடுக்கும் நேரம் மிகக் குறைவானதாக இருந்தாலும், புதியதாக ஒரு வரியை உள்ளீடு செய்து சேமிப்பது அதிக நேரமெடுக்கும். எனவே தான் ஒரே நேரத்தில் அதிக நடவடிக்கைகளைப் பதிவதற்கு இந்த தரவுத்தளம் உகந்ததல்ல. நடவடிக்கைகள் முடிந்த பின்னர் அத்தகவல்களைப் பகுப்பாய்வு செய்வதற்கான நோக்கத்துடன் (Online Analytical Processing - OLAP) அடுக்கி வைக்கப்படும் தரவுக்கிட்டங்கிகளில் முக்கியமாக இந்த நெடுவரிசைத் தரவுத்தளங்கள் பயன்படுத்தப்படுகின்றன.

தற்போதைய மைக்ரோசாப்ட் ஆஃபீஸ் 365 பதிப்பில் தரவு பகுப்பாய்வுக்கு இந்த வகை தரவுத்தளம் பயன்படுத்துவதால், எக்செல் மென்பொருளின் 1048576 வரிகள் என்ற குறைபாடில்லாமல் எத்தனை கோடி வரிகள் கொண்ட தரவாக இருந்தாலும் மிகச் சிறப்பாக பகுப்பாய்வு செய்ய முடிகின்றது.

சென்ற பகுதியில் ஒரு முக்கியமான சொற்றொடரைக் கண்டோம். அதுபற்றிய விளக்கமாகவே இந்தப் பகுதி அமைந்துள்ளது. தரவு பகுப்பாய்வில் மிகவும் பிரபலமானதும் அனைவரும் சரளமாகப் பயன்படுத்தப்படுவதும் இச்சொற்றொடரின் முக்கியத்துவத்தைப் பறைசாற்றும்.

ஆங்கிலத்தில் *Online Analytical Processing* என்று முழுமையாகவும் *OLAP* என்று சுருக்கமாகவும் அழைக்கப்படுகின்றது இச்சொற்றொடர். நிகழ்வுகளின் தரவை நடவடிக்கையின் அடிப்படையில் பார்க்காமல் அல்லது பார்த்து முடித்தவுடன் அதை மூட்டை கட்டி வைத்து விடாமல், அதைத் தரவுக் கிட்டங்கியில் தகுந்த முறையில் அடுக்கி வைத்துப் பின் அதைத் தொடர் பகுப்பாய்வுக்கு உட்படுத்துவதே இவ்வாறு அழைக்கப்படுகின்றது.

செய்தி எப்போது தரவாகின்றதோ அப்போதே அது மதிப்பு மிக்கதாகி விடுகின்றது என்று முன்பே கண்டோம். எப்போதைய தரவாக இருந்தாலும் அது எப்போதும் மதிப்பு மிக்கது தான். தினந்தோறும் நடவடிக்கைகள் நடைபெற்று முடிந்ததும் அதை உதாசீனப்படுத்தாமல் தொடர்ந்து பகுப்பாய்வு செய்வதன் மூலம் வியாபாரத்தில் பல அறிவார்ந்த முடிவுகள் எடுக்க முடியும் என்பது எந்த வித சந்தேகத்துக்கும் இடமின்றி நிரூபணமாகி இருக்கின்றது.

மாதந்தோறும் அல்லது குறிப்பிட்ட கால அடிப்படையில் வியாபார அறிக்கைகளைக் கண்காணித்து அதன்படி முடிவுகள் எடுக்கப்பட்டு வியாபாரம் நடைபெற்று வரும் காலம் மலையேறி வருகின்றது. குறிப்பிட்ட காலம் முடியும் வரை நாம் காத்திருக்க முடியாத அவசர அவசிய காலகட்டத்தில் இப்போது இருக்கின்றோம்.

ஒரு காரில் ஓட்டுநருக்கு வண்டி ஓட்டும் போதே முடிவெடுக்கத் தகுந்தவாறு காரின் வேகம், உள் வெப்பநிலை, மின்கல மின்னளவு, எரிபொருள் அளவு என்று அனைத்தும் உடனுக்குடன் காட்டுகின்ற கட்டுப்பாட்டுப் பலகம் *(Dashboard)* இருக்கின்றது. அதே போல் தரவுலகிலும் நிகழ்நிலையில் பகுப்பாய்வு செய்து உடனுக்குடன் முடிவெடுக்கத் தகுந்தவாறு இயங்குவதால் இந்த நிகழ்தரவு பகுப்பாய்வு முறையின் முக்கிய கருவியும் அதே பெயருடனே கணினி உலகிலும் அவ்வாறே அழைக்கப்படுகின்றது .

உண்மையில் இந்த நிகழ்தரவு பகுப்பாய்வு முறையில் அப்படி என்ன தான் நடக்கின்றது என்று கேள்வி கேட்போம். இங்கு காலங்காலமாக சேர்ந்து கொண்டே இருக்கும் பெருந்தரவு அல்லது தற்போது நடைபெற்றுக் கொண்டிருக்கும் நிகழ்தரவு இவற்றைப் பல்பரிமாண நோக்கில் ஆய்வு செய்தல் நடக்கின்றது என்பதே விடையாகக் கிடைக்கின்றது.

நாம் முதல் பகுதியில் கண்ட தரவைப் பல பரிமாணங்களில் காண முயற்சி செய்யும் கன தரவு *(Data Cube)* முறையும் இந்த ஆய்வில் தான் பிறக்கின்றது. ஆக, இந்த நிகழ்தரவு பகுப்பாய்வு முறை என்பது பாற்கடலைக் கடைந்து எடுப்பதைப் போல பல புதுமைகளையும் அமுதங்களையும் நமக்கு வாரி வாரி வழங்கிக் கொண்டே இருக்கின்றது.

இதன் முக்கியப் பயன்பாடுகளாக, நிகழ்நிலை அறிக்கை உருவாக்கம், எதிர்காலக் கணிப்பு, திட்டமிடல், தர ஒப்பீடு, ஒருங்கிணைத்தல் என அடுக்கிக் கொண்டே செல்லலாம்.

நமது இந்தத் தொடர்கட்டுரையின் நோக்கமே இந்த நிகழ்தரவு பகுப்பாய்வு முறை பற்றி விளக்குவதாகும். இதில் பயன்படும் கருவிகள், செயலாக்கும் திட்டம், தரவு மாதிரி கொண்டு எவ்வாறு கனதரவு உருவாக்கப்படுகின்றது, பலபரிமாணங்களைக் கண்ணுறுவது எவ்வாறு? தகவல் வரைகலையைக் கொண்டும், பல்வேறு தரவுக் காட்சிகளையும் கொண்டு எவ்வாறு கட்டுப்பாட்டுப் பலகம் உருவாக்கப்படுகின்றது என்றும் அறிந்து கொண்டு அதன் மூலம் அனைத்துத் தொழில்முனைவோரும் பயனடைய வேண்டும் என்பதைக் கருத்தில் கொண்டே நாம் இத்தொடர் கட்டுரையில் பயணிக்கின்றோம்.

இனி அடுத்த பகுதியில் இந்த நிகழ்தரவு பகுப்பாய்வு முறையை உருவாக்கிப் பயனுள்ளதாக ஆக்கிக் கொள்வதற்கான படிநிலைகளை விரிவாகக் காண்போம்.

தரவுச் சுரங்கம் – 12

தரவுச் சுரங்கமும் (Data Mining), நிகழ்தரவு பகுப்பாய்வு முறையும் (Online Analytical Processing) ஒன்றுக்கொன்று தொடர்புடையவை மட்டுமல்ல. ஒன்றையொன்று சார்ந்தும் இருப்பவையாகும். இரண்டுமே வியாபார நுண்ணறிவைப் பெறுதற்கு முக்கியப் பங்கு வகிக்கின்றன.

தரவுச்சுரங்கத்தின் முக்கியப் பயன்கள் அல்லது நோக்கங்கள்:

1. *தரவிலிருக்கும் பாங்குகள் அல்லது திரும்பத் திரும்ப வரும் வார்ப்புகளைக் கண்டறிதல். (Hidden Patterns)*
2. *பெருந்தரவினைப் பகுப்பாய்வு செய்தல்*
3. *எதிர்காலத்தில் என்ன நிகழக் கூடும் என்று கணித்தல்*
4. *தரவு சார்ந்த முடிவெடுத்தல்*

தரவுச் சுரங்கத்தின் உறுப்புகள்:

1. *தரவு குறித்த அறிவுத்தளம் (knowledge base)*
2. *தரவுச் சுரங்கப் பொறியமைப்பு*
3. *தரவிலிருக்கும் வார்ப்பு/மாதிரி/பாங்கு கண்டறியும் உறுப்பு*
4. *பயனர் இடைமுகம் (User Interface)*

தரவுச் சுரங்கத்தின் படிநிலைகளாவன:

1. *வியாபாரம் குறித்த புரிதல் (business understanding)*
2. *தரவு குறித்த புரிதல் (data understanding)*
3. *தரவு தயார்படுத்தல் (data preparation)*
4. *மாதிரி உருவாக்கம் (pattern designing)*
5. *அளவிடுதல்/சீரமைத்தல்/சரிபார்த்தல் (evaluation)*
6. *பயன்பாட்டுக்கு விடுதல் (deployment)*

தரவுச் சுரங்கத்தில் பயன்படுத்தும் உத்திகள்:

1. *பகுத்தல் (Classification)*
2. *தொகுத்தல் (Clustering)*
3. *கணித்தல் (Regression)*
4. *விளிம்பு கண்டறிதல் (Outliers)*

நிகழ்தரவு பகுப்பாய்வு முறையின் பயன்கள்:

1. *தரவில் பல்பரிமாணம் உணர்தல் மற்றும் பகுப்பாய்வு செய்தல்*
2. *நடந்து முடிந்த/நடக்கின்ற நடவடிக்கைகளை ஆராய்தல்*
3. *குறிப்பிட்ட கால அளவில் தரவு சார்ந்த முடிவெடுத்தல்*

நிகழ்தரவு பகுப்பாய்வு முறையில் கனதரவில் பயன்படுத்தும் செயல்பாடுகள்:

1. *அச்சு அமைப்பு (Pivoting)*
2. *பகடை/பாய்ச்சிகை (Dicing)*
3. *துண்டாடுதல்/வெட்டுதல் (Slicing)*
4. *துளையிடல் (Drilling Down)*
5. *உருட்டுதல் (Rolling Up)*

நிகழ்தரவு பகுப்பாய்வின் முக்கிய வகைகள்:

1. *பல்பரிமாணப் பகுப்பாய்வு (Multidimensional OLAP)*
2. *உறவுமுறை பகுப்பாய்வு (Relational OLAP)*
3. *கலப்பின பகுப்பாய்வு (Hybrid OLAP)*

அப்பப்பா! வியாபாரத் தகவல்களை ஆராய்ந்து முடிவெடுப்பதற்குத் தான் எத்தனைகருவிகள்? எத்தனை வழிமுறைகள்? எத்தனை படிநிலைகள்? இப்பெருந்தரவுக் காலத்தில் வியாபார நுண்ணறிவினைப் பெறுவது அவ்வளவு எளிதா என்ன? இருந்தாலும் அடிப்படைக் கணிதமும், புள்ளியியலும் எனும் உளியையும் கடப்பாறையையும் கொண்டு தான் இந்தத் தரவுச் சுரங்கத்தைப் பலரும் வெட்டி உள்ளிருக்கும் விலைமதிக்க முடியாத செல்வங்களைப் பயன்படுத்திக் கொண்டிருக்கின்றனர். முடிந்த வரையில் இதில் முக்கியமானவற்றை அடுத்தடுத்த பகுதிகளில் நாம் காண்போம்.

