OPEN SOURCE BUSINESS CONFERENCE

Building Your Big Data Future with Open Source

COMPUTERWORLD
OSBC
SAN FRANCISCO

The Person
Who Knows the Most Wins
How Data Is Changing the Nature of Business

Mike Olson

Chief Executive Officer

@mikeolson, mike.olson@cloudera.com

Background

- "Big Data" is so 2010.

- "Extreme Data" is a better name:
  - Volume
  - Variety
  - Velocity

Background

- "Big Data" is so 2010.

- "Extreme Data" is a better name:
  - Volume
  - Variety
  - Velocity

Forecast Growth in Data Volume
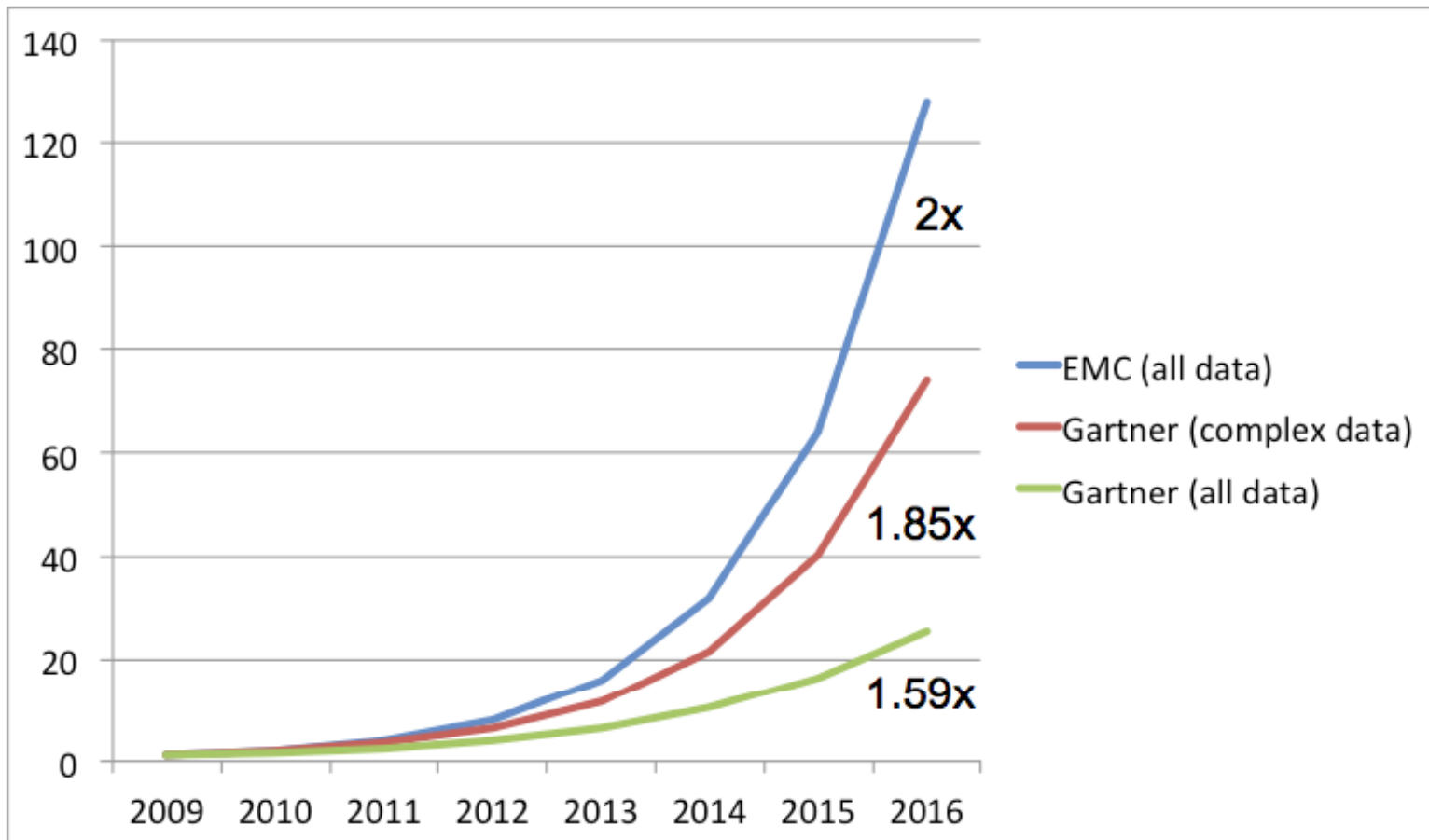
@biz gets our servers ready for the royal wedding
By @Twitter ★ Favorite

