

# Big Data, Analytics and the Mainframe

Paul Pendle  
EMC Corporation

March 15, 2012 7:00AM-8:00AM  
Session Number: 3278



# Big Data – What is it?

“Data not easily handled by the traditional data processing mechanisms of today”

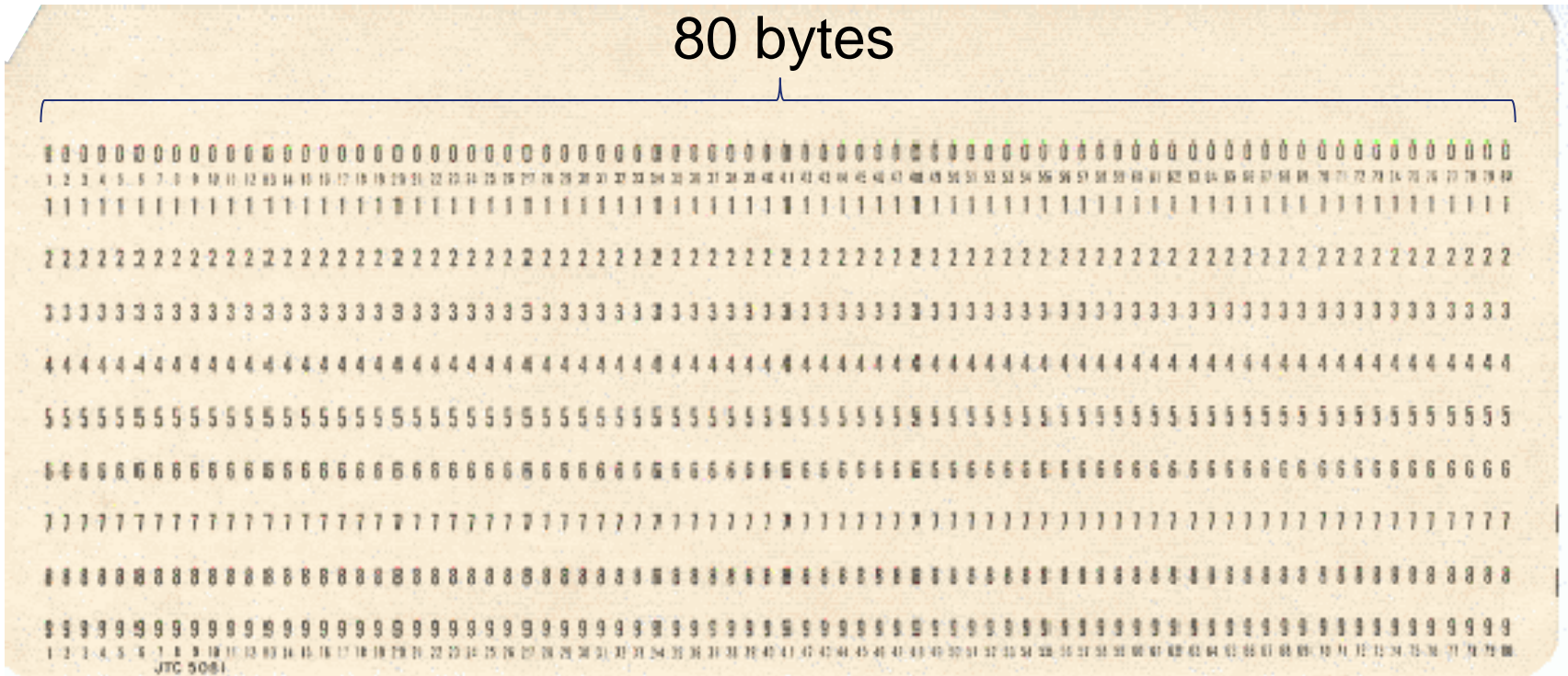
Characterized by:

- High volume and arrival rate
  - Billions of data events per hour
- Semi-structured or unstructured
  - Not easily analyzed or parsed
- Wide disparity
  - Text, video, graphics, sensor data, RFID, web, etc

