Hands on Introduction to IBM's
Data Science Experience

Power of data. Simplicity of design. Speed of innovation.

Joel Patterson

Agenda

Time	Description
7:30 AM - 8:00 AM	**Registration and Coffee**
8:00 AM - 8:30 AM	**Overview of the Watson Data Platform and IBM Data Science Experience (DSX)**
8:30 AM - 10:00 AM	**Lab 1 - Surviving the Titanic**
10:00 AM - 11:00 AM	**Lab 2 - Machine Learning with Spark ML**
11:00 AM – 11:15 AM	**Break**
11:15 AM – 11:45 AM	**Lab 3 - R, Shiny, and GUI Interfaces**
Extra time	**Lab 4 - Choose From Several Options**
11:45 AM – 12:00 PM	**Questions and Wrap Up**

Participant Background

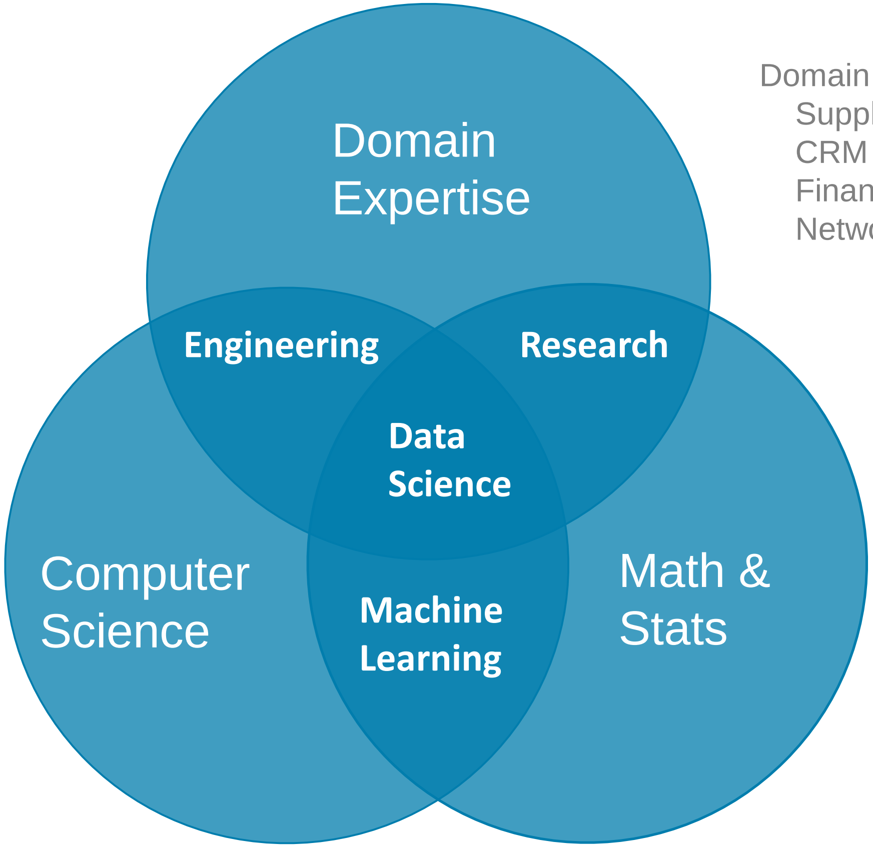
Open Source

- R/Python/Scala
- Jupyter Notebook
- Spark
- Hadoop

IBM

- Bluemix

What is a Data Scientist?



Domain Knowledge
Supply Chain
CRM
Financials
Networking

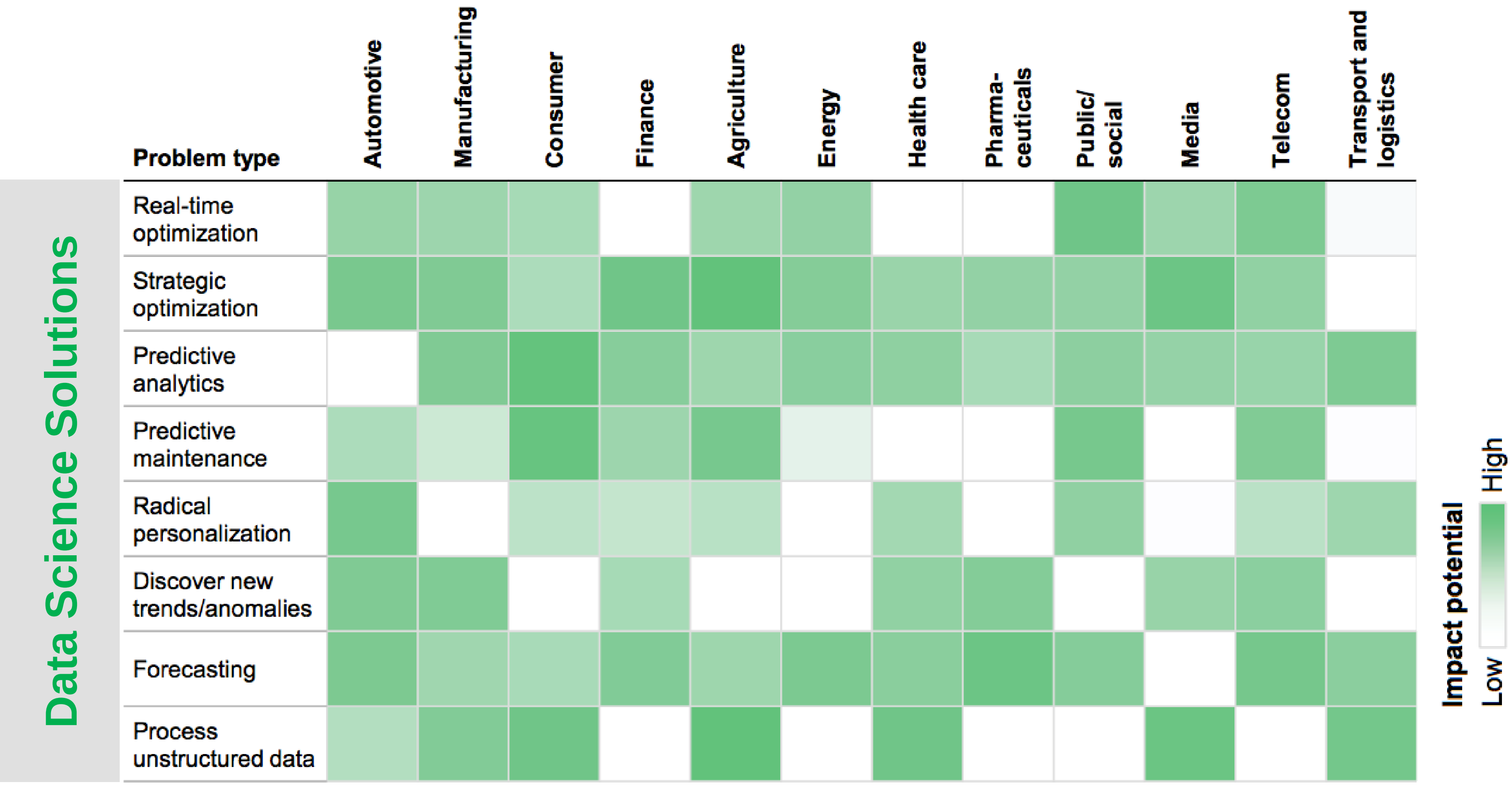
Scripting, SQL
Python, R Scala
Data Pipelines
Big Data/ Apache Spark

Mathematics
Computational

Data Science Projects Require Multiple Skills

Data Science Impact Across Industries and Use Cases

$10s of Billions in each industry and use case



SOURCE: McKinsey Global Institute analysis

Challenges in delivering value with Data Science

Data

- Data resides in silos and difficult to access
- Detailed data was never stored
- Unstructured and external data wasn't considered

Governance

- If the data isn't secure, self-service isn't a reality
- Understanding lineage and getting to a system of truth

Skills

- Data Science skills are in low supply and high demand
- Nurturing new data professionals is challenging

Infrastructure

- Need an environment that enables collaboration and deployment to production
- Discrete tools present barriers to progress

Watson Data Platform

IBM Watson Data Platform

Mission: Make Data Simple and Accessible to All

Platform.

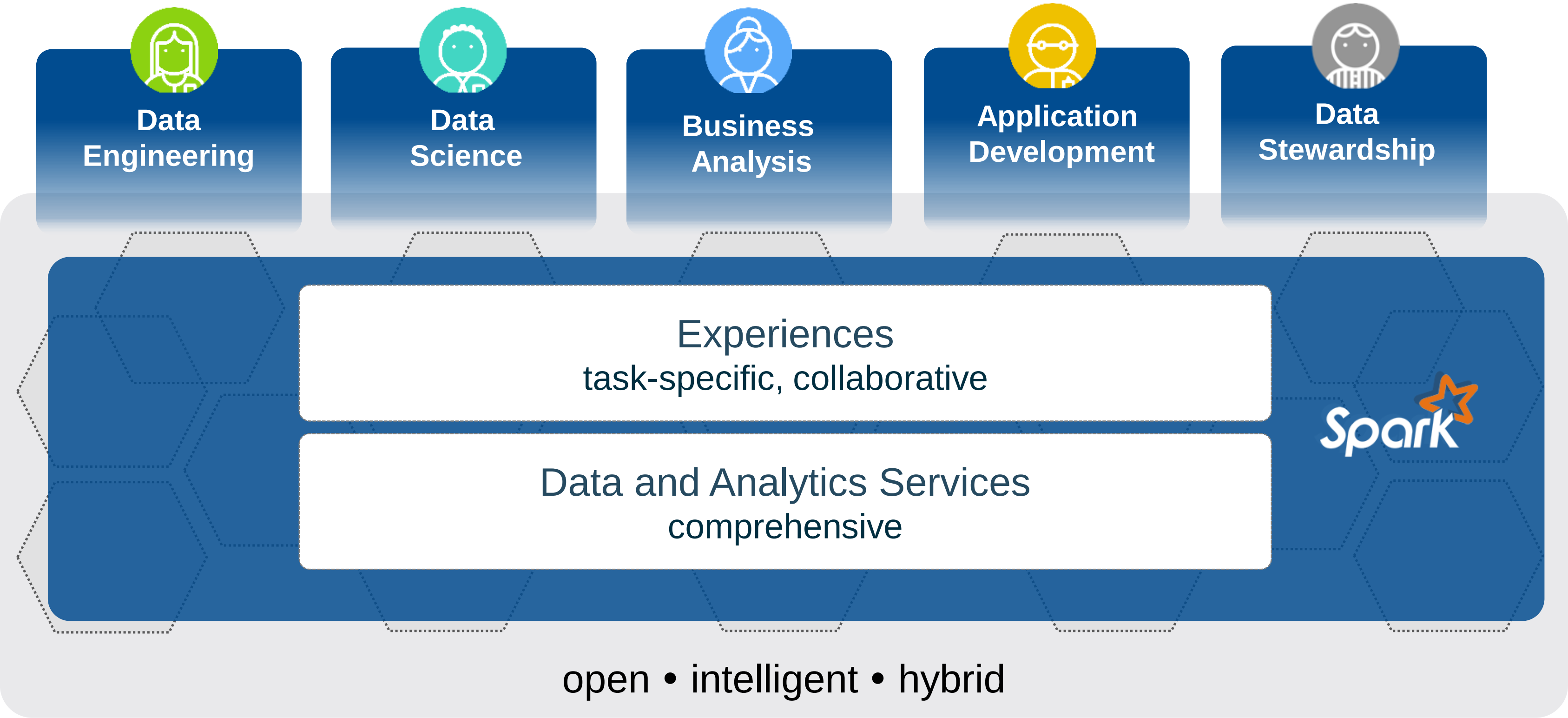
Method.