தரவுச்சுரங்கம் – 13:

இந்தப் பகுதியில் பல கருத்தாழமிக்க தகவல் தொழில்நுட்பச் சொல்லாடல்களையும் ஒன்றாகக் கட்டி வைத்திருக்கும் சங்கிலியைப் பற்றிக் காண்போம். எடுத்துக்காட்டாக வியாபார நுண்ணறிவு (Business Intelligence), தரவுத்தள மேலாண்மை (Database Management), தகவல் தொழில்நுட்பம் (Information Technology), தரவுச்சுரங்கம் (Data Mining), தரவுக் கிட்டங்கி (Data Warehouse), நிகழ்தரவு பகுப்பாய்வு (Online Analytical Processing), செயற்கை நுண்ணறிவு (Artificial Intelligence), இயந்திரக் கற்றல் (Machine Learning), ஆழக் கற்றல் (Deep Learning) போன்ற பல நவீனக் கணினித் தரவுச் சொற்களைத் தற்போது காண முடிகின்றது. இவை அனைத்தையும் ஒன்றிணைத்து ஒரு நவயுகத் துறையாக உருவெடுத்து வருவது ***தரவு அறிவியல்*** *ஆகும். தகவல் அறிவியல் வேறு. தரவு அறிவியல் வேறு. இன்று நாம் காண இருப்பது தரவு அறிவியல்.*

தரவு அறிவியல் என்றால் என்ன? விக்கிப்பீடியாவில் அறிவார்ந்த பெருமக்கள் என்ன சொல்கின்றார்கள் என்று பார்ப்போம்.

தரவு அறிவியல் கணிதம், புள்ளியியல், கணினி அறிவியல், தரவுச் செயலாக்கம் போன்ற பல்வேறு துறைகளிலிருந்து கோட்பாடுகளையும், தொழில்நுட்பங்களையும் பயன்படுத்துகிறது. தரவு அறிவியலின் ஒரு முக்கியமான பிரிவு பெருந் தரவு ஆகும்.

உழவு, சந்தைப்படுத்தலை மேம்படுத்தல், மோசடிகளைக் கண்டுபிடித்தல் போன்ற சிக்கல்களை ஆராய, தரவுத் தயார்ப்படுத்துதல், புள்ளியியல், இயந்திரக் கற்றல் போன்ற துறைகளைத் தரவு அறிவியல் பயன்படுத்துகிறது.

தரவு விஞ்ஞானிகள் தங்களின் திறனைப் பயன்படுத்தி தரவு மூலத்தை கண்டுபிடித்து விளக்கவும், வன்பொருள், மென்பொருள் மற்றும் அலைவரிசை தடைகள் இருந்தபோதிலும் அதிக அளவிலான தரவுகளை நிர்வகிக்கவும், தரவுகளுக்கான ஆதாரங்களை இணைக்கவும், தரவுகளின் நிலைத்தன்மையை உறுதிப்படுத்தவும், தரவுகளை புரிந்து கொள்வதற்கு உதவியாக காட்சிப்படங்களை உருவாக்கவும், தரவுகளைப் பயன்படுத்தி கணித மாதிரிகளை கட்டமைக்கவும் மற்றும் தரவுகளைப்பற்றிய கண்டுபிடிப்புகளை வெளிப்படுத்தவும் செய்கிறார்கள்.

என்ன அருமையான விளக்கம் பாருங்கள்? மேற்கண்ட பத்திகளைத் திரும்பத் திரும்ப இருமுறை வாசித்துப் பார்த்தாலே தரவு அறிவியல் பற்றி உணர்ந்து கொள்ளலாம்.

இன்றைய தகவல் தொழில்நுட்ப யுகத்தின் புதிய துறையாகவும் அதிக வேலைவாய்ப்பைத் தரக் கூடிய துறையாகவும் இந்தத் தரவு அறிவியல் துறையைக் காண முடிகின்றது. இந்தியர்கள் அதிலும் குறிப்பாக கணிப்பொறியாளர்கள் தங்கள் வாயால் உச்சரிக்கும் ஒரு மந்திரமாக இந்தச் சொல் இருப்பதைக் காணலாம்.

இதில் ஒரு கணிப்பொறியாளர் தம்மை ஒரு தரவு விஞ்ஞானியாக மேம்படுத்திக் கொள்வதைக் காணலாம். விஞ்ஞானிகள் என்பவர்கள் உலகில் ஆல்பர்ட் ஐன்ஸ்டீன் போல ஐசக் நியூட்டனைப் போல அபூர்வமாகப் பிறப்படுத்துத் தங்கள் கண்டுபிடிப்புகளைக் கொண்டு சாதனைகளையும் புதுவிதக் கருவிகளையும், இயற்கையின் விசித்திரமான புதிர்களுக்கான பதில்களையும் கண்டுபிடிப்பவர்களாகத் தான் இதுவரை நாம் உணர்ந்திருக்கின்றோம்.

இன்றோ யார் வேண்டுமானாலும் தரவு விஞ்ஞானி ஆக முடிகின்ற ஒரு வசதியையும் வாய்ப்பையும் இந்தக் கணினித் துறை உருவாக்கிக் கொடுத்திருக்கின்றது என்றால் அது மிகையாகாது.

மேலே குறிப்பிட்டபடி, கணிதம், புள்ளியியல், கணிப்பொறியியல், தரவு மேலாண்மை குறித்த அறிவு ஆகியவற்றை வளர்த்துக் கொண்டால் நீங்கள் நானும் கூட தரவு விஞ்ஞானிகளாக ஆகி விட முடியும்!

இன்றைய உலகம் காண்கின்ற மற்றும் காணப் போகின்ற பல்வேறு பிரச்னைகளுக்கும் கணினியைப் பயன்படுத்தி தீர்வு காணும் ஒரு துறை தரவு அறிவியல் துறை.

இதனை இதனால் இவன்முடிக்கும் என்றாய்ந்து அதனை அவன்கண் விடல். என்னும் திருக்குறளுக்கு ஏற்ப விண்மீண்களின் எண்ணிக்கையை ஒத்திருக்கும் பெருந்தரவை ஆராய்ந்து அதிலிருந்து காலப் புதிர்களை விடுவிக்கும் அறிவைப் பெறுவதற்குக் கணினியை எவ்வாறு பயன்படுத்துவது என்பதை இப்போது மனிதன் நன்றாகவே கற்றுக் கொண்டு விட்டான். தரவு அறிவியலில் இன்று பலரும் கற்றுத் தேர்ந்து கணினியின் கையில் சுக்கானைக் கொடுத்து பெருந்தரவுக் கடலில் கப்பல் ஓட்ட வைக்கின்றார்கள்.

தரவு விஞ்ஞானிகள் "தரவுகளை புரிந்து கொள்வதற்கு உதவியாக ***காட்சிப்படங்களை*** *(Visualizations) உருவாக்கவும், தரவுகளைப் பயன்படுத்தி* ***கணித மாதிரிகளை*** *(Mathematical Models) கட்டமைக்கவும் மற்றும் தரவுகளைப்பற்றிய கண்டுபிடிப்புகளை வெளிப்படுத்தவும் செய்கிறார்கள்." மேலே தடிப்பான எழுத்துகளில் கூறப்பட்டிருப்பதில் நம் கவனத்தைச் செலுத்துவோம்!*

தரவை எவ்வாறு காட்சிப்படுத்தலின் மூலம் பகுப்பாய்வு செய்ய இயலும் என்பதையும் அது தரவு அறிவியலின் மிக முக்கிய பங்கு வகிக்கின்றது என்பதையும் சென்ற பகுதியில் பார்த்தோம். இப்போது தரவைக் காட்சிப்படுத்தும் கலை பற்றிக் காண்போம். சமீபகாலங்களில் இதற்கு ஒரு கலைச்சொல் உருவாக்கப்பட்டு பயன்பாட்டில் உள்ளது. ஆங்கிலத்தில் *infographics* என்றழைக்கப்படும் "தகவல் வரைகலை" வருங்காலத்தில் தனிப்பட்ட ஒரு துறையாக மாறும் அளவுக்கு அதில் வல்லுநர்களும் புதிய உத்திகளும் உருவாகி வருகின்றது. தரவு காட்சிப்படுத்தல் என்பது ஒரு வரைபடத்தில் எவ்வாறு தரவு காட்டப்படுகின்றது என்பதாம். ஆனால் அதுவே தகவல் வரைகலை என்பது வரைபடத்தை எவ்வாறு காண்பது என்ற தொழில்நுட்பம் அறியாத பாமர மக்களும் அறிந்துய்ந்துணரும் வண்ணம் தகவலாக உருமாற்றம் அடைந்த தரவை எவ்வாறு குறைந்த சொற்களைக் கொண்டு வரைகலையில் கொண்டு வருவது எனும் கலையாம். இரண்டும் வெவ்வேறு என்றாலும் ஒன்றுக்கொன்று தொடர்புடையது.

தரவு காட்சிப்படுத்தல்:

தரவு காட்சிப்படுத்தல் என்பது தரவுகளை விரைவாக பார்ப்பதற்கு பயன்படுத்தப்படும் ஒரு பயன்பாடு. இது ஒரு காகிதத்தில் அல்லது ஒரு பக்கத்தில் பார்க்க முடியாத பெருந்தரவையும் பார்க்க, விளக்க, வாசிக்க, மற்றும் செயல்படுத்த முடிவுகளை ஏற்படுத்துவதற்கு உபயோகிக்கப்படும். பலவகை தரவு காட்சிப்படுத்தல் மென்பொருள் இப்போது கிடைக்கின்றது. தகவல் காட்சிப்படுத்தல் துறை மனிதர்-கணினி ஊடாட்டம், கணினியியல், வரைகலை, காட்சி வடிவமைப்பு, உளவியல், வணிக மாதிர்கள் ஆகிய துறைகளின் ஆய்வுகளில் இருந்து உருவாகி உள்ளது. இத்துறை அறிவியல் ஆய்வு, எண்ணிம நூலகங்கள், தரவு அகழ்வு, நிதித் தரவு பகுப்பாய்வு, சந்தையியல், உற்பத்திச் செயலாக்கக் கட்டுப்பாடு, மருந்துக் கண்டுபிடிப்பு போன்ற துறைகளில் முதன்மையாகப் பயன்படுகிறது.

தரவு காட்சிப்படுத்தலின் முக்கிய நோக்கம் பயனாளர்களுக்கு தகவல்களை படக் காட்சி முறையில் காண்பிப்பது. தரவை காட்சிப்படுத்துதல், புள்ளியியல் வரைகலை முறையைப் பயன்படுத்துகிறது. இம்முறையில் எண்வகை தரவு, புள்ளி, கோடு அல்லது பட்டையைக் கொண்டு, குறியாக்கப்பட்டு, அளவைக்குரிய செய்திகளை காட்சிப்படுத்துவதன் மூலம் அறிவிக்கலாம். (நன்றி: தமிழ்நாடு பாடநூல் நிறுவனம்)

தரவு காட்சிப்படுத்தலின் வகைகள்

• வரைப்படங்கள் *(Charts)*

• அட்டவணைகள் *(Tables)*

• வரைகலை *(Graphs)*

• நிலப்படங்கள் *(Maps)*

• தகவல் வரைகலை *(Infographics)*

• முகப்பலகம் – கட்டுப்பாட்டகம் *(Dashboards)*

தரவு காட்சிப்படுத்தலின் பயன்கள்:

• தரவு காட்சிப்படுத்தல் பயனர்கள் தரவுகளை எளிதாக கூர்ந்து ஆய்வு செய்யவும், உட்பொருளை வெளிப்படுத்த உதவுகிறது.

