

# Datameer Overview & Demonstration

John Mulhall  
HUG Community Manager

# Introduction

## About John Mulhall:

**Community Manager** – Hadoop User Group (HUG) Ireland

**Marketing Technologist** – Sonra Intelligence Ltd  
Software & Finance trained, 15 years commercial back office experience in 5 industry sectors



• **Twitter:** @sonra\_io | @johnmlhll

• **LinkedIn:** [ie.linkedin.com/in/johnmlhll](https://ie.linkedin.com/in/johnmlhll)

• **Email:** [john.mulhall@sonra.io](mailto:john.mulhall@sonra.io)



# Introduction



## What is Datameer?

Datameer is a self service tool for big data analytics that is designed to work on big data sets

Datameer is compatible with big data frameworks like MapR

It supports:

Structured data

Semi Structured (JSON/XML)

Unstructured Data

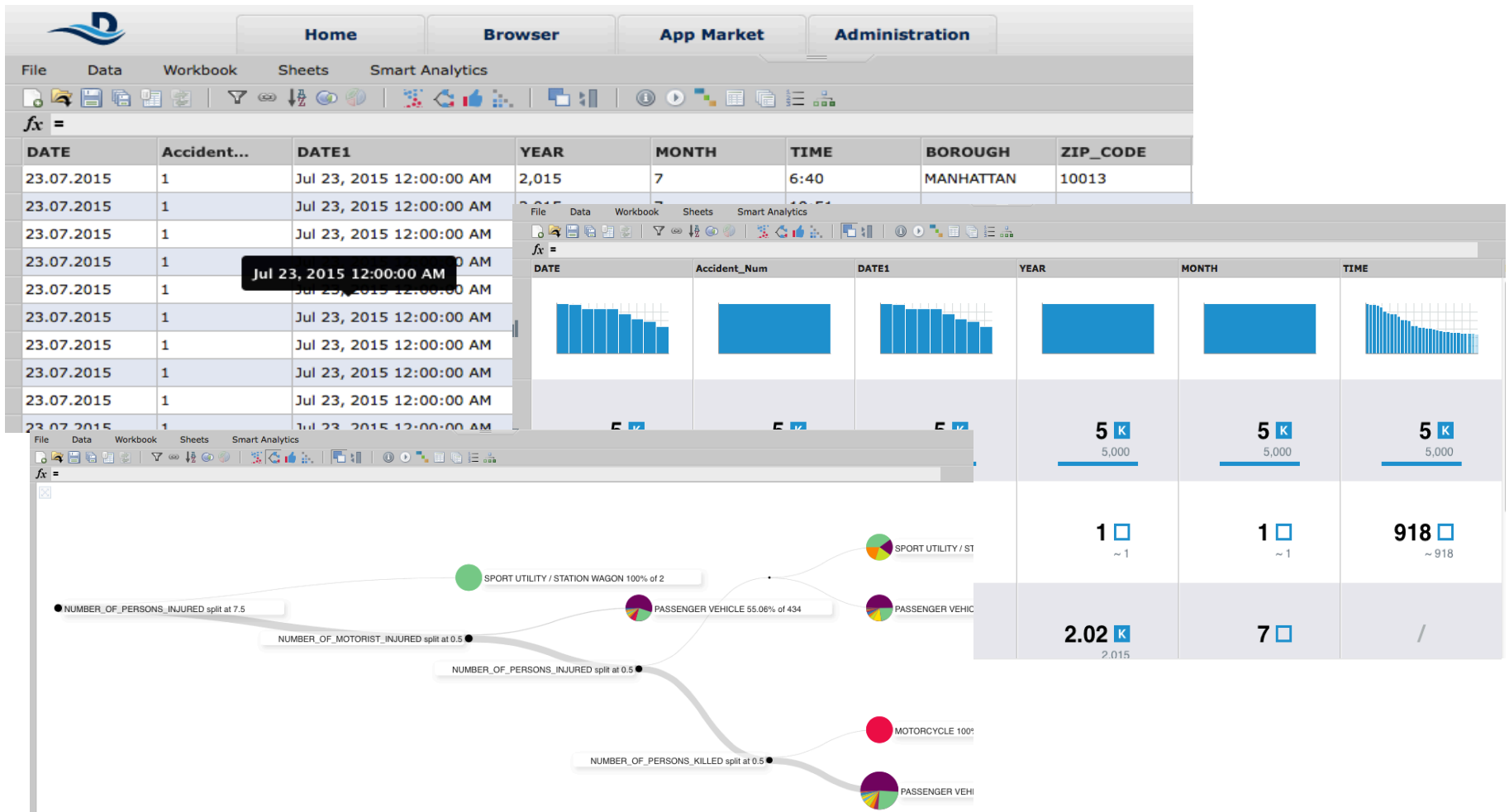