Ecosystem.

http://ibm.co/makedatasimple

IBM Watson Data Platform

Experience New Ways To Put Data To Work



Data Engineering

Data Science

Business Analysis

Application Development

Data Stewardship

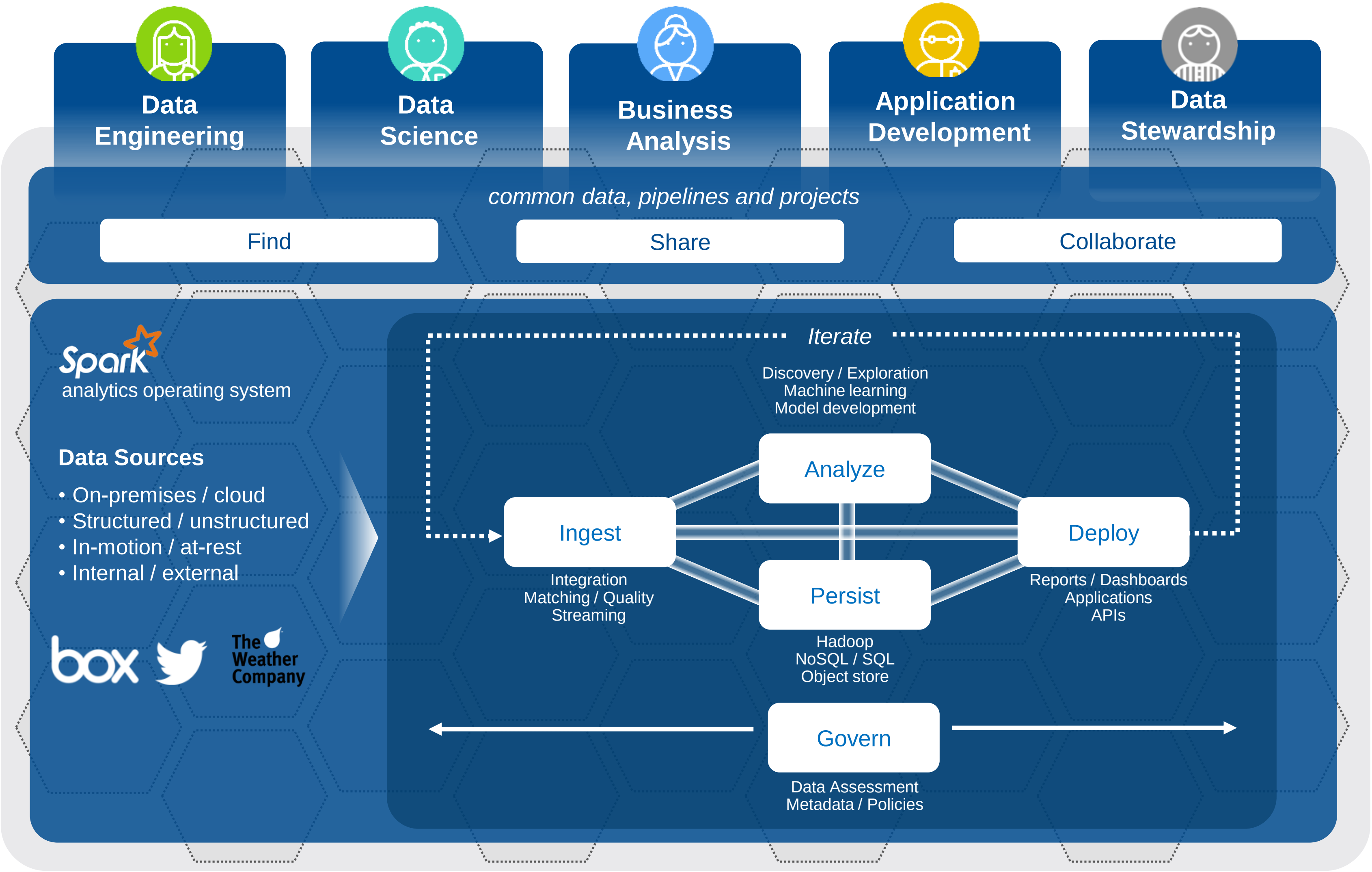
Experiences
task-specific, collaborative

Data and Analytics Services
comprehensive

open • intelligent • hybrid

IBM Watson Data Platform

Connects Users to Data and Analytics



Data Engineering

Data Science

Business Analysis

Application Development

Data Stewardship

common data, pipelines and projects

Find

Share

Collaborate

Spark

analytics operating system

Data Sources

- On-premises / cloud
- Structured / unstructured
- In-motion / at-rest
- Internal / external

box

The Weather Company

Iterate

Discovery / Exploration
Machine learning
Model development

Analyze

Ingest

Integration
Matching / Quality
Streaming

Persist

Hadoop
NoSQL / SQL
Object store

Deploy

Reports / Dashboards
Applications
APIs

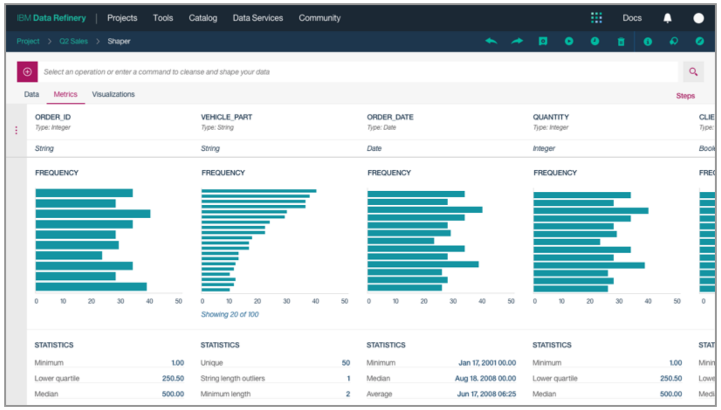
Govern

Data Assessment
Metadata / Policies

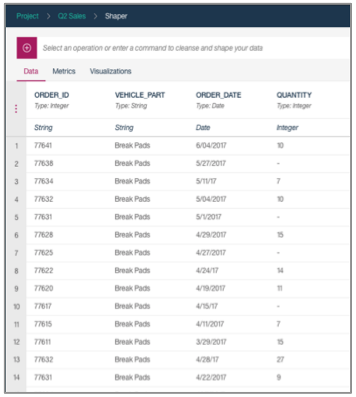
Data Refinery – Open Beta

Data Engineer

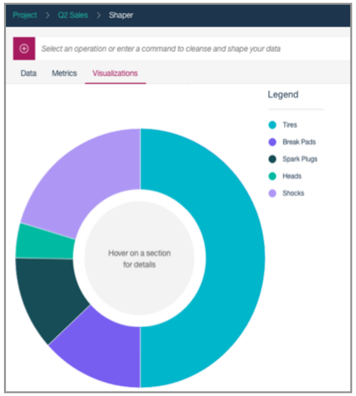
A Breakthrough Approach to Explore and Prepare Data



ORDER_ID Type: Integer		VEHICLE_PART Type: String		ORDER_DATE Type: Date		QUANTITY Type: Integer	
String		String		Date		Integer	
STATISTICS		STATISTICS		STATISTICS		STATISTICS	
Minimum	1.00	Unique	50	Minimum	Jan 17, 2001 00.00	Minimum	1.00
Lower quartile	250.50	String length outliers	1	Median	Aug 18. 2008 00.00	Lower quartile	250.50
Median	500.00	Minimum length	2	Average	Jun 17, 2008 06:25	Median	500.00


Project > Q2 Sales > Shaper
Select an operation or enter a command to cleanse and shape your data
Data
Metrics
Visualizations

	ORDER_ID Type: Integer	VEHICLE_PART Type: String	ORDER_DATE Type: Date	QUANTITY Type: Integer
	String	String	Date	Integer
1	77641	Break Pads	6/04/2017	10
2	77638	Break Pads	5/27/2017	-
3	77634	Break Pads	5/11/17	7
4	77632	Break Pads	5/04/2017	10
5	77631	Break Pads	5/1/2017	-
6	77628	Break Pads	4/29/2017	15
7	77625	Break Pads	4/27/2017	-
8	77622	Break Pads	4/24/17	14
9	77620	Break Pads	4/19/2017	11
10	77617	Break Pads	4/15/17	-
11	77615	Break Pads	4/11/2017	7
12	77611	Break Pads	3/29/2017	15
13	77632	Break Pads	4/28/17	27
14	77631	Break Pads	4/22/2017	9


Project > Q2 Sales > Shaper
Select an operation or enter a command to cleanse and shape your data
Data
Metrics
Visualizations
Legend
Tires
Break Pads
Spark Plugs
Heads
Shocks
Hover on a section for details

Wrangle

Interactively explore, resolve quality issues, enrich, classify, standardize, and summarize data.

Flow

Create data flows visually, schedule for repeatability, monitor and notify

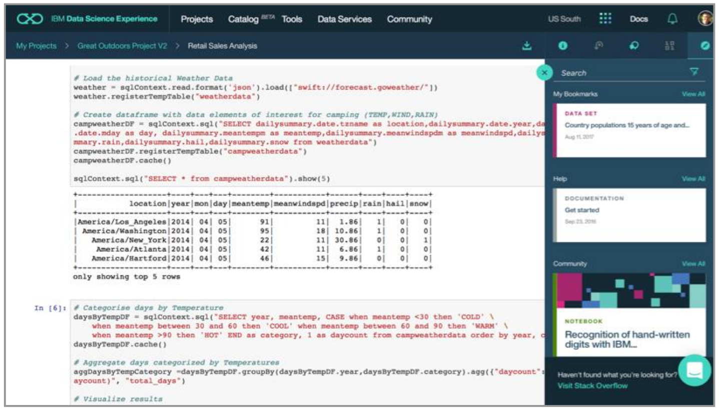
Adapt

Connect to 30+ cloud and on-premises stores and scale on demand with cataloging and governance

Data Science Experience

Data Scientist

Brings together everything a Data Scientist needs to be successful



```
# Load the historical Weather Data
weather = sqlContext.read.format('json').load(["swift://forecast.goweather/"])
weather.registerTempTable("weatherdata")

# Create dataframe with data elements of interest for camping (TEMP,WIND,RAIN)
campweatherDF = sqlContext.sql("SELECT dailysummary.date.tzname as location,dailysummary.date.year,da
.date.mday as day, dailysummary.meantempm as meantemp,dailysummary.meanwindspdm as meanwindspd,dailys
mmary.rain,dailysummary.hail,dailysummary.snow from weatherdata")
campweatherDF.registerTempTable("campweatherdata")
campweatherDF.cache()

sqlContext.sql("SELECT * from campweatherdata").show(5)
```

```
+-------------------+----+---+---+--------+-----------+------+----+----+----+
|           location|year|mon|day|meantemp|meanwindspd|precip|rain|hail|snow|
+-------------------+----+---+---+--------+-----------+------+----+----+----+
|America/Los_Angeles|2014| 04| 05|      91|         11|  1.86|   1|   0|   0|
| America/Washington|2014| 04| 05|      95|         18| 10.86|   1|   0|   0|
|   America/New_York|2014| 04| 05|      22|         11| 30.86|   0|   0|   1|
|    America/Atlanta|2014| 04| 05|      42|         11|  6.86|   1|   0|   0|
|   America/Hartford|2014| 04| 05|      46|         15|  9.86|   0|   0|   0|
+-------------------+----+---+---+--------+-----------+------+----+----+----+
only showing top 5 rows
```

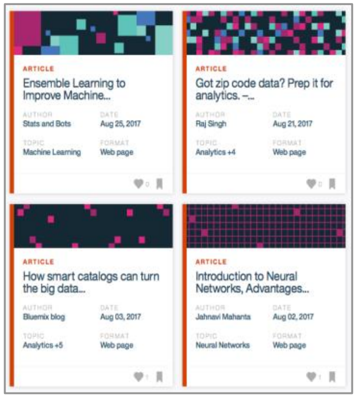
```
In [6]: # Categorise days by Temperature
daysByTempDF = sqlContext.sql("SELECT year, meantemp, CASE when meantemp <30 then 'COLD' \
    when meantemp between 30 and 60 then 'COOL' when meantemp between 60 and 90 then 'WARM' \
    when meantemp >90 then 'HOT' END as category, 1 as daycount from campweatherdata order by year, c
daysByTempDF.cache()

# Aggregate days categorized by Temperatures
aggDaysByTempCategory =daysByTempDF.groupBy(daysByTempDF.year,daysByTempDF.category).agg({"daycount":
aycount)", "total_days")

# Visualize results
```

DATA SET
Country populations 15 years of age and...
Aug 11, 2017

DOCUMENTATION
Get started
Sep 23, 2016

NOTEBOOK
Recognition of hand-written digits with IBM...

Haven't found what you're looking for?
Visit Stack Overflow

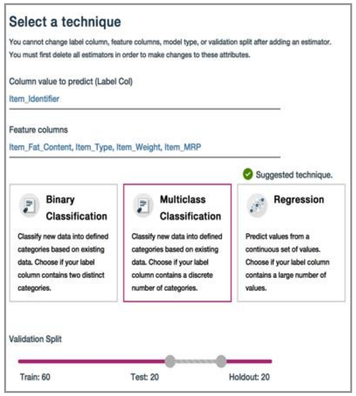


ARTICLE
Ensemble Learning to Improve Machine...
AUTHOR Stats and Bots
DATE Aug 25, 2017
TOPIC Machine Learning
FORMAT Web page
0

ARTICLE
Got zip code data? Prep it for analytics. –...
AUTHOR Raj Singh
DATE Aug 21, 2017
TOPIC Analytics +4
FORMAT Web page
0

ARTICLE
How smart catalogs can turn the big data...
AUTHOR Bluemix blog
DATE Aug 03, 2017
TOPIC Analytics +5
FORMAT Web page
1

ARTICLE
Introduction to Neural Networks, Advantages...
AUTHOR Jahnavi Mahanta
DATE Aug 02, 2017
TOPIC Neural Networks
FORMAT Web page
1



Select a technique

You cannot change label column, feature columns, model type, or validation split after adding an estimator. You must first delete all estimators in order to make changes to these attributes.

Column value to predict (Label Col)
Item_Identifier

Feature columns
Item_Fat_Content, Item_Type, Item_Weight, Item_MRP

Suggested technique.

Binary Classification
Classify new data into defined categories based on existing data. Choose if your label column contains two distinct categories.

