The D in CDO is for Disruption

Peter Wang

CTO, Co-founder
Continuum Analytics

pwang@continuum.io
THE DATA REVOLUTION
EVERYTHING
^ THE DATA REVOLUTION
DATA is Everywhere
EVERYTHING surrounding the data is changing
Why Not Just Use Faster Computers?

Microprocessor Transistor Counts 1971-2011 & Moore’s Law
Why Not Just Use Faster Computers?
3D Xpoint
• Persistent
• 1000x faster than SSDs
• 10x more dense than RAM
• Available in 2017
THE DATA REVOLUTION is just beginning

Technology innovation is accelerating

Every aspect of how we ingest, store, manage and compute on business data will be disrupted
THE CDO IN “INTERESTING TIMES”
Anti-Patterns

What you don’t want to do

- Sandbagging
- Making “safe bets”
- Outsourcing your core
- Reinforcing silos
‘D’ is for Disruption

No Silver Bullet  Team >> Tech  Coach  Agility
BE OPEN.
SET UP A BIG TENT.
INVITE EVERYONE
TO THE DATA PARTY
Data Science Is A Team Sport

**Biz Analyst**
- Spreadsheets
- Visualization
- Notebooks
- Analytic Development Environment

**Data Scientist**
- Hadoop / Spark
- Programming Languages
- Analytic Libraries
- IDE
- Notebooks
- Visualization

**Developer**
- Programming Languages
- Analytic Libraries
- IDE
- Notebooks
- Visualization

**Data Engineer**
- Database / Data Warehouse
- ETL

**DevOps**
- Database / Data Warehouse
- Middleware
- Programming Languages
Open Data Science
Vibrant and Growing Community

Python Community
30M+

R Community
16M+

Spark Python Usage
50%+

ANACONDA Downloads*
3M+
Open Source?

- Not just about source code
- Of the people, by the people
  - must have community buy-in
- Your goals:
  - Avoid lock-in
  - Harness innovation
OPEN DATA SCIENCE PLATFORM
ACCELERATE. CONNECT. EMPOWER.

Package Management  Operations Management  Enterprise Notebooks  Data Exploration
Interactive Visualization  Scalable Data Science  High Performance Hadoop

Deploying Data Science Models & Visualizations Easily
Building Interactive Visualizations & Web Applications
Collaboration & Sharing of Results Easily
Scaling Up & Out Data Science to Meet Performance Goals
Reproducibility & Governance for Data Science Assets
The Jupyter Notebook is a web application that allows you to create and share documents that contain live code, equations, visualizations and explanatory text.
Interactive Visualization
Gorgeous Data Narratives
Anaconda Mosaic

• Create PORTABLE transformations
• Interactively EXPLORE heterogeneous data
• Easily ANALYZE large flat file repositories
• ELIMINATE data movement and redundant storage
• CATALOG datasets and transformations
• ESTABLISH data lineage
Interactive Visual Statistics on Millions of Rows
Anaconda Fusion: Excel Data Science
OPEN DATA SCIENCE PLATFORM
ACCELERATE. CONNECT. EMPOWER.

Package Management  Operations Management  Enterprise Notebooks  Data Exploration
Interactive Visualization  Scalable Data Science  High Performance Hadoop

Deploying Data Science Models & Visualizations Easily
Building Interactive Visualizations & Web Applications
Collaboration & Sharing of Results Easily
Scaling Up & Out Data Science to Meet Performance Goals
Reproducibility & Governance for Data Science Assets