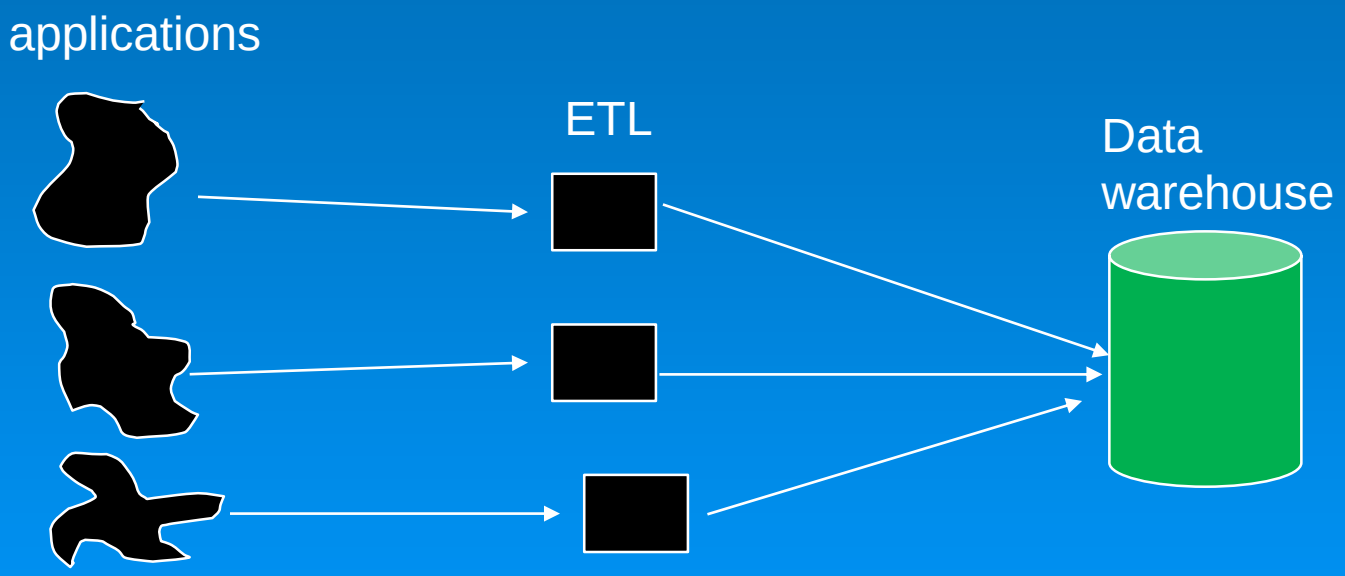
LAKEHOUSE AND WAREHOUSE – A COMPARISON

A presentation by
W H Inmon



The data warehouse infrastructure

FOREST RIM
TECHNOLOGY Inc

A fundamental transformation of data occurs

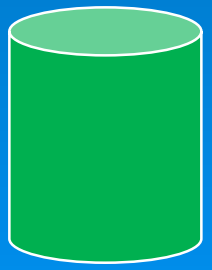
applications

Gender – m/f
Distance - inches

Gender – male/female
Distance - feet

Gender – 1/0
Distance - cms

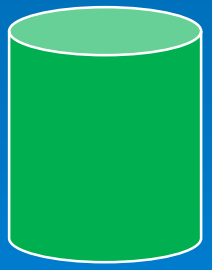
Data warehouse



Gender – m/f
Distance - feet

FOREST RIM TECHNOLOGY INC

Data warehouse



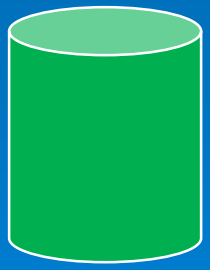
Data warehouse –
- subject oriented
- integrated
- non volatile
- time variant

collection of data for management decisions

The single version of the truth

FOREST RIM TECHNOLOGY

Data warehouse

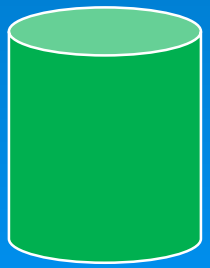


Historical data as well.