Volume, Variety, Velocity

1. Accurate modeling of risk in financial portfolios

2. Analysis of customer acquisition costs and churn

3. Product and content recommendations based on user behavior

4. Advertisement targeting based on preferences

5. Point-of-sale transaction analysis

6. Capture and analyze traffic data on a network to better manage bandwidth among devices

7. Real-time and forensic analysis of network activity to detect and respond to threats

8. Monitor financial trades for evidence of money laundering, fraud

9. Deliver better search results based on user behavior

10. Data exploration and experimentation

Volume, Variety, Velocity

1. Accurate modeling of risk in financial portfolios

2. Analysis of customer acquisition costs and churn

3. Product and content recommendations based on user behavior

4. Advertisement targeting based on preferences

5. Point-of-sale transaction analysis

6. Capture and analyze traffic data on a network to better manage bandwidth among devices

7. Real-time and forensic analysis of network activity to detect and respond to threats

8. Monitor financial trades for evidence of money laundering, fraud

9. Deliver better search results based on user behavior

10. Data exploration and experimentation

Optimizing Hotel Search

Behavioral analysis based on user activity on web site;

Clustering of users with similar behavior into cohorts;

Cohort behavior used to predict behavior and preferences of individuals.

San Jose International Airport, San Jose, California, United States | Check-in: Tue, Apr 26, 2011 | Check-out: Sat, Apr 30, 2011 | Nights: 4 | Room(s): 1 | Guest(s): 2

Call us to book!
1-800-733-1297
(toll free)

Why book on Orbitz? Price Assurance + Low Price Guarantee + No Orbitz hotel change or cancel fees Learn more

Stars
☑ Any
☐ ★★★★★
☐ ★★★★☆
☐ ★★★☆☆
☐ ★★☆☆☆
☐ ★☆☆☆☆

Reviewer score
1.0 to 5.0
1 ———————— 5

San Jose Airport Garden Hotel
★★★☆☆   Save | Review

Nightly rates from
$119 $83
Total Price $94
Select
Price Assurance

Overview | Description | Photos | Map | Amenities

Reviewer score 3.4 out of 5 66 reviews

1740 North First St., San Jose, CA 95112-4584
1.2 miles North from the center of San Jose International Airport

More hotel details | Virtual tour

• Book Now and Save 30%

Special Offer

View all available promotions

Amenities
☑ Any
☐ Airport Shuttle (7)
☐ Free parking (123)
☐ Pool (134)
☐ Spa services (4)
☐ Wireless Internet (59)
See all

Hotel chain
☑ Any
☐ All Seasons Europe (0)
☐ Aloft Hotels (0)

Arena Hotel
★★★☆☆   Save | Review

Nightly rates from
$66
Total Price $73
Select
Price Assurance

Overview | Description | Photos | Map | Amenities

Reviewer score 3.6 out of 5 77 reviews

817 The Alameda, San Jose, CA 95126
1.6 miles Southeast from the center of San Jose International Airport

Source: Orbitz, "Extending the Data Warehouse with Hadoop",
http://www.cloudera.com/resource/chicago-data-summit-extending-the-enterprise-data-warehouse-with-hadoop-orbitz-worldwide

Recommendations

Predictive analytics present users with choices that are more likely to appeal to them.

More appealing choices translates into more rooms booked.

Further refinements: A/B testing on displayed options to increase the transaction rates and prices.

The Whitehall Hotel Rooms

Superior King
This room offers one king bed with cable TV, air conditioning, in room safe, hair dryer, iron and ironing board.

Pinnacle King
One king bed. Restored antique furnishings, crown molding, original artworks. Premium bedding with 390-thread-count sheets. Turndown service. Oversized desk with leather desk chair. Bathrooms with marble-and-tile countertops and floors, bathrobes, designer toiletries. Cable TV with premium channels. Hair dryer, in-room safe, minibar. High-speed Internet access (surcharge) and wireless printing to business center (surcharge). Two dual-line telephones (one in the bathroom). Umbrella, clock radio. Complimentary ice service and 24-hour room service.

