



# Big Data: Fire, Ready Aim

Dr Jerry Rosenbaum  
The Rose Tree Group  
[jsr314159@gmail.com](mailto:jsr314159@gmail.com)



# Some Statistics

- The WSJ has reported that the average ROI for a big data project is 55 cents on the dollar
- Various reports have the failure rate for big data projects as high as 80%
- Only 13% of companies have big data projects in full scale production
- 27% of executives say their big data projects were successful

The goal of this talk is to help you improve your odds.

Dwight D. Eisenhower,  
President of the US,  
and a 5 star general

“In preparing for battle I have always found that plans are useless, but planning is indispensable”

- Every Big Data project is a “battle”
- It is the up front work that is the key to success
- The 7P approach rarely succeeds
- Remember, Eisenhower was responsible for D-Day and Custer was responsible for Little Big Horn

# The Big Picture of a Big Data Project

- Up front preparations and planning
  - WHO, WHAT WHERE, WHEN, WHY, HOW
- Analysis of the data and findings
  - Test ones hypotheses
  - Determine how to best present the results
- Post analysis activities
  - Now what

# First Things First

- Identify a use case
  - Start small (modest), you can always have follow on projects
  - Avoid scope creep (especially for the first few projects)
  - Have a way to measure success (hypotheses)
  - Have a way to know when you are done
  - What you will do with the results
- Identify the project team
  - Who are the business people
    - Senior exec, business sponsor, business expert
  - Who are the technical people
  - Did you include a data scientist or an actuary
  - Who are the security people
  - Has the team ever worked together before
- Select tools (analysis, database, etc.)

# Data Questions

- Identify the data sources
  - Internal databases and files
  - External data
    - What is its structure
    - one time or periodic downloads
    - Legal considerations
  - Any PII or confidential data
  - Does data have to be de-identified
- Where will the data be stored
  - How much data
  - Do we have enough disk storage
  - For how long will be keep the data
  - Can older data be moved to a cheaper storage mechanism

# More Data Questions

- Are our databases (e.g. Hadoop) set up to handle the new data we either acquire or generate
- Do we have metadata for the data
  - Usable clear definitions
  - Does the metadata from various sources (internal and external) agree with each other
  - Does the metadata from external sources conform to your internal metadata
  - What are the units of measure
  - For external data, how was it acquired and aggregated
- What about data governance

# Still More Data Questions

- Will we need to use any ETL processes
  - What do we do about duplicate data, unloadable data, missing data
  - Does the Transform step lose any data precision
- Data quality
  - Accuracy, validity, completeness, timeliness, understandability, etc.
  - Is any data cleansing needed
- Is the data really suitable
  - Acquire samples of the data
  - Run queries or tests to check out actual data



# Few More First Things

- Develop Training Plan
- Budget – Be realistic, but not extravagant
- Project Plan - Make sure time line is realistic
- Sell the project to management (if not previously done)
- Deploy all the pieces
  - Hardware, network, software, tools
  - Access Permissions and security
  - Monitoring
  - Check out all the pieces
- Automate whenever reasonable

# Finally, Do the Work

- Acquire the needed data
  - Do an integrated check out of all the pieces
- Use tools to help analyze hypotheses and find insights
  - Do the insights make sense
  - Can the insights be sharpened
  - How can these be of benefit to the business
- Visualize the results
- Measure potential ROI
- Determine possible next steps
- Present the results to management

# Afterthoughts

- Should this be put into production
- Suggest the next big data project
- Lessons learned

# Thank You Questions

Dr Jerry Rosenbaum  
The Rose Tree Group  
[jsr314159@gmail.com](mailto:jsr314159@gmail.com)