• இது சிக்கலான தரவுகளை புரிந்துக் கொண்டு அவற்றை பயன்படுத்திக் கொள்ள வழி செய்கிறது.

• தரவு காட்சிப்படுத்தல் பல்வேறு வரைப்படங்களைக் கொண்டு தரவு மாறிகளுக்கு இடையே உள்ள உறவு நிலையை வெளிப்படுத்துகிறது.

வரைபட வகைகள்

வரைபடங்களில் பல்வேறு வகைகள் உள்ளன. குறிப்பிட்ட வகைத் தரவுக்கு குறிப்பிட்ட வகை வரைபடமே ஏற்றதாக இருக்கும். ஏராளமான வரைபட வகைகள் இருக்கின்றன. அவற்றுள் மிகப் பரவலாகப் பயன்படுத்தக் கூடிய முக்கியமான சிலவற்றைப் பார்ப்போம்:

பட்டை வரைபடம் *(Bar Chart):*

காலம், இடம் அல்லது வகைப்பாட்டு வாரியான மதிப்புகளை ஒப்பிட ஏற்றது. மாத வாரியான விற்பனை, மாவட்ட வாரியான மழை, மாத வாரியாக வரவு செலவுக்கான ஒப்பீடு போன்ற புள்ளி விவரங்களுக்குப் பயன்படுத்தலாம். அளவீட்டுப் பட்டைகள் செங்குத்தாகவோ கிடைமட்டமாகவோ அமையலாம். பட்டைகள் செங்குத்தாக அமையும் வரைபடத்தை 'நெடுக்கை வரைபடம்' *(Column Chart)* என்று கூறுவதுமுண்டு.

அடுக்குப் பட்டை வரைபடம் *(Stacked Bar Chart):*

காலம், இடம் அல்லது வகைப்பாட்டு வாரியாக ஒன்றுக்கு மேற்பட்ட கூறுகளின் கூட்டு மதிப்பை ஒப்பிட ஏற்றது. மாத வாரியாக இரண்டு மூன்று பொருள்களின் மொத்த விற்பனை போன்ற ஒப்பீடுகளுக்குப் பயன்படுத்தலாம்.

வட்ட வரைபடம் *(Pie Chart):*

முழுமை மதிப்பில் ஒவ்வொரு கூறுகளும் எவ்வளவு பங்கு என்பதைத் தெளிவுபடுத்த ஏற்றது. வரவு-செலவுத் திட்டத்தில் மொத்த வருமானத்தில் பல்வேறு பணிகளுக்கான செலவு மதிப்பீடு, மக்கள் தொகையில் வெவ்வேறு மதங்களைச் சார்ந்தோரின் விழுக்காடு, நாடாளுமன்றத் தில் பல்வேறு கட்சிகள் வகிக்கும் இடங்கள் - இதுபோன்ற விவரங்களை வெளியிட மிகவும் உகந்தது.

கோட்டு வரைபடம் *(Line Chart):*

ஏற்றம் இறக்கம் கொண்ட ஒரு நிகழ்வின் தொடர்ச்சியான போக்கினைச் சுட்டிக்காட்ட ஏற்றது. மாதந்தோறும் விலைவாசிப் புள்ளி உயர்வு, பங்குச் சந்தையில் சென்செக்ஸ் குறியீட்டின் ஏற்ற இறக்கம், கிரிக்கெட் போட்டியில் இந்தியாவும் பாகிஸ்தானும் ஒவ்வோர் ஓவரிலும் எடுத்த ரன்கள் - இத்தகைய விவரங்களுக்கு உகந்தது.

வரைபடம் உருவாக்கும்போது, தரவுக்கு ஏற்ற வரைபட வகையை நாம்தான் முடிவுசெய்ய வேண்டும். வகையை முடிவு செய்துவிட்டால் வரைபடத்தை எளிதில் உருவாக்கி விடலாம். *(நன்றி: தமிழ் மெய்நிகர் பல்கலைக்கழகம்)*

அப்படியே தகவல் வரைகலை பற்றியும் காணலாம். தகவல் வரைகலை நாம் ஏற்கனவே கண்டது போல் தகவலையும் சொற்களையும் படங்களையும் கொண்டு நேர்த்தியான புதுமையான வரைகலையையும் சேர்த்து அதை யாவரும் உணரும் வண்ணம் உருவாக்கும் கலையாகும். எடுத்துக்காட்டாக, கொரொனா பெருந்தொற்றின் போது நமது அரசாங்கம் அனைவரும் அறிந்து கொள்ளும் வகையில் கைகளை எவ்வாறு சுத்தம் செய்ய வேண்டும் என்று படத்தொகுப்பின் மூலம் தெரிவித்தது அல்லவா? அதைக் குறிப்பிடலாம்.

தகவல் வரைகலை குறைந்தபட்சம் எட்டு வகைப்படும்.

பட்டியல் தகவல் வரைகலை *List infographics.*

புள்ளியியல் தகவல் வரைகலை *Statistical infographics.*

வழிமுறை தகவல்வரைகலை *How-to infographics.*

காலவரிசை தகவல் வரைகலை *Timeline infographics.*

ஒப்பீடு தகவல் வரைகலை *Comparison infographics.*

வரைபடம் மற்றும் இடம் சார்ந்த தகவல் வரைகலை *Map and location infographics.*

பாய்வுப்பட தகவல் வரைகலை *Flowchart infographics.*

செயல் வழிமுறை தகவல் வரைகலை *Process description infographics.*

தரவு காட்சிப்படுத்தலும் தகவல் வரைகலையும் இவ்வளவு தான் என்று வரையறைப்படுத்த முடியாத அளவுக்கு நூதனமான பல அமைப்புகளைக் கொண்டு திகழ்கின்றது. தினமும் புதுமைகளைக் கொண்டு வந்து சேர்க்கின்றது. இதனால் வியாபார உலகிற்குக் கிடைக்கும் நன்மைகள் சொல்லிலடங்கா! அடுத்த பகுதியில் தரவுக்காட்சிப்படுத்தலுக்கும், தகவல் வரைகலைக்கும், முகப்பலகத்திற்கும் பயன்படுத்தப்படும் பல மென்பொருட்கள் பற்றி விரிவாகக் காணலாம்.

இன்றைய தகவல் தொழில்நுட்பச் சந்தையில் வியாபார நுண்ணறிவு தொடர்பான மென்பொருளின் பங்கு மிக அதிகமாகவே உள்ளதைக் காண்கின்றோம். சென்ற பகுதியில் நாம் கண்ட தகவல் வரைகலை மற்றும் தரவு காட்சிப்படுத்தலின் தொடர்ச்சியாக, அதற்கு உதவிகரமாக இருக்கும் சில மென்பொருட்களின், சேவைகளின் பட்டியலைத் தற்போது காண்போம். இவற்றில் பல மென்பொருட்கள்/வசதிகள் இலவசமாகக் கிடைக்கின்றன என்பது வியாபாரத் துறைக்கு ஒரு வரப்பிரசாதம் என்றே கொள்ளலாம்.

இந்த மென்பொருட்களின் பட்டியல் என்பது ஒரு முடிவான ஒன்று அல்ல. அது ஒரு மிகப் பெரிய பட்டியலாகும். இந்த மென்பொருட்களைப் பட்டியலிடும் போது சில விஷயங்களைக் கருத்தில் கொள்ளலாம்.

தரவு காட்சியமைப்பு மென்பொருட்கள் அனைத்துமே அனைவரும் பயன்படுத்தும் வண்ணம் எளிமையாக உருவாக்கப்பட்டுள்ளன. அனைவருமே செயற்கை நுண்ணறிவையும், இயந்திரவழிக் கற்றலையும் எவ்வாறு இதில் புகுத்தலாம் என்பதில் மிகுந்த முனைப்புடன் இருக்கின்றார்கள் என்பதும் தெரிகின்றது.

இந்தத் துறையில் இருப்பவர்கள் இவற்றைக் குறைந்த நேரம் செலவிட்டுக் கற்றுக் கொள்ளவியலும். மேலும் ஒன்றைப் போலவே மற்ற மென்பொருட்கள் இருப்பதாலும் ஒன்றுக் கொன்று அதிக வித்தியாசமில்லாத வசதிகளைக் கொண்டிருப்பதாலும் ஒரு மென்பொருளைக் கற்றுக் கொண்டாலே மற்றவற்றைக் கற்றுக் கொள்ளத் தேவையில்லை என்பதையும் உணரலாம். மேலும் ஒன்றை ஒன்று மாறுபடுத்தும் வசதிகளே அவற்றை நோக்கி அவற்றின் வாடிக்கையாளர்கள் செல்வதற்கான காரணமாக அமைகின்றது. எது எப்படியிருப்பினும் எப்பேர்ப்பட்ட பெருந்தரவையும் உள்வாங்கி அதைப் பகுப்பாய்வு செய்வதை இம்மென்பொருட்கள் அனைத்துமே அடிப்படை வசதியாகக் கொண்டுள்ளன.

பகுதி இலவசமாகவும், முக்கிய வசதிகளைச் செயல்படுத்த வேண்டியிருக்கும் போது அதிக விலையும் கொண்ட மென்பொருட்களையும் காண்கின்றோம். அப்படியான மென்பொருட்களில் பெரும்பாலும் ஆலையில்லா ஊருக்கு இலுப்பைப்பூ சர்க்கரை எனும் பழமொழிக்கேற்ப இலவச வசதிகளை மட்டுமே பயன்படுத்தி திருப்தியடையும் பலரையும் காணமுடிகின்றது.

இப்போது சில மென்பொருட்களின் பட்டியலைக் காணலாம்:

1. Tableau
2. Power BI
3. Zoho Reports
4. Google Charts
5. Visual.ly
6. IBM Watsom
7. Plotly
8. Fusioncharts
9. Qlikview
10. Infogram
11. ChartBlocks
12. Chart.js
13. KNIME
14. Grafana
15. Data Wrapper
16. JupyteR

மேற்கண்ட பட்டியலைக் காண்பவர்கள் நாம் அன்றாடம் பயன்படுத்தும் மின்விரிதாள் மென்பொருளான எக்செல் இந்தப் பட்டியலில் இடம் பெறாதா என்று கேட்பது வழக்கம். அடுத்த பகுதியில் இதற்கான விடையையும், எவ்வாறு எக்செல் அல்லது எந்த ஒரு மின்விரிதாள் மென்பொருளைப் பயன்படுத்தியும் தகவல் வரைகலையையும் தரவு காட்சிப்படுத்தலையும் செய்யலாம் என்பதைக் காணலாம்.

மேலே குறிப்பிட்ட மென்பொருட்களில் சிலவற்றைப் பயன்படுத்திப் பார்க்கவும்.

பெருந்தரவைப் பகுப்பாய்வு செய்வதற்கும், அதைக் காட்சியாக வரைபடத்தில் உள்ளுணர்ந்து காண்பதற்கும் எக்செல் போன்ற மின்விரிதாள் மென்பொருளைப் பயன்படுத்த இயலாதா? அதிலேயே அத்தனை வசதிகளும் இருக்கத் தானே செய்கின்றன என்ற கேள்வி எழுகின்றது.

