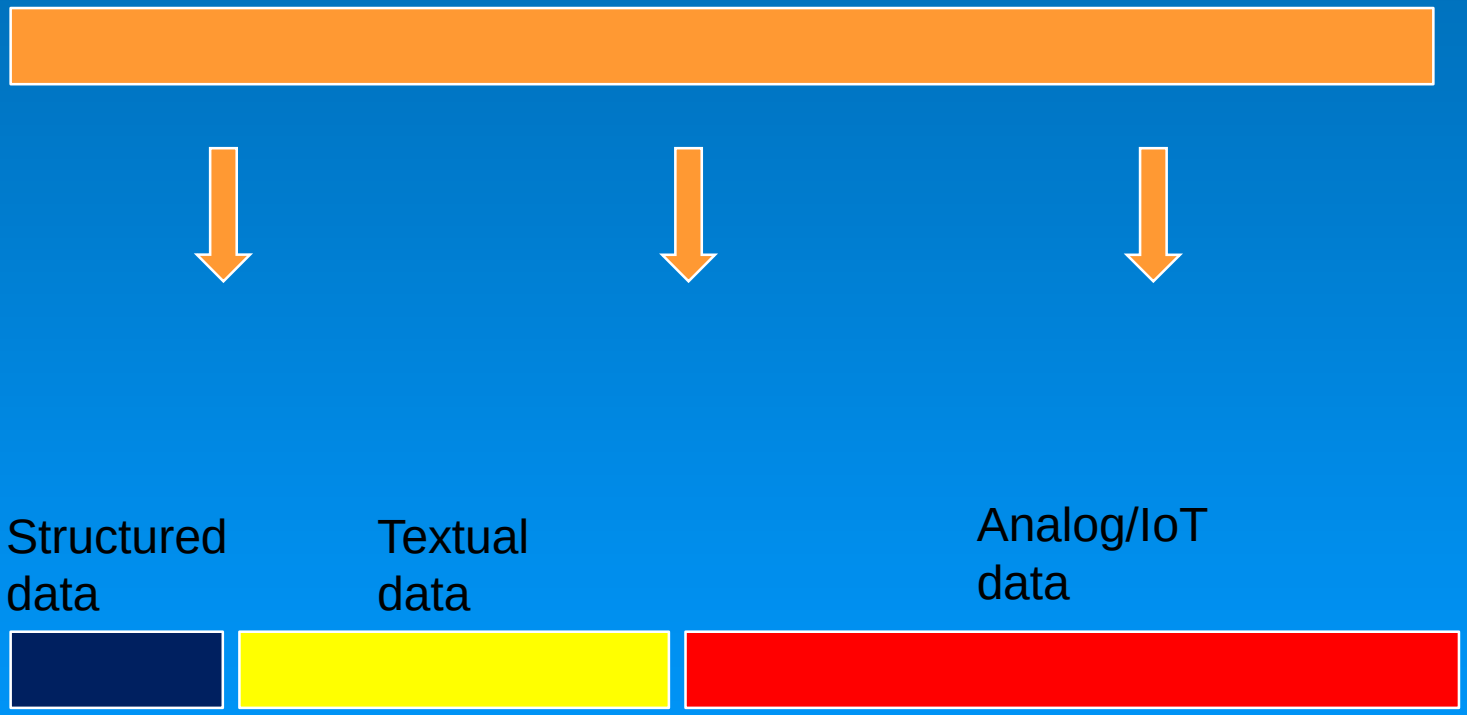
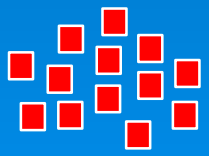
DATA LAKEHOUSE – THE BASIC ELEMENTS

A presentation by
W H Inmon

All data in the corporation



FOREST RIM TECHNOLOGY



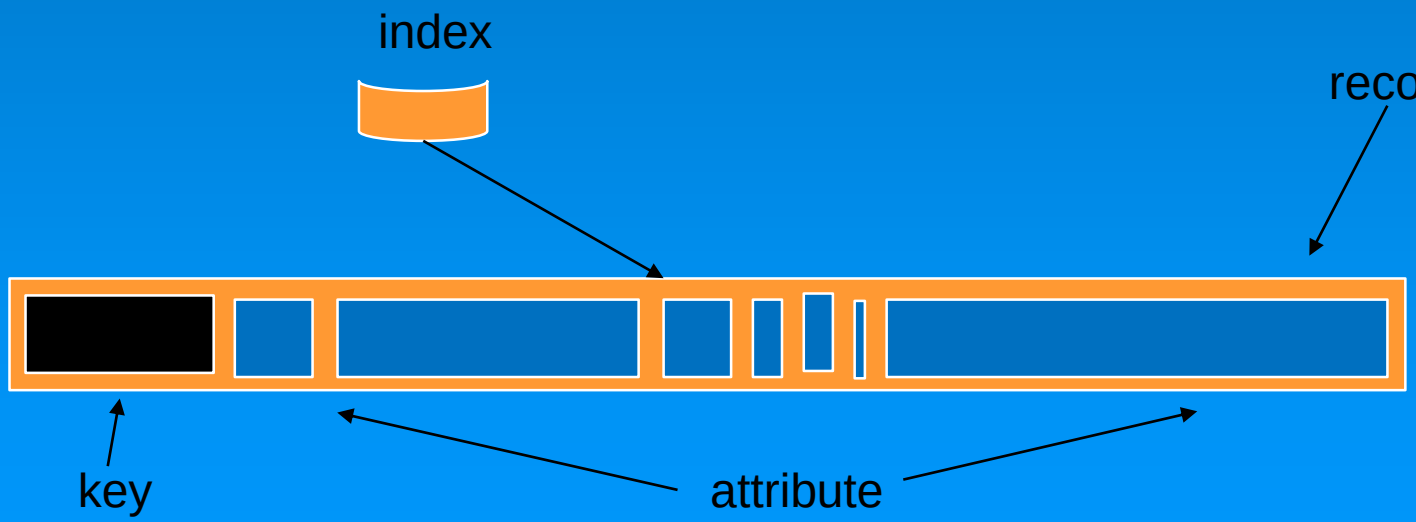
Each of the different types of data have their own unique characteristics