Show all room descriptions

Recommendations for you
Customers who viewed The Whitehall Hotel also booked these hotels

★★★★☆
The Talbott Hotel

Reviewer score 4.6 out of 5

★★★☆☆
Ambassador East Hotel

Reviewer score 3.8 out of 5

★★★☆☆
The Tremont Chicago Hotel at Magnificent Mile

Reviewer score 3.7 out of 5

Source: Orbitz, "Extending the Data Warehouse with Hadoop",
http://www.cloudera.com/resource/chicago-data-summit-extending-the-enterprise-data-warehouse-with-hadoop-orbitz-worldwide

Volume, Variety, Velocity

1. Accurate modeling of risk in financial portfolios

2. Analysis of customer acquisition costs and churn

3. Product and content recommendations based on user behavior

4. Advertisement targeting based on preferences

5. Point-of-sale transaction analysis

6. Capture and analyze traffic data on a network to better manage bandwidth among devices

7. Real-time and forensic analysis of network activity to detect and respond to threats

8. Monitor financial trades for evidence of money laundering, fraud

9. Deliver better search results based on user behavior

10. Data exploration and experimentation

VoIP Packetization and Transmission

VoIP call quality depends on quality of service – if data packets cross the network at different speeds, users will hear static or drops in their calls.
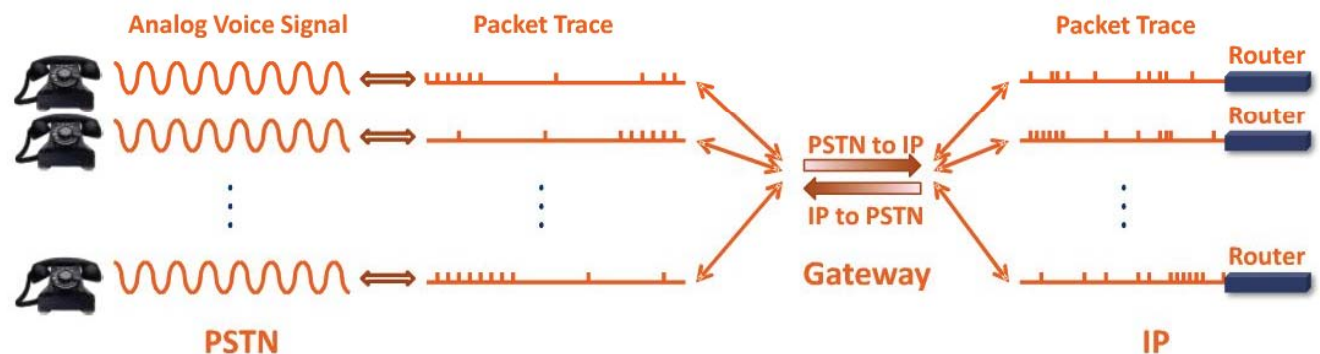
Analyzing the distribution of packet transmission rates and delays, and correlating with QoS guarantees in the network, allows carriers to provide high-quality service cheaply.
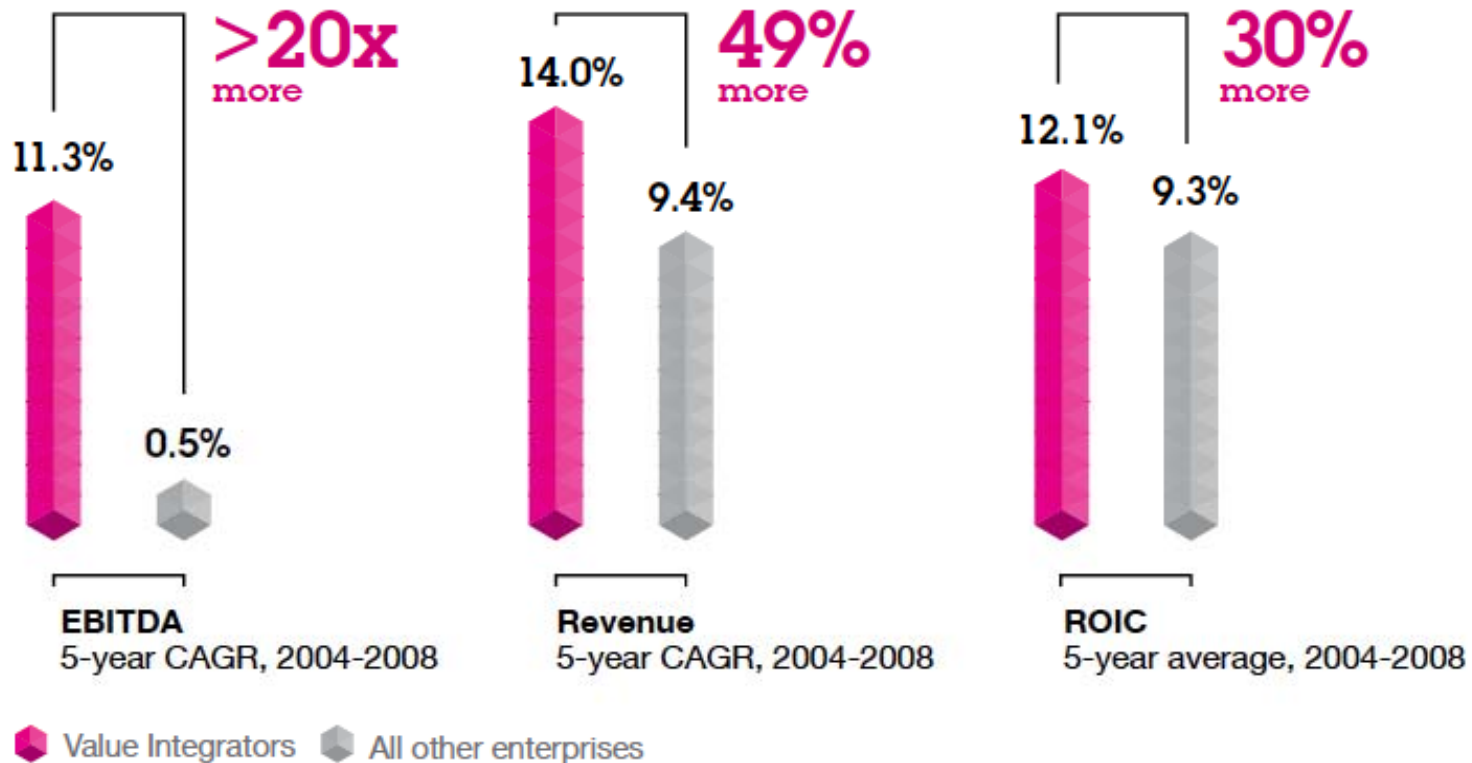
- Semi-call Packetization

