UK Data Service Open Data Platform for Social Science

Aidan Condron & Deirdre Lungley
Big Data Network Support
UK Data Archive
University of Essex

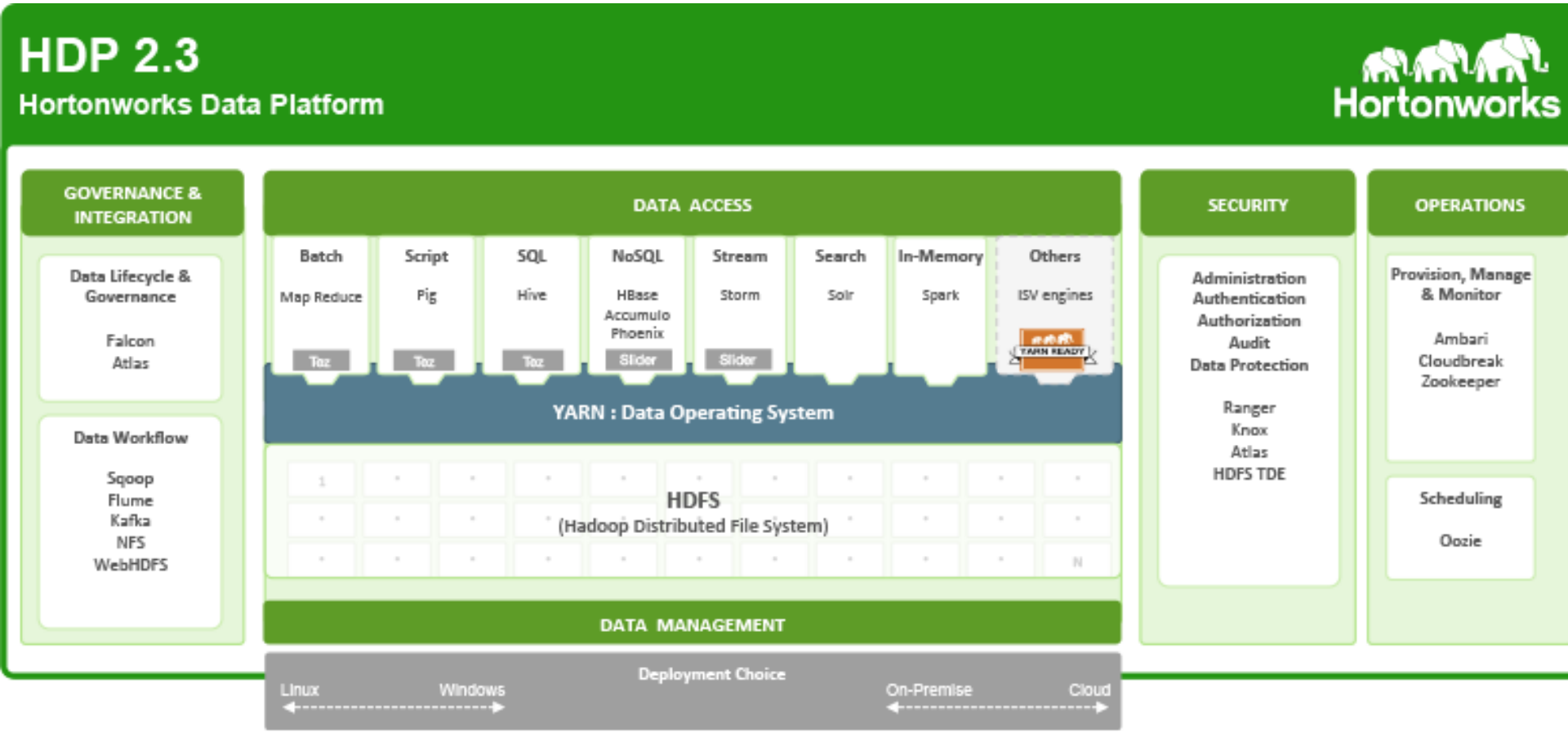
UK Data Service

BDN2 Goal

‘The overarching aim of the ESRC Big Data Network is to build infrastructure which enables researchers to exploit the richness of the big data landscape in the UK to advance research in the social sciences and economics and to enhance its societal impact’

