DATA IS IN OUR DNA

Citizen Data Science initiative at Seagate
and where KNIME fits in

Seagate Technology Operations & Technology Advanced Analytics Group

Allan Luk, Analytics Business Solutions Director
Dr. Eric Lin, Sr Staff Analytics Business Solutions Lead

March 8, 2018
About the Speakers

Allan Luk
Operations & Technology Advanced Analytics Group
Analytics Business Solutions Director
Seagate Technology (US)

- Engineering background (MS, ME). Certified PMP, CSQE.
- Career chapters in R&D, Product Development, Quality & Advanced Reliability, Customer Experience Engineering and Analytics
- Manage Seagate Technical Data Science & Data Analytics projects.
- Manage US and Asia Analytics Business Solutions teams to deliver advanced analytics business solutions, apps development & deployment and conduct training within the company.
- Lead Seagate Citizen Data Science Initiative & Guided Analytics Software (KNIME) Development & Deployment.
- Build Citizen Data Scientist Community inside Seagate; Changing Culture.

Dr. Eric Lin
Operations & Technology Advanced Analytics Group
Sr. Staff Analytics Business Solutions Lead
Seagate Technology (US)

- Engineering background (PhD).
- Career chapters in Product Performance Engineering, Supply Chain, Reliability, Cloud, Firmware and Analytics.
- Passionate about data, analytics tools and automation.
Presentation Overview

Share our story and journey…

• Background

• Building a Citizen Data Science Community at Seagate

• What do we need?

• Where KNIME fits in?

• KNIME use case examples & Live demo

• Summary
You May Know Seagate as a Hard Drive Manufacturer...

- #1 OEM storage
- 1st to build and ship a collective zettabyte to the world
- Stores more than 40% of the world’s data
But We’re also a Company that:

- Serves many types of customers and businesses
- Delivers deep expertise and unique IP in storage & data management
- Combines UX, software & design capabilities to create new categories of storage solutions
- Ranks as one of the top 25 companies worldwide in supply chain operations
Analytics Types

- **Descriptive**
  - Insight
  - What happened?

- **Diagnostic**
  - Foresight
  - When will it happen again?

- **Predictive**
  - Insight and Experience
  - Do this when it happens

- **Prescriptive**
  - Oversight
  - Learn and react

- **Cognitive**
  - Oversight
  - Learn and react

Value vs. Difficulty graph.
Gartner defines the citizen data scientist (CDS) as:

“A person who creates or generates models that use advanced diagnostic analytics or predictive and prescriptive capabilities, but whose primary job function is outside the field of statistics and analytics.”

As data scientists are increasingly taxed with requests to make a business more data-driven, citizen data scientists can help their businesses in two key ways:

• They can lighten the load of data scientists by using the right business intelligence software to do the simpler data science tasks.

• They can bring an outsider’s, business-side perspective to data science.

Source: https://blog.capterra.com/4-easy-steps-to-become-a-citizen-data-scientist/
Quotes from Gartner Inc.

Advanced Analytics and Data Science

Advanced analytics and data science are fast becoming mainstream solutions and competencies in most organizations, even supplanting traditional BI and analytics resources and budgets. They allow more types of knowledge and insights to be extracted from data. To become and remain competitive, enterprises must seek to adopt advanced analytics, and adapt their business models, establish specialist data science teams and rethink their overall strategies to keep pace with the competition. "Predicts 2017: Analytics Strategy and Technology" offers advice on overall strategy, approach and operational transformation to algorithmic business that leadership needs to build to reap the benefits.

- By 2018, deep learning (deep neural networks [DNNs]) will be a standard component in 80% of data scientists' tool boxes.
- By 2020, more than 40% of data science tasks will be automated, resulting in increased productivity and broader usage by citizen data scientists.
- By 2019, natural-language generation will be a standard feature of 90% of modern BI and analytics platforms.
- By 2019, 50% of analytics queries will be generated using search, natural-language query or voice, or will be auto-generated.
- By 2019, citizen data scientists will surpass data scientists in the amount of advanced analysis produced.

Who are our Citizen Data Scientist candidates?