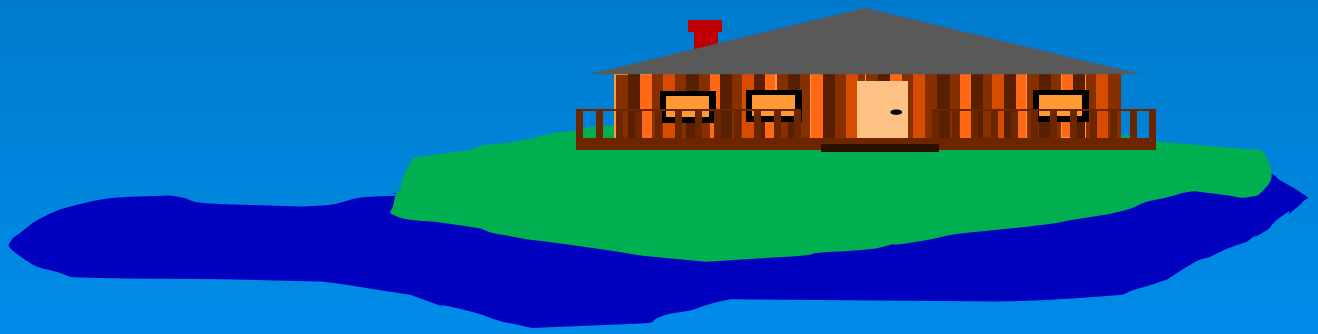
FOREST RIM TECHNOLOGY INC

Structured, transaction based data

Data warehouse

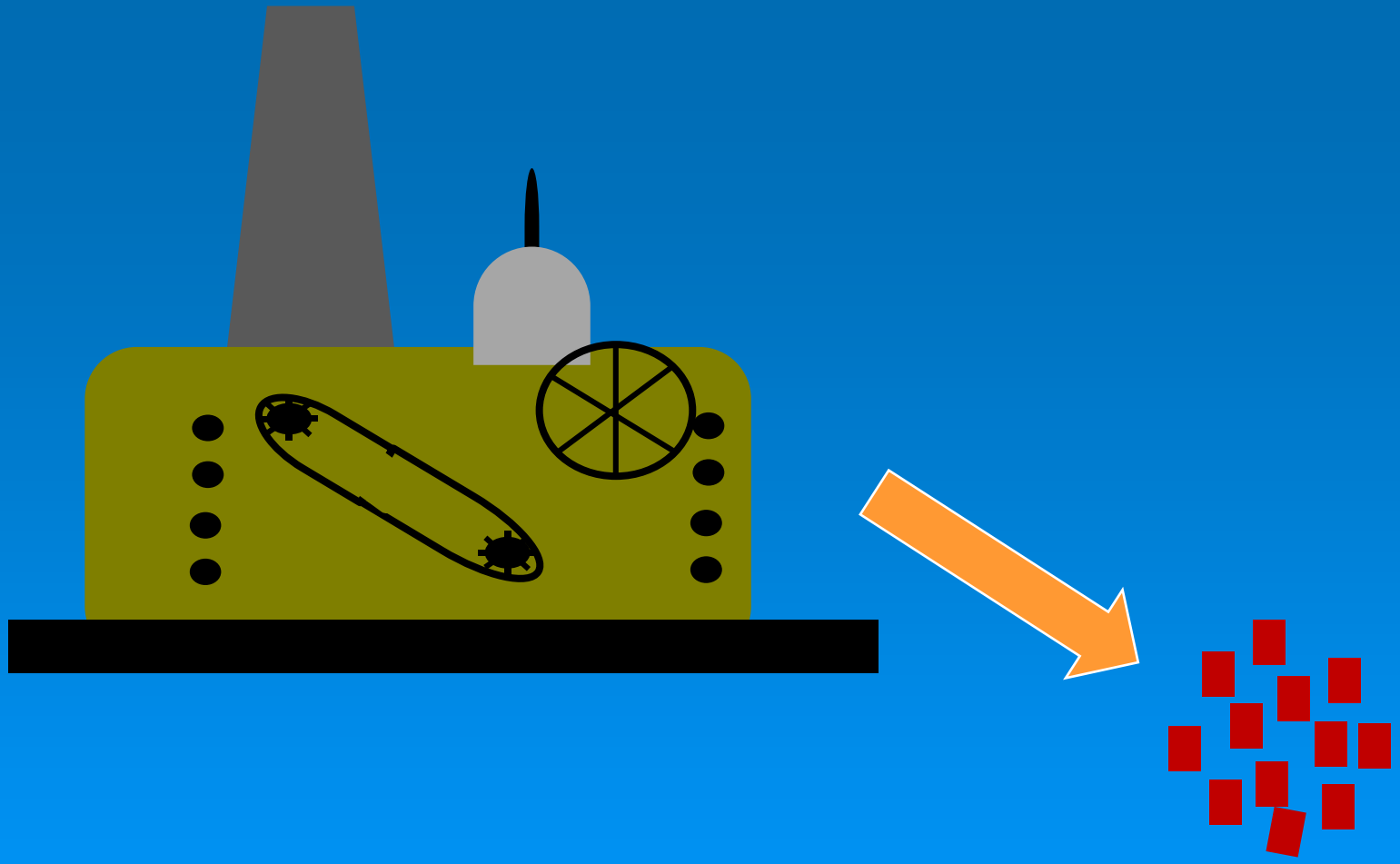