பொதுவாக மின்விரிதாள் மென்பொருள் என்பது பெருந்தரவுக்கு உகந்ததல்ல. ஏனெனில் அதன் வடிவமைப்பும் இயல்பான தன்மையும் அவ்வாறுள்ளது. அடிக்கடி மா(ற்)றக் கூடிய சிக்கலான கணக்குகளை எளிதான சூத்திரங்களின் மூலம் சாத்தியமாக்கும் மென்பொருளாக அது விளங்குகின்றது. சிறிய வகை தரவைப் பகுப்பாய்வு செய்கின்ற வசதி இருந்தாலும் கட்டமைப்பில்லா இடத்தில் கட்டமைவுள்ள தரவைப் பயன்படுத்த முடியும் என்றாலும் அதன் அமைப்பு பெருந்தரவுக்கு ஒத்ததாக இல்லை.

மொத்தமே 1048576 வரிகளே வைக்க முடியும். ஒவ்வொரு நிரலுக்கும் சுட்டி (Index) உருவாக்க முடியாது. ஒரு அட்டவணைக்கும் இன்னொரு அட்டவணைக்குமான உறவுமுறையை உறவுமுறைத் தரவமைப்பைப் போல் நிர்ணயித்துப் பொருத்தி ஒட்டு மொத்தமாக ஒரு தரவுப் படிமத்தை (Data Model) உருவாக்க இயலாது. இன்றைய நவீனப் பகுப்பாய்வுக் களமான நெடுவரிசைத் தரவுத்தள (Columnar Database) அடிப்படையில் மின்விரிதாளில் தரவைப் பேண முடிவதில்லை. தற்போதைய தகவல் பலகத்தில் (Dashboard) உள்ளது போல் ஒரு வரைபடத்தில் சுட்டி அதில் குறிப்பிட்ட இடத்தை மட்டும் வெட்டினால் மற்ற வரைபடங்களிலும் அது எதிரொலிப்பது போன்ற வசதிகள் (Interactive Slicing) இல்லை என்று மின்விரிதாள் மென்பொருளின் பகுப்பாய்வுக் குறைகளை அடுக்கிக் கொண்டே செல்லலாம்.

இருந்த போதிலும், மின்விரிதாளிலும் மேலே கண்ட வசதிகளைக் கொண்டு வந்து விட்டால் பகுப்பாய்வுக்கு மட்டும் என்று வேறொரு மென்பொருளை நாட வேண்டிய அவசியம் இல்லையே என்பதே அனைவரின் எதிர்பார்ப்பாக இருந்தது. எனவே தற்போதைய மின்விரிதாள் மென்பொருள் கட்டமைப்பாளர்களில் முதன்மையாகக் கருதப்படும் மைக்ரோசாஃப்ட் எக்செல் மற்றும் கூகுள் ஷீட்ஸ் ஆகிய இரு பெரும் நிறுவனங்களும் இதில் ஈடுபட்டுள்ளன. மற்ற மென்பொருள் வல்லுநர்களும் இதே பாதையில் தான் செல்ல முற்பட்டுள்ளனர். இந்தப் போட்டியில் முன்னணியில் தொடர்ந்து சென்று மைக்ரோசாஃப்ட் எக்செல்லின் சமீபத்திய பதிப்புகள் யாவும் பெருந்தரவுப் பகுப்பாய்வை முன்வைத்தே அனைத்து புதுமைகளையும் கண்டு வருகின்றது. எனவே எக்செல்லின் சமீபத்திய பதிப்புக்கு மாறுவதும், அதன் புதிய அம்சங்களைக் கற்றுணர்வதும் அனைத்து மின்விரிதாள் மென்பொருள் பயனர்களின் பொறுப்பாகின்றது.

எக்செல் கொண்டு வந்துள்ள சில புதிய வசதிகள்:

1. அணிகளில் புதுமை – இயங்கு அணி (Dynamic Array)
2. டேபுலர் என்ற உள்ளமைந்த நெடுவரிசைத் தரவுத்தளம் (Tabular – in memory Columnar Database)
3. எந்த அளவு, விதம், அமைவும் உடைய எத்தகைய தரவாக இருந்தாலும் அதைச் சீரமைத்து தரவுப் படிவத்திற்கு ஏற்றவாறு மாற்றும் பவர் க்வரி எனும் அமைப்பு (Power Query for ETL)
4. அவ்வாறு தரவுப்படிவத்திலிருக்கும் தரவில் இருந்து பகுப்பாய்வுக்கு ஏற்றவாறு பல்பரிமாணக் கண்ணோட்டத்தில் காண வைக்கும் பவர் பைவட் என்னும் அமைப்பு (Power Pivot with DAX)
5. காலத்துக்குத் தகுந்த மாதிரியான பலவித தரவுசார் வசதிகள் (எ.கா. தரவு வகை) (Data Types)
6. கால பருவ மாற்றங்களைக் கணக்கில் கொள்ள Forecasting with Exponential Smoothing
7. புள்ளியியல் வரைபடங்கள்

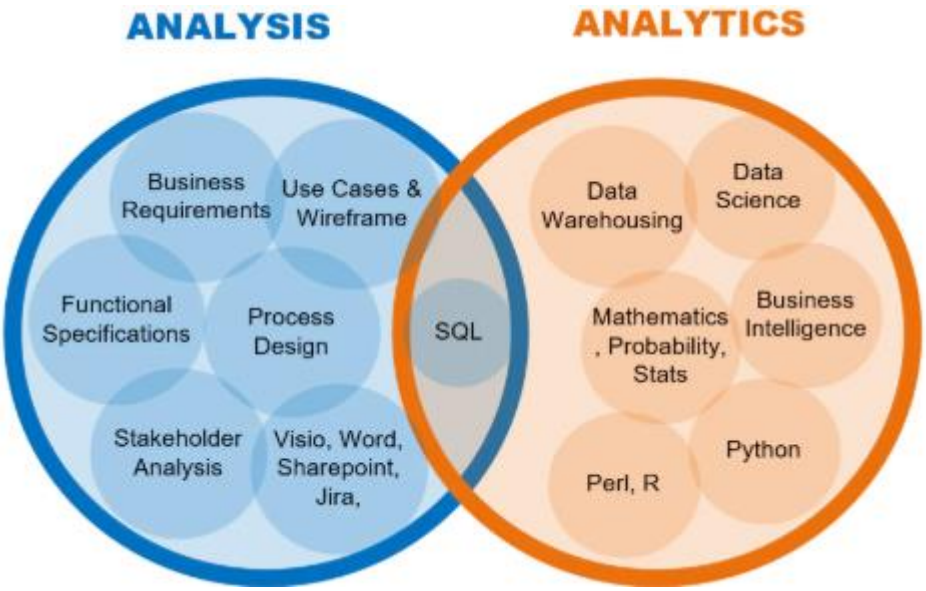
இவ்வாறு நிறைய வசதிகளைக் கொண்டு வரும் மின்விரிதாள் மென்பொருளைப் பயன்படுத்த ஆரம்பிக்கும் போது நம்மால் பெருந்தரவைப் பகுப்பாய்வு செய்வதற்கும் காட்சியாகக் காண்பதற்கும் ஏதுவாகின்றது. இருந்த போதும் மின்விரிதாள் மென்பொருளின் தனித் தன்மை இதனால் இழந்து போகும் என்று வாதிடும் வல்லுனர்களும் உள்ளனர். எனினும் அடுத்த தலைமுறை தரவு பகுப்பாய்வுக்கான துவக்கப் புள்ளியாக மின்விரிதாளின் பங்கு மகத்தானது.

இன்று கணினித் தரவுச் சந்தையில் கோலோச்சியிருக்கும் தரவு பகுப்பாய்வின் பல வகைகளை இந்தப் பகுதியில் காண்போம். ஆங்கிலத்தில் இரண்டு சொற்களைக் காண்கின்றோம். *Analysis – Analytics.* இவையிரண்டு சொற்களும் ஒன்றுக்கொன்று மாற்றாகவோ அல்லது ஒன்றையொன்று வேறுபடுத்துவதாகவோ கையாளப்படுகின்றது. இரண்டும் ஒன்றே என்று பலரும், இரண்டும் வெவ்வேறே என்று பலரும் பல இணையதளங்களிலும் புத்தகங்களிலும் உரையாடுவதையும் விவாதம் செய்வதையும் காணமுடிகின்றது. வேறுபடுத்திப் பார்க்கும் பலரும் விவாதிக்கும் குறிப்பிடத்தக்க வித்தியாசங்கள் இவை:

1. *Analysis* ஐ விட *Analytics* என்பது உயர்தரமானது.
2. முதலாவது முடிந்ததும் இரண்டாவது ஆரம்பமாகின்றது.
3. முதலாவது இறந்த காலத்தை ஆய்வு செய்வதாகவும், இரண்டாவது எதிர்காலத்தை ஊகிக்க வல்லதாகவும் உள்ளது.
4. முதலாவது பொதுவானது. இரண்டாவது குறிப்பிட்ட வகையானது என்போரும் உள்ளனர்.
5. முதலாவது குறிப்பிட்ட வகையானது, இரண்டாவது பொதுவானது என்போரும் உள்ளனர்!
6. முதலாவது படைப்பு என்றால் இரண்டாவது அதைப் படைக்கும் கருவித் தொகுப்பு என்றும் சிலர் கூறுகின்றனர்.
7. ஆங்கில அகராதியோ *Analysis* என்பது தீர ஆய்வது *(a detailed examination of anything complex in order to understand its nature or to determine its essential features : a thorough study)* என்றும், *Analytics* என்பது *(the method of logical analysis)* அவ்வாறு ஆய்வதற்கான முறையான வழிமுறை என்றும் குறிப்பிடக் காண்கின்றோம்.

ஆக, ஒன்று மட்டும் நிச்சயமாகின்றது. இரண்டும் ஒன்றாகவே இருந்தாலும் இன்றைய காலகட்டத்தில் இவை இரண்டையும் ஒன்றாய்ப் பார்ப்பதற்கான வாய்ப்பைக் கடந்து விட்டதாகவே நாம் கருதலாம். ஏனென்றால் கணினி உலகில் இவ்விரு சொற்களின் பயன்பாடும் வெவ்வேறு இடங்களில் வெவ்வேறு விதமாகவே உள்ளது. கடந்த இருபது ஆண்டுகளில் தரவுப் பகுப்பாய்வில் ஏற்பட்டிருக்கும் முன்னேற்றத்தைச் சுட்டிக் காட்டுவதாக *Analytics* என்ற சொல்லே பயன்படுத்தப்படுகின்றது. எனவே பழையன கழிதலும் புதியன புகுதலும் வழுவல எனும் சொல்லுக்கேற்ப நானும் இந்தத் தொடரில் இவ்விரண்டும் வேறெனக் கருதித் தொடர்கின்றேன்.

இரண்டுக்கும் முக்கிய வேறுபாடாகப் பலரும் குறிப்பிடுவது இவற்றில் பயன்படுத்தும் கருவிகளையும், முறைகளையும், பயன்பாடுகளையும் பற்றியதாகவே உள்ளது. எடுத்துக்காட்டாக *Analytics*ல் அதிகப்படியான புள்ளியியல் கோட்பாடுகளையும், கணித சூத்திரங்களைப் பயன்படுத்துவதும், பெருந்தரவில் உள்ளமைந்துள்ள பாங்குகளைக் கண்டறிந்து அதை எதிர்காலத்தில் எவ்வாறு பயன்படுத்தி வியாபார முடிவுகளை எடுக்க முடியும் என்பதையும் காண முடிகின்றது. கீழ்க்கண்ட படத்தைக் காணுங்கள். நீங்களே முடிவெடுத்துக் கொள்ளுங்கள். இணையத்தில் இது போன்ற வித்தியாசமான கருத்துள்ள பல்லாயிரக்கணக்கான படங்களைக் காண முடிகின்றது.