Multiclass Classification
Classify new data into defined categories based on existing data. Choose if your label column contains a discrete number of categories.

Regression
Predict values from a continuous set of values. Choose if your label column contains a large number of values.

Validation Split
Train: 60
Test: 20
Holdout: 20

Learn

Built-in learning to get started or go the distance with advanced tutorials

Create

The best of open source and IBM value-add to create state-of-the-art data products

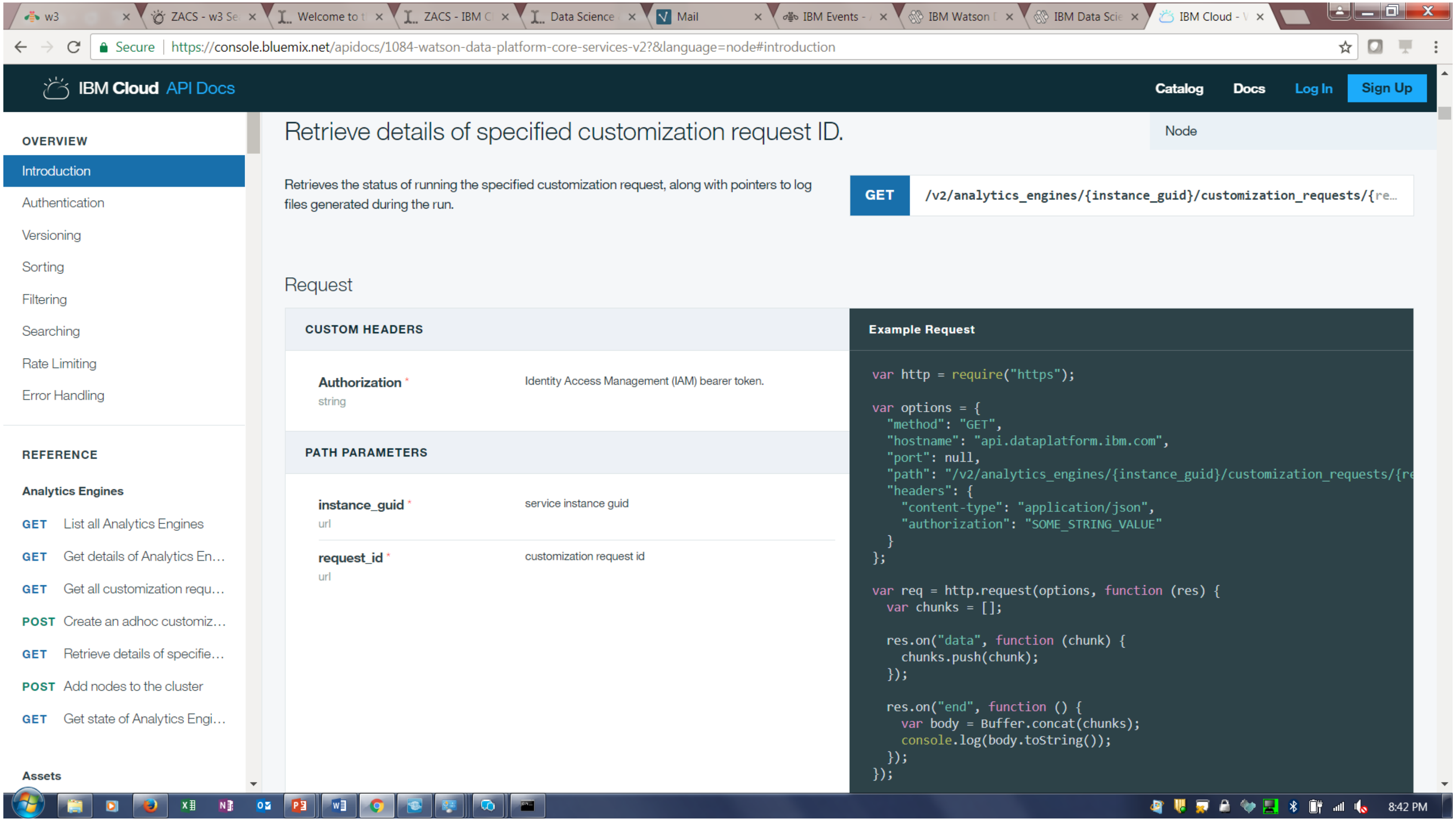
Collaborate

Community and social features that provide meaningful collaboration

IBM Cloud PaaS

Application Developer

Rich Platform and Service APIs for your developers



OVERVIEW

- Introduction
- Authentication
- Versioning
- Sorting
- Filtering
- Searching
- Rate Limiting
- Error Handling

REFERENCE

Analytics Engines

- GET List all Analytics Engines
- GET Get details of Analytics En…
- GET Get all customization requ…
- POST Create an adhoc customiz…
- GET Retrieve details of specifie…
- POST Add nodes to the cluster
- GET Get state of Analytics Engi…

Assets

Retrieve details of specified customization request ID.

Retrieves the status of running the specified customization request, along with pointers to log files generated during the run.

GET /v2/analytics_engines/{instance_guid}/customization_requests/{re…

Request

CUSTOM HEADERS

Authorization * string	Identity Access Management (IAM) bearer token.

PATH PARAMETERS

instance_guid * url	service instance guid
request_id * url	customization request id

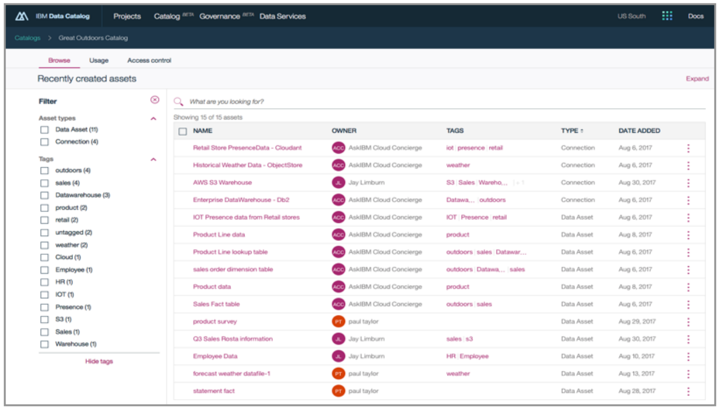
Example Request

```
var http = require("https");

var options = {
  "method": "GET",
  "hostname": "api.dataplatform.ibm.com",
  "port": null,
  "path": "/v2/analytics_engines/{instance_guid}/customization_requests/{re
  "headers": {
    "content-type": "application/json",
    "authorization": "SOME_STRING_VALUE"
  }
};

var req = http.request(options, function (res) {
  var chunks = [];

  res.on("data", function (chunk) {
    chunks.push(chunk);
  });

  res.on("end", function () {
    var body = Buffer.concat(chunks);
    console.log(body.toString());
  });
});
```

Data Catalog – Open Beta

Data Steward

Unlock tribal knowledge to unleash your data professionals



Recently created assets

Filter

Asset types

- □ Data Asset (11)
- □ Connection (4)

Tags

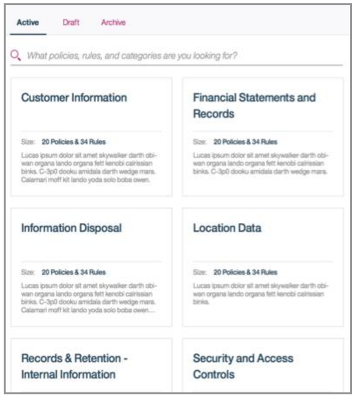
- □ outdoors (4)
- □ sales (4)
- □ Datawarehouse (3)
- □ product (2)
- □ retail (2)
- □ untagged (2)
- □ weather (2)
- □ Cloud (1)
- □ Employee (1)
- □ HR (1)
- □ IOT (1)
- □ Presence (1)
- □ S3 (1)
- □ Sales (1)
- □ Warehouse (1)

Hide tags

What are you looking for?

Showing 15 of 15 assets

□ NAME	OWNER	TAGS	TYPE	DATE ADDED
Retail Store PresenceData - Cloudant	ACC AskIBM Cloud Concierge	iot presence retail	Connection	Aug 6, 2017
Historical Weather Data - ObjectStore	ACC AskIBM Cloud Concierge	weather	Connection	Aug 6, 2017
AWS S3 Warehouse	JL Jay Limburn	S3 Sales Wareho... +1	Connection	Aug 30, 2017
Enterprise DataWarehouse - Db2	ACC AskIBM Cloud Concierge	Datawa... outdoors	Connection	Aug 6, 2017
IOT Presence data from Retail stores	ACC AskIBM Cloud Concierge	IOT Presence retail	Data Asset	Aug 6, 2017
Product Line data	ACC AskIBM Cloud Concierge	product	Data Asset	Aug 8, 2017
Product Line lookup table	ACC AskIBM Cloud Concierge	outdoors sales Datawar...	Data Asset	Aug 6, 2017
sales order dimension table	ACC AskIBM Cloud Concierge	outdoors Datawa... sales	Data Asset	Aug 6, 2017
Product data	ACC AskIBM Cloud Concierge	product	Data Asset	Aug 8, 2017
Sales Fact table	ACC AskIBM Cloud Concierge	outdoors sales	Data Asset	Aug 6, 2017
product survey	PT paul taylor		Data Asset	Aug 29, 2017
Q3 Sales Rosta information	JL Jay Limburn	sales s3	Data Asset	Aug 30, 2017
Employee Data	JL Jay Limburn	HR Employee	Data Asset	Aug 10, 2017
forecast weather datafile-1	PT paul taylor	weather	Data Asset	Aug 13, 2017
statement fact	PT paul taylor		Data Asset	Aug 28, 2017



What policies, rules, and categories are you looking for?

Customer Information

Size: 20 Policies & 34 Rules

Lucas ipsum dolor sit amet skywalker darth obi-wan organa lando organa fett kenobi calrissian binks. C-3p0 dooku amidala darth wedge mara. Calamari moff kit lando yoda solo boba owen.

Financial Statements and Records

Size: 20 Policies & 34 Rules

Lucas ipsum dolor sit amet skywalker darth obi-wan organa lando organa fett kenobi calrissian binks. C-3p0 dooku amidala darth wedge mara.

Information Disposal

Size: 20 Policies & 34 Rules

Lucas ipsum dolor sit amet skywalker darth obi-wan organa lando organa fett kenobi calrissian binks. C-3p0 dooku amidala darth wedge mara. Calamari moff kit lando yoda solo boba owen....

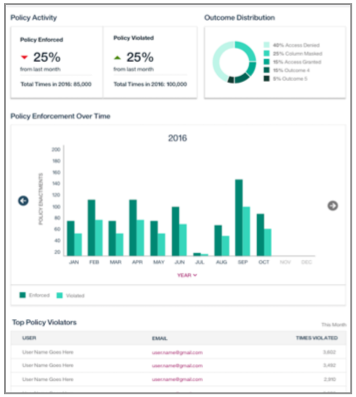
Location Data

Size: 20 Policies & 34 Rules

Lucas ipsum dolor sit amet skywalker darth obi-wan organa lando organa fett kenobi calrissian binks.

Records & Retention - Internal Information

Security and Access Controls



Policy Activity

Policy Enforced
▼ 25%
from last month
Total Times in 2016: 85,000

Policy Violated
▲ 25%
from last month
Total Times in 2016: 100,000

Outcome Distribution

Policy Enforcement Over Time

Top Policy Violators

This Month

USER	EMAIL	TIMES VIOLATED
User Name Goes Here	user.name@gmail.com	3,602
User Name Goes Here	user.name@gmail.com	3,492
User Name Goes Here	user.name@gmail.com	2,910

Discover

Intelligent discovery of data, advanced classification and profiling to provide context

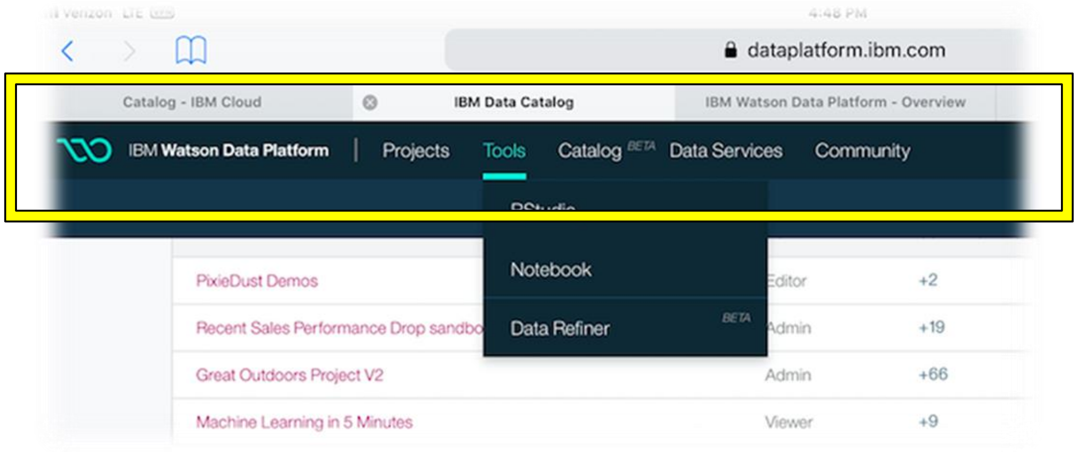
Catalog

A rich metadata index of all data, with social collaboration and enhanced findability

Govern

Powerful governance policy tools to control and protect access to data with visibility to data use

Intelligent data fabric provides consistent platform experience



PixieDust Demos	Editor	+2
Recent Sales Performance Drop sandbo	Admin	+19
Great Outdoors Project V2	Admin	+66
Machine Learning in 5 Minutes	Viewer	+9

This fabric remains consistent throughout the Watson Data Platform experience – regardless if you are ingesting data, shaping data, building algorithms, deploying models and more…

How does WDP help fulfill the promise of your data?