20 ms 2 s
Caller → Internet ← Transmission interval Callee

- Transmission through IP network and PSTN via gateway

Analog Voice Signal — Packet Trace — PSTN to IP / IP to PSTN — Gateway — Packet Trace — Router

PSTN IP

Source: Revolution Analytics, "Voice Over IP: Studying Traffic Characteristics for Quality of Service using R and Hadoop",
http://www.cloudera.com/resource/hw10_voice_over_ip_studying_traffic_characteristics_for_quality_of_service

Source: IBM, "The New Value Integrator,"
http://www-935.ibm.com/services/us/cfo/cfostudy2010/

Building Your Big Data Future with Open Source

Should a CFO Care?

Driving integration of information across the enterprise
Measuring / monitoring business performance
Developing your people in the Finance organization
Providing inputs into enterprise strategy
Executing continuous Finance process improvements
Supporting / managing / mitigating enterprise risk
Driving Finance cost reduction
Driving enterprise cost reduction
Strengthening compliance programs and internal controls

0% 20% 40% 60% 80% 100% 120%

Percent more effective than baseline
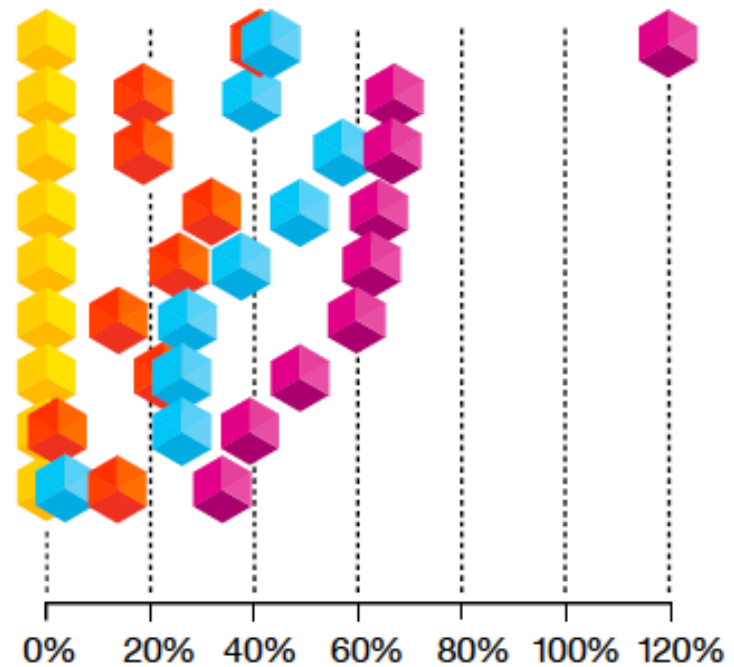
- Scorekeepers (baseline)
- Disciplined Operators
- Constrained Advisors
- Value Integrators

Source: IBM, "The New Value Integrator,"
http://www-935.ibm.com/services/us/cfo/cfostudy2010/

Should You Care?

- In business, medicine, agriculture, energy and other domains, data presents us with transformative opportunities.

- The stakes have never been higher.

- Winning is imperative.