DATA IS POTENTIAL

Scott Morrison – Data Analytics Engineer
Brendan Doherty – Data Analytics Engineer

KNIME Spring Summit - 2019, Berlin
Data Science in Action
Outline

- Introduction to Seagate
  - Who we are within Seagate
  - Background information on the product
  - Data Challenges

- Embedding KNIME in a manufacturing environment
  - Equipment Tool Root Cause Diagnostic
  - Pathway Taken – Springtown N.Ireland

- Q&A
Who are we?

Seagate Analytics Group

People come from many different organizations within Seagate as well as the outside.

We have individuals located in most of the time zones where Seagate operates.
- Longmont (USA)
- Minneapolis (USA)
- Derry/Londondery – N.Ireland (Springtown)
- Wuxi (China)
- Singapore
- Korat and Teparuk (Thailand)

Languages
- English (American and Irish)
- Chinese (Mandarin and Cantonese)
- Thai
- Spanish
- Amharic (Ethiopia)
- Hindi, Telugu (India)
- Filipino

Actively participating in new college grad & intern programs to ensure we stay up to date on the latest advances in analytics.

Partnering with IT teams for data and technology delivery.
Hard Drive Scale Analogy

- Boeing 747 flying 1mm above the ground, Circling the earth once every 25 seconds
  - Counting every blade of grass along the way
  - Shrink that down so it fits in the palm of your hand
  - You have something equivalent to a modern day hard drive

Sourced from Ted.com, 2015
Disk Drive Components

Recording Heads

Wafer

R/W Head

Head Gimbal Assembly (HGA)

Head Stack Assembly (HSA)

Media

Disk Drive

Vertically integrated company
Complexity of Read/Write head

- Many components
- Many measurements
- Lots of data!
The Challenge

1 Wafer

X 1xxx’s HDD Drives

Wafer Process facts
- 100,000 complex micro-electronic devices per wafer
- Up to 1600 steps to build one wafer
- 100+ different types of chemical, compounds and elements

Critical Process Metrics
- Zero excursions
- Downstream customer performance/yield
- Process cycle time
- Process yields

Process Data
- 150 million Oracle records per day
- 1.3M individual data points per week – structured & unstructured data sets, image files, equipment data
- Over 400 million device electrical measurements per week

- Bottom line – High margin for error, early containment is vital
Embedding KNIME in a Manufacturing Environment
Equipment Tool Root Cause Diagnostic
Problem:

Electrical Measurement

Tool Sensor Trace

DATE

222
192

Sensors

X 398
KNIME Workflow

**Data Extraction**
- KPIV – Time series tool sensor data (Model inputs)
- KPOV – Electrical test data (Model target)

**Diagnostics**
- Data cleanse
- Feature extraction
- Classification
- Variable influence
Feature extraction captures the characteristics of a sensor trace via:

- **Statistical Features**
- **Time Domain Features**
- **Freq Domain Features**

Taking what's visible to the human eye and making it visible to the machine.
Diagnostics Analysis Meta-node

Feature 1

Feature 2

Data Cleanse

Feature Extraction

Classification

Python Script (1-1)

Joiner

Table to R

Add Table to R

R to Table

Extract top vars data

Node 205

Node 206

Node 207

Node 201

Node 10

Node 109

Node 12

Node 13

Node 126

Sensor names

Missing Value

Column Filter

Column Filter

Missing Value

Low Variance Filter

Join features to xPDI

Row Filter

Report

Load data

Best features per sensor

Rank best features

Seagate Technology LLC

Open for Innovation

KNIME

technology

python

R

technology

Seagate Technology LLC
Output - Root cause results

Results

Benefits
• Get to root cause of issue in minutes, rather than analysing volumes of data
• Quicker decisions on defective material
• Can hand off complex programming projects to non native programmers

Saving
• Multimillion dollar saving from mitigation of scrap events
Embedding KNIME in a Manufacturing Environment
Pathway taken
Embedding KNIME in a High Volume Semi-Conductor Factory

Cross Pollination and Synergy between groups

Data Analytics

Failure Analysis  Quality  Manufacturing

R & D  Industrial Eng  Process Eng
Embedding KNIME in a High Volume Semi-Conductor Factory

Pathway of KNIME for Data Analytics and Task Automation in Seagate Springtown
Product Performance Use Case
ARIMA Modelling using Product Performance Data

Problem: Long manufacturing cycle time for recording heads, feedback loop on how the product performs can be in the order of months.

Benefit: Predicted trends of product performance provide quicker feedback. ARIMA modelling used to highlight single anomalous data points months before the actual physical product arrives at the point of measurement.

