

**The Key!:**

**Integrating diverse data**

# After this video you will be able to..

- Explain what data integration is and how it makes big data sources more valuable even before they are analyzed

# Getting Value from Big Data

**Value comes from  
integrating different  
types of data  
sources**

# Who's Ready For Some Big Data Success Stories?



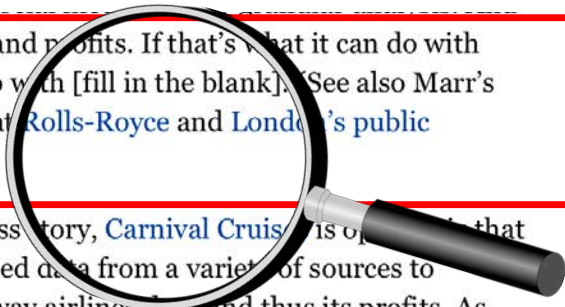
**Howard Baldwin**  
FOLLOW ON FORBES (7)  
Opinions expressed by Forbes

Enough of this lighthearted *Laugh-In* (oh, man, am I short on data success stories, just to s

I've been referencing Bernardi some really interesting stuff. *Dickey's*, the barbecue chain big data – near-real-time, out all tied back to increasing efficiency and profits. If that's what it can do with barbecue, imagine what you could do with [fill in the blank]. (See also Marr's earlier *Forbes* pieces about big data at *Rolls-Royce* and *London's public transport* system.)

Although it's not yet an official success story, *Carnival Cruises* is optimistic that it can take structured and unstructured data from a variety of sources to improve its pricing – not unlike the way airlines do – and thus its profits. As writer Kim Nash noted in the *Wall Street Journal*, "At Carnival, the number [of passenger cruise days] is 80 million across its fleet of 100 ships and nine

Although it's not yet an official success story, *Carnival Cruises* is optimistic that it can take structured and unstructured data from a variety of sources to improve its pricing – not unlike the way airlines do – and thus its profits. As writer Kim Nash noted in the *Wall Street Journal*, "At Carnival, the number [of passenger cruise days] is 80 million across its fleet of 100 ships and nine brands. To CEO Arnold Donald, that means that if every passenger spent just \$1 more per day aboard ship, Carnival would see an extra \$80 million in revenue for the year." Talk about turning small data turning into big business: Donald has also told financial analysts that "small tweaks add up to real dollars."