FOREST RIM TECHNOLOGY

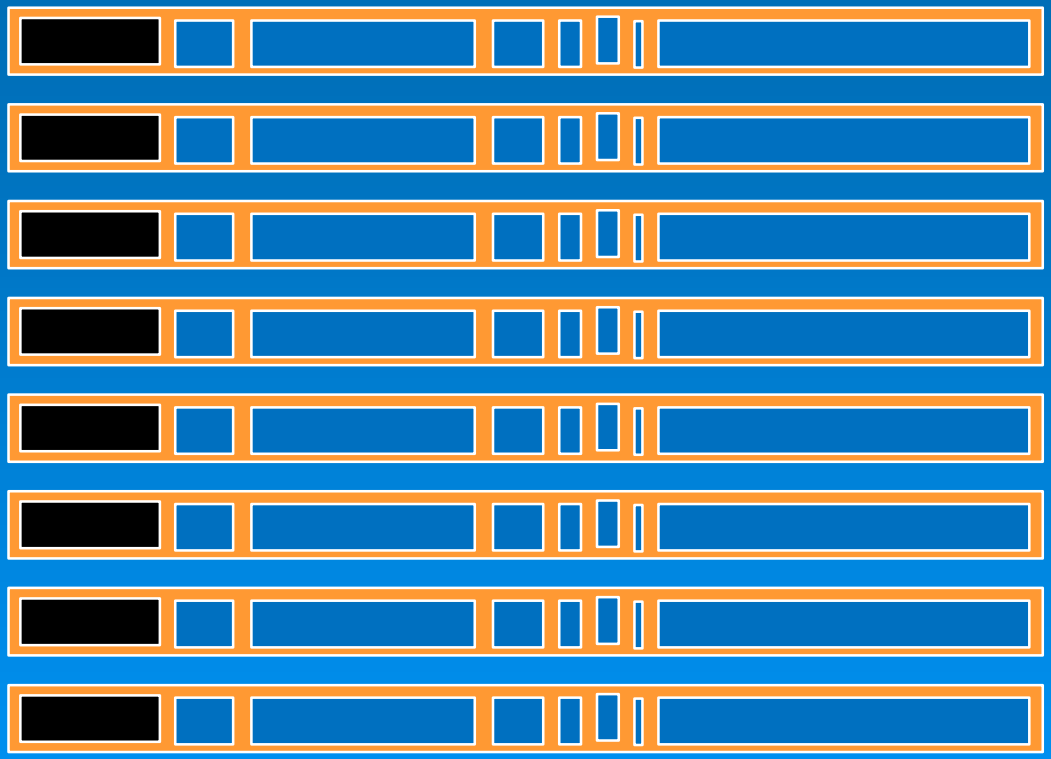
Structured data

Usually transaction based

Bank transactions
Point of sale
Telephone call
Payments made
Payments received
……



Structured data



The same record type is repeated
Each record has different contents

FOREST RIM TECHNOLOGY Inc

Textual data

Text is found everywhere

Medical records
Contracts
Internet
Call centers
Warranty claims
Insurance claims
Email
……..

Voice
Written
Internet
Video
……..

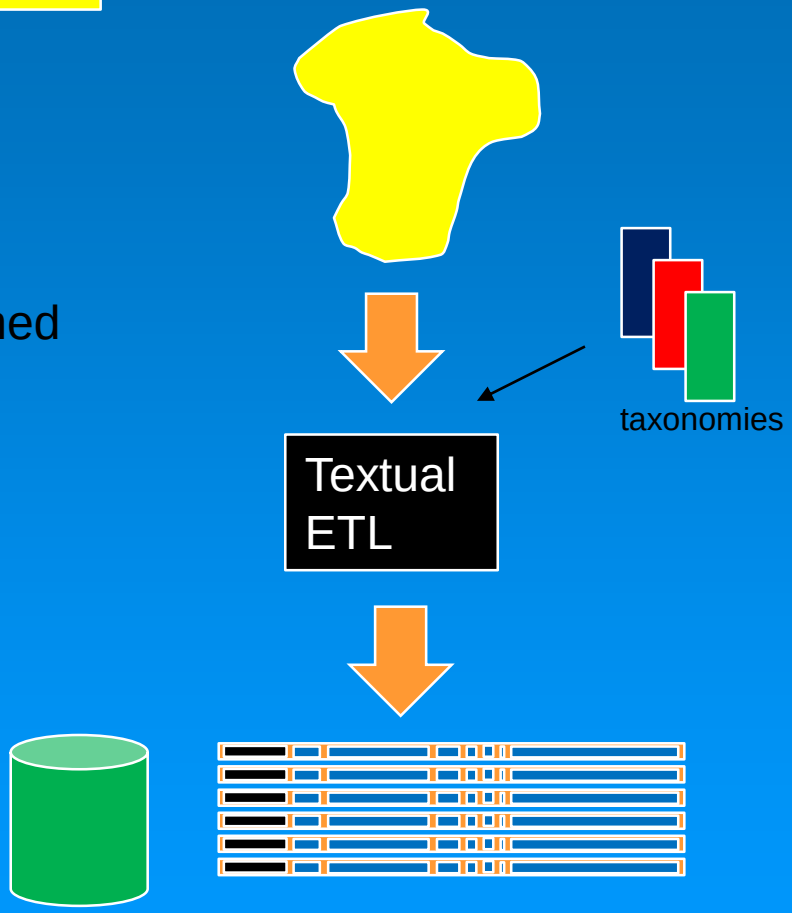
English
Spanish
Portuguese
French
Mandarin
Korean
German

Formal language
Slang
Acronyms
……..