எதிலும் நமக்கென்று ஒரு கருத்தும் இருக்கும் அல்லவா? அதையும் இங்கே குறிப்பிடுகின்றேன். ஆனால் இது எனது தனிப்பட்ட சொந்தக் கருத்தாகும். என்னைப் பொருத்தவரை *Analysis* என்பது நாம் என்ன காண்கின்றோம் என்று அறிந்து அதைக் காண்பதாகும். *Analytics* என்பது நாம் என்ன காண்கின்றோம்

என்பதை அறியாமல் அதைப் புதியதாகக் கண்டறிந்து வெளியிடுவதாகும்! அதாவது ஆய்வுக்கும் ஆராய்ச்சிக்கும் இருக்கும் வேறுபாடை நான் இவ்விரு சொற்களின் பயன்பாட்டின் மூலம் உணர முடிகின்றது. மற்றவர்கள் இதுவரை கண்டிராத ஒரு கண்ணோட்டத்தில் தரவைப் பயன்படுத்திக் கண்டறிந்து அதைச் சரியான நேரத்தில் சரியான முறையில் பயன்படுத்திப் பலரும் பயனடைவதே சிறப்பு. அதையே பகுப்பாய்வு செய்ய வேண்டும்.

கூகுள் நிறுவனம் தரவுப் பகுப்பாய்வின் படிநிலைகளை இவ்வாறு வரையறுக்கின்றது. சுருக்கமாக *APPASA* என்றும் அழைக்கப்படுகின்றது.

Ask - கேள்வி

Prepare - தயாரித்தல்

Process - தகவமைத்தல்

Analyze - பகுப்பாய்வு

Share - பகிர்வு

Act – செயல்பாடு

எந்த ஒரு விடையின் தராதரமும் கேள்வியின் தராதரத்தை ஒட்டியே இருக்கின்றது. ஒரு புத்திசாலித்தனமான கேள்விக்கு ஒரு புத்திசாலித்தனமான பதிலே விடையாகக் கிடைக்கக்கூடும். எனவே தகுந்த வித்தியாசமான கோணத்தில் கேள்வி கேட்பதில் ஆரம்பித்து, அதற்கான தரவுகளை எங்கெங்கு கிடைத்தாலும் அதனைப் பக்குவமாகச் சேகரித்து, பின்னர் நாம் ஏற்கனவே கண்ட பதப்படுத்தும் முறைகளைக் கொண்டு தகவமைத்து அதனை மீண்டும் மீண்டும் ஆய்வுக்கு உட்படுத்தி நாம் மட்டுமே மஞ்சள் கண்ணாடி கொண்டு உலகைக் காண்பது போல் காணாமல் மற்றவர்களுக்கும் அதைப் பகிர்ந்து அவர்களின் கருத்துகளையும் உள்வாங்கிப் பின் தேவையான முடிவுகளை எடுத்து அதைச் செயல்பாட்டில் சாதித்துக் காட்டுவதே கூகுளின் தரவுப் பகுப்பாய்வுப் படிநிலையின் உன்னதம் ஆகும். உலகின் முன்னணி நிறுவனமான கூகுளால் முடியுமென்றால் நம்மாலும் முடியும் தானே?

தரவுச் சுரங்கம் – 18

சென்ற பகுதியில் தரவுப் பகுப்பாய்வு குறித்து விளக்கமாகக் கண்டோம். அடுத்ததாக பகுப்பாய்வின் வகைகளைக் குறித்து காண்போம். நோக்கங்கள், வழிமுறைகள், செயல்பாடுகள், அதற்கான கருவிகள் ஆகியவற்றைக் கருத்தில் கொண்டு பகுப்பாய்வினைப் பலவகைகளாகப் பிரிக்கின்றார்கள்.

1. ***விளக்கப் பகுப்பாய்வு*** *(Descriptive Analytics): இருக்கின்ற தரவிலிருந்து "அப்படி என்ன தான் நடந்திருக்கின்றது?" என்ற கேள்விக்கான விடை தேட முற்படுவது விளக்கப் பகுப்பாய்வாகும். இதற்கு முக்கியமாக அடிப்படை புள்ளியியல் அளவைகளும் சூத்திரங்களும் பயன்படுத்தப்படுகின்றன. எவ்வாறு ஒரு துணியைத் தைப்பதற்கு முன்னர் தையல் வல்லுநர் ஒருவரை அளந்து பார்த்துச் சில எண்களை மட்டும் குறித்துக் கொண்டு பின்னர் தேவையான நேரத்தில் அதைப் பயன்படுத்திக் கொள்கின்றாரோ அதே போல் நாமும் ஒட்டு மொத்தத் தரவிற்கும் சில எண்களைக் கண்டுபிடித்து அது குறித்து ஆராய முற்படுகின்றோம். அனைத்துத் தரவுகளையும் ஒரே நேரத்தில் அலசி ஆராய்வது என்பது கடற்கரையில் மணலை எண்ணுவது போல ஆகிவிடும். அது நம்மால் சாத்தியமாகாது. இந்தப் பகுப்பாய்வில் பயன்படுத்தப்படும் அளவைகள்:*

a. *சராசரி* (Average)
b. *முகடு* (Mode)
c. *இடைநிலை* (Median)
d. *கால்மானங்கள்* (Quartiles)
e. *சதமானங்கள்* (Percentiles)
f. *திட்டவிலக்கம்* (Standard Deviation)
g. *சிறுமம்* (Minimum)
h. *பெருமம்* (Maximum)
i. *நிகழ்வெண் பரவல்* (Frequency Distribution)
j. *கோட்டம் – கோணல் தன்மை* (Skewness)
k. *முகட்டளவு – தட்டை அளவு* (Kurtosis)
l. *தண்டு இலை படம்* (Stem-Leaf picture)
m. *கட்ட வரைபடம்* (Box Plot)
n. *அலைவெண் செவ்வகப்படம்* (Histogram)

மேற்கண்ட அளவைகளுடன் மொத்தம், சதவீதம் போன்ற பல கணித சூத்திரங்களையும் பயன்படுத்துவதன் மூலம் என்ன நடந்து கொண்டிருக்கின்றது என்று அறிந்து கொள்ள முடியும்.

2. ***பரிசோதனைப் பகுப்பாய்வு*** *(Diagnostic Analytics) அடுத்த கட்டக் கேள்வியாக ஏன் நடக்கின்றது அல்லது எப்படி நடக்கின்றது என்று முன்னேறிச் செல்வது அடுத்த வகை பகுப்பாய்வு ஆகும். இதற்கும் சில புள்ளியியல் அளவைகள் பயன்படுத்தப்படுகின்றன. அவற்றுள் சில:*

a. *ஒட்டுறவு – காரணம்* (Correlation – Causation)
b. *கருதுகோள்* (Hypothesis testing)
c. *போக்குப் பகுப்பாய்வு* (Trend Analysis)
d. *தொடர்புப்போக்கு* (Regression)
e. *வேற்றுமைப் பகுப்பாய்வு* (Analysis of Variance -ANOVA)
f. *விகிதம்* (Ratios)

இதன்மூலம் விளைவுகளையும் நடவடிக்கைகளையும் ஆராய்ந்து ஏன், எதனால் இப்படி நடக்கின்றது என்று கண்டுகொள்ளலாம். மேலும் தரவுகளுக்கு இடையேயான உறவுகளையும் கண்டறிய முடிகின்றது.

3. ***முன்கணிப்புப் பகுப்பாய்வு****: (Predictive Analytics) இது வரை என்ன நடந்தது, ஏன் நடந்தது என்று கண்டோம். இனி முக்கியமாக அடுத்ததாக என்ன நடக்கும் என்று கணிப்பது தான் இந்தப் பகுப்பாய்வாகும்.*

எதிரதாக் காக்கும் அறிவினார்க்கு இல்லை

அதிர வருவதோர் நோய்

எனும் வள்ளுவர் குறளுக்கிணங்க அடுத்து என்ன வரப்போகின்றது என்று ஊகித்து அதற்கேற்றாற்போல் முடிவெடுக்க வல்லோர் வெற்றியை நோக்கிய பயணத்தின் அடுத்த கட்டத்தை எட்டுகின்றார்கள். இதில் பாங்கு கண்டறிதல் (Pattern Identification) முக்கியமானதாகும். பயன்படுத்தப்படும் புள்ளியியல் அளவைகள்:

முன்னறிவிப்பு/முன்கணிப்பு, *(Forecasting)*

இயல்பரவல் *(Normal Distribution)*

நிகழ்தகவு *(Probability)*

கணிச மட்டம் *(Level of Significance)*

நம்பிக்கை எல்லை/இடைவெளி *(Confidence Limit/Interval)*

வழிமொழி பகுப்பாய்வு *(Prescriptive Analytics)* இது ஒரு முக்கியமான பகுப்பாய்வாகும். இதில் அடுத்து நாம் என்ன செய்ய வேண்டும் என்று வழிமொழிய முனைகின்றோம். இது பகுப்பாய்வின் உச்சகட்டம் எனலாம். என்ன தான் நடக்கும் நடக்கட்டுமே என்று இருந்து விட முடியாதல்லவா? நாம் அதற்கு என்ன செய்யவேண்டும் என்று முடிவெடுத்து அதன்படி நடக்கவேண்டும். யூடியூப், கூகுள் மேப் ஆகியவை அடுத்து நாம் எந்த அசைபடத்தைப் பார்க்கவேண்டும் என்றோ அடுத்து எந்த வழி திரும்பவேண்டுமென்றோ தானியங்கியாகச் சொல்ல என்ன மாதிரி பகுப்பாய்வை மேற்கொள்ள வேண்டும் என்று எண்ணிப் பார்க்க வேண்டும். இதற்குப் பல மாதிரிகள் *(Models)*, கணித/புள்ளியியல் சமன்பாடுகள், கோட்பாடுகள் பயன்படுத்தப்படுகின்றன. எக்செல் மென்பொருளில் இருக்கும் தீர்வியை *(Solver)* இங்கே குறிப்பிடலாம். செய்யறிவும் *(AI)*, இயந்திரக்கற்றலும் *(Machine Learning)*, ஆழக்கற்றலும் *(Deep Learning)* இதை நோக்கியே பயணிக்கின்றன.

இத்தனை பாகங்களாக தரவுச்சுரங்கம் பற்றிய ஒரு நீண்ட அறிமுகத்தை வழங்கியிருக்கின்றோம். இனி சுரங்கம் தோண்டுவதற்கான பணியினை ஆரம்பிக்கலாம். ஒரு நிறுவனத்தில் தரவுகளின் பயன்பாடு கீழ்க்கண்ட இரண்டு வகைகளில் ஒன்றாக இருக்கின்றது.