# Structured + Unstructured Data



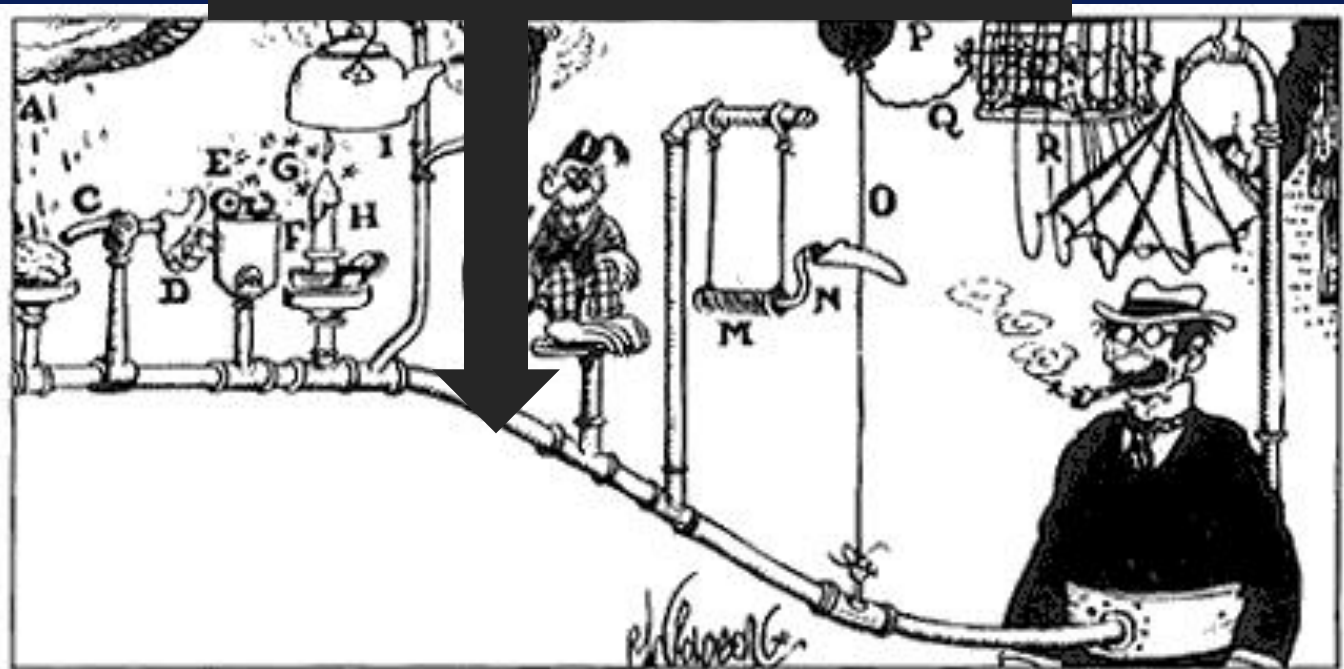
## Price optimization



## Increased revenue!

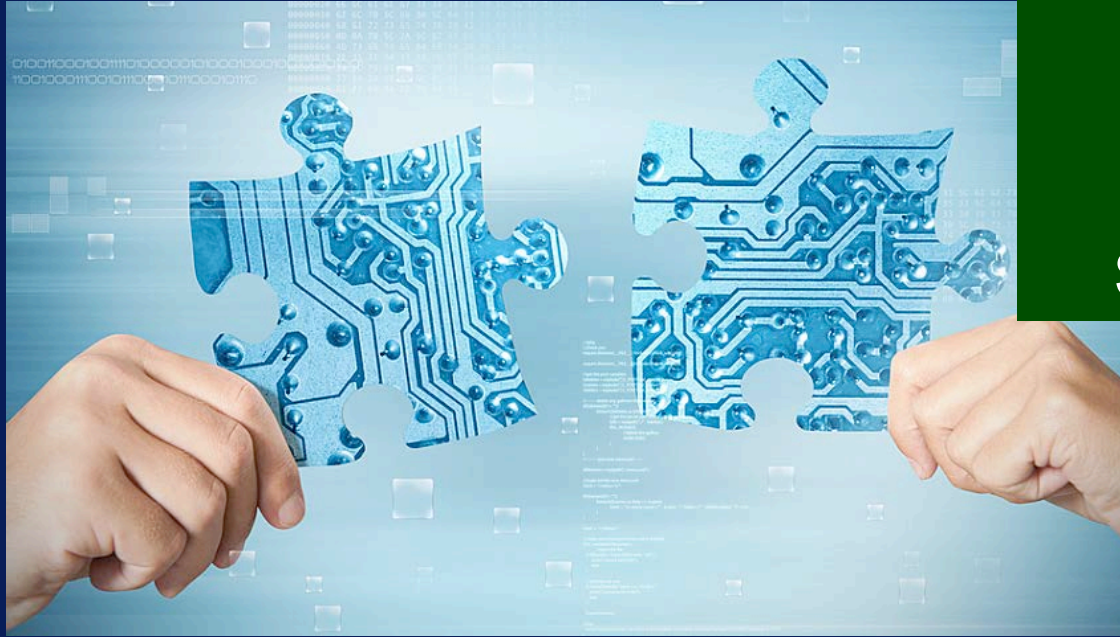
Although it's not yet an official success story, [Carnival Cruises](#) is optimistic that it can take structured and unstructured data from a variety of sources to improve its pricing – not unlike the way airlines do – and thus its profits. As writer Kim Nash noted in the Wall Street Journal, “At Carnival, the number [of passenger cruise days] is 80 million across its fleet of 100 ships and nine brands. To CEO Arnold Donald, that means that if every passenger spent just \$1 more per day aboard ship, Carnival would see an extra \$80 million in revenue for the year.” Talk about turning small data turning into big business: Donald has also told financial analysts that “small tweaks add up to real dollars.”