FOREST RIM TECHNOLOGY Inc

Textual
data

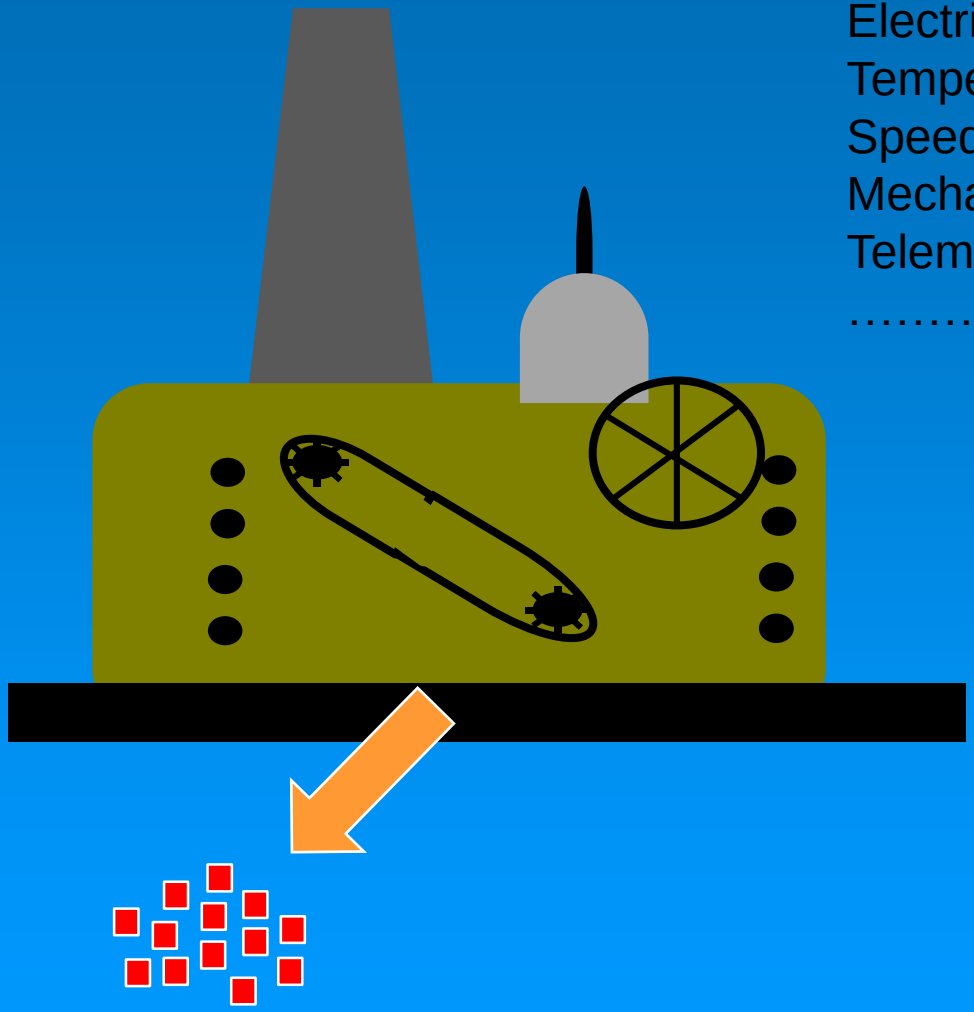
Text is transformed
Into a structured
format



FOREST RIM
TECHNOLOGY

Analog/IoT

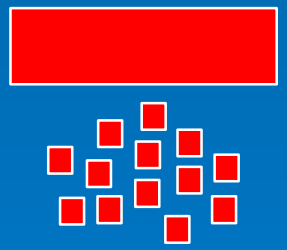
Machine generated



Drones
Electric eye
Temperature gauge
Speed
Mechanical
Telemetry
……..

FOREST RIM TECHNOLOGY Inc

Analog/IoT



telemetry

Telemetry data is generated as the rocket is launched and is measured throughout the flight