Data

Puts every important data source at the fingertips of the teams that need it wherever resides

Governance

Enforces your policies without getting in the way of delivering insights

Skills

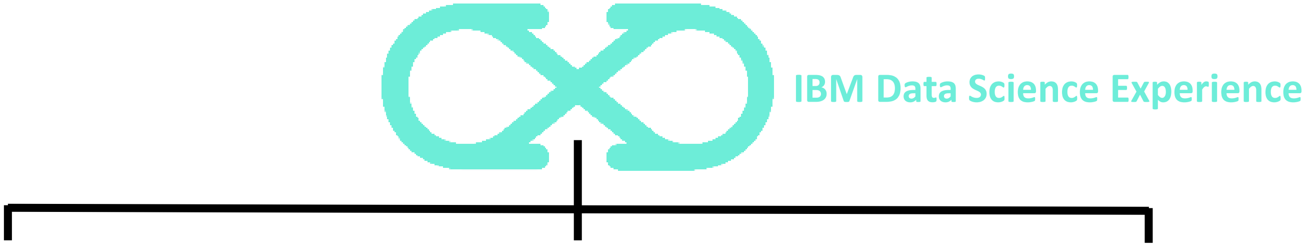
Makes the most of the data professionals you have and helps them grow and learn from each other as a team

Infrastructure

Delivers the foundation for your first data project through to the complete transformation of your business

Data Science Experience

Core Attributes of the Data Science Experience



IBM Data Science Experience

Community

- Find tutorials and datasets
- Read articles and papers
- Connect with Data Scientists
- Share comments
- Copy and share notebooks

Open Source

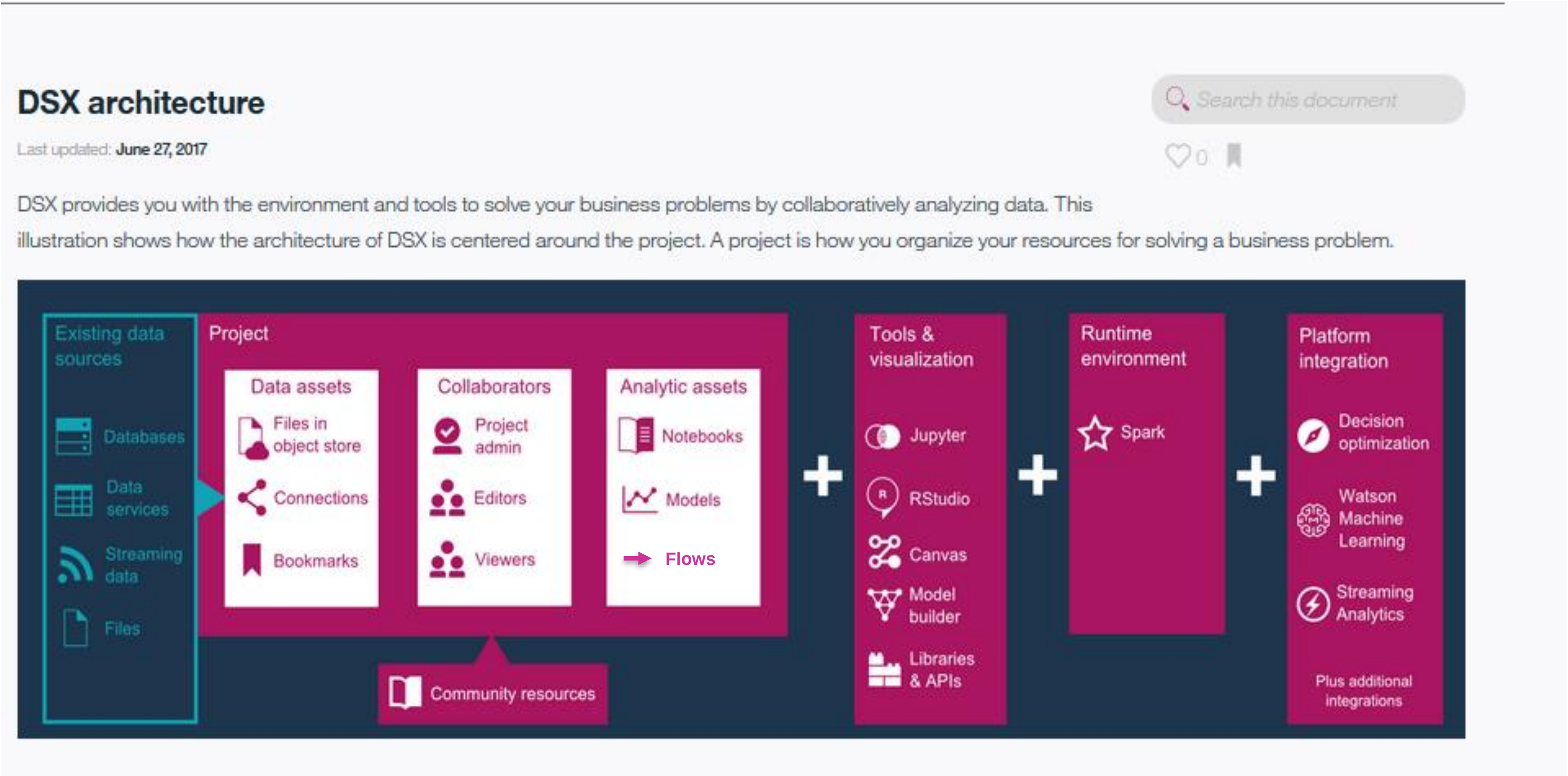
- Code in Scala/Python/R/SQL
- Jupyter Notebooks
- RStudio IDE and Shiny
- Apache Spark
- Your favorite libraries

IBM Added Value

- IBM Machine Learning
- SPSS Modeler Canvas
- Prescriptive Analytics - DOcplexcloud
- Projects and Version Control
- Managed Spark Service

Powered by IBM **Watson Data Platform**

DSX Architecture



DSX architecture

Search this document

Last updated: **June 27, 2017**

0

DSX provides you with the environment and tools to solve your business problems by collaboratively analyzing data. This illustration shows how the architecture of DSX is centered around the project. A project is how you organize your resources for solving a business problem.

Existing data sources

Databases

Data services

Streaming data

Files

Project

Data assets

Files in object store

Connections

Bookmarks

Collaborators

Project admin

Editors

Viewers

Analytic assets

Notebooks

Models

Flows

Community resources

+

Tools & visualization

Jupyter

RStudio

Canvas

Model builder

Libraries & APIs

+

Runtime environment

Spark

+

Platform integration

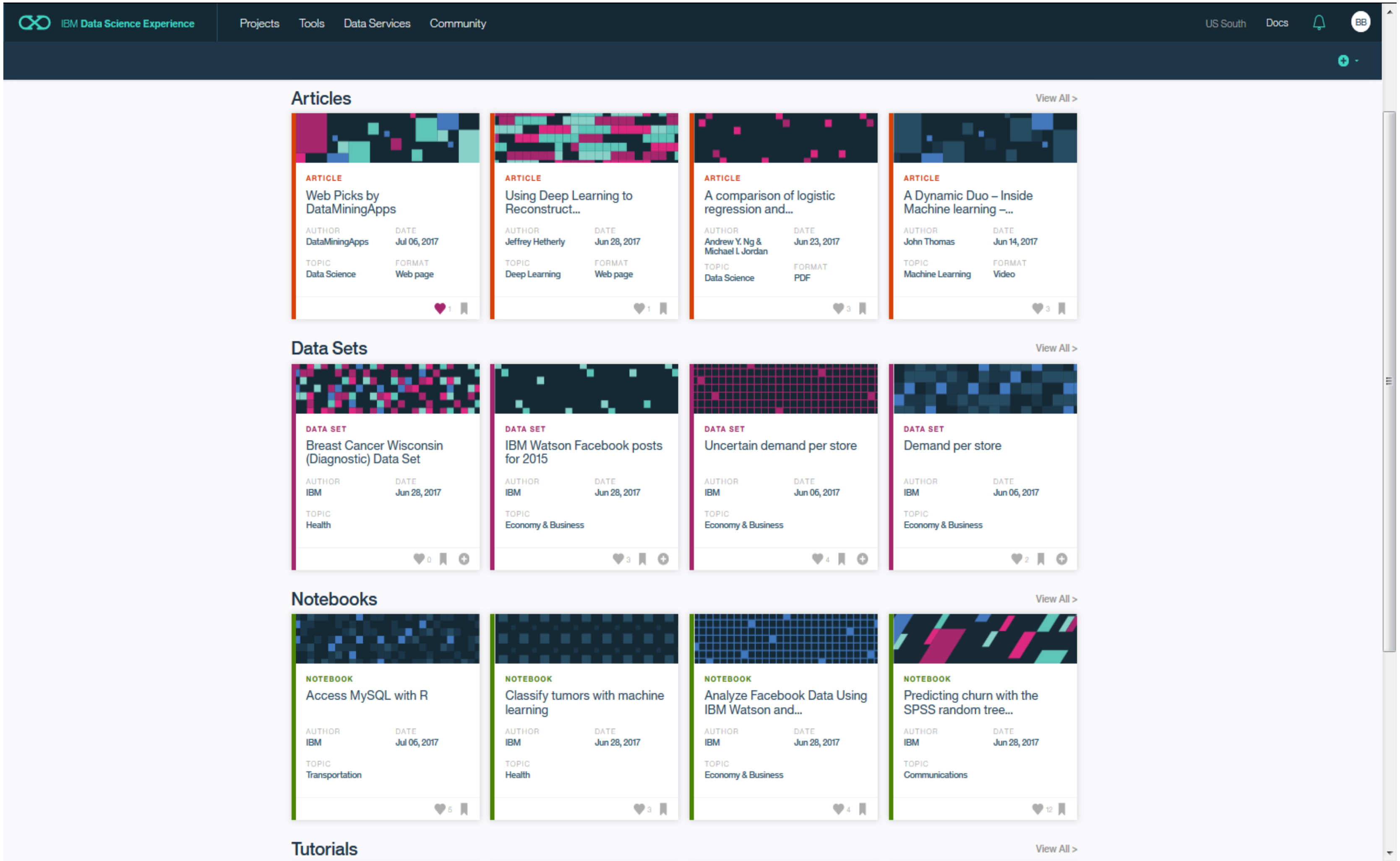
Decision optimization

Watson Machine Learning

Streaming Analytics

Plus additional integrations

Community Cards provide in-context learning



Articles

View All >

ARTICLE
Web Picks by DataMiningApps
AUTHOR DataMiningApps · DATE Jul 06, 2017
TOPIC Data Science · FORMAT Web page

ARTICLE
Using Deep Learning to Reconstruct...
AUTHOR Jeffrey Hetherly · DATE Jun 28, 2017
TOPIC Deep Learning · FORMAT Web page

ARTICLE
A comparison of logistic regression and...
AUTHOR Andrew Y. Ng & Michael I. Jordan · DATE Jun 23, 2017
TOPIC Data Science · FORMAT PDF

ARTICLE
A Dynamic Duo – Inside Machine learning –...
AUTHOR John Thomas · DATE Jun 14, 2017
TOPIC Machine Learning · FORMAT Video

Data Sets

View All >

DATA SET
Breast Cancer Wisconsin (Diagnostic) Data Set
AUTHOR IBM · DATE Jun 28, 2017
TOPIC Health

DATA SET
IBM Watson Facebook posts for 2015
AUTHOR IBM · DATE Jun 28, 2017
TOPIC Economy & Business

DATA SET
Uncertain demand per store
AUTHOR IBM · DATE Jun 06, 2017
TOPIC Economy & Business

DATA SET
Demand per store
AUTHOR IBM · DATE Jun 06, 2017
TOPIC Economy & Business

Notebooks

View All >

NOTEBOOK
Access MySQL with R
AUTHOR IBM · DATE Jul 06, 2017
TOPIC Transportation

NOTEBOOK
Classify tumors with machine learning
AUTHOR IBM · DATE Jun 28, 2017
TOPIC Health