DISAC 03/14, November 2014

Open Source Data Platform



Big Data Management and Exploratory Data Analysis (EDA)
Smarter Household Energy

- Processing
- Descriptive Statistics
- Linking
- Dynamic Aggregation / Data ‘Slicing and Dicing’
- Analytics
- Visualisation
- Geographical Visualisation

Smart Meter Data

Data Service
cover

Discover

Variable and question bank

QualiBank

About us | Get data | Use data | Manage data | Deposit data | News and events

Discover > Catalogue

Catalogue

SHARE

UK Data Service data catalogue record for:

Energy Demand Research Project: Early Smart Meter Trials, 2007-2010

Documentation | Publications | Download/Order | DDI XML

TITLE DETAILS

SN:	7591
Title:	Energy Demand Research Project: Early Smart Meter Trials, 2007-2010
Alternative title:	EDRP
Persistent identifier:	10.5255/UKDA-SN-7591-1
Depositor:	Department of Energy and Climate Change
Principal investigator(s):	AECOM Building Engineering
Data collector(s):	Centre for Sustainable Energy
Sponsor(s):	Department of Energy and Climate Change
Other acknowledgements:	Energy Suppliers: EDF Energy, E.ON UK, Scottish Power Energy Retail and SSE Energy Supply. Ofgem.

UK Data Service

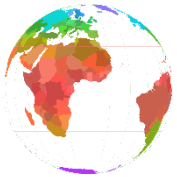
Data

- Geodemograpic Data - 14621 Households Throughout UK
  - area
  - socioeconomic indicators
  - energy tariffs
- Electricity and Gas Smartmeters
  - Anonymized Household Identifier
  - 30 Min Energy Usage
  - Date and Time
- Meteorological
  - Temperature
  - Precipitation

Exploratory Data Analysis (EDA) – Mapping ‘small’ data to ‘big’ energy data

- ODP Apache Hadoop Components
  - Ambari
  - Hive
  - Spark
  - Kylin
  - Zeppelin
- Mapped
  - Geographical areas
  - Socioeconomic groups
  - Time periods – days, months, years
  - Individual households