1. ***நிகழ்நிலை நடப்புச் செயல்முறை (Online Transaction Processing – OLTP)***

நடவடிக்கைகள் நடந்தவுடன் நமக்குத் தேவைப்படும் அறிக்கைகளுக்காக இந்தச் செயல்முறை இருக்கின்றது. எடுத்துக்காட்டாக ஒரு வாடிக்கையாளர் சரக்கு வாங்கும் போது அவருக்குத் தருவதற்காக ஒரு பட்டி அல்லது ரசீது தரவேண்டியிருக்கின்றது. அதன் பின்னர் அதைக் கணக்குப் பதிவு செய்ய வேண்டியிருக்கின்றது. அன்றாடம் நமது நிறுவனம் செவ்வனே நடைபெறுவதற்காக இந்தச் செயல்முறை நடைமுறையில் இருக்கின்றது. இது உறவுமுறைத் தரவுத்தள மேலாண்மையைக் *(Relational Database Management System)* கொண்டு இயங்குகின்றது. உருபொருட்களை *(Entities)* மையப்படுத்தியும் அவற்றின் பண்புகளைச் *(Attributes)* சுற்றிலும் அமையுமாறு உள்ளது. இதற்கு நிரைகளை அடிப்படையாகக் கொண்டுஇயங்கும் உறவுமுறைத் தரவுத்தளம் *(Row based Databases)* பொருத்தமானதாக உள்ளது. அதாவது ஒரு ரசீதை நாம் அச்சிட வேண்டுமானால் நமக்கு வாடிக்கையாளரின் பெயர், முகவரி, வரிவிதிப்பு எண் என்று ஒட்டு மொத்தத் தரவும் தேவைப்படுகின்றது.

2. ***நிகழ்நிலை பகுப்பாய்வுச் செயல்முறை (Online Analytical Processing – OLAP)***

நடவடிக்கைகள் நடந்து முடிந்தவுடன் எதிர்காலத் தேவைக்காக இரண்டாம் வகை தரவு வடிவமைப்பு அவசியமாகின்றது. ஏனெனில் பகுப்பாய்வு செய்வதற்கு நமக்கு ஒவ்வொரு வாடிக்கையாளரின் முகவரி, வரிவிதிப்பு எண் போன்ற துல்லியமான தரவுகள் தேவைப்படாது. நடவடிக்கைகள் குறித்த செய்தித் தரவுகள் மட்டுமே தேவைப்படும். எனவே இங்கே செய்தியை *(Facts)* நடுவில் வைத்து உருபொருட்களைச் *(Entities)* சுற்றிலும் அமைக்கும் வழிமுறை தேவைப்படுகின்றது. இவ்வுருபொருட்களும் பின்னால் பரிமாணங்களில் *(Dimensions)* சிலவாக உருவெடுக்கும் விந்தையும் காணப் போகின்றோம்.

இரண்டு வகைகளுமே ஒரு நிறுவனம் செவ்வனே இயங்குவதற்குத் தேவை என்றாலும் அதே தரவையே அடிப்படையாகக் கொண்டு இரண்டுமே வடிவமைக்கப்படுகின்றது என்றாலும் அடிப்படையில் இவற்றுக்கிடையே மலைக்கும் மடுவுக்குமான வேறுபாடுகளைக் காண முடிகின்றது. அவற்றை இங்கே பட்டியலிடுவோம்.

வ. எண்	***நிகழ்நிலை நடப்புச் செயல்முறை (OLTP)***	***நிகழ்நிலை பகுப்பாய்வுச் செயல்முறை (OLAP)***
1	அன்றாட நடவடிக்கைகளுக்காகப் பயன்படுத்தப்படுகின்றது.	எதிர்காலத் திட்டமிடலுக்கும் பகுப்பாய்வுக்கும் பயன்படுத்தப்படுகின்றது.
2	அன்றாடத் தரவுகளை மட்டுமே பயன்படுத்துகின்றது.	தேவைப்பட்டால் நிறுவனம் தொடங்கிய நாளில் இருந்து இன்று வரையான வரலாற்றுத் தரவுகளும் பயன்படுத்தப்படுகின்றது.
3	அறிக்கைகளுக்குப் பொதுவாக தரவுத்தள மேலாண்மையைப் பயன்படுத்துகின்றது. *(DBMS)*	அறிக்கைகளுக்கு தரவுக்கிட்டங்கிகள் தேவைப்படுகின்றன. *(Data warehouses)*
4	இதன் தேவை அன்றாட மென்பொருள் பயன்பாட்டுக்கானதாகும். *(Application Oriented)* தொழில் நடைமுறைக்குப் பயன்படுகின்றது.	இது புலம் சார்ந்த மேலான *(Subject Oriented)* பயன்பாட்டுக்காகும். திட்டமிடல், முடிவெடுத்தல், பகுப்பாய்வு செய்தல் ஆகியவற்றுக்குப் பயன்படுகின்றது.
5	தரவுத் தள இயல்பாக்கம் *(Database Normalization)* அவசியமாகும்.	இயல்பாக்கம் இருக்க வேண்டிய அவசியமில்லை.
6	இன்றைய தொழிலின் நிலையை அறிய உதவும் ஒரு கண்ணாடி போன்றது..	இது பல்பரிமாணத் தரவுப் பெட்டகமாகத் திகழ்கின்றது.
7	எப்போதுமே *100* சதவீதத் தரவுகளைப் பயன்படுத்தப்போவதில்லை ஆதலால் மிகவும் வேகமாக இயங்கும்*!*	அதிகப்படியான தரவுகளைப் பயன்படுத்துவதால் சற்று மெதுவாகவே இயங்கும். மணிக்கணக்கில் கூட ஆகலாம்*!*
8	இதைப் பயன்படுத்துபவர்கள் நிறுவனத்தின் அடிமட்டப் பணியாளர்கள்.	இதைப் பயன்படுத்துபவர்கள் திட்டமிடும் அளவில் இருக்கின்ற மேல்மட்டப் பணியாளர்கள்.
9	தரவில் வாசித்தல் மற்றும் எழுதல் *(Read and Write)* ஆகிய இரு பணிகளும் அடிக்கடி நடக்கும்.	இதில் எழுத வேண்டிய அவசியமில்லை. வாசித்தல் மட்டுமே நடைபெறும்.

10	நிரைவரிசைத் தரவுகள் பொருத்தமானதாகும். *(Row based Database)*	நெடுவரிசைத் தரவுகள் *(Columnar Databases)* பொருத்தமானதாகும்.
11	உருபொருளை மையமாகக் கொண்டு, பண்புகளைச் சுற்றிலும் கொண்ட வடிவமைப்பில் இயங்குகின்றது. *(Entity Centered.)*	நடவடிக்கைச் செய்திகளை மையமாகக் கொண்டு பரிமாணங்களைச் சுற்றிலும் கொண்ட வடிவமைப்பில் இயங்குகின்றது. *(Fact Centered)*
12	இயல்பாக்கத்தினால் உருபொருள் உறவுமுறை வரைபடத்தின் மூலம் வரையப்படுகின்றது. *(Entity Relationship Diagrams)*	எளிய நட்சத்திர வடிவமைப்பிலிருந்து பனித்துகள் வடிவமைப்பு மற்றும் சிக்கலான நட்சத்திரத் திரள் வடிவமைப்பு வரை நீண்டு கொண்டே செல்லக்கூடும். *(Star/Snowflake/Cluster Schema)*

மேற்கண்ட வேறுபாடுகளிலிருந்து நிகழ்தரவு பகுப்பாய்வின் சிறப்பம்சங்களை அறிந்து கொள்ளலாம். மேலும் பல ஆண்டுகளின் தரவுகளையும் சேர்க்கும் போது சில சிக்கல்களும் ஏற்படுகின்றன. அவைபற்றிய விளக்கங்களை அடுத்தடுத்த பகுதிகளில் காணலாம்.

ஓராண்டுத் தரவை மட்டும் ஆய்வு செய்யும் போது எந்தப் பெரிய குழப்பமும் நேர்வதில்லை. ஆனால், ஓராண்டுத் தரவை மட்டும் ஆய்வு செய்தால் ஒரு துணியின் தன்மையைப் பற்றி அறிந்து கொள்ள அதன் ஒரு நூலை மட்டும் எடுத்து அறிந்து கொள்ள முனைவது போலாகும். ஒரு துணியின் சீரான வடிவமைவும் பாங்கும் *(Pattern)*, நிறங்களும் கோடுகளும் பற்றி அறிந்து கொள்ள வேண்டுமாயின் ஒரு நூலை மட்டும் வைத்து முடிவுக்கு வர முடியாது. ஓராயிரம் நூற்களை இணைத்துப் பார்த்தால் உடனே அறிந்து கொள்ள முடியும். இதே முறையில் தான் பெருந்தரவும்*(Big data)* இயங்குகின்றது. எனவே பல ஆண்டுகளாக இயங்கும் ஒரு நிறுவனம், அது துவங்கிய நாளில் இருந்து தன் வசம் இருக்கும் தரவைச் சரியாகப் பயன்படுத்திக் கொள்ள வேண்டும் என்பது தான் தரவுச்சுரங்கத்தின் அடிப்படையாகும்.

பிரபல நிகழ்நிலை அசைபட இணையத் தளமான யூடியூபில் எவ்வாறு அவரவர்களுக்கு இயைந்தவாறு அசைபடம் பரிந்துரைக்கப்படுகின்றது? கோடிக்கணக்கான மக்களின் விருப்பங்களையும் ஆராய்ந்து அவ்வாறு முடிவெடுக்க எத்தனை ஆய்வுகள் பின்னால் நடைபெற்றிருக்க வேண்டும்? இதற்கெல்லாம் தேவை ஒரு தரவுக்கிட்டங்கி *(Data Warehose)*. அதில் தரமாக்கப்பட்ட தரவு *(Standardized Data)* சீராய் தேவைக்குத் தகுந்தவாறு அடுக்கி வைக்கப்பட்டிருக்க வேண்டும். அது கோடிக்கணக்கான வரிகளைக் கொண்டு பல கணினிகளின் நினைவகங்களிலும் விரவி வைக்கப்பட்டிருக்கலாம். எனவே தான் அதைப் பெருந்தரவு என்று அழைக்கின்றோம். அதற்குண்டான வன்பொருள் *(Hardware)*, மென்பொருள் *(Software)*, முறைமை *(Logic)* அனைத்தும் தினந்தோறும் வளர்ந்து கொண்டே இருப்பதைக் கண்கூடாகக் காண முடிகின்றது. சாமான்னியர்களும் இந்தப் பெருந்தரவிலிருந்து தரவைத் தேவைக்கேற்ப எடுத்துக் கொள்ள ஏதுவாக பயன்பாட்டு நிரல் இடைமுகங்கள் *(API – Application Program Interface)* அனைத்து பிரபலமான சமூக ஊடகங்களுக்கும் கிடைக்கின்றன. எடுத்துக்காட்டாக ஒரு நிறுவனத்தைப் பற்றி பேஸ்புக்கில் என்ன பேசுகின்றார்கள் என்பதை ஒரு சிறு பைத்தான் மொழி நூலகப் பொதியின் மூலம் நமது கணினிக்கு எளிதாக நிகழ்நிலையில் கொண்டு வரமுடியும்.

இவ்வாறு பெருந்தரவைத் தொகுக்கும் போது அல்லது உருவாக்கும் போது நமக்கு ஏற்படும் இடையூறுகளை இந்தப் பகுதியில் காண்போம்.

