What is Big Data?

![Big Data Infographic]

**Volume**: Scale of Data
- 40 Zettabytes: estimated data will be created by 2020, an increase of 300x times from 2005.
- 6 billion people have cell phones.
- 500 billion photos uploaded to Facebook.

**Velocity**: Analysis of Streaming Data
- 1TB of Oil Information processed during most trading sessions.
- Most cars have close to 100 sensors.

**Variety**: Different Forms of Data
- 30 billion pieces of content shared on Facebook every month.
- 6 billion hours of video are watched on YouTube each month.
- 400 million tweets are sent per day by about 500 million monthly active users.

**Veracity**: Uncertainty of Data
- 27% of respondents don’t trust the information they used to make decisions.
- The U.S. economy lost $12 billion a year.

**The Four V’s of Big Data**
- Volume
- Variety
- Velocity
- Veracity

By 2020, it is expected there will be 1.8 billion network connections, amounting to 2.5 connections per person on earth.
The Four V’s of Big Data

- Volume – Yottabyte, Zettabyte, Petabyte, Terabyte, Gigabyte, Megabyte
- Velocity – Real Time, Near Real Time, Periodic, Batch
- Veracity – Anomaly, Biased, Incomplete, Noise
- Variety – Unstructured, Social, Sensor, Mobile, Video, Photo, Audio, Database, Table
Big Data Landscape
How fast is data growing?

Every 60 seconds:
- 98,000+ tweets
- 695,000 status updates
- 11 million instant messages
- 698,445 Google searches
- 168 million+ emails sent
- 1,820TB of data created
- 217 new mobile web users
How fast is data growing?

Data growing forecast

1 zettabyte = $10^{21}$ bytes
1,000,000,000,000,000,000,000,000 bytes

- 2012: 23 billion global users
- 2017: 3.8 billion global users
- 2012: 12 billion global networked devices
- 2017: 19 billion global networked devices
- 2012: 11.3 Gbps global broadband speed
- 2017: 39 Gbps global broadband speed
- 2012: 0.5 zettabytes global traffic
- 2017: 1.4 zettabytes global traffic

What Devices / Application generate data?
Top Big Data Use Cases

- Customer Analytics: 48%
- Experience Analytics: 45%
- Risk Analysis: 37%
- Threat Analysis: 30%
- Regulatory Compliance Analysis: 28%
- Location-based Targeting: 23%
- Campaign Optimization: 26%
- Fraud Analysis: 22%
- Brand Sentiment Analysis: 16%
- Drug Discovery: 16%
- Product Placement Optimization: 16%
- Other: 9%
- Other: 1%

Other: 1%
Big Data and Horse Racing

Volume – Velocity – Veracity – Variety

- Tote companies
- Equibase
- TrackMaster
- Jockey Club
- Daily Racing Form
- Brisnet
- InCompass

- Equineline
- CHRIMS \ PGSI
- Regulators
- RCN
- Video
- Universities and Colleges
Data Touch Points

- Past Performances – Coupled with wagering data
- Weather – Track conditions and speed related to weather conditions
- Track conditions – Temperature and moisture throughout day
- Morning workouts – Real time results from tracks and training centers
- Scratches – Analysis of scratches per horse
- Equipment changes – Analysis of equipment changes
- Jockey changes – Analysis of Jockey changes
Data Touch Points

- Advanced Deposit Wagering – Segment where and how wagers are being placed
- Morning Line odds – Overlay over MLO to Opening and Current
- Stop betting times – Histories and analysis of wagering cycles by horse and pool
- Race times by horse - Comparing race times by horse, track, conditions, etc.
- Odds changes – Timing intervals of odds changes by race, horse, pool, conditions etc.
- Over Lays - That novices can understand
Data Touch Points

- Percentage odds – Graphics that show both
- Analysis of betting patterns – Analysis of wagering patterns
- Horse imaging – Workout, Paddock, Post Parade, Race
- Paddock – Temp., Humidity, Horse’s Condition
- Horses’ Running Analysis - Overlays of horses running styles
- Jockeys’ Ridding Analysis - Overlays of Jockeys’ ridding styles
Total Handle $1,167,432,027

**NETAMOUNT BY POOL**

- 2013: 15k, WP6
- 2014: 30k, WP6
- 2015: 10k, WP6

**NETAMOUNT BY EVENT**

- Derby Day: 10k, WP6
- Preakness Day: 30k, WP6
- Belmont Day: 15k, WP6
Total Handle $1,315,587,302
THANK YOU!