- Professional staff who conduct analysis currently with Excel or other descriptive statistics software
- With a passion for data. Driven to pull together the right data.
- Iterative workflow – one question leads to the next.
- Not a trained data scientist or developer. May or may not have good hacking/coding skills.

Analytics Types

- Identify a software for these areas
Who are our Citizen Data Scientist candidates? What will help them?

- What will help them?
  - Guided analytics software
    - Easy to use
    - Able to build complex data preparation & analytics workflows with drag and drop comments
    - Can be enhanced with R & Python through inclusion of script modules
    - Power users to build standalone applications & share with others
    - Ability to do push-down analytics

but beware of the **Danger Zone**

Data Prep Tools

Drag and Drop Analytics Tools

Programming / Hacking

Math & Statistics

Descriptive Analytics

Subject Expertise

Machine Learning

Data Science

© Seagate Technology LLC, 2018
Citizen Data Science Initiative - Building a Citizen Data Science Community

Citizen Data Science (CDS) Initiative @ Seagate

Turning Data into INSIGHTS

- CDS & Guided Analytics Software Workshop
- Guided Analytics Software Deployment
- CDS Certification
- CDS Community Building

Awareness ➔ Software & Tools ➔ Capability ➔ Ecosystem: Community Contributions

© Seagate Technology LLC, 2018
Building a Seagate Citizen Data Scientist Community
• The Seagate Citizen Data Scientist (CDS) Community site launched in Oct 2017 is a platform for exchanging ideas and views about Citizen Data Science and KNIME topics.

• This site includes a link to a repository where the KNIME workflows can be saved. [Link]

• It also has a Discussion Forum page and a KNIME Tips page.

General Data Science / KNIME Articles from Forbes, Gartner, Kaggle etc

Practical tips and best practices on using KNIME like ETL, Visualization, Connection to db and etc.

Compilation of actual use cases from KNIME users in R&D, Operations, PLM and etc.

Collection of KNIME workflows shared by the users like Text Mining, Machine Learning, Auto email and etc.

Q&A, Discussions about Citizen Data Science & KNIME like issues, errors, new discoveries etc

© Seagate Technology LLC, 2018
KNIME Applications at Seagate - Examples

KNIME Functionality & Usage

Prediction
- Image Analytics
- Text Analytics
- Drive Yield Sensitivity
- Field Return modeling

ETL
- Hive/SQL Connectors
- KNIME to Google Sheets
- Other data query & connections (ODS/EDW, MySQL, SQLite)
- Data Cleaning & Processing

Integration Platform
- D3 Javascripting Visualization
- R & Python Integration
- Shell Command & JMP Integration
- Emailing & Scheduling
- Web portal
KNIME streamlines Excel reporting with cell coloring

KNIME integrates analytics in manufacturing production lines

KNIME simplifies HR analytics

KNIME automates key Drive parametric performance monitoring

Use Case Examples

PHM

Reliability

Production Lines

R & D

Quality monitoring

HR

© Seagate Technology LLC, 2018
KNIME Applications at Seagate - Examples

Thousands of Processing & Machine Learning nodes to achieve your analytics needs.
KNIME Applications at Seagate - Examples

Retrieve & combine data from any Seagate database and from any google sheet for your ML workflow integration!
KNIME Applications at Seagate - Examples

- **Integration Platform**
  - Integration of R & Python
  - Shell Command & JMP Integration
  - Emailing & Scheduling

- **Visualization**
  - D3 Javascripting

- **Webportal**

- **Capability of integrating Open Source / External Tool**
  - Enabling analytics flexibility & possibility.
KNIME Live Demo
Seagate Citizen Data Science Community Vision
Summary

- Seagate Citizen Data Science initiative deployment
  - Forming a community of Seagate Citizen Data Scientists
  - Cultural change
  - Education on Analytics ‘Danger Zone’

- Early phase of KNIME adoption & deployment
  - Still on the learning curve
  - Use case evaluation from different functional groups underway
  - Create Seagate specific nodes, features & workflows to enhance the software

- Guided analytics software (KNIME) to bridge the gap
  - Requirements:
    - Three E’s - Easy to access, Easy to learn, Ease of use … [open source]
    - Integrate with other data sources and existing tools
    - Capabilities: Data query, mining, predation, prediction, push-down analytics, visualization, reporting, integration platform and etc.