FOREST RIM TECHNOLOGY INC



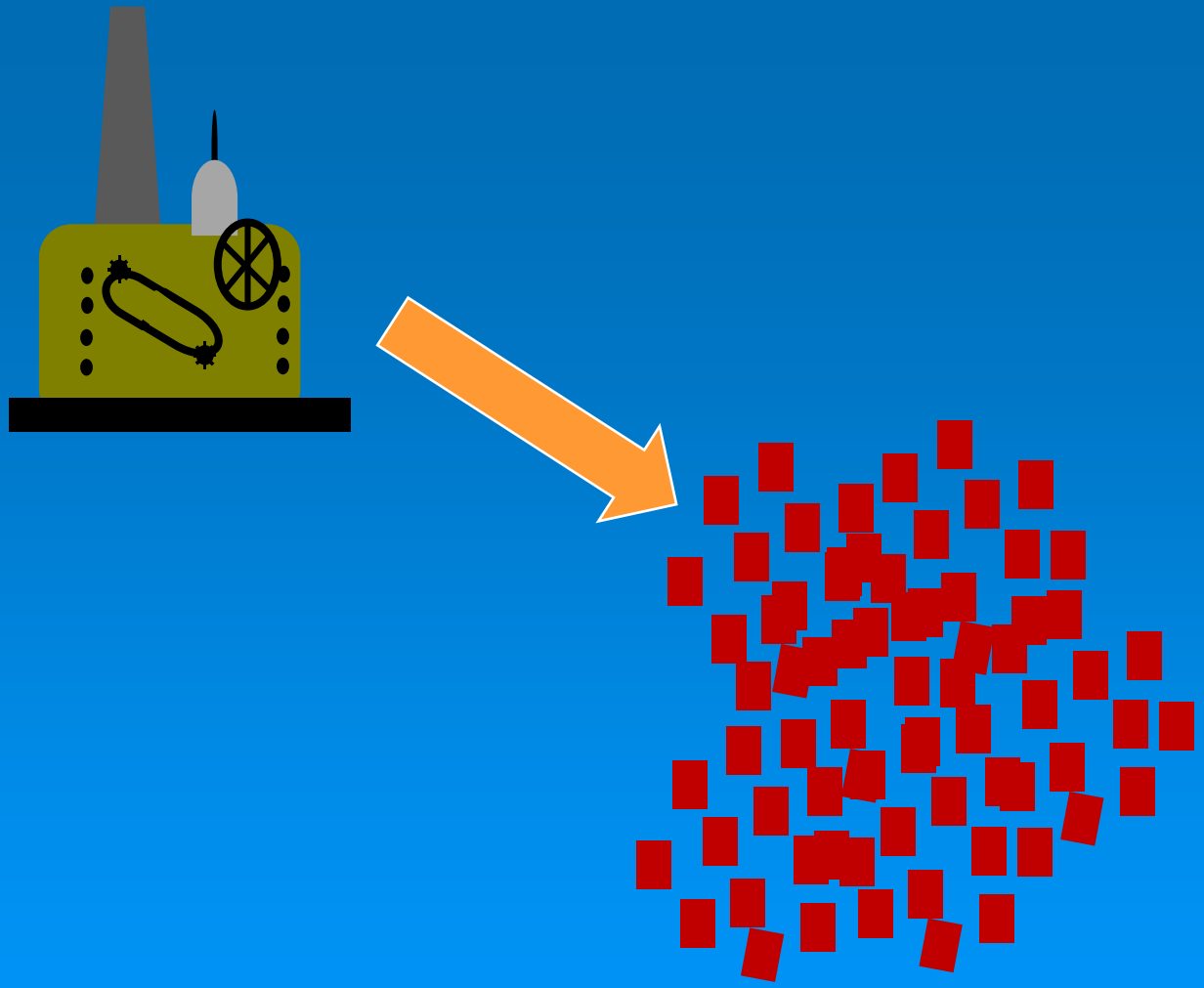
The data lakehouse

FOREST RIM TECHNOLOGY



Machine generated data