Saving: Cost and resource benefit saving from being able to highlight under performing products much earlier in the manufacturing cycle.
Process Engineering Use Case

Engineering Holds Report Automation

Problem: ~3 hours per day spent by Process Engineering compiling a WIP (work-in-process) hold report. Very time consuming process. Report was not standardised and not always available first thing in the morning when required.

Benefits: Automated process saves daily manual engineering input. Workflow runs in seconds every morning before engineers arrive at work allowing them to tackle WIP in a hold state faster and through a standardised format.

Saving: The use of KNIME to automate the process has standardised the format and saved ~90 hours of engineering time per month allowing them to concentrate on other tool issues and projects. Improved cycle time through engineers tackling WIP on hold faster.
Industrial Engineering Use Case

Cycle Time Improvement through Data Wrangling

Problem: Very difficult to identify wafers which exceeded particular attribute criteria using either JMP or SQL for users in Industrial Engineering group.

Benefit: Makes data cleansing and data manipulation easier for non-programmers. Integration with JMP to provide familiar charts which are easily interpreted by engineers.

Saving: Improved cycle times through the identification of tools and locations when lots are spending an increased time and the next processing step is not located close by.

Photo Lots Exceed 30 Minutes In Transit Time where next step is not expose
Manufacturing Use Case
Undo Track In Analysis Automation and Reporting

**Problem:** Tool sets have higher cycle times due to products being scanned into them erroneously and then scanned out again by operators without the tool processing the product.

**Benefit:** Project by an intern after some training on KNIME was able to provide ROI in terms of highlighting tools with higher incidents of issue occurring, in turn improving cycle time. Intern was able to provide a real ROI to the business and their own skill set after short amount of training time.

**Saving:** Improved cycle times and ROI in a short space of time from new employees.
# Sample Use Cases

<table>
<thead>
<tr>
<th>Department</th>
<th>Use Case</th>
<th>Project Theme</th>
<th>Benefit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quality</td>
<td>ARIMA Modelling on time series predicted data trends to highlight anomalous point where product quality does not meet criteria</td>
<td>Data Modelling</td>
<td>Product Quality</td>
</tr>
<tr>
<td>Process Engineering</td>
<td>Automated identification and notification of products which surpassed defined transit times and criteria</td>
<td>Data Engineering</td>
<td>Improved cycle time</td>
</tr>
<tr>
<td>Process Engineering</td>
<td>Automation of complicated manual data wrangling tasks to analyse and visualize data over 2 year time span from a Hadoop database</td>
<td>Data Engineering, Visualisation and Automation</td>
<td>Product Quality</td>
</tr>
<tr>
<td>Process Engineering</td>
<td>Automation of Process Engineering WIP on hold process document for using Google Sheets and Oracle</td>
<td>Data Engineering, Automation</td>
<td>Increased engineering time for other tasks / Improved cycle time</td>
</tr>
<tr>
<td>Process Engineering</td>
<td>Prediction of processing metrology times for Process Eng Toolsets to reduce the amount of Scrap</td>
<td>Data Modelling</td>
<td>Reduction in Scrap Material</td>
</tr>
<tr>
<td>Industrial Engineering</td>
<td>Analysis and visualization of Industrial Engineering data based around products tracked and then untracked into tools in order to improve cycle times</td>
<td>Data Engineering, Automation</td>
<td>Improved cycle time</td>
</tr>
<tr>
<td>Industrial Engineering</td>
<td>Automation and modelling of manufacturing transition data to identify when wafers are exceeding time limits for intransit between processing stages. Integration with JMP for visualization</td>
<td>Data Engineering, Visualisation and Automation</td>
<td>Improved cycle time</td>
</tr>
<tr>
<td>R &amp; D</td>
<td>Automation of R &amp; D cycltime scorecard. Use of Google Sheets and Oracle databases for Cycle Time Reporting. Wide range of reports to analyse and improve cycle time for R &amp; D wafers.</td>
<td>Automation</td>
<td>Improved cycle time / Product Quality</td>
</tr>
<tr>
<td>R &amp; D</td>
<td>Automation of the shift handover document creation using data from Oracle databases and Google Sheets</td>
<td>Automation</td>
<td>Improved cycle time</td>
</tr>
<tr>
<td>Yield</td>
<td>Visualisation of Yield Data using Heat Maps and Integration with Tableau</td>
<td>Data Engineering, Visualisation and Automation</td>
<td>Product Quality</td>
</tr>
</tbody>
</table>
Summary

• Rapid adoption by non-native programmers
  Can solve complex business problems much more easily

• Allows users to quickly provide a return on investment from training

• Cross pollination and collaboration from users helped generate new projects