

The 47th International Conference
APPLICATIONS OF MATHEMATICS IN ENGINEERING AND ECONOMICS
(AMEE 2021)

Biomedical Data Integration and Innovations Concept

Stella Vetova

Primary funding for the presented work was provided by the National Science Fund, Ministry of Education and Science, Republic of Bulgaria under contract **KP-06-N37/24**, research project “Innovative Platform for Intelligent Management and Analysis of Big Data Streams Supporting Biomedical Scientific Research”.

Essence of Large-Scale Data

1. Generated heterogeneous amount of data used for future processing.

2. Features of Large-Scale Data :

complexity

variety

volume

application opportunities

Application of Large-Scale Data

-science, business, medicine, biomedicine,
bioinformatics, agriculture, biology

-in the field of medicine big data is gained on the base of
laboratory results, clinical test, patients' exams,
symptoms data captured by the means of telemedicine,
etc.

The 47th International Conference
APPLICATIONS OF MATHEMATICS IN ENGINEERING AND ECONOMICS
(AMEE 2021)

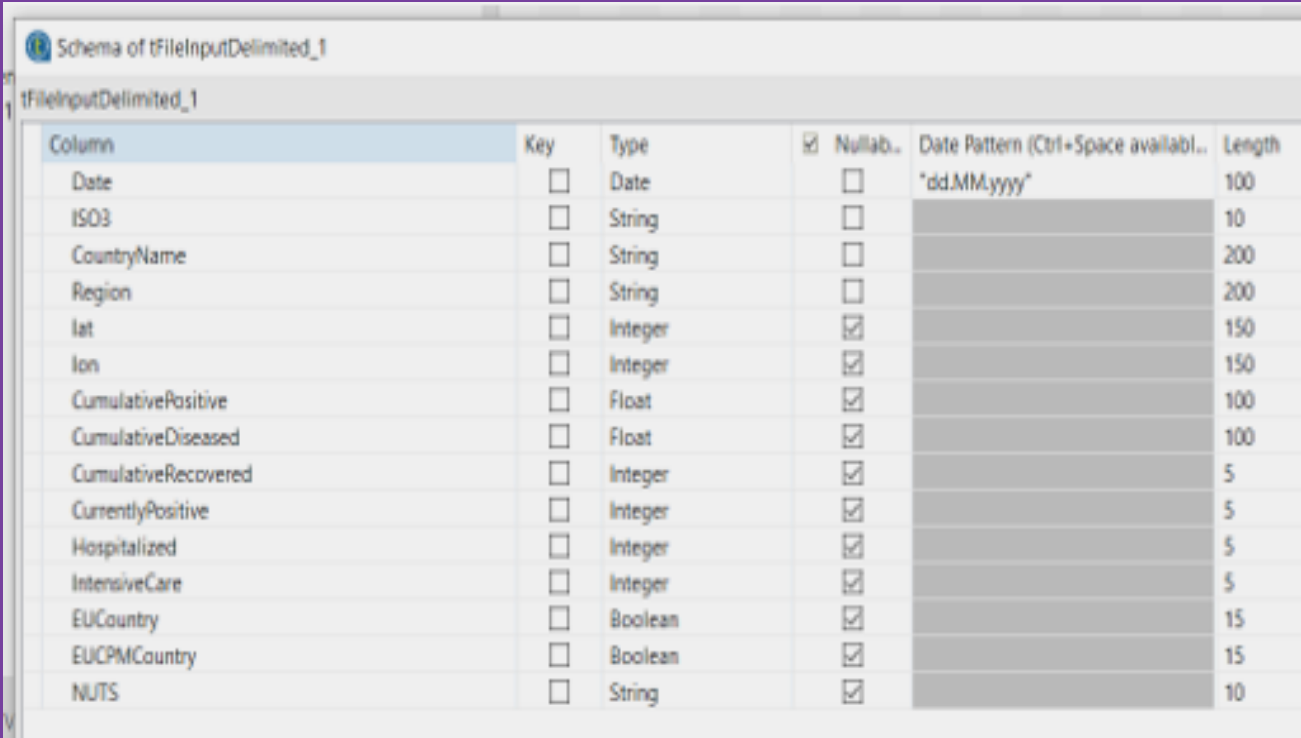
Integration - the process of data capture from heterogeneous sources, combining various types of data, data filtering to eliminate duplicates, format conversion and management on the base of established rules thus, creating a dataset which is accessible by users.