**“The three Vs”**

# Mainframe Data Feeds

80 bytes



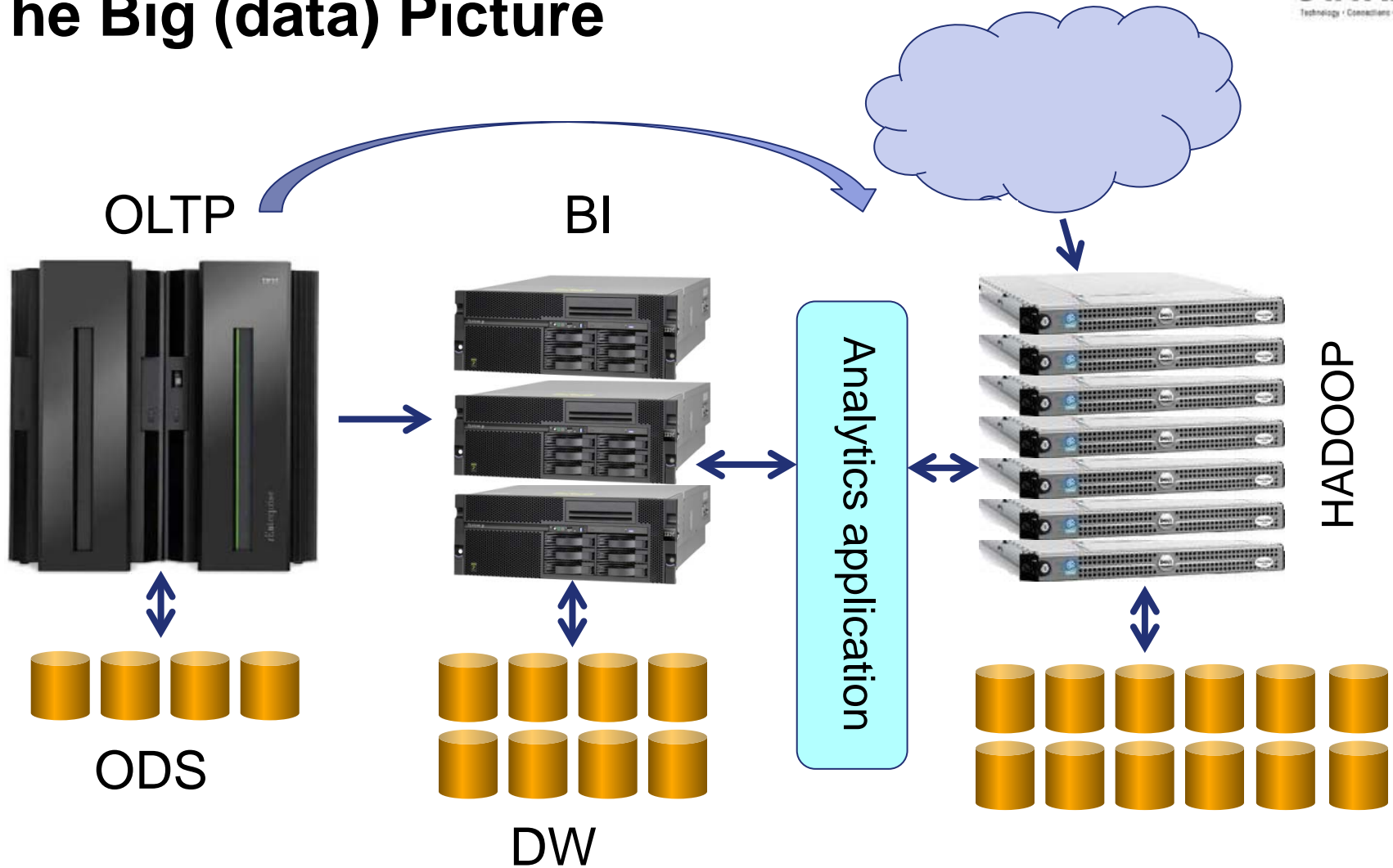
# Mainframe data feeds (3270)