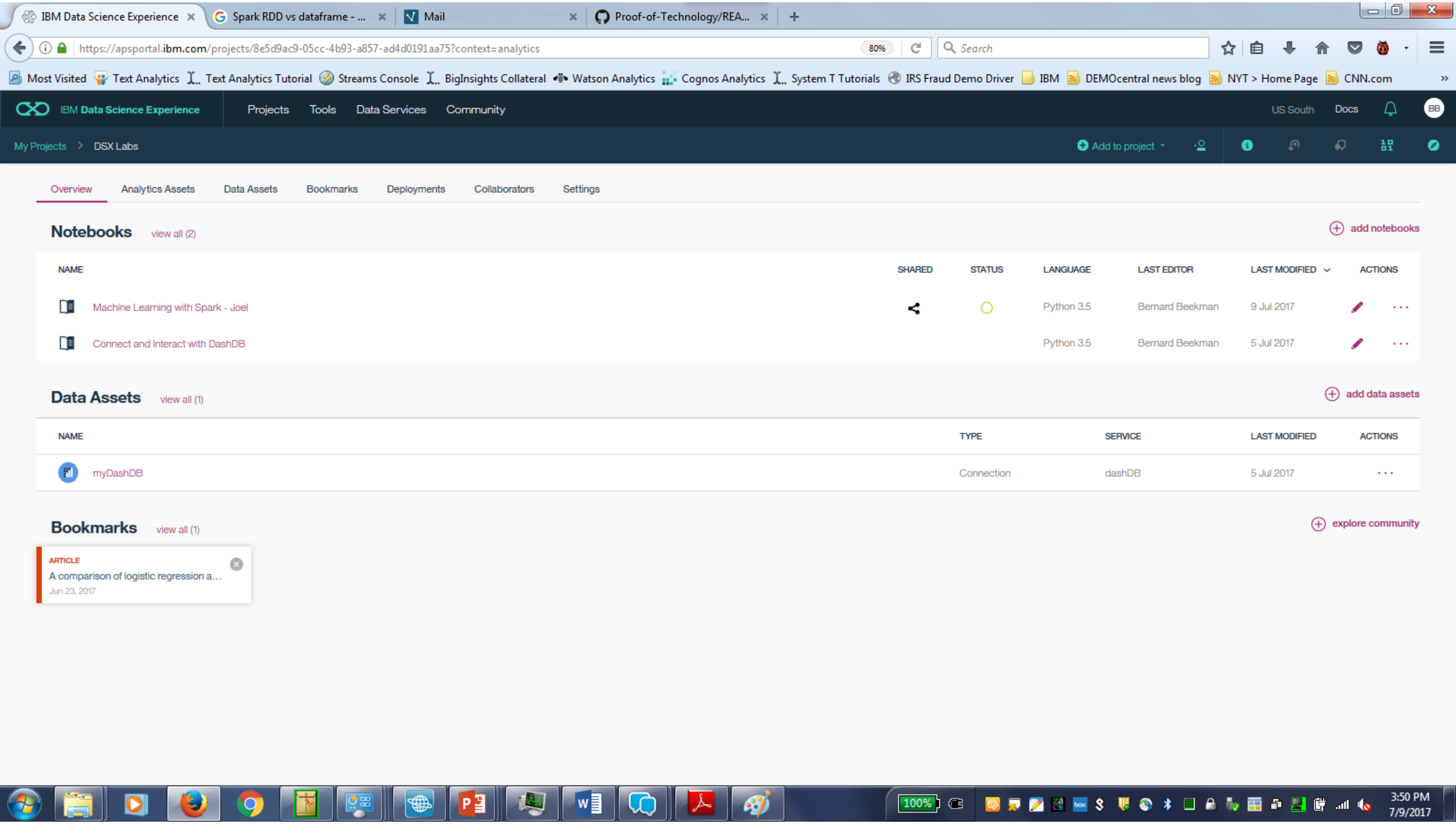
NOTEBOOK
Analyze Facebook Data Using IBM Watson and...
AUTHOR IBM · DATE Jun 28, 2017
TOPIC Economy & Business

NOTEBOOK
Predicting churn with the SPSS random tree...
AUTHOR IBM · DATE Jun 28, 2017
TOPIC Communications

Tutorials

View All >

Collaborate Using Projects



Notebooks

view all (2)

add notebooks

NAME	SHARED	STATUS	LANGUAGE	LAST EDITOR	LAST MODIFIED	ACTIONS
Machine Learning with Spark - Joel	✓	○	Python 3.5	Bernard Beekman	9 Jul 2017	···
Connect and Interact with DashDB			Python 3.5	Bernard Beekman	5 Jul 2017	···

Data Assets

view all (1)

add data assets

NAME	TYPE	SERVICE	LAST MODIFIED	ACTIONS
myDashDB	Connection	dashDB	5 Jul 2017	···

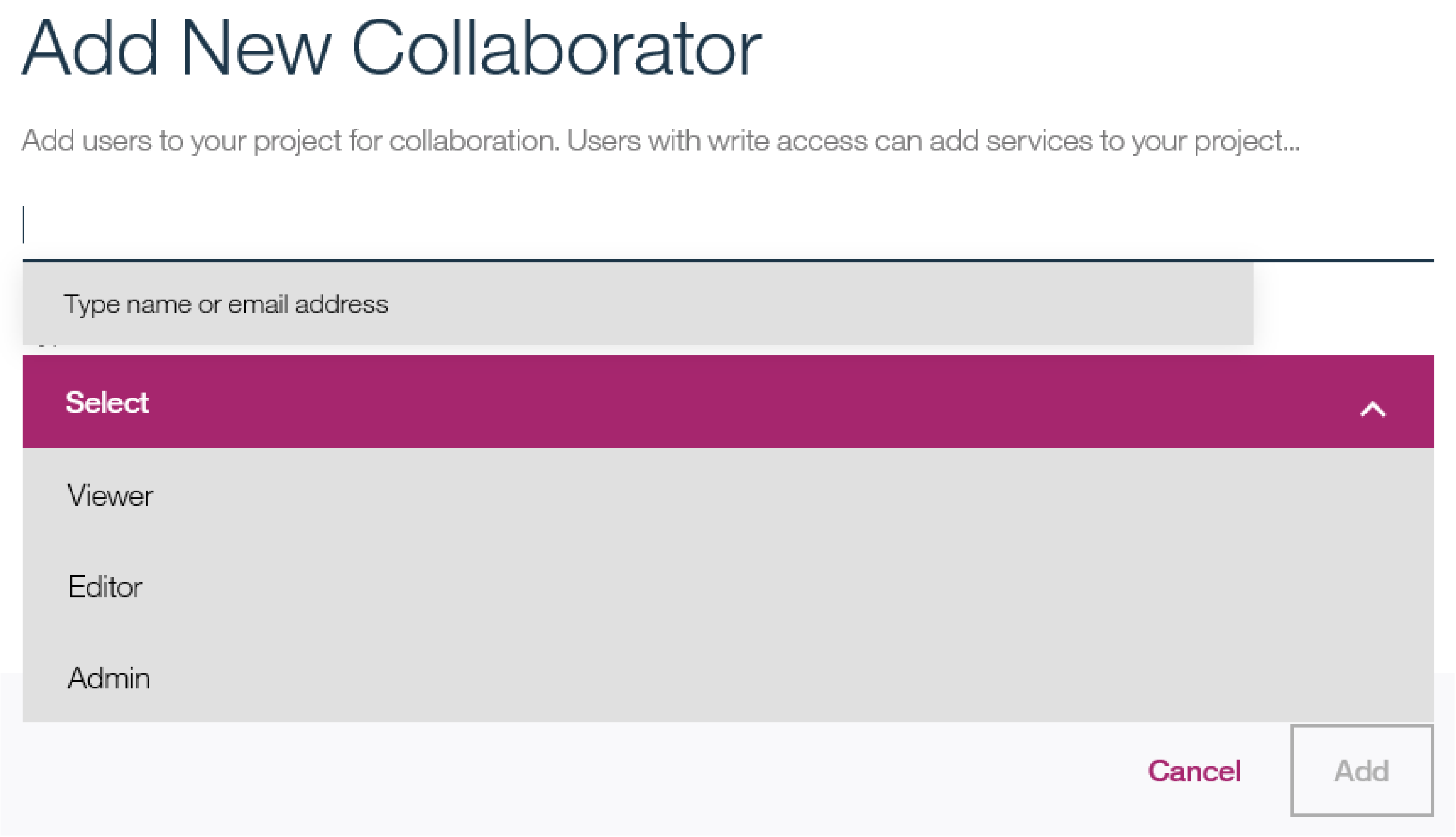
Bookmarks

view all (1)

explore community

ARTICLE
A comparison of logistic regression a...
Jun 23, 2017

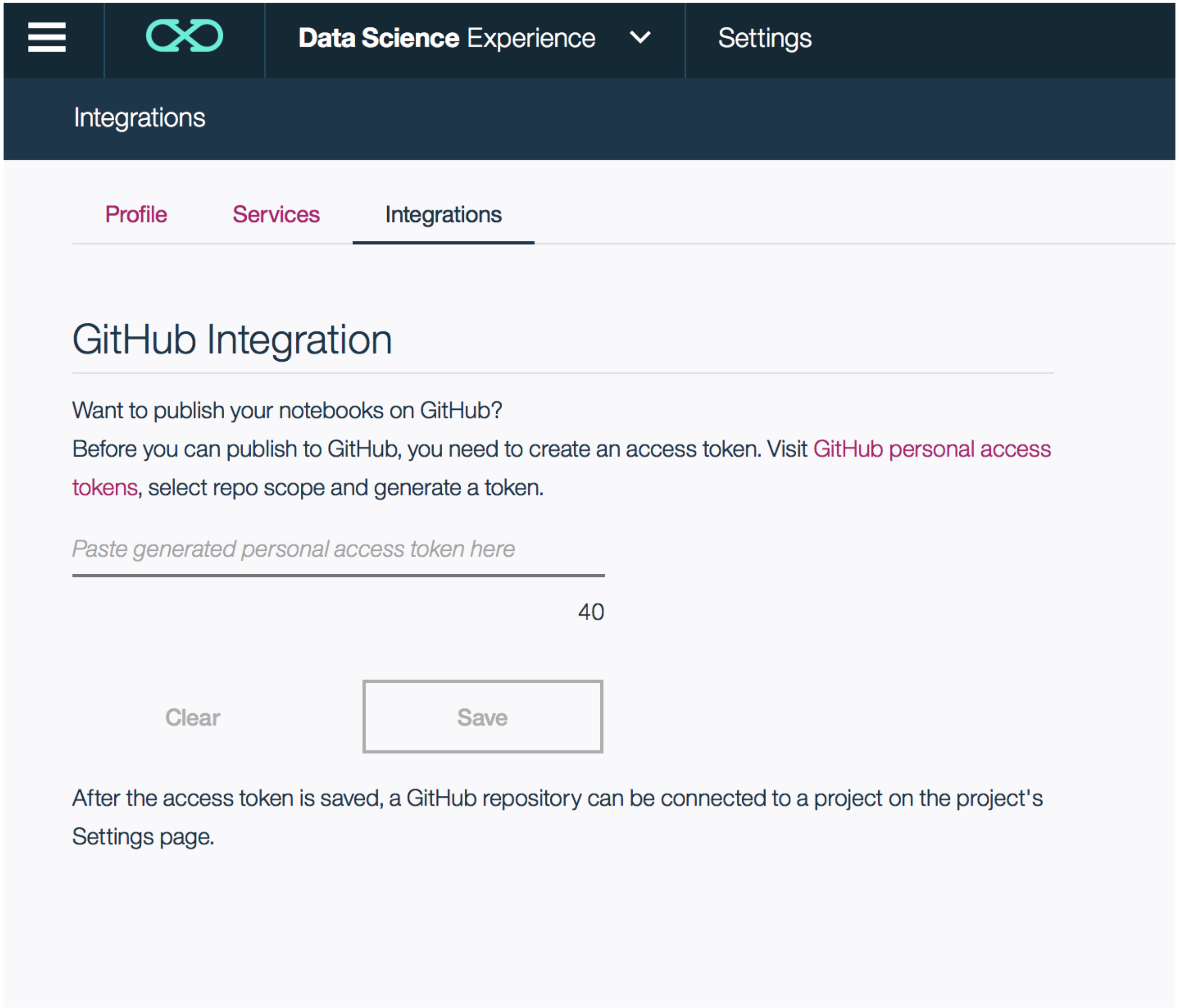
Add Collaborators to a Project



Add New Collaborator

Add users to your project for collaboration. Users with write access can add services to your project...

GitHub Integration



GitHub Integration

Want to publish your notebooks on GitHub?