Open Data Platform:
Smarter Household Energy Workflow



Geographical

Smart Meters

Meteorological

WinSCP

ODP

Spark

Hive

Descriptives/Statistics

Kylin

Analytics

Zeppelin

Visualisations

UK Data Service

```
16/01/19 14:52:11 INFO TaskSetManager: Finished task 81.0 in stage 3.0 (TID 84) in 18980 ms on sandbox.hortonworks.com (82/97)
16/01/19 14:52:31 INFO TaskSetManager: Starting task 83.0 in stage 3.0 (TID 86, sandbox.hortonworks.com, NODE_LOCAL, 1415 bytes)
16/01/19 14:52:31 INFO TaskSetManager: Finished task 82.0 in stage 3.0 (TID 85) in 20284 ms on sandbox.hortonworks.com (83/97)
16/01/19 14:52:49 INFO TaskSetManager: Starting task 84.0 in stage 3.0 (TID 87, sandbox.hortonworks.com, NODE_LOCAL, 1415 bytes)
16/01/19 14:52:49 INFO TaskSetManager: Finished task 83.0 in stage 3.0 (TID 86) in 18373 ms on sandbox.hortonworks.com (84/97)
16/01/19 14:53:08 INFO TaskSetManager: Starting task 85.0 in stage 3.0 (TID 88, sandbox.hortonworks.com, NODE_LOCAL, 1415 bytes)
16/01/19 14:53:08 INFO TaskSetManager: Finished task 84.0 in stage 3.0 (TID 87) in 18327 ms on sandbox.hortonworks.com (85/97)
16/01/19 14:53:28 INFO TaskSetManager: Starting task 86.0 in stage 3.0 (TID 89, sandbox.hortonworks.com, NODE_LOCAL, 1415 bytes)
16/01/19 14:53:28 INFO TaskSetManager: Finished task 85.0 in stage 3.0 (TID 88) in 20216 ms on sandbox.hortonworks.com (86/97)
16/01/19 14:53:47 INFO TaskSetManager: Starting task 87.0 in stage 3.0 (TID 90, sandbox.hortonworks.com, NODE_LOCAL, 1415 bytes)
16/01/19 14:53:47 INFO TaskSetManager: Finished task 86.0 in stage 3.0 (TID 89) in 18805 ms on sandbox.hortonworks.com (87/97)
16/01/19 14:54:06 INFO TaskSetManager: Starting task 88.0 in stage 3.0 (TID 91, sandbox.hortonworks.com, NODE_LOCAL, 1415 bytes)
16/01/19 14:54:06 INFO TaskSetManager: Finished task 87.0 in stage 3.0 (TID 90) in 18930 ms on sandbox.hortonworks.com (88/97)
16/01/19 14:54:27 INFO TaskSetManager: Starting task 89.0 in stage 3.0 (TID 92, sandbox.hortonworks.com, NODE_LOCAL, 1415 bytes)
16/01/19 14:54:27 INFO TaskSetManager: Finished task 88.0 in stage 3.0 (TID 91) in 21168 ms on sandbox.hortonworks.com (89/97)
16/01/19 14:54:46 INFO TaskSetManager: Starting task 90.0 in stage 3.0 (TID 93, sandbox.hortonworks.com, NODE_LOCAL, 1415 bytes)
16/01/19 14:54:46 INFO TaskSetManager: Finished task 89.0 in stage 3.0 (TID 92) in 18812 ms on sandbox.hortonworks.com (90/97)
16/01/19 14:55:05 INFO TaskSetManager: Starting task 91.0 in stage 3.0 (TID 94, sandbox.hortonworks.com, NODE_LOCAL, 1415 bytes)
16/01/19 14:55:05 INFO TaskSetManager: Finished task 90.0 in stage 3.0 (TID 93) in 19062 ms on sandbox.hortonworks.com (91/97)
16/01/19 14:55:26 INFO TaskSetManager: Starting task 92.0 in stage 3.0 (TID 95, sandbox.hortonworks.com, NODE_LOCAL, 1415 bytes)
16/01/19 14:55:26 INFO TaskSetManager: Finished task 91.0 in stage 3.0 (TID 94) in 21300 ms on sandbox.hortonworks.com (92/97)
16/01/19 14:55:45 INFO TaskSetManager: Starting task 93.0 in stage 3.0 (TID 96, sandbox.hortonworks.com, NODE_LOCAL, 1415 bytes)
16/01/19 14:55:45 INFO TaskSetManager: Finished task 92.0 in stage 3.0 (TID 95) in 18939 ms on sandbox.hortonworks.com (93/97)
16/01/19 14:56:04 INFO TaskSetManager: Starting task 94.0 in stage 3.0 (TID 97, sandbox.hortonworks.com, NODE_LOCAL, 1415 bytes)
16/01/19 14:56:04 INFO TaskSetManager: Finished task 93.0 in stage 3.0 (TID 96) in 18885 ms on sandbox.hortonworks.com (94/97)