மெதுவாய் மாறும் பரிமாணங்கள்: *(Slowly Changing Dimensions - SCD)* தரவை ஆய்வு செய்வதற்கு அதை நான்கு அம்சங்களாகப் பிரித்து அதில் ஒன்று தான் பரிமாணம் என்று பார்த்தோம். அந்தப் பரிமாணம் தான் நிலையானது என்றும். தரவு ஒரு நடவடிக்கையின் செய்திகளைச் சேகரிப்பது என்றால், அதற்கு அர்த்தம் உண்டாக்குவது பரிமாணம் என்றும் பார்த்தோம். ஆனால், இப்பூவுலகில் எதுவும் நிரந்தரமன்று என்பதை நாம் உணர்வோம்! அது போல் பரிமாணங்கள் நிரந்தரமாக இருப்பதில்லை. எடுத்துக்காட்டாக இன்று சென்னையில் இருக்கும் நமது வாடிக்கையாளர், அடுத்த ஆண்டு மதுரையில் குடியேறிவிடலாம். அதனால் அவர் நமது வாடிக்கையாளராக இல்லாது போக மாட்டார். இப்போது அவர் தொடர்பான தரவை சென்னையில் சேர்ப்பதா, மதுரையில் சேர்ப்பதா எனும் குழப்பம் நேரிடுகின்றது அல்லவா? இதைத் தான் மெதுவாய் மாறும் பரிமாணங்கள் என்கின்றனர். இதைச் சரி செய்ய நிறைய வழிமுறைகள் உள்ளன. வகை *0* லிருந்து வகை *7* வரை புதியதாக பல முறைகள் இருக்கின்றன. அதைச் சரி செய்யவில்லை என்றால் நமது ஆய்வு முடிவுகள் சரியானதாக இருக்காது.

தரவு நகல்கள்: *(Duplicate data)* ஒரு நிறுவனத்தின் மேலாளர் ஒவ்வொருவரும் அவர்களுக்கேற்றவாறு தரவுக் கிட்டங்கியினை வடிவமைத்தோ அல்லது பயன்படுத்தவோ முயலும் போது இயல்பாகவே ஒரே தரவு பல நகல்கள் எடுக்கப்பட்டு நிறுவனத்தின் கணினியை ஆக்கிரமிப்பு செய்து கொள்கின்றது. அதுவும் ஏற்கனவே பெருந்தரவே ஒரு பிரச்னை தான். அதை நகல்கள் வேறு எடுத்தால்? சொல்லவே தேவையில்லை!

தேவையா இல்லையா?: *(Redundant Data)* ஒரு சமயத்தில் தேவை என்று நினைக்கும் ஒன்று இன்னொரு சமயத்தில் தேவையில்லாததாகத் தோன்றுவதைக் காண்கின்றோம். தேவையான தரவு என்று தேவையில்லாததைச் சேர்த்து வைத்தாலும் பிரச்னை தான். தேவையில்லாத தரவு என்று தேவையானதை ஒதுக்கி வைத்து விட்டாலும் பிரச்னை தான். எனவே தேவை அறிந்து தரவைத் தரமாக்கம் செய்தல் வேண்டும்.

செலவினம்: ஒரு தரவுக்கிட்டங்கியை உருவாக்கி அதை நிர்வகித்து அதிலிருந்து பலன் பெறுவது என்பது அதிகமான செலவாகும்.எந்த ஒரு முயற்சியிலும் பலன் அறிந்து, அந்தப் பலனானது செலவை விட அதிகமாக இருக்கின்றதா என்று அறிந்து முன்னெடுப்பது காலத்தின் கட்டாயமாகும்!

சிக்கல்: ஒரு பெரு நிறுவனத்தின் அத்தனை தரவையும் தரமாக்க முற்படுவது அத்தனை எளிதானதல்ல. பல சிக்கல்களை வழிவகுக்கும். அனைத்திற்கும் ஏற்றவாறு வளைந்து கொடுத்துக் கொண்டே இருந்தால் பிள்ளையார் பிடிக்கக் குரங்காகவும் மாறிவிடக்கூடும்.

நேரவிரையம்: *ஏற்கனவே குறிப்பிட்டது தரவு பகுப்பாய்வு அதிக நேரம் எடுத்துக் கொள்ளும். அதற்கான பொறுமை மிகவும் முக்கியமாகும். அதே போன்று தான் தரவுக்கிட்டங்கியில் தரவை நேர்த்தியாக அடுக்கி வைப்பதற்கும் நேரமாகும்.*

பொருந்தாத் தரவு: *(Irrelevant Data) பல்லாண்டுத் தரவுகளைச் சேமித்து ஆய்வு செய்யும் போது மிகவும் பழமையான தரவு தேவைப்படாமலும் பொருந்தாலும் போகலாம். 'பழையன கழிதலும் புதியன புகுதலும் வழுவல' என்பது தரவுத் தரமாக்கத்துக்கும் பொருந்தும்!*

பாதுகாப்பு: *ஒரு நிறுவனத்தின் ஒட்டுமொத்தத் தரவும் ஓரிடத்தில் சேர்த்து வைப்பதென்றால் அதை எவ்வாறு பாதுகாக்க வேண்டும் என்னும் பிரச்னையும் உடனே வந்து விடுகின்றது. தேவையானவர்கள் மட்டுமே தேவையான தரவைப் பார்க்க முடியவேண்டும் என்பதும், தனியுரிமையைப் பேண வேண்டும் என்பதும் மிகப் பெரிய சவாலாக அமைகின்றது.*

இவ்வாறு பல இடையூறுகள் வந்தாலும் தரவுக்கிடங்கின் நன்மைகளைக் கணக்கில் கொண்டால், தரவுக்கிடங்கின் மேன்மை நமக்குத் தெரியவரும். அதை அடுத்த பகுதியில் காண்போம்!

தரவுச்சுரங்கம் – 21

தரவுச் சுரங்கத்தில் ஒரு முக்கியமான பங்கு தரவுக்கிடங்குக்கு உண்டு. அதனைப் பராமரிப்பதில் இருக்கும் சிக்கல்களை முந்தைய பகுதியில் பார்த்தோம். இப்போது அந்தத் தடைகளைத் தாண்டி அதைச் சரியாகப் பயன்படுத்தினால் ஏற்படும் நன்மைகளைக் காண்போம். இந்த நன்மைகளைக் காணும் போது பல்லாண்டுகளாகக் கணினியைப் பயன்படுத்தி நிர்வகிக்கும் ஒவ்வொரு நிறுவனமும் தமக்கேற்ற தரவுக் கிடங்கைப் பராமரிப்பது அவசியம் என்பது புரியும்.

தரவுக் கிடங்கின் நன்மைகள்:

1. ***தரவு சார்ந்த முடிவெடுத்தல் (Data Driven Decision Making):*** *ஒரு தரவுக் கிடங்கு நிறுவனத்தின் அனைத்து தரவுகளுக்கும் ஒரே உண்மை மூலமாக செயல்படுகிறது, இது மேலாளர்கள் காண்பதற்கு ஏதுவாக வழக்கமான அறிக்கைகள் மற்றும் பார்வைகளை உருவாக்க உதவுகிறது. இது வணிகங்கள் தயாரிப்பு மேம்பாட்டிலிருந்து சந்தைப்படுத்தல் நுகர்வோரைச் சென்று சேர்வது வரை அனைத்திலும் சிறந்த முடிவுகளை எடுக்க உதவும்.*

2. ***திறன் மேம்பாடு (Increased Efficiency):*** *தரவு சேகரிப்பு, செயலாக்கம் மற்றும் பகுப்பாய்வு ஆகியவற்றில் ஈடுபட்டுள்ள பல பணிகளை ஒரு தரவுக் கிடங்கு தானியக்கப்படுத்த முடியும். இது ஊழியர்களின் நேரத்தை மிகவும் சேமித்து அவர்களது நேரத்தை மேலும் சமயோசிதமான பணிகளில் கவனம் செலுத்த அனுமதிக்கும்.*

3. ***குறைந்த செலவு:*** *ஒரு தரவு கிடங்கு நிறுவனங்களுக்கு பல்வேறு வழிகளில் செலவைக் குறைக்க உதவும். எடுத்துக்காட்டாக, நிறுவனங்கள் திறனற்ற தன்மைகளை அடையாளம் காண்பதற்கும் அவற்றை நீக்குவதற்கும், அவற்றின் விநியோக சங்கிலிகளை மேம்படுத்துவதற்கும் உதவும்.*

4. ***வாடிக்கையாளர் சேவை மேம்பாடு:*** *வாடிக்கையாளர்களின் தேவைகள் மற்றும் விருப்பங்களை நன்கு புரிந்துகொள்ள ஒரு தரவுக் கிடங்கு நிறுவனங்களுக்கு உதவும். இது மேம்படுத்தப்பட்ட வாடிக்கையாளர் சேவை மற்றும் வாடிக்கையாளர் திருப்திக்கு வழிவகுக்கும். மீண்டும் மீண்டும் நம்மை நோக்கியே வருபவர் தானே வாடிக்கையாளர்? ஒவ்வொரு வாடிக்கையாளரிடமிருந்தும் பல்லாண்டுகள் தொடர்பை உருவாக்கி மேம்படுத்தி சிறந்த சேவையளிப்பதன் மூலம் வாடிக்கையாளரையும் லாபத்தையும் தக்க வைத்துக் கொள்ளலாம்.*

5. ***சந்தைச் சவால்கள் சமாளிப்பு:*** *மிகுந்த சவால் மிக்க சந்தை கொண்ட இன்றைய உலகில், சிறந்த முடிவுகளை விரைவாக எடுக்க உதவுவதன் மூலம் தரவுக் கிடங்கு நிறுவனங்களுக்கு முன்னோக்கிச் செல்ல வழிகாட்டியாக விளங்கும்.*

6. ***இடர் மேலாண்மை:*** *இன்றைய உலகில் ஏமாற்றுபவர்களின் எண்ணிக்கை பல மடங்கு அதிகரித்து இருக்கின்றது. நமக்கான இக்கட்டு நிறுவனத்தின் உள்ளே, வெளியே என்று பல கோணங்களிலிருந்து ஏற்பட வாய்ப்பிருக்கின்றது. தரவுக் கிடங்கின் மூலம் ஒட்டு மொத்தத் தகவல்களையும் ஒரே இடத்தில் காண முடிவதால் பல கோணங்களிலும் பணியாளர்கள், வாடிக்கையாளர்கள், நமக்கு சரக்கு மற்றும் சேவையளிப்போரைக் கண்காணித்து அறிந்து கொள்ள முடியும். இதனால் மோசடிகளை மிக விரைவில் கண்டறிந்து களவையும் கழிவையும் குறைக்க இயலும்.*

7. ***ஆராய்ச்சியில் சிக்கனம்:*** *எப்போதும் புதிய கண்டுபிடிப்புகளையே நாடி நிற்கும் நிறுவனங்கள் (எடுத்துக்காட்டு மருத்துவம், வாகன உற்பத்தி) தங்கள் ஆராய்ச்சிக்குச் செலவிடும் தொகையைக் குறைக்க இந்தத் தரவுக் கிடங்கிகள் உதவி புரிகின்றன. பெருந்தரவைப் பயன்படுத்துவதன் மூலம் தேவையற்ற செலவுகளைக் குறைக்கலாம்.*

8. ***தரவின் தரம்:*** *தரவென்றாலே அது தரப்படுத்தப்பட்டது என்று முன்பே அறிவோம். இருந்தாலும் அதை இன்னும் துப்புரவு செய்தால் இன்னும் பொலிவாகத் திகழும் அல்லவா? ஒரு சாதாரண பொருள் சேமிப்பு அறைக்கும் ஒரு பெரிய அறிவியல் சார்ந்த கிட்டங்கிக்கும் இருக்கும் மலைக்கும் மடுவுக்குமான வித்தியாசத்தை நாம் உணர வேண்டும். சாதாரண அறையில் மூட்டை மூட்டையாக அடுக்கி வைக்க இயலும். ஆனால் ஒரு துறைமுகத்திலிருக்கும் ஆயிரக்கணக்கான கப்பல்களில் ஏற்றும் வகையில் பொருட்களை அடுக்கி வைக்க மூட்டைகள் காணாது அல்லவா? மிகப்பெரிய இரும்புக் கொள்கலன்களில் கொண்டு செல்வதைக் காண்கின்றோம். அதே போன்று தரவமைப்பிலும் செய்து தரப்படுத்துவதால் தரக்கிடங்குகளில் இருக்கும் தரவுகளின் தரம் உயர்ந்ததாக நம்பகமானதாக இருக்கின்றது.*

9. ***மேம்பட்ட பாதுகாப்பு:*** *தரவுக் கிடங்கில் இருக்கும் தரவு சேமித்துப் பின்னர் பயன்படுத்துவதற்கு மட்டுமே என்பதால் முதலில் "காண்பதற்கு மட்டும்" என்று கணினியில் சேமித்து யாரும் மாற்ற*

முடியாதவாறு செய்கின்றோம். மேலும் அனைத்து தரவையும் ஒரே இடத்தில் குவித்து வைத்து தேவையானவர்களுக்கு தேவையான தரவுக்கு உண்டான அனுமதியை மட்டும் வழங்குவதால் பாதுகாப்பு எளிதாகின்றது.

