

Opportunities and Challenges

Evolving Big Data Use Cases and Technologies



Mark Davis

Distinguished Engineer, Dell Software Group

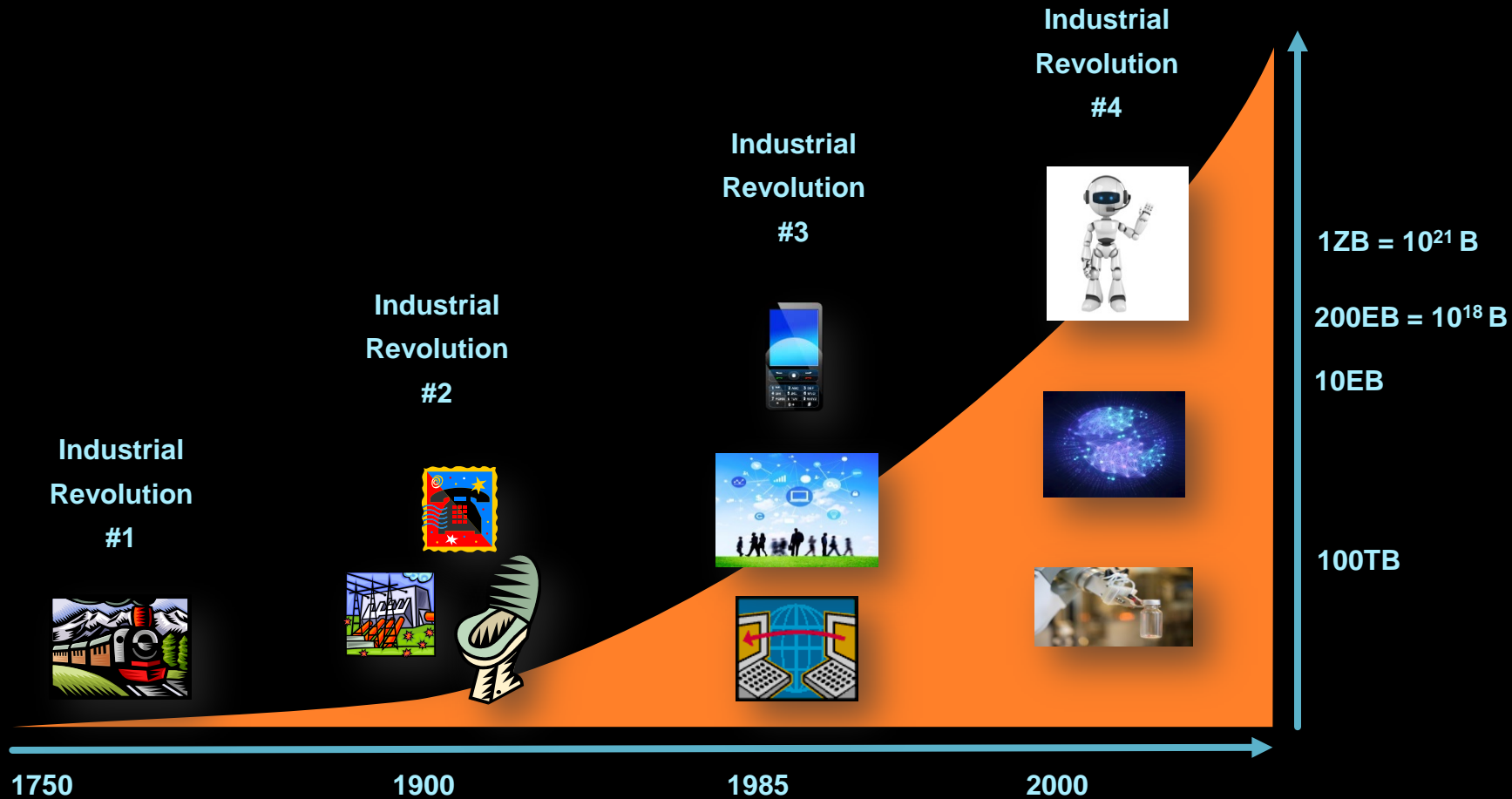
IEEE Cloud Computing Initiative

IEEE Intercloud Interoperability Testbed

Mark.Davis@software.dell.com

The Paleohistory and Protofutures of Big Data





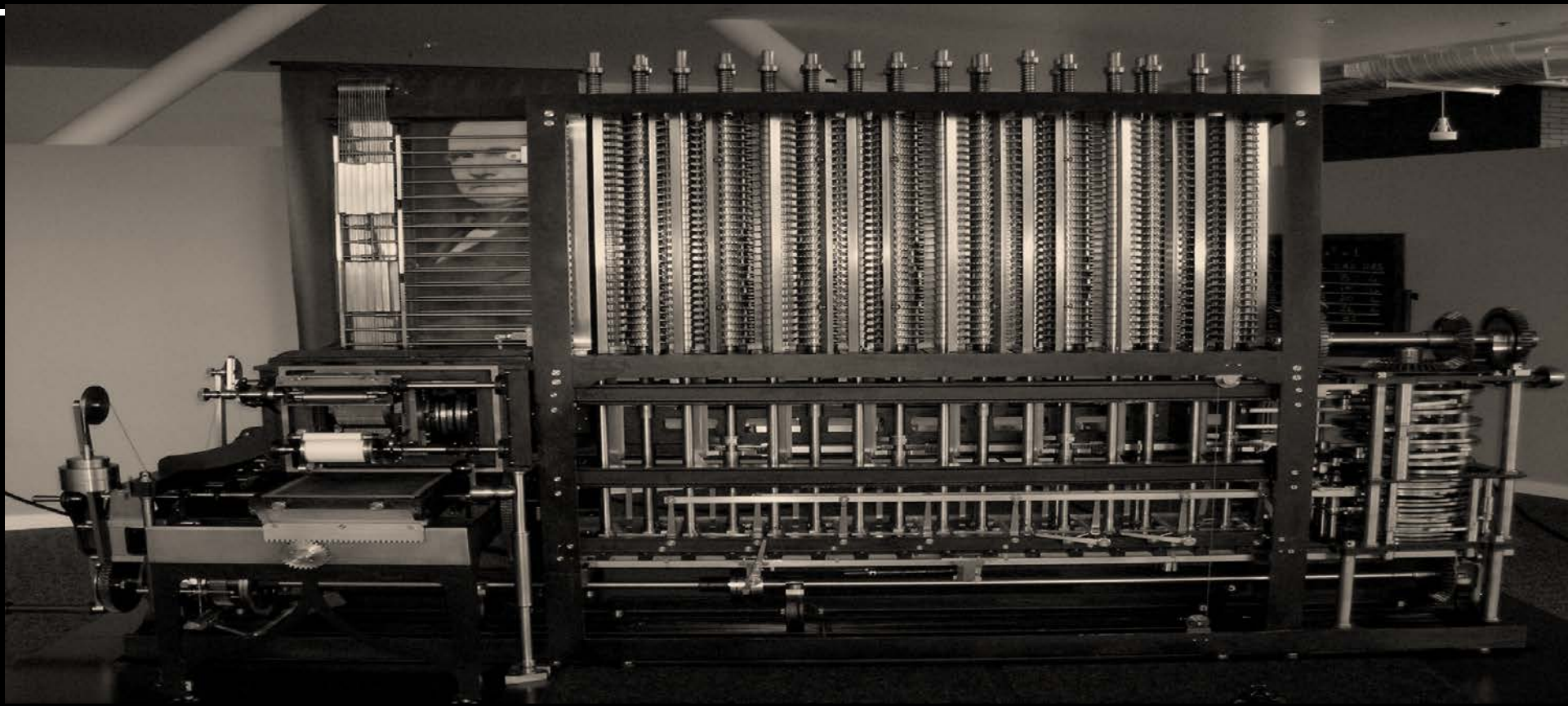
R. J. Gordon: Is US economic growth over? Faltering innovation confronts the six headwinds. CEPR Policy Insight No 63

Krugman, P. Is Growth Over? New York Times, 12 December 2012.

I visualize a time when we will be to robots what dogs are to humans, and I'm rooting for the machines.

Claude Shannon, IEEE Medal of Honor 1966

The Modern History of Big Data



Information Management 1994-2004

The Oracle logo, consisting of the word "ORACLE" in a white, sans-serif font with a registered trademark symbol (®) at the end, set against a red rectangular background.

ORACLE®

The DBA Crisis

- Faster reporting
- Interactive visualization
- \$\$\$
- Data warehousing
- Advanced analytics
- Just-in-time decision making
- The Perl Scripting option



The Google Indexing Crisis

4TB of data on disk

4 channels of I/O at 100Mb/s

2.7 hours



The Social Media Startup Crisis (2006)

- Server farms are too expensive
- Oracle is too expensive
- We just need key-value stores
- VCs say no CAPEX
- Be lean, young entrepreneurs, be lean