Large-scale Integration stages:

1. data is to be downloaded from its original data source
2. to be formatted according to the new warehouse requirements
3. to be transferred to the warehouse

The 47th International Conference
APPLICATIONS OF MATHEMATICS IN ENGINEERING AND ECONOMICS
(AMEE 2021)

Covid-19 data fields
configuration



Schema of tFileInputDelimited_1

Column	Key	Type	<input checked="" type="checkbox"/> Nullab..	Date Pattern (Ctrl+Space availabl..	Length
Date	<input type="checkbox"/>	Date	<input type="checkbox"/>	"dd.MM.yyyy"	100
ISO3	<input type="checkbox"/>	String	<input type="checkbox"/>		10
CountryName	<input type="checkbox"/>	String	<input type="checkbox"/>		200
Region	<input type="checkbox"/>	String	<input type="checkbox"/>		200
lat	<input type="checkbox"/>	Integer	<input checked="" type="checkbox"/>		150
lon	<input type="checkbox"/>	Integer	<input checked="" type="checkbox"/>		150
CumulativePositive	<input type="checkbox"/>	Float	<input checked="" type="checkbox"/>		100
CumulativeDiseased	<input type="checkbox"/>	Float	<input checked="" type="checkbox"/>		100
CumulativeRecovered	<input type="checkbox"/>	Integer	<input checked="" type="checkbox"/>		5
CurrentlyPositive	<input type="checkbox"/>	Integer	<input checked="" type="checkbox"/>		5
Hospitalized	<input type="checkbox"/>	Integer	<input checked="" type="checkbox"/>		5
IntensiveCare	<input type="checkbox"/>	Integer	<input checked="" type="checkbox"/>		5
EUCountry	<input type="checkbox"/>	Boolean	<input checked="" type="checkbox"/>		15
EUCPMCountry	<input type="checkbox"/>	Boolean	<input checked="" type="checkbox"/>		15
NUTS	<input type="checkbox"/>	String	<input checked="" type="checkbox"/>		10

The 47th International Conference
APPLICATIONS OF MATHEMATICS IN ENGINEERING AND ECONOMICS
(AMEE 2021)

Schema of tSortRow_1

Column	K...	Type	<input checked="" type="checkbox"/> N.	Date Patt...	Leng...	Preci...	De...	Com...
Date		Date	<input type="checkbox"/>	"dd.MM.y...	100			
ISO3		String	<input type="checkbox"/>		10			
CountryName		String	<input type="checkbox"/>		200			
Region		String	<input type="checkbox"/>		200			
lat		Inte...	<input checked="" type="checkbox"/>		150			
lon		Inte...	<input checked="" type="checkbox"/>		150			
CumulativePo...		Float	<input checked="" type="checkbox"/>		100			
CumulativeDi...		Float	<input checked="" type="checkbox"/>		100			
CumulativeRe...		Inte...	<input checked="" type="checkbox"/>		5			
CurrentlyPositi...		Inte...	<input checked="" type="checkbox"/>		5			
Hospitalized		Inte...	<input type="checkbox"/>		5			
IntensiveCare		Inte...	<input checked="" type="checkbox"/>		5			
EUCountry		Bool...	<input checked="" type="checkbox"/>		15			
EUCPMCountry		Bool...	<input checked="" type="checkbox"/>		15			
NUTS		String	<input checked="" type="checkbox"/>		10			