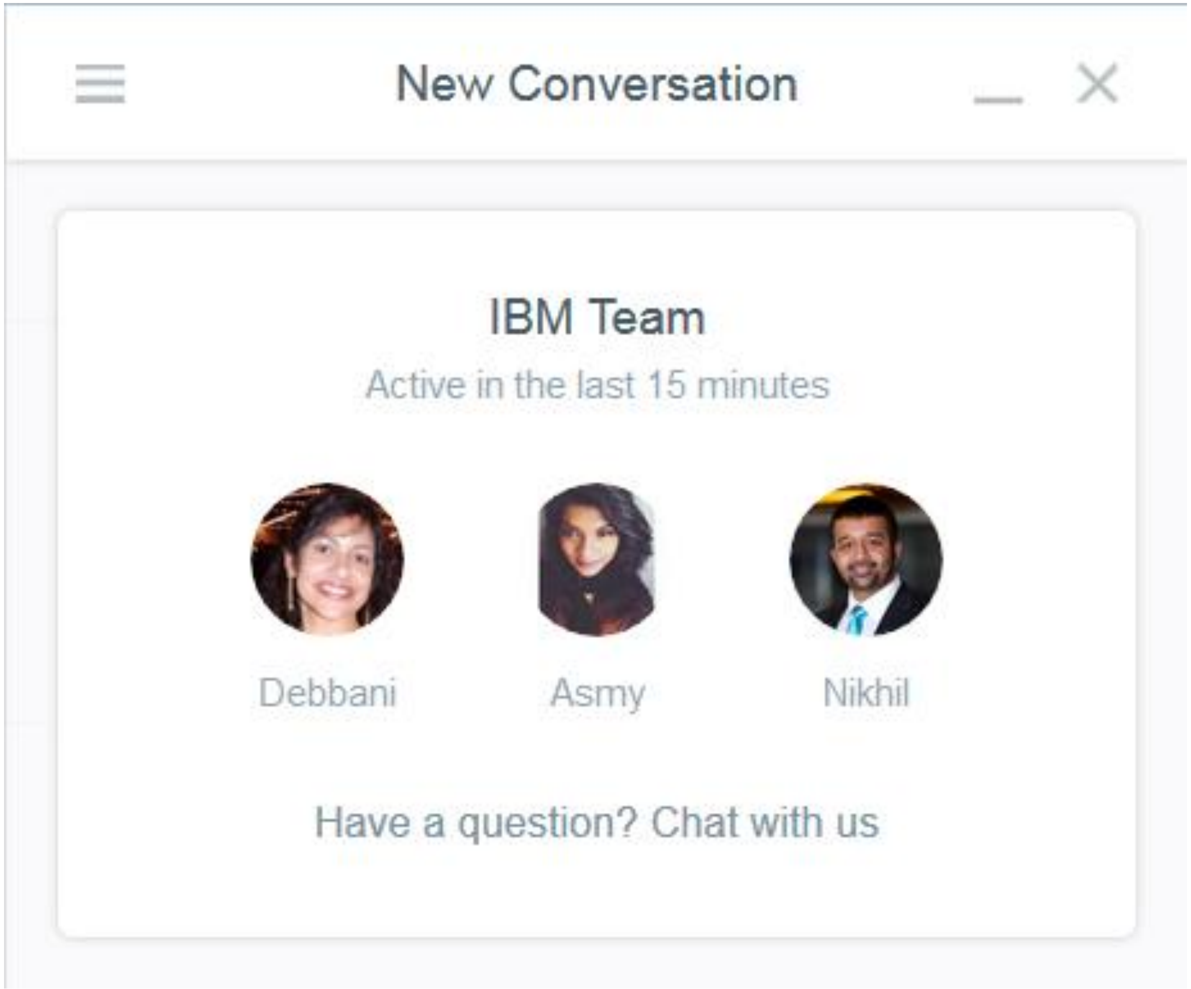
Before you can publish to GitHub, you need to create an access token. Visit GitHub personal access tokens, select repo scope and generate a token.

Paste generated personal access token here

40

After the access token is saved, a GitHub repository can be connected to a project on the project's Settings page.

Live chat on Intercom for support from the IBM team and to provide your feedback on how we can improve



New Conversation

IBM Team

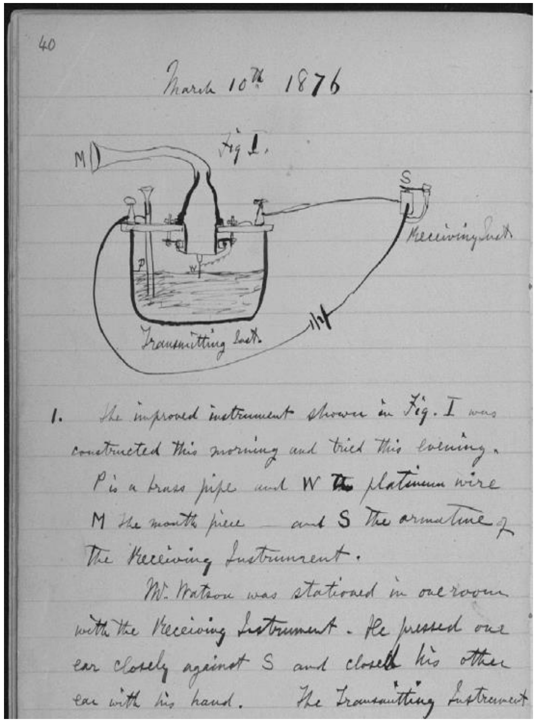
Active in the last 15 minutes

Have a question? Chat with us

What is a “Notebook”?

Pen and Paper

- Pen and paper has long provided the rich experience that scientists need to document progress through notes and drawings:
  - Expressive
  - Cumulative
  - Collaborative



40

March 10th 1876

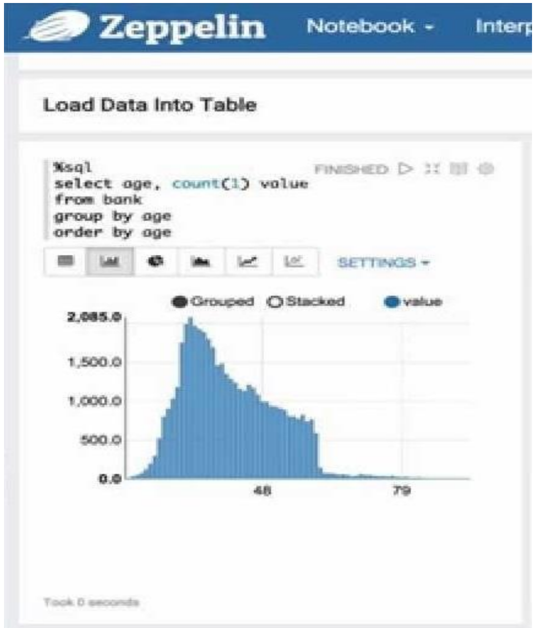
Fig 1.
M
S
Receiving Inst.
P
W
Transmitting Inst.

1. The improved instrument shown in Fig. I was constructed this morning and tried this evening. P is a brass pipe and W ~~the~~ the platinum wire M the mouth piece and S the armature of The Receiving Instrument.

Mr. Watson was stationed in one room with the Receiving Instrument. He pressed one ear closely against S and closed his other ear with his hand. The Transmitting Instrument

Notebooks

- Notebooks are the digital equivalent of the “pen and paper” lab notebook, enabling data scientists to document reproducible analysis:
  - Markdown and visualization
  - Iterative exploration
  - Easy to share

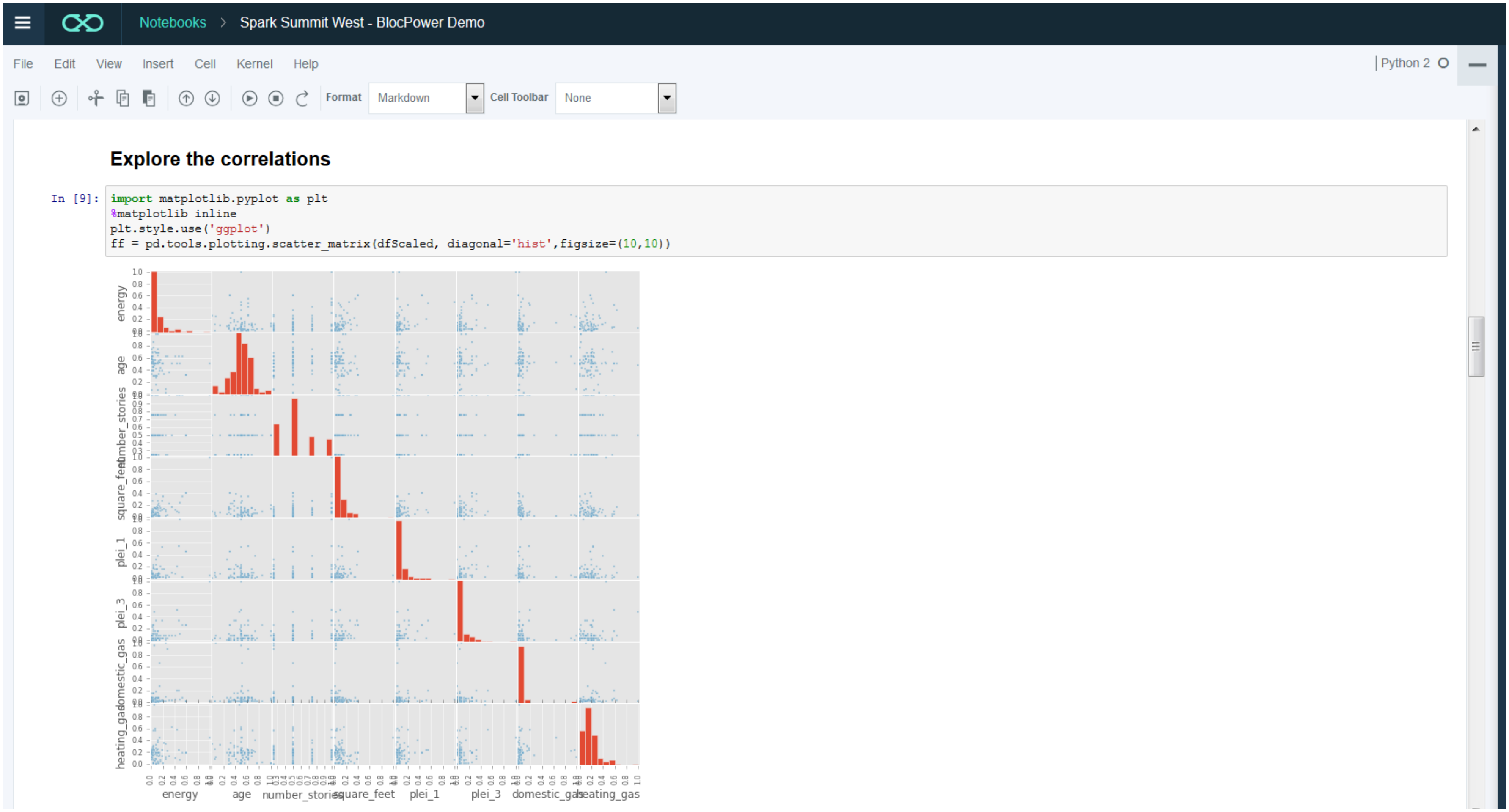


Load Data Into Table

```
%sql
select age, count(1) value
from bank
group by age
order by age
```

Took 0 seconds

Integrated Jupyter Notebooks for interactive and collaborative development - seamless execution on Spark

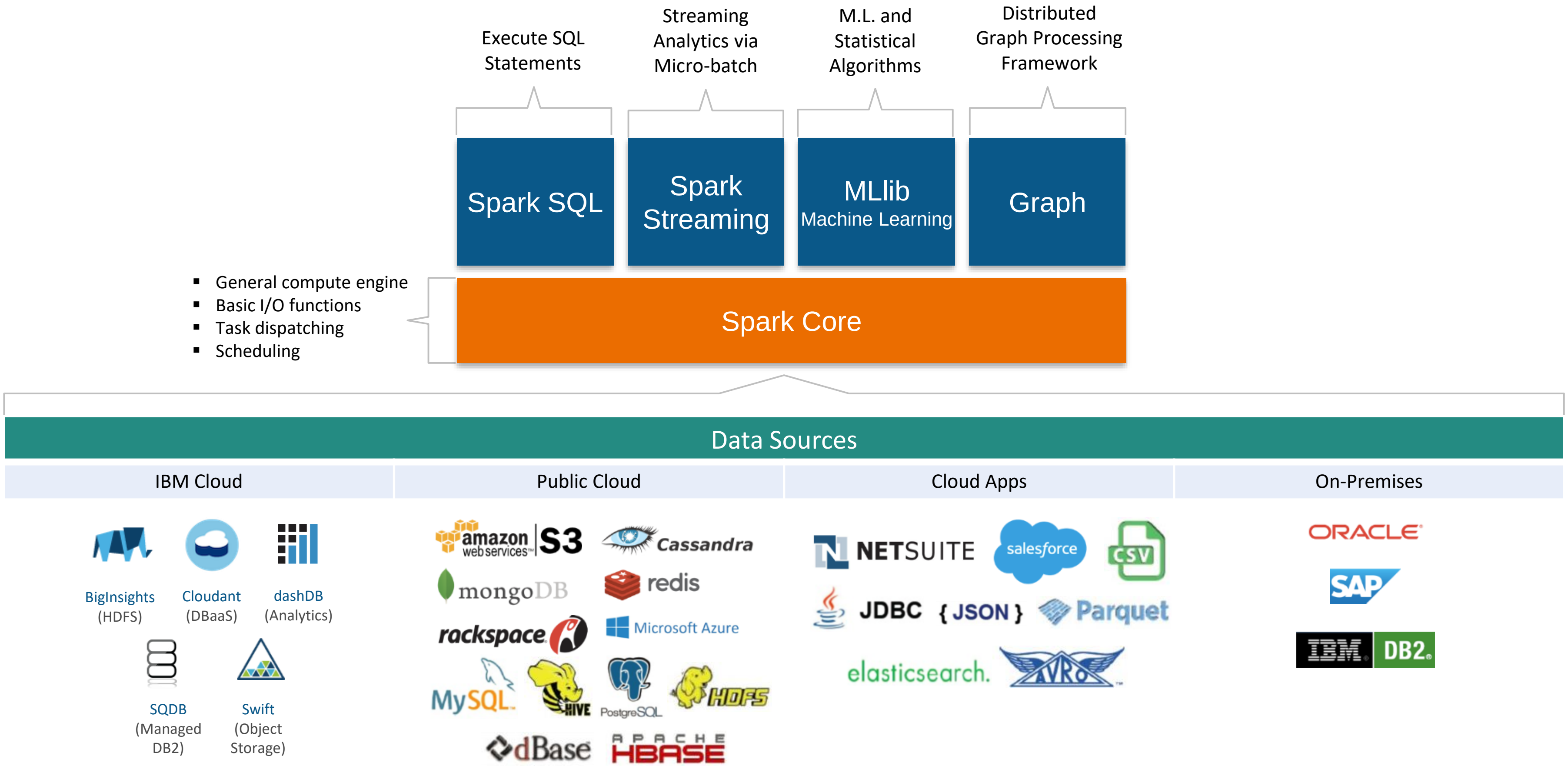


Explore the correlations

In [9]:

```
import matplotlib.pyplot as plt
%matplotlib inline
plt.style.use('ggplot')
ff = pd.tools.plotting.scatter_matrix(dfScaled, diagonal='hist',figsize=(10,10))
```