Date Sept 2, 2021
Time 11:21 am
Location from Denver
Location to Co Spgs

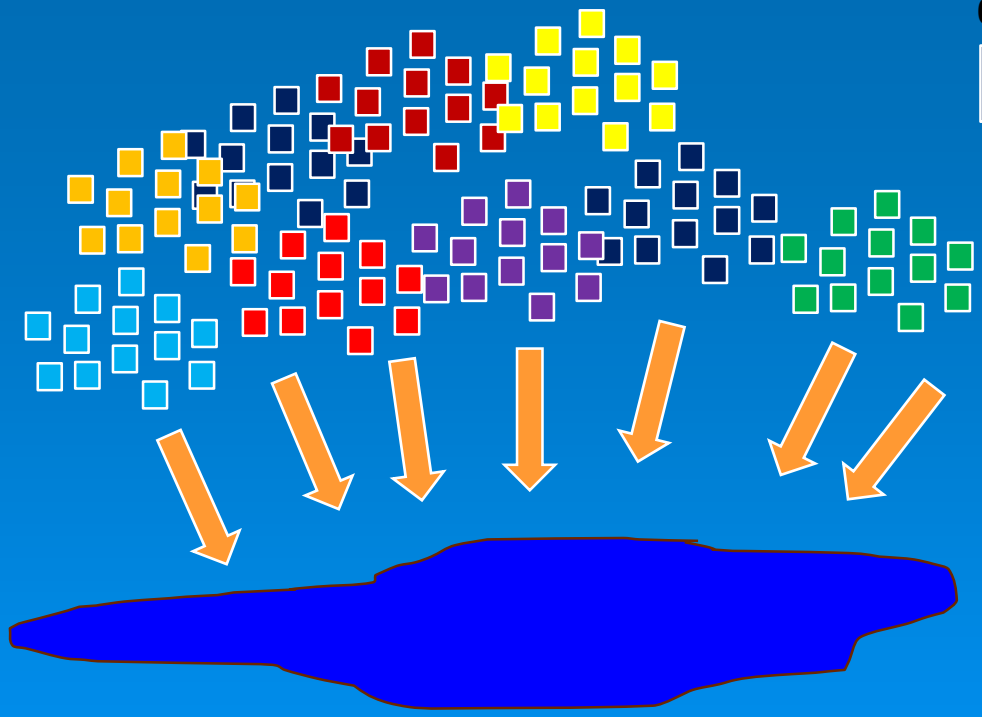
Elevation	Speed
786	0
792	35
812	79
854	124
901	197
978	276
1012	367
1256	416
1469	521
1672	702
2018	835
2259	915
2871	…..
……..	

FOREST RIM TECHNOLOGY INC

Analog/IoT

Structured data

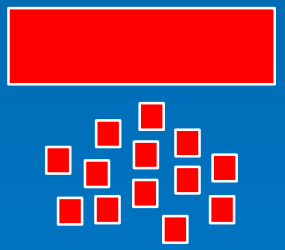
Textual data

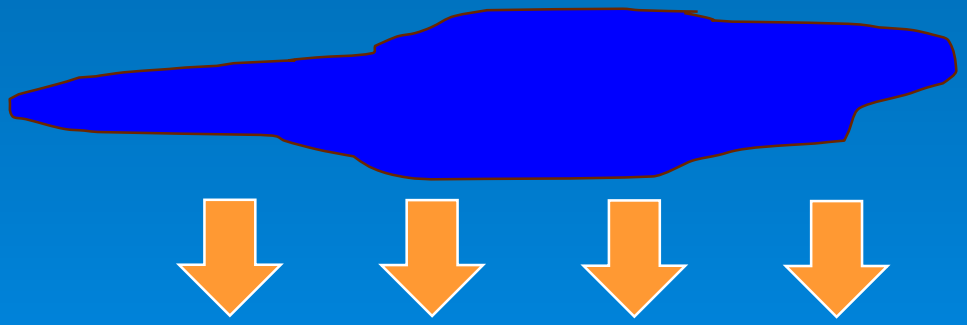


The data lake is created by throwing data all the data into the lake

FOREST RIM TECHNOLOGY

Analog/IoT

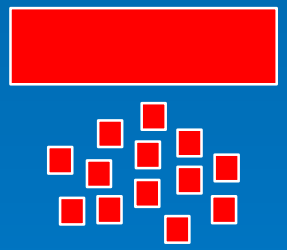




Soon the data lake
turned into a swamp

FOREST RIM TECHNOLOGY Inc

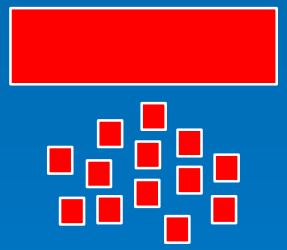
Analog/IoT

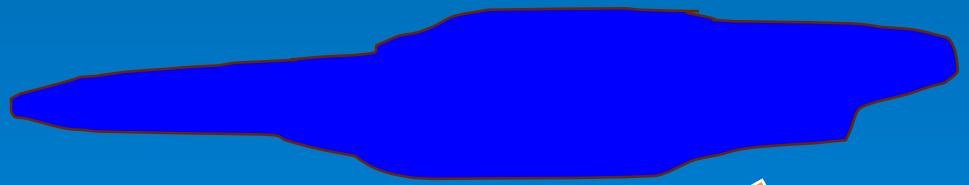


The data swamp was not good for anyone….

FOREST RIM TECHNOLOGY Inc

Analog/IoT

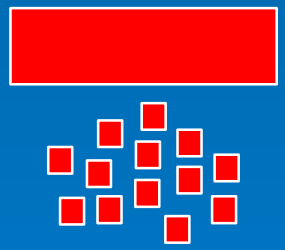


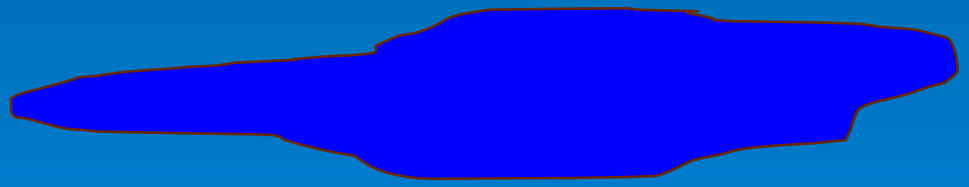


The data lake needs to be turned into a lakehouse

FOREST RIM TECHNOLOGY Inc

Analog/IoT



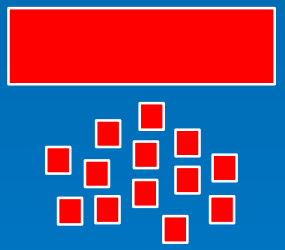


Data scientist

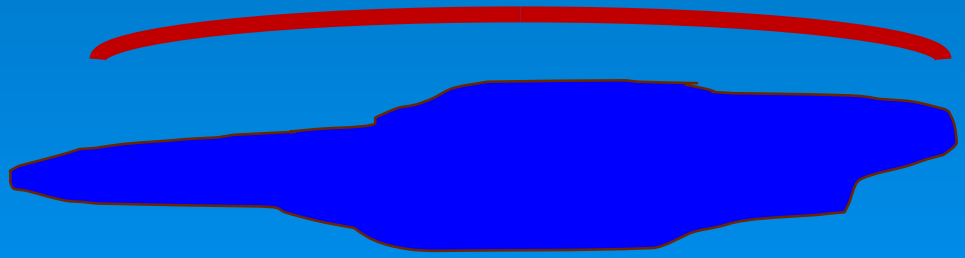
All this education and 95% of my job is being a data garbageman