Column	K...	Type	<input checked="" type="checkbox"/> N.	Date Patt...	Leng...	Preci...	De...	Com...
Date		Date	<input type="checkbox"/>	"dd.MM.y...	100			
ISO3		String	<input type="checkbox"/>		10			
CountryName		String	<input type="checkbox"/>		200			
Region		String	<input type="checkbox"/>		200			
lat		Inte...	<input checked="" type="checkbox"/>		150			
lon		Inte...	<input checked="" type="checkbox"/>		150			
CumulativePo...		Float	<input checked="" type="checkbox"/>		100			
CumulativeDi...		Float	<input checked="" type="checkbox"/>		100			
CumulativeRe...		Inte...	<input checked="" type="checkbox"/>		5			
CurrentlyPositi...		Inte...	<input checked="" type="checkbox"/>		5			
Hospitalized		Inte...	<input type="checkbox"/>		5			
IntensiveCare		Inte...	<input checked="" type="checkbox"/>		5			
EUCountry		Bool...	<input checked="" type="checkbox"/>		15			
EUCPMCountry		Bool...	<input checked="" type="checkbox"/>		15			
NUTS		String	<input checked="" type="checkbox"/>		10			

OK Cancel

Configuration
schema for data
sorting

The 47th International Conference
APPLICATIONS OF MATHEMATICS IN ENGINEERING AND ECONOMICS
(AMEE 2021)

Sorting result of Covid-19 data

Job SortingFile

Basic Run
 Debug Run
 Advanced settings
 Target Exec
 Memory Run

Execution

Run Kill Clear

```

08.03.2021|BGR|Bulgaria|Razgrad||2493.0|0.0|0|2493|||true|true|BG ^
08.03.2021|BGR|Bulgaria|Sofia Region||7455.0|0.0|0|7455|||true|tr
08.03.2021|BGR|Bulgaria|Targovishte||2406.0|1.0|0|2405|||true|tru
08.03.2021|BGR|Bulgaria|Yambol||4444.0|14.0|18|4412|||true|true|B
08.03.2021|DEU|Germany|NOT SPECIFIED|||2310900|-2310900|||true|t
08.03.2021|UKR|Ukraine|NOT SPECIFIED|||1406800.0|27128.0|1198254|1
07.03.2021|ALB|Albania|Dibër||2342.0|45.0|1933|364|11|1|false|fa
07.03.2021|ALB|Albania|Gjirokastrë||3816.0|42.0|2565|1209|16|1|f
07.03.2021|ALB|Albania|NOT SPECIFIED|||0|||false|false|
07.03.2021|BGR|Bulgaria|NOT SPECIFIED||0.0|10448.0|211790|-222238 v
  
```

Job SortingFile

Basic Run
 Debug Run
 Advanced settings
 Target Exec
 Memory Run

Execution

Run Kill Clear

```

07.03.2021|RUS|Russian Fed.|Nenets||1045.0|3.0|951|91|||false|fal ^
07.03.2021|SVK|Slovakia|NOT SPECIFIED|||7836.0||-7836|3691|358|tr
07.03.2021|UKR|Ukraine|NOT SPECIFIED|||1401228.0|27022.0|1196520|1
07.03.2021|GBR|United Kingdom|England||35258.0||0|35258|||false|t
06.03.2021|ALB|Albania|Dibër||2332.0|45.0|1925|362|11|1|false|fa
06.03.2021|ALB|Albania|Gjirokastrë||3778.0|42.0|2538|1198|14|1|f
06.03.2021|ALB|Albania|NOT SPECIFIED|||0|||false|false|
06.03.2021|BEL|Belgium|NOT SPECIFIED||14648.0|0.0||14648|0|0|true
06.03.2021|BGR|Bulgaria|NOT SPECIFIED||0.0|10426.0|211206|-221632
06.03.2021|BGR|Bulgaria|Razgrad||2475.0|0.0|0|2475|||true|true|BG v
  
```


The 47th International Conference
APPLICATIONS OF MATHEMATICS IN ENGINEERING AND ECONOMICS
(AMEE 2021)

Thank You!

Stella Vetova

vetova.bas@gmail.com

June 7-13, 2021