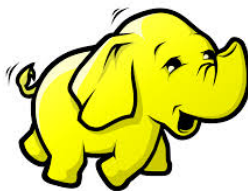
VERTICA
An HP Company



Greenplum



Google



APACHE
H
BASE

Apache
STORM

ORACLE®



Spark

 **mongoDB**



Cassandra



 **riak**

Big Data Use Cases

ia



Analyze log files from web servers

Group users based on behavior patterns

Build machine learning models of behavior

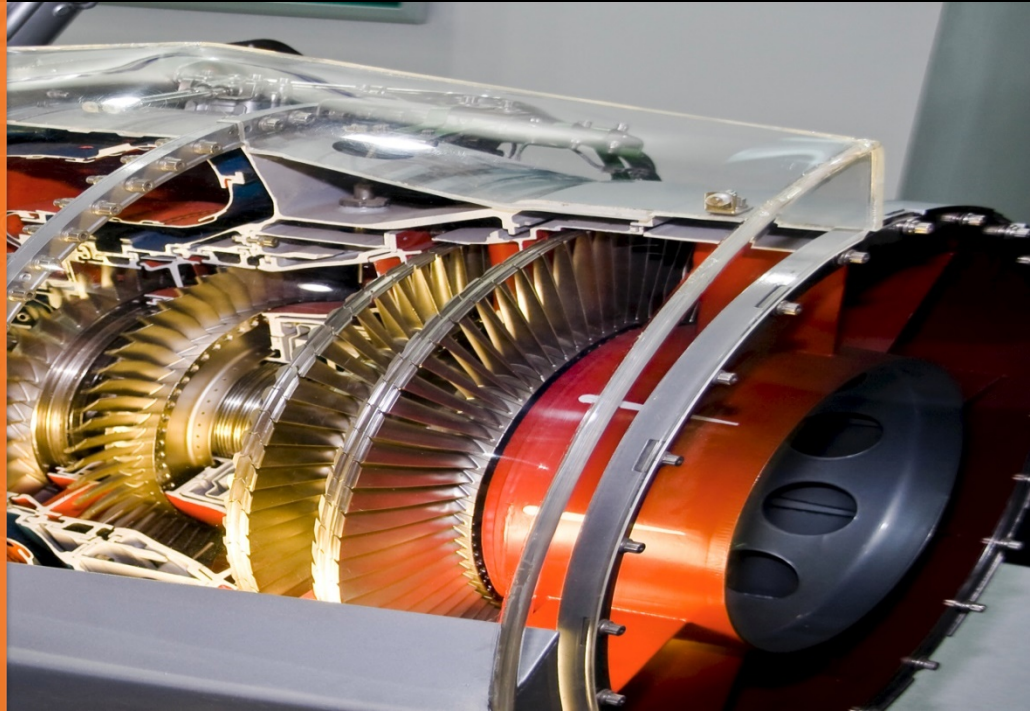
Recommend news and information

Collect millions of data
points from sensors

Analyze failure modes

Proactively improve
products

Optimize systems





Defense Intelligence

Anomaly Detection



Scientific Analysis and Visualization



Collective Intelligence



Social Network Analysis





Defense
Intelligence



Distribution
Network
Optimization

Anomaly
Detection



Spam
Filtering



Scientific
Analysis and
Visualization

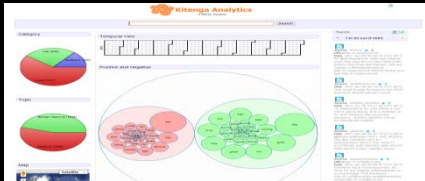


Collective
Intelligence

Internet of
Things



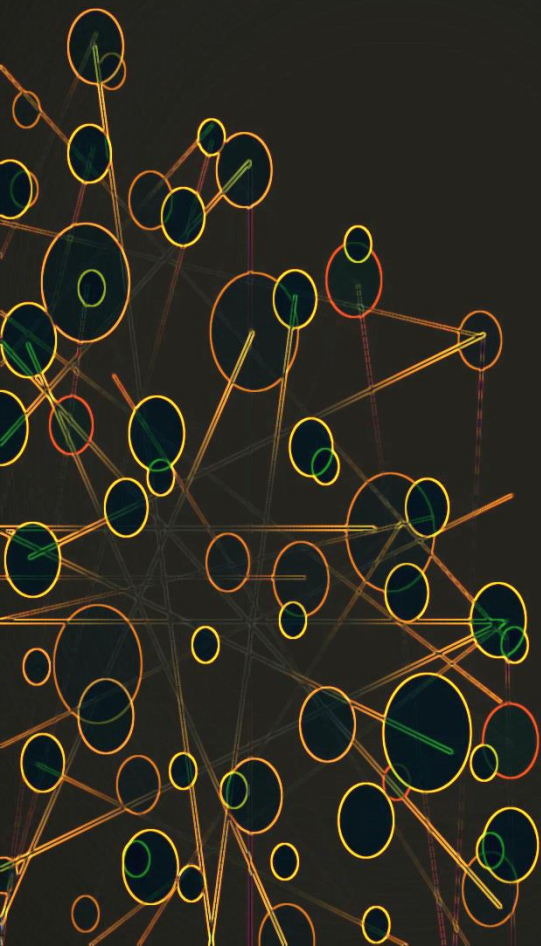
Search



Social Network
Analysis



Product
Design



Scalable
Unstructured
Schema-on-read
Fault-tolerant
Cluster Computing
MapReduce
Graph database
NoSQL
Machine Learning
Streaming Database
Eventually Consistent

Apache
Hadoop
Spark
HDFS
Storm
Hive
Pig
ETL
Solr
Cloudera
Hortonworks



The (Near) Future of Big Data

Education

Data Scientist
Organizational
Leadership

Technology

Simplification
Integration
Use Case
Alignment

Data Availability

Public APIs
Organizational
Culture

Q/A