FOREST RIM TECHNOLOGY Inc

Analog/IoT



infrastructure



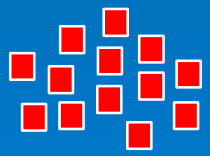
Data scientist

Ah, that's more like it

FOREST RIM TECHNOLOGY

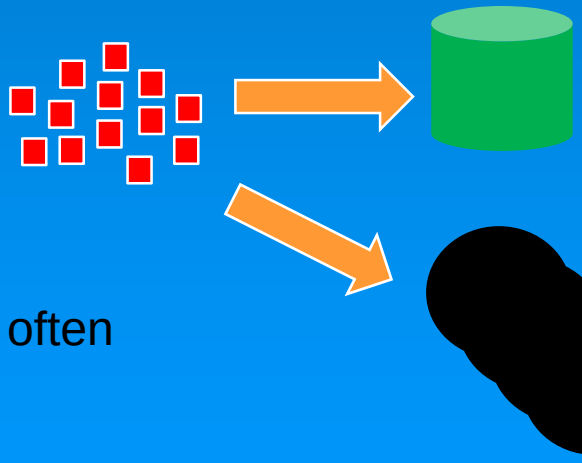
Analog/IoT

Machine generated



Basic, raw measurements

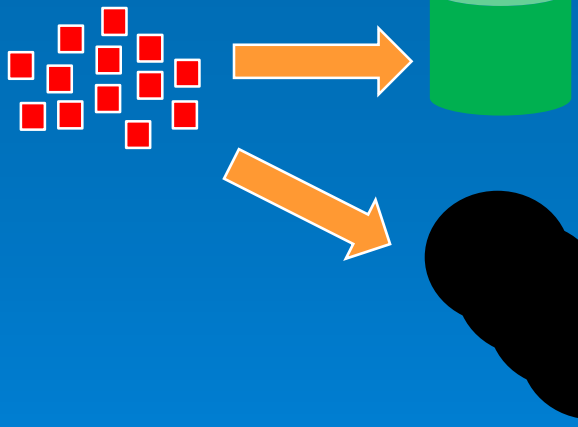
Time – 0912
Time – 0916
Time – 1002
Time – 1008
Time – 1017
…….