16/01/19 14:56:24 INFO TaskSetManager: Starting task 95.0 in stage 3.0 (TID 98, sandbox.hortonworks.com, NODE_LOCAL, 1415 bytes)
16/01/19 14:56:24 INFO TaskSetManager: Finished task 94.0 in stage 3.0 (TID 97) in 20511 ms on sandbox.hortonworks.com (95/97)
16/01/19 14:56:44 INFO TaskSetManager: Starting task 96.0 in stage 3.0 (TID 99, sandbox.hortonworks.com, NODE_LOCAL, 1415 bytes)
16/01/19 14:56:44 INFO TaskSetManager: Finished task 95.0 in stage 3.0 (TID 98) in 19942 ms on sandbox.hortonworks.com (96/97)
16/01/19 14:56:55 INFO TaskSetManager: Finished task 96.0 in stage 3.0 (TID 99) in 10737 ms on sandbox.hortonworks.com (97/97)
16/01/19 14:56:55 INFO DAGScheduler: ShuffleMapStage 3 (count at <console>:40) finished in 1889.296 s
16/01/19 14:56:55 INFO YarnScheduler: Removed TaskSet 3.0, whose tasks have all completed, from pool
16/01/19 14:56:55 INFO DAGScheduler: looking for newly runnable stages
16/01/19 14:56:55 INFO DAGScheduler: running: Set()
16/01/19 14:56:55 INFO DAGScheduler: waiting: Set(ResultStage 4)
16/01/19 14:56:55 INFO DAGScheduler: failed: Set()
16/01/19 14:56:55 INFO DAGScheduler: Missing parents for ResultStage 4: List()
16/01/19 14:56:55 INFO DAGScheduler: Submitting ResultStage 4 (MapPartitionsRDD[18] at count at <console>:40), which is now runnable
16/01/19 14:56:55 INFO MemoryStore: ensureFreeSpace(16960) called with curMem=748275, maxMem=278302556
16/01/19 14:56:55 INFO MemoryStore: Block broadcast_7 stored as values in memory (estimated size 16.6 KB, free 264.7 MB)
16/01/19 14:56:55 INFO MemoryStore: ensureFreeSpace(7283) called with curMem=765235, maxMem=278302556
16/01/19 14:56:55 INFO MemoryStore: Block broadcast_7_piece0 stored as bytes in memory (estimated size 7.1 KB, free 264.7 MB)
16/01/19 14:56:55 INFO BlockManagerInfo: Added broadcast_7_piece0 in memory on 10.0.2.15:56213 (size: 7.1 KB, free: 265.3 MB)
16/01/19 14:56:55 INFO SparkContext: Created broadcast 7 from broadcast at DAGScheduler.scala:874
16/01/19 14:56:55 INFO DAGScheduler: Submitting 1 missing tasks from ResultStage 4 (MapPartitionsRDD[18] at count at <console>:40)
16/01/19 14:56:55 INFO YarnScheduler: Adding task set 4.0 with 1 tasks
16/01/19 14:56:55 INFO TaskSetManager: Starting task 0.0 in stage 4.0 (TID 100, sandbox.hortonworks.com, PROCESS_LOCAL, 1165 bytes)
16/01/19 14:56:55 INFO BlockManagerInfo: Added broadcast_7_piece0 in memory on sandbox.hortonworks.com:40250 (size: 7.1 KB, free: 265.4 MB)
16/01/19 14:56:55 INFO MapOutputTrackerMasterEndpoint: Asked to send map output locations for shuffle 0 to sandbox.hortonworks.com:58993
16/01/19 14:56:55 INFO MapOutputTrackerMaster: Size of output statuses for shuffle 0 is 192 bytes
16/01/19 14:56:55 INFO TaskSetManager: Finished task 0.0 in stage 4.0 (TID 100) in 272 ms on sandbox.hortonworks.com (1/1)
16/01/19 14:56:55 INFO DAGScheduler: ResultStage 4 (count at <console>:40) finished in 0.272 s
16/01/19 14:56:55 INFO YarnScheduler: Removed TaskSet 4.0, whose tasks have all completed, from pool
16/01/19 14:56:55 INFO DAGScheduler: Job 3 finished: count at <console>:40, took 1889.650503 s
res2: Long = 413836038

scala>
```

Further Contact

- Data Curation and Technical

  Deirdre Lungley dmlung@essex.ac.uk
  +44 1206 873574

- Engagement

  Aidan Condron – acondron@essex.ac.uk

  + 44 1206 874254

  Nathan Cunningham njcunna@essex.ac.uk

  +44(0)1206 872269

- Training

  Sarah King-Hele – sarah.king-hele@manchester.ac.uk
  + 44 161 275 0279