From a Notebook in DSX you can use IBM's managed Spark Service to blend multiple data types, sources, and workloads



Execute SQL Statements

Streaming Analytics via Micro-batch

M.L. and Statistical Algorithms

Distributed Graph Processing Framework

Spark SQL

Spark Streaming

MLlib Machine Learning

Graph

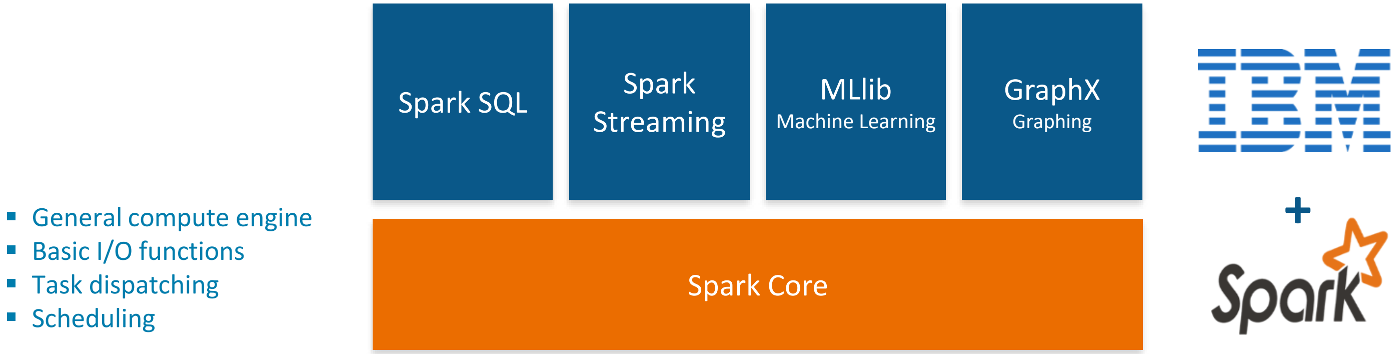
- General compute engine
- Basic I/O functions
- Task dispatching
- Scheduling

Spark Core

Data Sources

IBM Cloud	Public Cloud	Cloud Apps	On-Premises
BigInsights (HDFS) Cloudant (DBaaS) dashDB (Analytics) SQDB (Managed DB2) Swift (Object Storage)	amazon web services S3 Cassandra mongoDB redis rackspace Microsoft Azure MySQL HIVE PostgreSQL HDFS dBase APACHE HBASE	NETSUITE salesforce CSV JDBC { JSON } Parquet elasticsearch AVRO	ORACLE SAP IBM DB2

Benefits of Spark for Data Science



- General compute engine
- Basic I/O functions
- Task dispatching
- Scheduling

- Allows Data Scientists to code at scale
  - In-Memory processing that scales in a distributed architecture
- Supports multiple programing interfaces (Scala, Python, Java and R)
- Provides unified APIs (SQL, Streaming, Machine Learning, etc.)

The Spark service uses Bluemix Object Storage as its preferred data store for building performant applications

- **Object storage provides inexpensive, scalable and self-healing retention of massive amounts of unstructured data**

- **Every object exists at the same level in a flat address space**

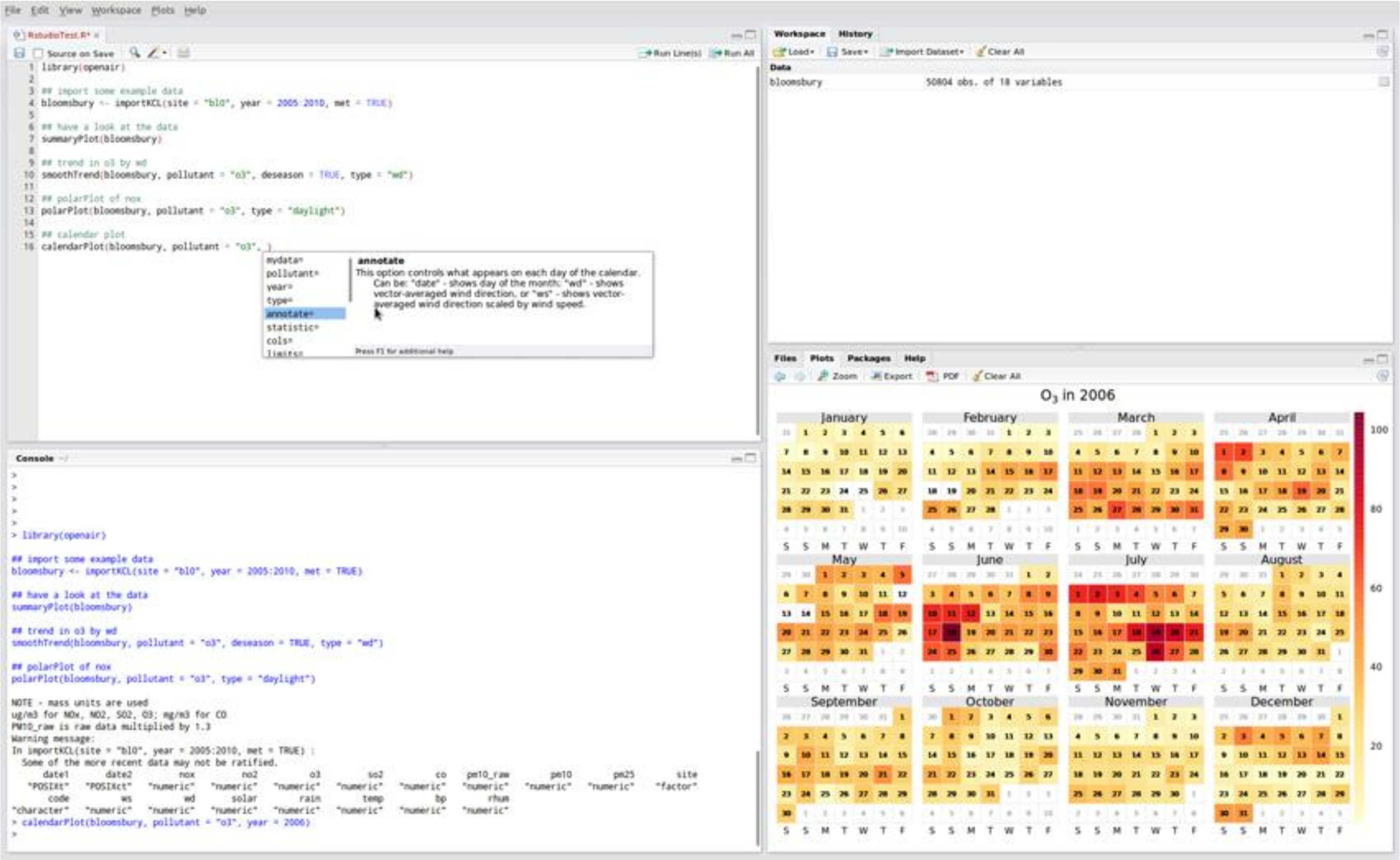
- **Bluemix Object Storage has a drag-and-drop upload and Swift API for programmatic access**

Supported Data Sources/Targets for DSX via on- premises and cloud Connectors

Cloud Sources	On-Premises Sources	Cloud Targets	On-Premises Targets
Amazon Redshift	Apache Hive	Amazon S3	IBM DB2® LUW
Amazon S3	Cloudera Impala	Bluemix Object Storage	IBM Pure Data for Analytics®
Apache Hive	IBM DB2® LUW	IBM Cloudant™	Teradata
Bluemix Object Storage	IBM Informix®	IBM dashDB	
IBM BigInsights™ on Cloud *	IBM Pure Data for Analytics®	IBM BigInsights™ on Cloud *	
IBM Cloudant™	Microsoft SQL Server	IBM DB2® on Cloud	
IBM dashDB	MySQL Enterprise Edition	IBM SQL Database	
IBM DB2® on Cloud	Oracle	IBM Watson™ Analytics	
IBM SQL Database	Pivotal Greenplum	PostgreSQL on Compose	
Microsoft Azure	PostgreSQL	SoftLayer Object Storage	
PostgreSQL on Compose	Sybase		
Salesforce	Sybase IQ		
SoftLayer Object Storage	Teradata		

All of the supported targets are compatible with each source

DSX has RStudio built into the experience thanks to our strategic partnership



```
library(openair)

## import some example data
bloomsbury <- importKCL(site = "bl0", year = 2005:2010, met = TRUE)

## have a look at the data
summaryPlot(bloomsbury)

## trend in o3 by wd
smoothTrend(bloomsbury, pollutant = "o3", deseason = TRUE, type = "wd")

## polarPlot of nox
polarPlot(bloomsbury, pollutant = "o3", type = "daylight")

## calendar plot
calendarPlot(bloomsbury, pollutant = "o3", )
```

```
mydata=
pollutant=
year=
type=
annotate=
statistic=
cols=
limits=
```

annotate

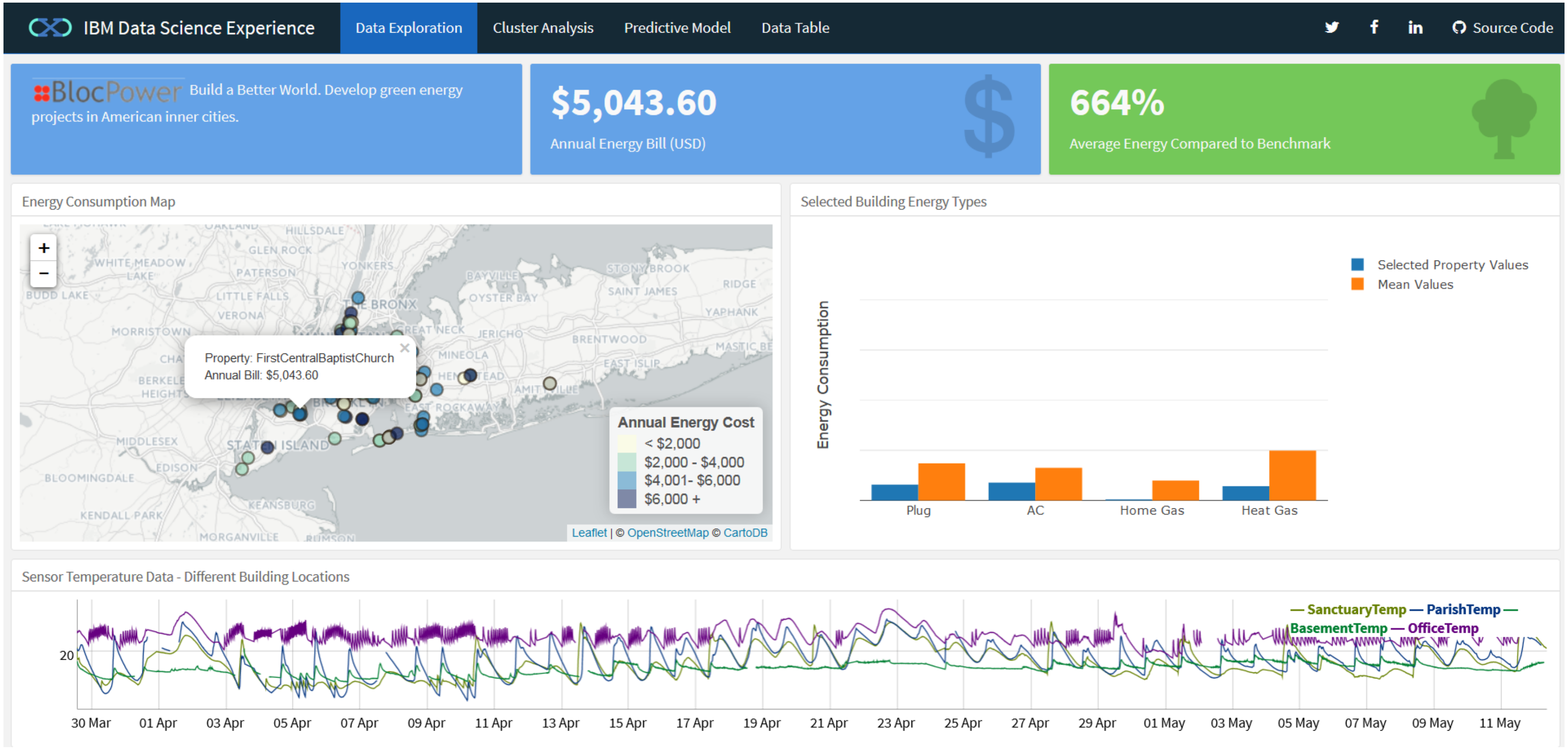
This option controls what appears on each day of the calendar. Can be: "date" - shows day of the month; "wd" - shows vector-averaged wind direction, or "ws" - shows vector-averaged wind direction scaled by wind speed.