FOREST RIM TECHNOLOGY INC

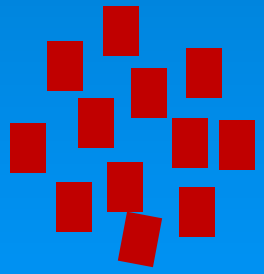


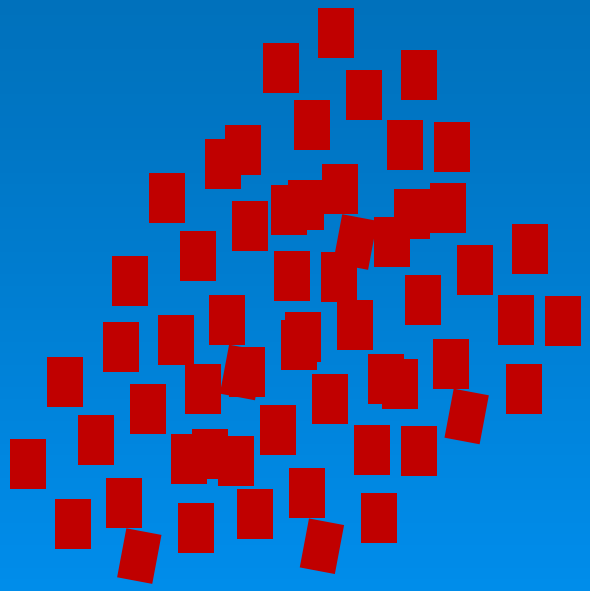
A lot of data generated

FOREST RIM TECHNOLOGY Inc.