# Basic Features - Connectivity

Name	Type	Status	Last Processed	Records	Size	License
Vorkbo...	.wbk	✓	6 days, 20 hrs ...	-----	271.1 KB	---
ted	.pdf		none		---	---
MyFirst	.ifg		none	---	---	---
	.wbk	✓	41 days, 20 hrs...	-----	394 Bytes	---
Crash...	.upl	✓	16 days, 3 hrs ...	-----	29.5 MB	144.9 ME
Crash...	.upl	✓	9 days, 2 hrs ago	-----	29.3 MB	29.3 MB
Crash...	.wbk	✓	7 days, 21 hrs ...	-----	38.0 MB	---

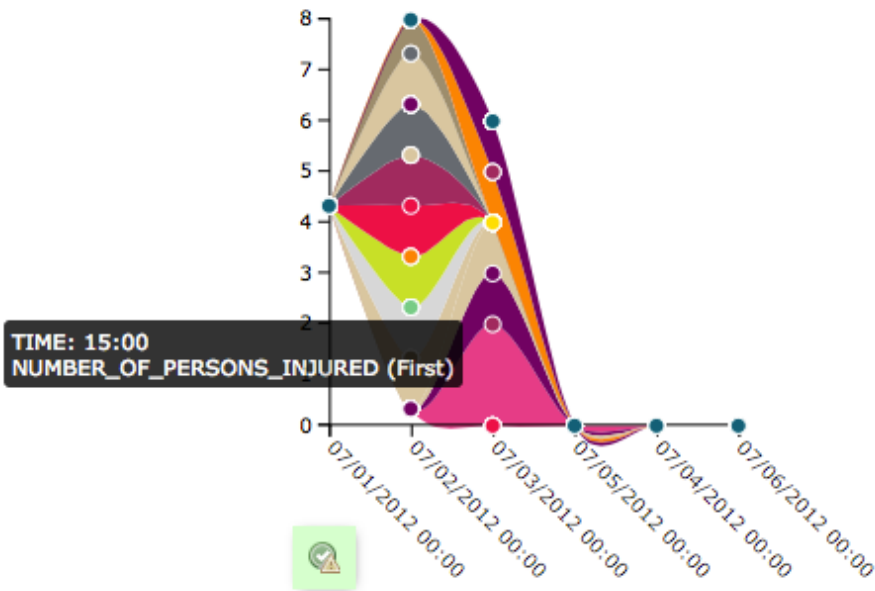
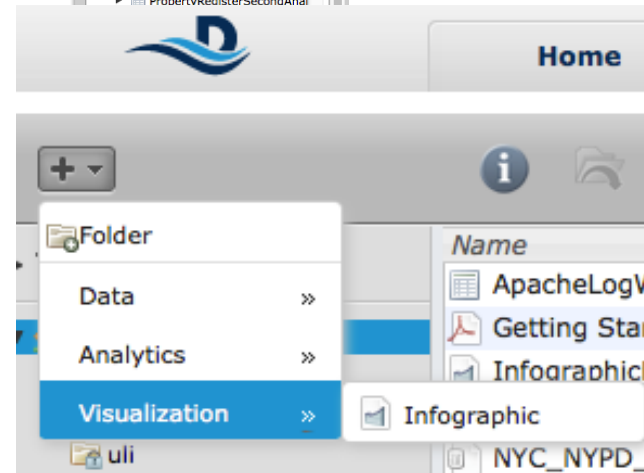
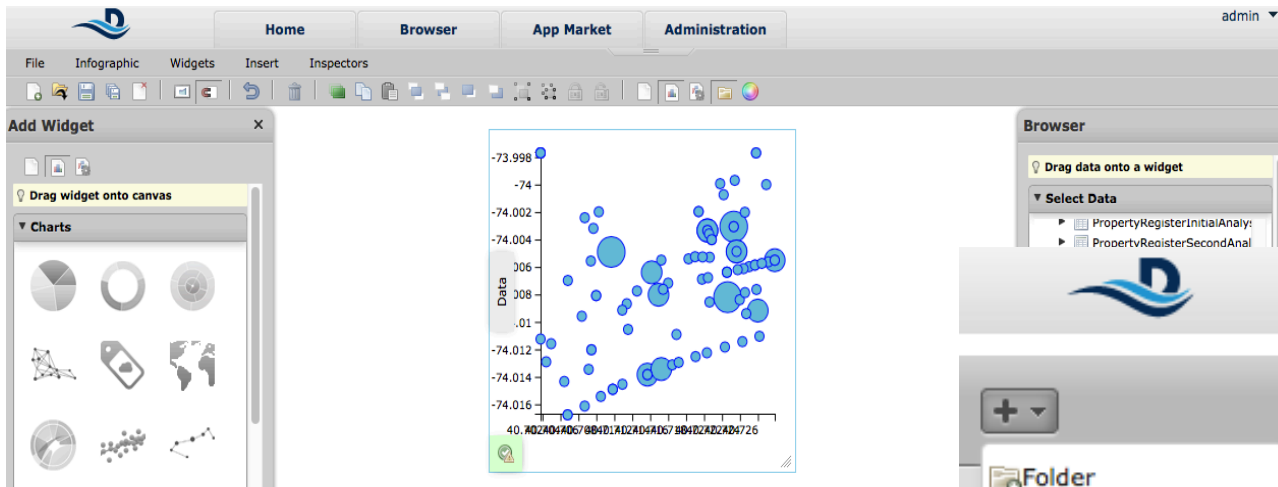
Feature	Count	Examples
Web APIs	18	Google Adsense, Facebook, Twitter, Github, Crunchbase
File Types	19	Apache Log, Avro, Cobol Copybook, CSV/TSV, Excel, IIS, JSON, K.V., Neo4j, Unstructured Files
Databases/Other	9 DB, 2 OT	HSQL_file/HSQL_http (ready to go), Redshift, Oracle, MySQL, PostgreSQL and DB2. Other: Tableau & Email