10. தற்போதைய மேக் கணிமைத் தொழில்நுட்பத்தில் சரியாக வடிவமைத்து விட்டால் மிகச் சிறிதான ஒரு கிடங்கிலிருந்து ஆரம்பித்து நிறுவனம் வளர வளர தரவுக் கிடங்கினையும் தேவைக்கேற்ப பெரிதாக்கிக் கொள்ள இயலும் என்பதால் குறைந்த செலவில் நிறைந்த பயனை அடைய முடியும்.

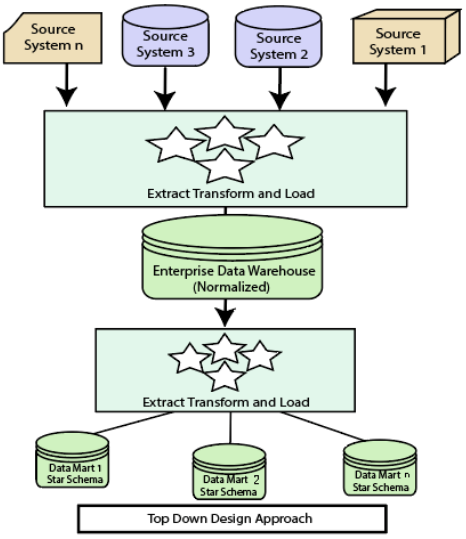
சற்றே கூர்ந்து கவனித்தால் எதையெல்லாம் குறையென்று சென்ற பகுதியில் குறிப்பிட்டோமோ அதையே நிறைகளாக இங்கே காண்கின்றோம். எனவே தரவுக் கிடங்கானது பூங்கொத்தாக மாறுவதும் குரங்கு கை பூமாலை ஆவதும் அதைத் திட்டமிட்டுப் பயன்படுத்துவோரின் கையில் உள்ளது என்தை அறிந்து கொள்ளலாம்!

தரவுச் சுரங்கத்தின் ஒரு முக்கியமான அங்கமான தரவுக் கிட்டங்கியைப் பற்றிச் சற்று விளக்கமாகக் காண்கின்றோம். முந்தைய பகுதிகளில் தரவுக் கிட்டங்கியின் நன்மை தீமைகளைப் பட்டியலிட்டோம். ஒரு நிறுவனத்தின் அனைத்து வரலாற்றுத் தரவுகளின் ஒட்டு மொத்த மூலமாக விளங்கி அதைப் பயன்படுத்தும் அனைவருக்கும் வழங்கத் தயாராக இருக்க வேண்டும் அல்லவா? எனவே அதை உருவாக்கும் முறைகள் பற்றி இந்தப் பகுதியில் காணலாம்.

தரவுக் கிட்டங்கிகளை இரண்டு முறைகளில் உருவாக்கலாம். பில் இன்மானின் (Bill Inmon) மேலிருந்து கீழ் அணுகுமுறை மற்றும் ரால்ஃப் கிம்பாலின் (Ralph Kimball) கீழிருந்து மேல் அணுகுமுறை ஆகிய இரண்டு முறைகளையும் ஒவ்வொன்றாகக் காண்போம்.

பில் இன்மானின் (Bill Inmon) மேலிருந்து கீழ் அணுகுமுறை:

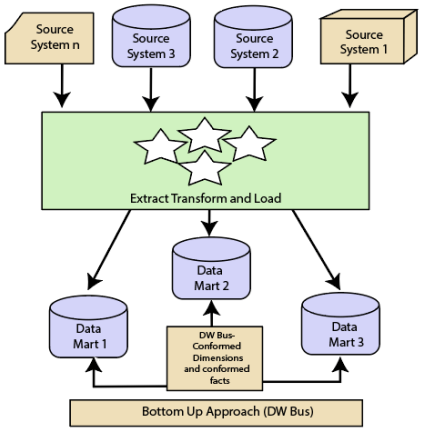
இந்த அணுகுமுறையின் படி, தரவுகளின் மூல வழங்கியிலிருந்து தரவுக் கிட்டங்கியின் வடிவமைப்பு ஆரம்பமாகின்றது. பல்வேறு தரவு மூலங்களிலிருந்தும் பிரித்தெடுத்தல், உருமாற்றம், ஏற்றுதல் ஆகிய மூன்று படிநிலைகளுக்கு உட்படுத்தப்பட்டு ஒட்டு மொத்த நிறுவனத்திற்கான தரவுக் கிட்டங்கி முதலில் வடிவாக்கம் செய்யப்படுகின்றது. பின்னர் அதிலிருந்து மீண்டும் தேவையான தரவு அங்காடிகள் (Data Marts) தமக்குத் தேவையான வகையில் மீண்டும் மேலே கண்ட பிரித்தெடுத்தல், உருமாற்றம், ஏற்றுதல் ஆகிய படிநிலைகளுக்கு உட்படுத்தப்பட்டு நட்சத்திர திட்டங்களாக தரவு மாதிரியைப் (Data Models) பெற்றுக் கொள்கின்றன. இங்கே கொடுக்கப்பட்டிருக்கும் படத்தைப் பார்த்தால் எளிதாகப் புரியும். இந்த அணுகுமுறையின் படி ஒட்டு மொத்த தரவுக் கிட்டங்கியை உருவாக்கிய பின்னர் தேவையான தரவு அங்காடிகள் தமக்குத் தேவையான தரவுகளைப் பெற முடிகின்றது.



Top Down Design Approach

ரால்ஃப் கிம்பாலின் கீழிருந்து மேல் அணுகுமுறை:

இந்த அணுகுமுறையின் படி முதலில் தேவைக்குத் தகுந்தவாறு தரவு அங்காடிகள் உருவாக்கப்படுகின்றன. அவ்வாறு உருவாக்கப்படும் போதே தரவு மாதிரிகளும், பரிமாணங்களும், தேவையான செய்திகளும் நிர்ணயிக்கப்படுகின்றன. அதன் பின்னர், அந்தச் செய்திகளும் பரிமாணங்களும் எந்தத் தரவு மூலங்களில் இருக்கின்றன என்று கண்டறியப்பட்டு பின்னர் பிரித்தெடுத்தல், உருமாற்றம், ஏற்றுதல் ஆகிய படிநிலைகளுக்கு உட்படுத்தப்பட்டு தரவு அங்காடிகளால் பெறப்படுகின்றன. இந்த முறையில் ஒட்டு மொத்த தரவுக் கிட்டங்கி மேலே இல்லாமல் கீழிருந்து சிறு சிறு அங்காடிகளின் தொகுப்பாக அமைவதைக் காணலாம். கீழிருந்து மேலான இந்த அணுகுமுறையில் தரவுக் கிட்டங்கியின் அமைப்பு நடுவில் இல்லாமல் கீழே இருக்கின்றது.



Bottom Up Design Approach

வழக்கம் போல் இந்த இரண்டு அணுகுமுறைகளிலும் நன்மை தீமைகளுண்டு. இரண்டில் எதைத் தேர்ந்தெடுப்பது என்ற கேள்விக்குப் பதிலாக, எந்த ஒரு நிறுவனம் பரிணாம வளர்ச்சியடைந்து முதிர்ச்சியடைந்து அதன் தரவுத் தேவைகளைப் பற்றி நன்கு அறிந்தும் தரவு அமைப்புகளைப் பூர்த்தியாகவும் வைத்திருக்கின்றதோ அவ்வகை நிறுவனத்திற்கு மேலிருந்து கீழ் அணுகுமுறை சிறப்பாக இருக்கும். (எ.கா. இரயில் போக்குவரத்து) அதே சமயத்தில் தற்போது பரிணாம வளர்ச்சி கண்டு வரும், அடிக்கடி மாறிக் கொண்டிருக்கும் தரவு அமைப்பைக் கொண்டிருக்கும் நிறுவனத்திற்கு கீழிருந்து மேலான அணுகுமுறை சிறப்பாக இருக்கும். (எ.கா. தொலைதொடர்பு மற்றும் கணினி) இவ்விரு அணுகுமுறைகளையும் ஒன்றிணைத்து கலப்பின அணுகுமுறைகளும் உருவாகின்றன. எனவே நமது நிறுவனத்திற்கு எது பொருத்தம் என்று உணர்ந்து அதைச் செயல்படுத்துவது நல்லது.

தரவுக் கிட்டங்கிகள் மற்றும் அங்காடிகளுக்கான சூழல் பின்வருவனவற்றை உள்ளடக்கியது. எனவே தரவுக் கிட்டங்கியை உருவாக்கும் போது கவனத்தில் கொள்ள வேண்டும்.

- கிட்டங்கி அல்லது அங்காடிக்குத் தரவை வழங்கும் மூல அமைப்புகள்;
- தரவு ஒருங்கிணைப்பு தொழில்நுட்பம் மற்றும் பயன்பாட்டிற்குத் தரவைத் தயாரிக்கத் தேவையான செயல்முறைகள்;
- ஒரு நிறுவனத்தின் தரவுக் கிட்டங்கி அல்லது தரவு அங்காடிகளில் தரவைச் சேமிப்பதற்கான பல்வேறு கட்டமைப்புகள்;
- பல்வேறு பயனர்களுக்கான பல்வேறு கருவிகள் மற்றும் பயன்பாடுகள்;
- கிட்டங்கி அல்லது அங்காடி அதன் நோக்கங்களைப் பூர்த்திசெய்கிறதா என்பதை உறுதிப்படுத்த, மீதரவு *(Meta Data)*, தரவுத் தரம் மற்றும் நிர்வாகச் செயல்முறைகள் இருக்க வேண்டும்.

மீதரவு எனப்படுவது தரவு பற்றிய தரவு ஆகும். மேனிலைத்தரவு என்றும் அழைக்கப்படுகின்றது. எடுத்துக்காட்டாக ஒரு புகைப்படம் என்பது தரவு ஆனால், அது எந்தப் புகைப்படக் கருவி கொண்டு எடுக்கப்பட்டது என்பது மீதரவு ஆகும். ஒரு தரவுப் பட்டியலின் நெடுவரிசைத் தலைப்புகளை மேனிலைத் தரவு எனலாம்.

அடுத்த பகுதியில் தரவுக் கிட்டங்கியின் படிநிலைகளைக் காணலாம்.