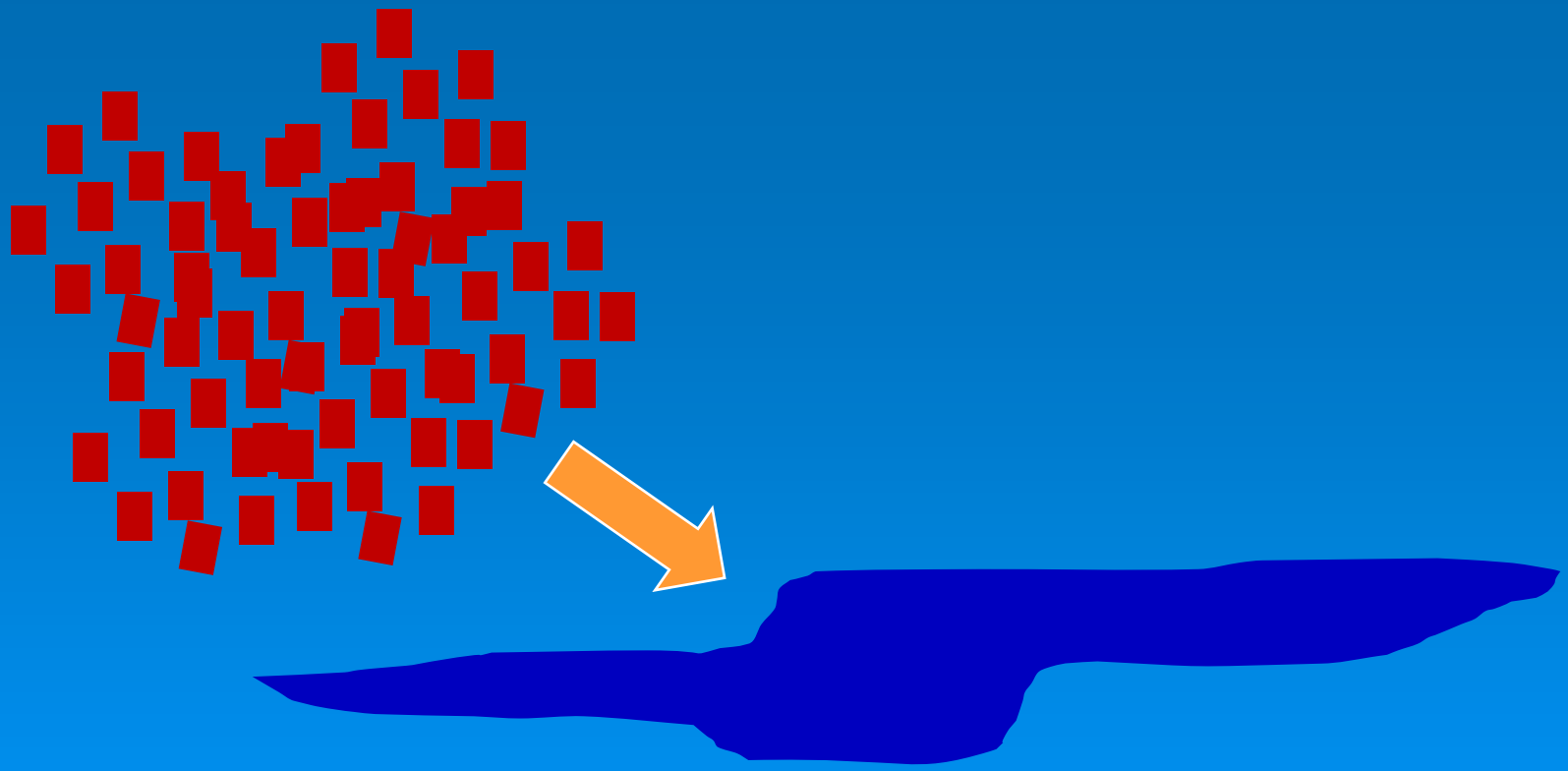
Some data has high business value,
much data has little business value

High business value





Low business value



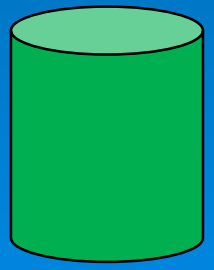
The data is put into a data lake

FOREST RIM TECHNOLOGY Inc

An analytical infrastructure is placed over the data lake

FOREST RIM TECHNOLOGY Inc.

The data lakehouse

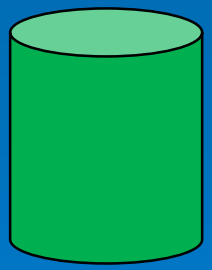


The data warehouse

How are the data lakehouse and the data warehouse
The same? How are they different?

FOREST RIM
TECHNOLOGY INC

The data lakehouse



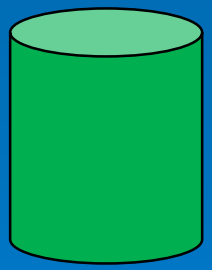
The data warehouse

The same -

Both exist to serve analytical processing
Both have vetted data – accurate, reliable
Both can look at data over a long period of time
Both have tame variant data
Both have data that once recorded cannot be changed

FOREST RIM TECHNOLOGY INC

The data lakehouse



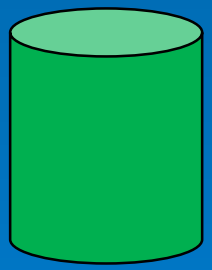
The data warehouse

Different -

There is MUCH MORE data in the lakehouse
The data is fundamentally different –
warehouse data is transaction based data
lakehouse data is machine generated data
The data structures are very different
Relating the two types of data is difficult to do

FOREST RIM TECHNOLOGY INC

The data lakehouse



The data warehouse

Like cousins -

Related yet still different

FOREST RIM TECHNOLOGY