# Basic Features - Workbook

## “Excel” like Workbook - Data Profiling Visual



# Basic Features - Infographic



Visualisations  
23 Chart Types  
Exportable Workbooks

# Machine Learning

The screenshot displays a software interface with a top navigation bar containing 'Home', 'Browser', 'App Market', and 'Administration'. Below this is a menu bar with 'File', 'Data', 'Workbook', 'Sheets', and 'Smart Analytics'. A toolbar with various icons is visible above a spreadsheet. The spreadsheet has columns for 'DATE', 'Accident...', 'DATE1', 'TIME', 'BOROUGH', and 'ZIP\_CODE'. Rows 1-4 contain data for accidents on 23.07.2015 in Manhattan with ZIP code 10013. Three 'Formula Builder' dialog boxes are open, showing a list of functions including 'ANALYZE\_POLARITY'. The dialog boxes are titled 'Apply Formulas in Your Sheets' and provide instructions on how to use the selected function.

	DATE	Accident...	DATE1				TIME	BOROUGH	ZIP_CODE
1	23.07.2015	1	Jul 23, 2015				6:40	MANHATTAN	10013
2	23.07.2015	1	Jul 23, 2015				10:51		
3	23.07.2015	1	Jul 23, 2015	12:00:00 AM	2,015	7	9:06	MANHATTAN	10013
4	23.07.2015	1	Jul 23, 2015	12:00:00 AM	2,015	7	17:30	MANHATTAN	10013

Smart Analytics (Separate Licence)