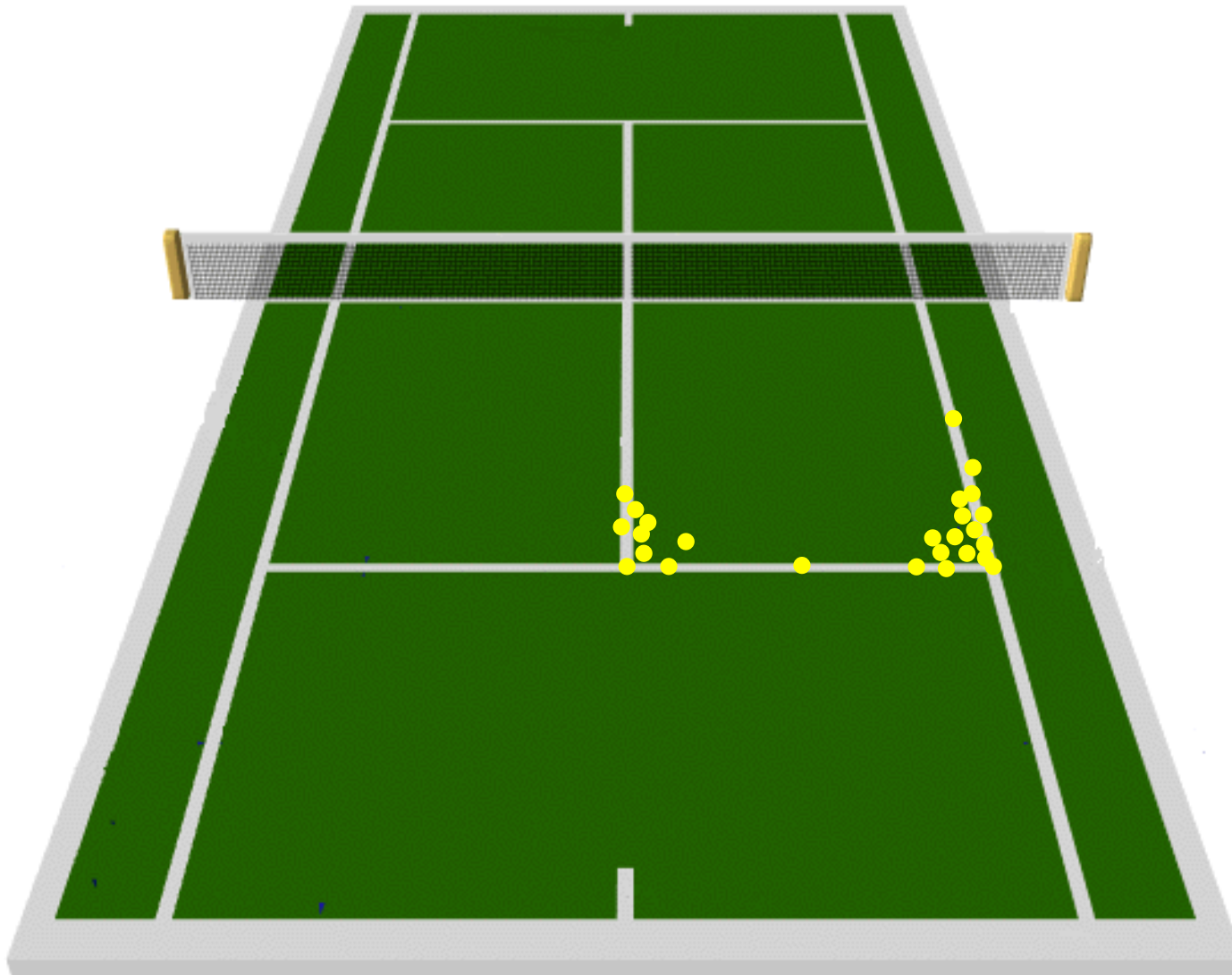
# Mainframe Today

- Still highly structured
- Built for low-cost per transaction
- Small transactions, high volume
- RDBMS
  - Limited capability for unstructured data
  - XML/LOB not fully assimilated in DB2

# The Big (data) Picture



# Australian Open Tennis



# A lot of unanswered questions

- Maturity of new software (Hadoop)?
  - Reliability security
- Integration with existing systems
- How reduce “time to analysis”
- How to get business value from the data
  - Competitive advantage
  - Fraud detection
  - Risk analysis/avoidance