High probability
High performance

Low probability
Bulk storage

Analog/IoT data is often segmented



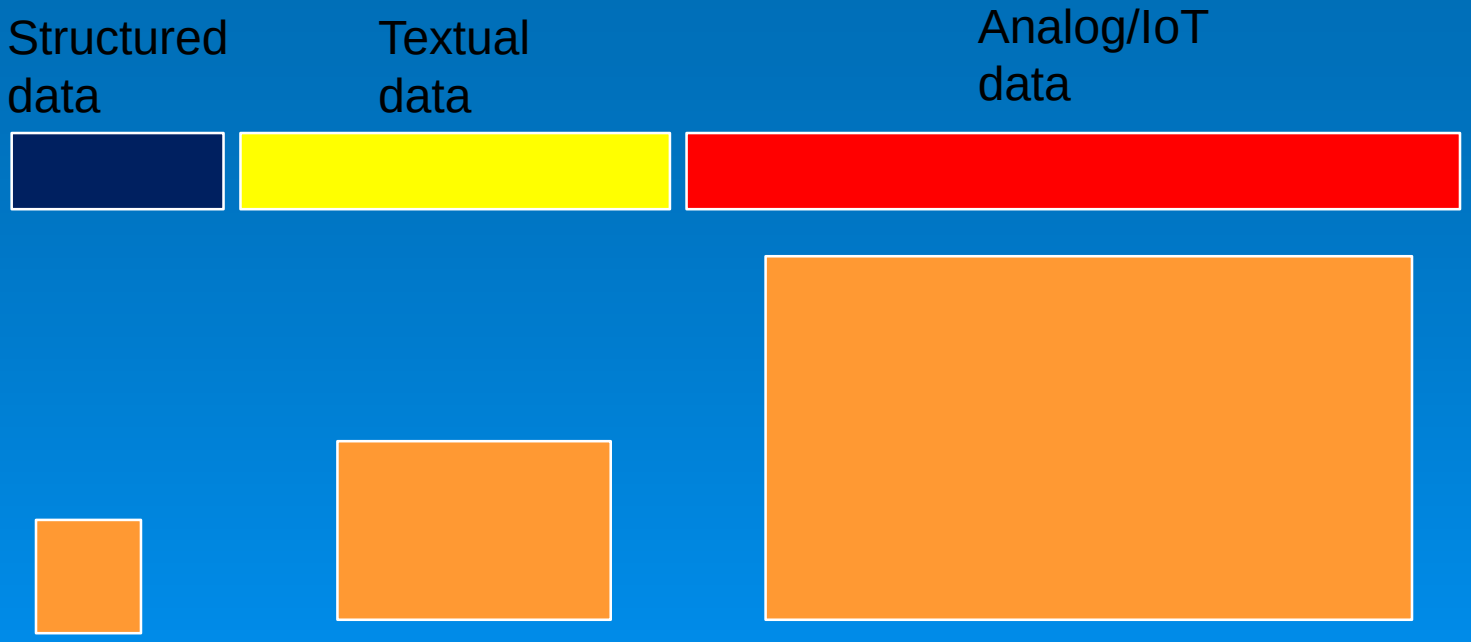
High probability
High performance

Low probability
Bulk storage

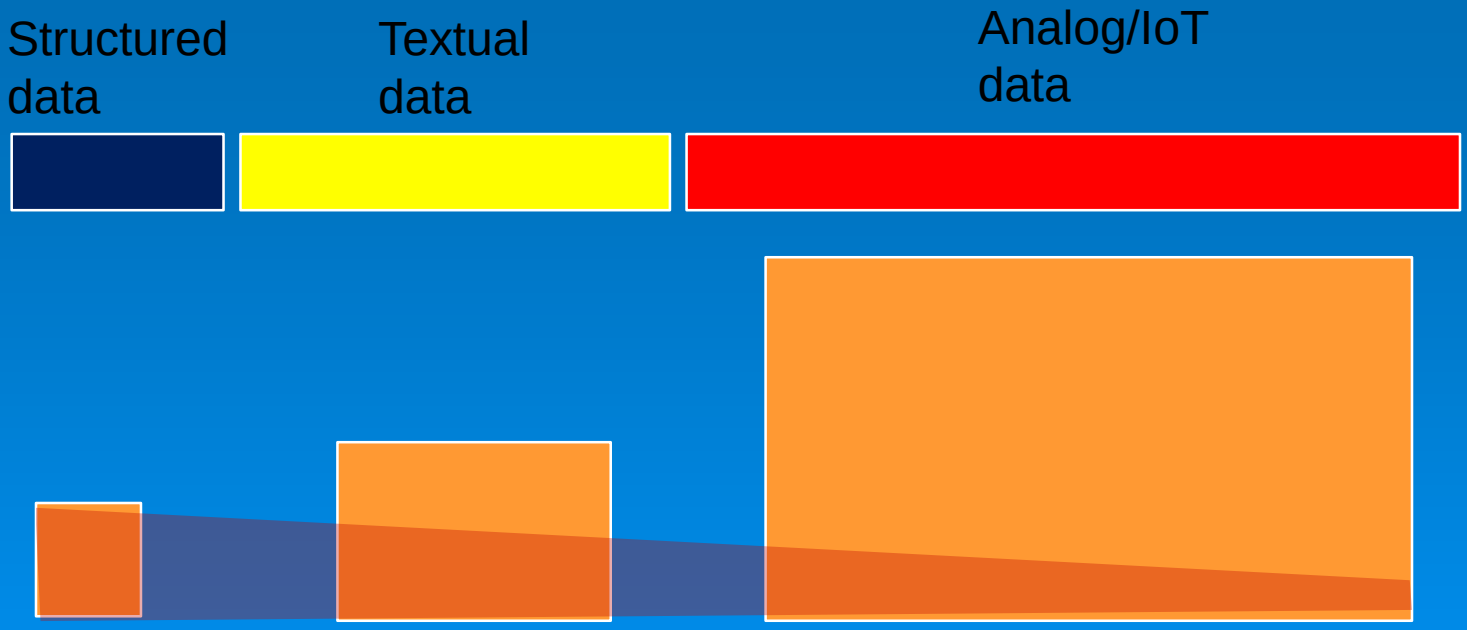
Date of launch
Ultimate speed
Ultimate height
Final landing point

Second by second measurements

FOREST RIM TECHNOLOGY Inc



Relative volumes of data in each sector



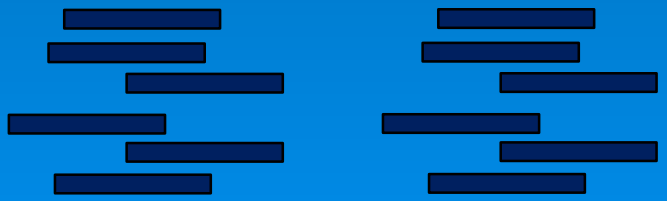
Business value and the volumes of data

Format compatibility

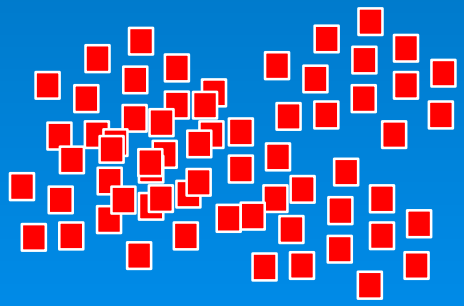
Structured data

Textual data

Analog/IoT data



Relational format



Raw data format

From a format standpoint, the structured and the textual environments are very different from the analog/IoT environment