3<sup>rd</sup> Party Plugins - Text Analytics

# Advanced Features – SDK

The screenshot displays the Eclipse IDE with a Java project named 'plugin-hello'. The main editor shows the source code for 'HelloFunction.java', which extends 'BaseSimpleFunctionType' and implements the 'compute' method. The code includes imports for 'BaseSimpleFunctionType', 'FieldType', and 'TextArgumentType'. The terminal window shows the build process, including compilation and packaging of the plugin.

```
1 package my.organization.datameer.hello;
2
3 import datameer.dep.sdk.function.BaseSimpleFunctionType;
4 import datameer.dep.sdk.function.FieldType;
5 import datameer.dep.sdk.function.argument.TextArgumentType;
6
7 @SuppressWarnings({ "deprecation", "serial" })
8 public class HelloFunction extends BaseSimpleFunctionType {
9
10 public HelloFunction() {
11     super("My Functions", "HELLO", "Greet a person.", new TextArgumentType());
12 }
13
14 @Override
15 protected Object compute(Object... input) {
16     String name = (String) input[0];
17     return null;
18 }
```

```
frankrosner-lan@plugin-hello frankrosner$ ant zip
Buildfile: /Users/frankrosner/Documents/workspace/Datameer-4.3.8-apache-1.8.3/plugin-sdk-4.3.8/plugin-hello/build.xml
setupArtifactClasspath:
compile:
[javac] /Users/frankrosner/Documents/workspace/Datameer-4.3.8-apache-1.8.3/plugin-sdk-4.3.8/plugin-hello/build-plugin-hello: warning: 'includeruntime' was not set, defaulting to build.sysclasspath:ignore
[javac] Compiling 1 source file to /Users/frankrosner/Documents/workspace/Datameer-4.3.8-apache-1.8.3/plugin-sdk-4.3.8/plugin-hello/build/classes-main
API:
[javac] Note: Recompile with -Xlint:deprecation for details.
jar:
[echo] Building jar build/artifacts/plugin-hello-1.8.jar
[jar] Building jar: /Users/frankrosner/Documents/workspace/Datameer-4.3.8-apache-1.8.3/plugin-sdk-4.3.8/plugin-hello/build/artifacts/plugin-hello-1.8.jar
zip:
[zip] Building zip: /Users/frankrosner/Documents/workspace/Datameer-4.3.8-apache-1.8.3/plugin-sdk-4.3.8/plugin-hello/build/plugin-hello-1.8.zip
BUILD SUCCESSFUL
Total time: 7 seconds
frankrosner-lan@plugin-hello frankrosner$ ./../bin/conductor.sh start
```

User Defined (Aggregate) Functions  
Custom Input Adaptor for File Types

Custom Authenticator  
Import/Export Adaptor for Custom connections/API



# Administration Governance

The screenshot displays the Administration Governance interface. At the top, there is a navigation bar with tabs for Home, Browser, App Market, and Administration. The left sidebar contains a menu with items: System Dashboard, Job History, Authentication, Groups, Roles, Users, Mail Server, Hadoop Cluster, Cluster Health, Database Drivers, Plug-ins, and License. The main content area is titled 'System Dashboard' and features a 'Job Scheduler' section. Under 'Job Scheduler', there is a 'User Management' section with a 'New user' button. Below this is an 'Authentication' section with a 'Use External:' label and two radio button options: 'Internal Datameer User Management' (selected) and 'Remote Authentication System'. A '\* required' note is present below the options. At the bottom right of the configuration area, there are 'Cancel' and 'Next' buttons.

Users Individual/  
Group  
Authentication  
internal/remote

Groups/Roles Setup/  
management  
Mail Server

# Now to the Demo... Over to Uli...



---

p: +353 1 254 2897  
t: @sonra\_io

e: hello@sonra.io  
w: www.sonra.io

**sonra**  
Unleash the Value of your Data

# Datameer – LicenseTable



Home

Browser

App Market

Administration

License Name	How many Nodes	User Limit	Apache Hadoop	Storage Limit
Professional	Cloud Based	25	HaaS	5TB
Enterprise	No Limit	No Limit	Any Distribution	No Limit
Workgroup	Single Server	50	Embedded	500GB
Personal	Single Computer	1	Embedded	5GB



**Load Data**  
Watch Tutorial ▶



**Analyze Data**  
Watch Tutorial ▶



**Visualize Data**  
Watch Tutorial ▶

p: +353 1 254 2897  
t: @sonra\_io

e: hello@sonra.io  
w: www.sonra.io

**sonra**  
Unleash the Value of your Data

# Next Steps

- **Sonra can:**

Consult on your Big Data Projects

Consult on Datameer implementations

Provide Training on Datameer

Help you unleash the value of your data



# Summary

- Datameer is:
  - Function** driven
  - Reasonably** simple to use
  - Processes** smaller big data sets faster than most tools
  - Easy** to integrate
  - Is highly customisable
  - Can** be used by an analyst with no coding experience



Thank you for your time...

Twitter @johnmlhll, @ubethke,  
@sonra\_io

#HUGIreland

**sonra**  
Unleash the Value of your Data