# Insert Big Data Integration Here



# Data Integration → Knowledge

Key:  
Turning complex  
data into  
something usable



# Data Integration Process

Discovering

Modeling

Accessing

Transforming

Monitoring





# **Why do we need Data integration?**

# Data comes in all shapes and sizes!

## Flat file

	Route No.	Miles	Activity
Record 1	I-95	12	
Record 2	I-495	05	
Record 3	SR-301	33	

## Relational Model

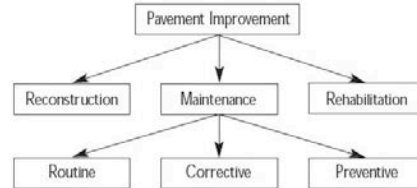
Activity Code	Activity Name
23	Patching
24	Overlay
25	Crack Sealing

Key = 24

Activity Code	Date	Route No.
24	01/12/01	I-95
24	02/08/01	I-66

## Relational

## Hierarchical Model



## Object-Oriented Model

### Object 1: Maintenance Report

### Object 1 Instance

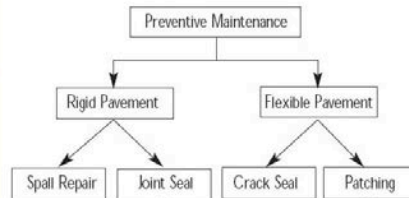
Date	
Activity Code	
Route No.	
Daily Production	
Equipment Hours	
Labor Hours	

01-12-01
24
I-95
2.5
6.0
6.0

### Object 2: Maintenance Activity

Activity Code	
Activity Name	
Production Unit	
Average Daily Production Rate	

## Network Model



# Data comes in all shapes and sizes!

## Flat file

	Route No.	Miles	Activity
Record 1	I-95	12	
Record 2	I-495	05	
Record 3	SR-301	33	

## Relational Model

Activity Code	Activity Name
23	Patching
24	Overlay
25	Crack Sealing

Key = 24

Activity Code	Date	Route No.
24	01/12/01	I-95
24	02/08/01	I-66

## Object-Oriented Model

### Object 1: Maintenance Report

### Object 1 Instance

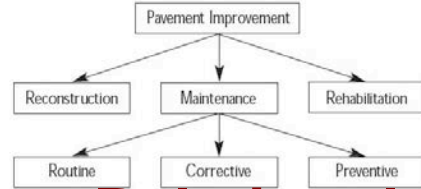
Date	
Activity Code	
Route No.	
Daily Production	
Equipment Hours	
Labor Hours	

01-12-01
24
I-95
2.5
6.0
6.0

### Object 2: Maintenance Activity

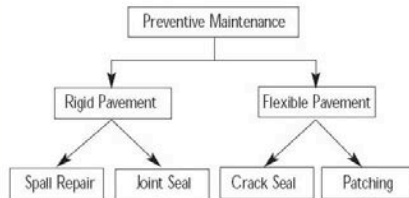
Activity Code	
Activity Name	
Production Unit	
Average Daily Production Rate	

## Hierarchical Model



## Relational

## Network Model



## Client Application NETCONF Server

```
<rpc message-id="messageID">
  <edit-config>
    <target>
      <candidate/>
    </target>
    <config>
      <configuration>
        <system>
          <syslog>
            <file>
              <name>messages</name>
              <contents>
                <name>any</name>
                <warning/>
              </contents>
              <contents>
                <name>authorization</name>
                <info/>
              </contents>
            </file>
          </syslog>
        </system>
        <configuration>
          <msg>
            <config>

```

# XML

```
<rpc-reply xmlns="urn:ietf:params:xml:ns:netconf:base:1.0">
  <ok/>
</rpc-reply>
]]>]]>
```

# Data comes in all shapes and sizes!

## Flat file

	Route No.	Miles	Activity
Record 1	I-95	12	
Record 2	I-495	05	
Record 3	SR-301	33	

## Relational Model

Activity Code	Activity Name
23	Patching
24	Overlay
25	Crack Sealing

Key = 24

Activity Code	Date	Route No.
24	01/12/01	I-95
24	02/08/01	I-66

## Object-Oriented Model

### Object 1: Maintenance Report

### Object 1 Instance

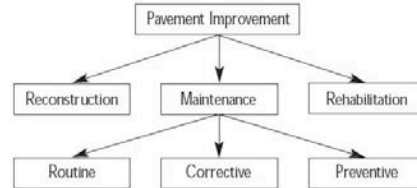
Date	
Activity Code	
Route No.	
Daily Production	
Equipment Hours	
Labor Hours	