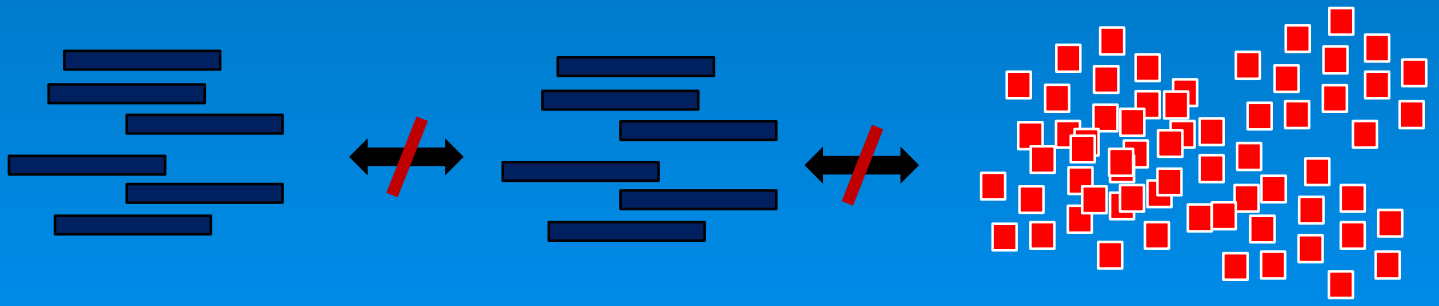
FOREST RIM TECHNOLOGY Inc

Content compatibility

Structured data

Textual data

Analog/IoT data

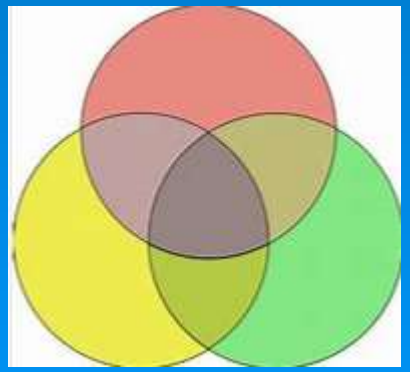


Key compatibility – very unintegrated

Structured data

Textual data

Analog/IoT data



In order to do analytics, there must be some common data on which to do a comparison

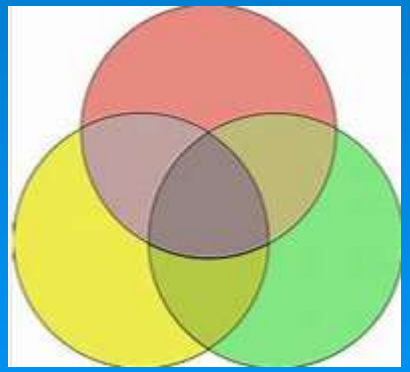
Without common data it is very difficult to do a meaningful comparison

FOREST RIM TECHNOLOGY

Structured data

Textual data

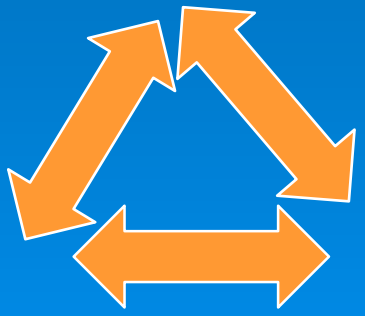
Analog/IoT data



The problem is that there may be no obvious, easy way to isolate common identifiers

FOREST RIM TECHNOLOGY

Textual
data



Structured
data

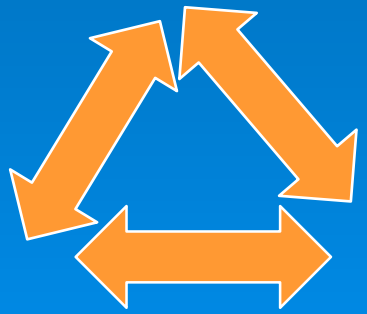
Analog/IoT
data

Fortunately there are such things as
universal common connectors

FOREST RIM
TECHNOLOGY

Textual data

Structured data



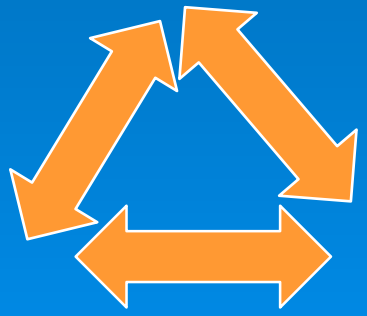
Analog/IoT data

Universal common connectors exist regardless of the way that data has been collected

FOREST RIM TECHNOLOGY

General common connectors

Textual
data



Structured
data

Analog/IoT
data

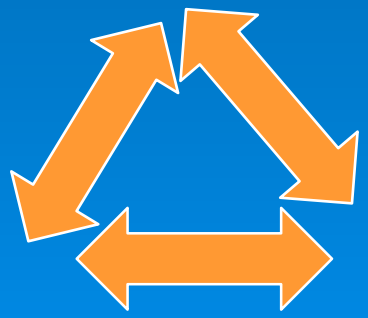
Universal common connector for anything

geography
time
dollar amount

FOREST RIM
TECHNOLOGY Inc

Common connectors for humans

Textual
data



Structured
data

Analog/IoT
data

Universal common connector for humans

gender
age
race

Common connectors for objects

Textual
data

Structured
data

Analog/IoT
data

Universal common connector for physical objects

weight
color
cost
size
shape

FOREST RIM
TECHNOLOGY

SOME EXAMPLES

Universal common connector

FOREST RIM TECHNOLOGY Inc.

Healthcare – outcomes analysis

FOREST RIM TECHNOLOGY Inc.

Outcome analysis

Did the medicine work?
Did the vaccination work?
Did the operation have the right effect?