Press F1 for additional help

```
bloomsbury                    50804 obs. of 18 variables
```

```
>
>
>
>
>
> library(openair)

## import some example data
bloomsbury <- importKCL(site = "bl0", year = 2005:2010, met = TRUE)

## have a look at the data
summaryPlot(bloomsbury)

## trend in o3 by wd
smoothTrend(bloomsbury, pollutant = "o3", deseason = TRUE, type = "wd")

## polarPlot of nox
polarPlot(bloomsbury, pollutant = "o3", type = "daylight")

NOTE - mass units are used
ug/m3 for NOx, NO2, SO2, O3; mg/m3 for CO
PM10_raw is raw data multiplied by 1.3
Warning message:
In importKCL(site = "bl0", year = 2005:2010, met = TRUE) :
  Some of the more recent data may not be ratified.
      date1       date2         nox         no2          o3         so2          co    pm10_raw        pm10        pm25        site
   "POSIXt"   "POSIXct"   "numeric"   "numeric"   "numeric"   "numeric"   "numeric"   "numeric"   "numeric"   "numeric"    "factor"
       code          ws          wd       solar        rain        temp          bp        rhum
"character"   "numeric"   "numeric"   "numeric"   "numeric"   "numeric"   "numeric"   "numeric"
> calendarPlot(bloomsbury, pollutant = "o3", year = 2006)
>
```

With RStudio you can create Shiny web applications to make your analysis accessible to the business



BlocPower Build a Better World. Develop green energy projects in American inner cities.

$5,043.60

Annual Energy Bill (USD)

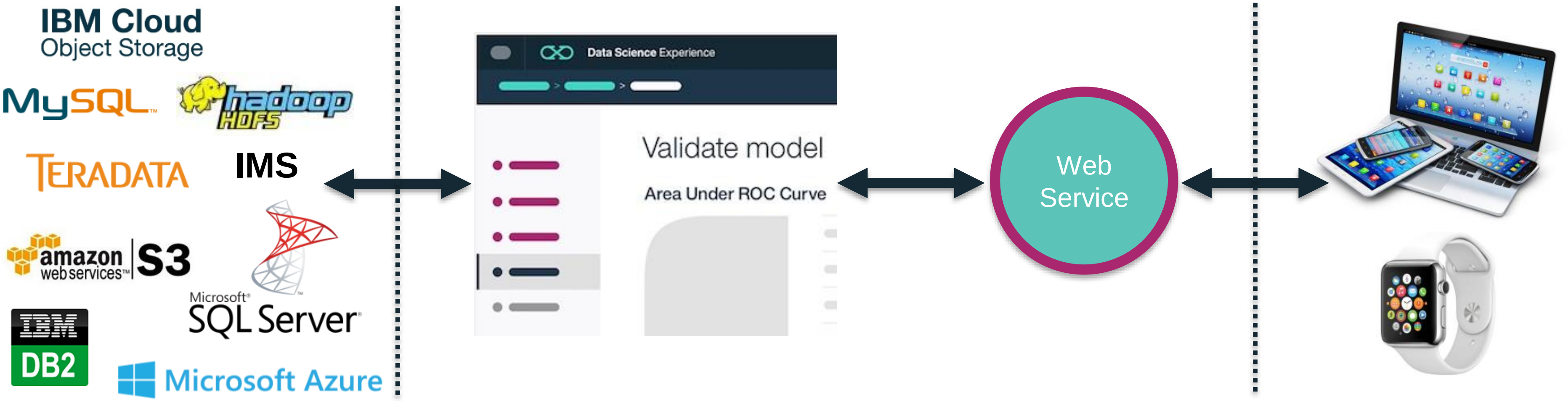
664%

Average Energy Compared to Benchmark

Property: FirstCentralBaptistChurch
Annual Bill: $5,043.60

Operationalize insights with IBM Machine Learning

IBM Machine Learning



Data Access:

- Easily connect to Behind-the-Firewall and Public Cloud Data
- Catalogued and Governed Controls through Watson Data Platform

Creating Models:

- Single UI and API for creating ML Models on various Runtimes
- Auto-Modeling and Hyperparameter Optimization

Web Service:

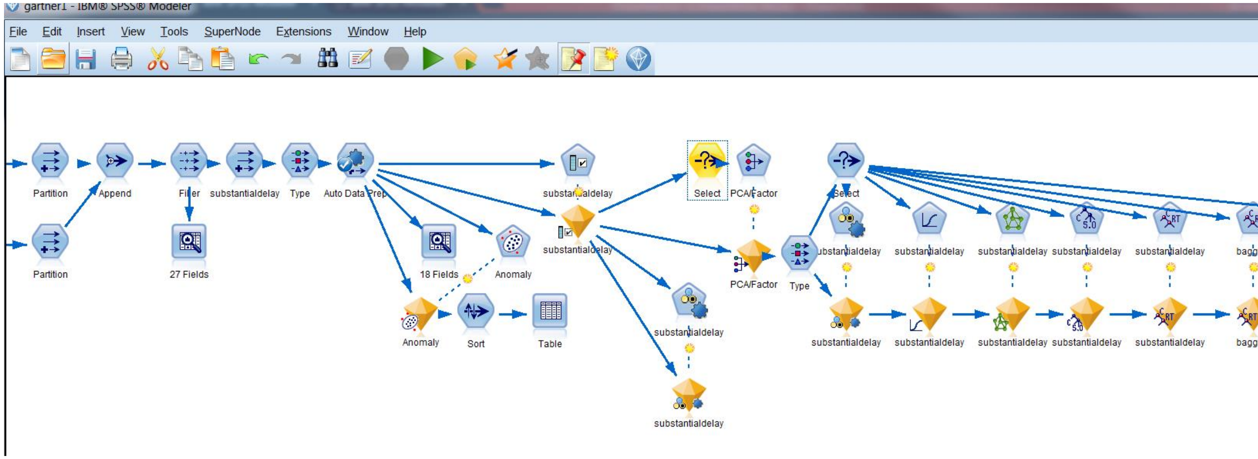
- Real-time, Streaming, and Batch Deployment
- Continuous Monitoring and Feedback Loop

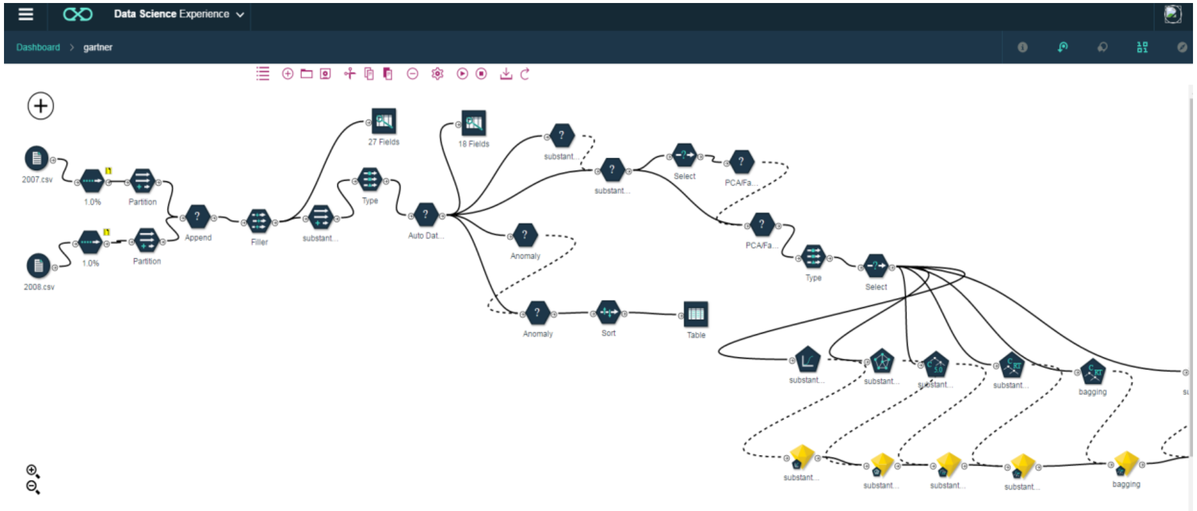
Intelligent Apps:

- Integrate ML models with apps, websites, etc.
- Continuously Improve and Adapt with Self-Learning

DSX Canvas

- DSX Canvas will have compatibility with legacy SPSS Modeler streams
- Multiple execution runtimes: SPSS Modeler, SparkML
- Planned support for R/Python/SQL code

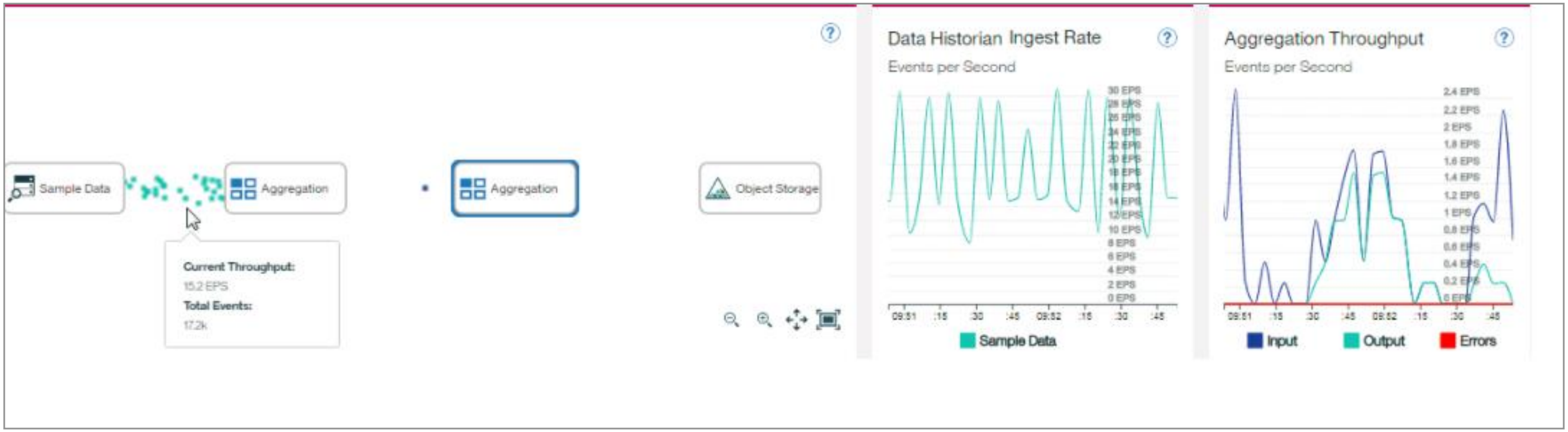




- Pipeline deployment from DSX Canvas (left) via IBM Machine Learning

Stream Designer – Open Beta

- Characteristics of Stream Processing
  - Continuous processing
  - Multiple varied data sources
  - High data rates/ data volumens
  - Near-real time action
- DSX
  - Design stream flow with new Stream Designer
  - Executes in Streaming Analytics Service (based on IBM Streams)
  - Can invoke stream within Jupyter notebooks using Stream API



Current Throughput:
15.2 EPS
Total Events:
17.2k

DSX Local

- **Very similar to the public cloud version of DSX**
- **Runs on hardware that is provided by the customer**
  - The DSX Local software and hardware are managed by the customer
- **DSX Local comes with all the software it needs to run, although it can integrate with existing customer systems such as**
  - Databases and HDFS storage
  - LDAP servers for authentication

Get Started with Data Science Experience Today!

Calling all Data Scientists!

- Our mission is to win the **hearts and minds** of Data Scientists

- IBM Data Science Experience is a **freemium model** with value-add features, pricing and up-sell in development

- **Sign up** and encourage your colleagues to do so at **datascience.ibm.com**

IBM Data Science Experience
https://www.youtube.com/watch?v=1HjzkLRdP5k&t=29s