01-12-01
24
I-95
2.5
6.0
6.0

### Object 2: Maintenance Activity

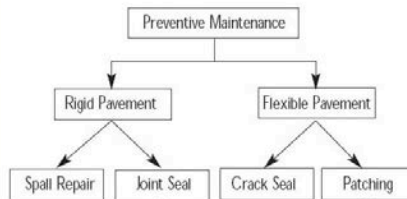
Activity Code	
Activity Name	
Production Unit	
Average Daily Production Rate	

## Hierarchical Model



## Relational

## Network Model



## Client Application NETCONF Server

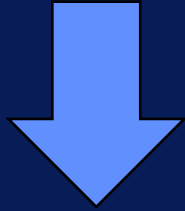
```
<rpc message-id="messageID">
  <edit-config>
    <target>
      <candidate/>
    </target>
    <config>
      <configuration>
        <system>
          <syslog>
            <file>
              <name>messages</name>
              <contents>
                <name>any</name>
                <warning/>
              </contents>
            </file>
            <name>authorization</name>
            <info/>
          </syslog>
        </system>
        <configuration>
          <msg>
            <config>
```

## XML

```
{
  "empid": "SJ011MS",
  "personal": {
    "name": "Smith Jones",
    "gender": "Male",
    "age": 28,
    "address": {
      "streetaddress": "7 24th Street",
      "city": "New York",
      "state": "NY",
      "postalcode": "10038"
    }
  },
  "profile": {
    "designation": "Deputy General",
    "department": "Finance"
  }
}
```

[www.kodingmadesimple.com](http://www.kodingmadesimple.com)

# Data Integration



# Richer Data

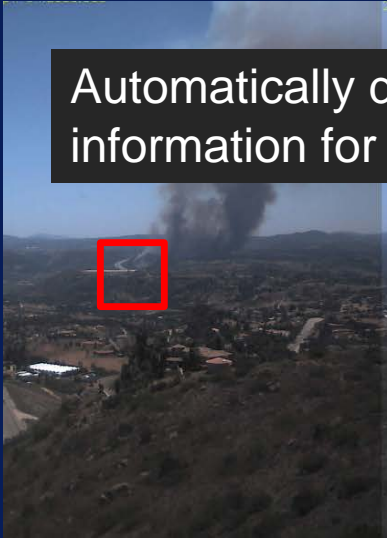
# Fire Detection



Spatial  
capabilities  
with  
Non-Spatial  
data

# Fire Detection

Automatically detected location  
information for fire ignition point



Spatial  
capabilities  
with  
Non-Spatial  
data

# Fire Detection



Automatically detected location information for fire ignition point



# Fire Simulation



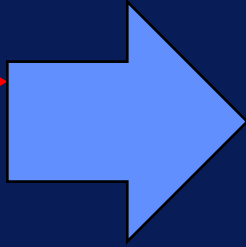
Spatial capabilities with Non-Spatial data



# Fire Detection



Automatically detected location  
information for fire ignition point



# Fire Simulation



Spatial  
capabilities  
with  
Non-Spatial  
data

More accessible data

Data integration



Reduce data complexity

Data integration



Reduce data complexity

Increase data availability

Data integration



Reduce data complexity

Increase data availability

Unify your data system

# Data integration



```
graph TD; A[Data integration] --> B[Reduce data complexity]; A --> C[Increase data availability]; A --> D[Unify your data system]; B --- E[Increase data collaboration]; C --- E; D --- E;
```

Reduce data complexity

Increase data availability

Unify your data system

Increase data  
collaboration

Data integration



```
graph TD; A[Data integration] --> B[Reduce data complexity]; A --> C[Increase data availability]; A --> D[Unify your data system]; B & C & D --> E[Increase data collaboration]; E --> F[Add value to your big data!]
```



Reduce data complexity

Increase data availability

Unify your data system



Increase data collaboration



Add value to  
your big data!