FOREST RIM TECHNOLOGY Inc

Textual data

Doctor's notes
- tests
- diagnosis
- procedure
- medication
- history
- ……

Structured data

Sales of -
- Prolia
- Estrogen
- Vitamin D
- Algaecal
- Calcitonin

Analog/IoT data

X rays
- date
- location
- patient age
- examination results

FOREST RIM TECHNOLOGY Inc

Structured data

What medicines have been purchased

By state
By age
By gender

Textual data

What medicines have been prescribed and/or discussed with doctors

By state
By age
By gender

Analog/IoT data

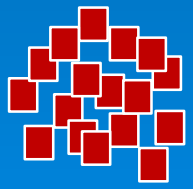
What outcomes have been achieved

By state
By age
By gender

FOREST RIM TECHNOLOGY Inc

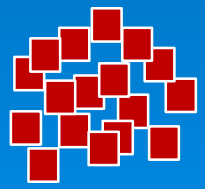
What medicines have been purchased

By state
By age
By gender



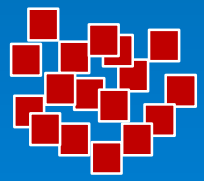
What medicines have been prescribed and/or discussed with doctors

By state
By age
By gender



What outcomes have been achieved

By state
By age
By gender



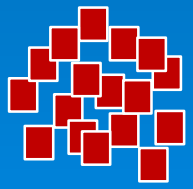
Analyses –

how does treatment in Utah vary from treatment in Oregon?
is Prolia more effective than estrogen?
when patients are treated with Algaecal, what other side effects are noticed?
do women have better results than men?
how much does age affect –
- the types of treatment for osteoporosis
- the effectiveness of treatment
- whether men react differently than women

FOREST RIM TECHNOLOGY Inc

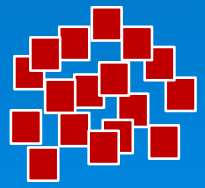
What medicines have been purchased

By state
By age
By gender



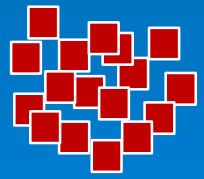
What medicines have been prescribed and/or discussed with doctors

By state
By age
By gender



What outcomes have been achieved

By state
By age
By gender



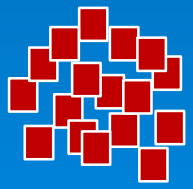
When you have both treatment and outcome data together, you can answer – for the first time – important questions about treatment, medication, dosage, side, effects, demographics of treatment

You can match outcome with treatment

FOREST RIM TECHNOLOGY Inc

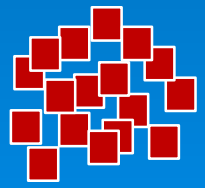
What medicines have been purchased

By state
By age
By gender



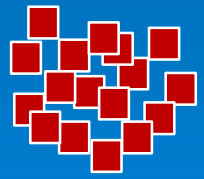
What medicines have been prescribed and/or discussed with doctors

By state
By age
By gender



What outcomes have been achieved

By state
By age
By gender



The result is healthier people and longer life and better quality of life

FOREST RIM TECHNOLOGY Inc

Manufacturing

FOREST RIM TECHNOLOGY Inc.

Textual
data

Warranty claims
- unit
- unit type
- defect
- severity
- in use desc

Structured
data

Sales data
- unit sold
- date of sale
- location of sale
- customer address

Analog/IoT
data

Manufacturing data
- unit id
- lot id
- date of manufacture
- machine used
- operator

FOREST RIM TECHNOLOGY Inc

Textual
data

Unit id
Defect description
Date of warranty

Unit id

Structured
data

Units sold
Date of sale
Location of sale

Unit id

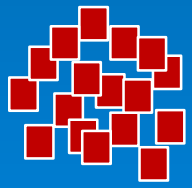
Analog/IoT
data

Unit id
Machine used for manufacture
Date of manufacture
Operator
Lot id
Manufacture telemetry

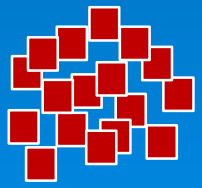
Unit id

FOREST RIM
TECHNOLOGY

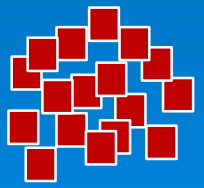
Units sold
Date of sale
Location of sale



Unit id
Defect description
Date of warranty



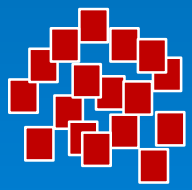
Unit id
Machine used for manufacture
Date of manufacture
Operator
Lot id
Manufacture telemetry



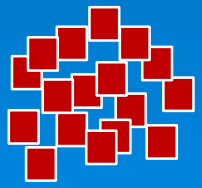
Analyses –

what manufacturing machines are producing defects
what manufacturing machines are not producing defects
what operators are producing defects
what operators are not producing defects
what telemetry needs to be adjusted
under what conditions are defects created
……

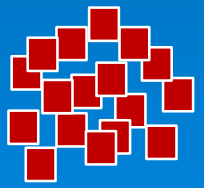
Units sold
Date of sale
Location of sale



Unit id
Defect description
Date of warranty



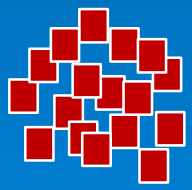
Unit id
Machine used for manufacture
Date of manufacture
Operator
Lot id
Manufacture telemetry



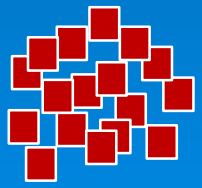
With all of this data together and able to be analyzed you can now tell what defects can be corrected and what conditions cause defects to occur. The manufacturing process can be materially improved

FOREST RIM TECHNOLOGY

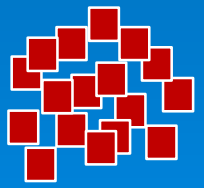
Units sold
Date of sale
Location of sale



Unit id
Defect description
Date of warranty



Unit id
Machine used for manufacture
Date of manufacture
Operator
Lot id
Manufacture telemetry



Now manufacturing can be done efficiently and in a cost effective manner

FOREST RIM TECHNOLOGY Inc

With analytics from the data lakehouse, you can improve the lives and livelihood of many people

FOREST RIM TECHNOLOGY