

Managing big data and analytics in multi-system environments

11th October 2016

Mandy Chessell CBE FREng CEng FBCS
Distinguished Engineer, Master Inventor
Analytics Group CTO Office



mandy_chessell@uk.ibm.com



Please Note:

IBM's statements regarding its plans, directions, and intent are subject to change or withdrawal without notice at IBM's sole discretion.

Information regarding potential future products is intended to outline our general product direction and it should not be relied on in making a purchasing decision.

The information mentioned regarding potential future products is not a commitment, promise, or legal obligation to deliver any material, code or functionality. Information about potential future products may not be incorporated into any contract.

The development, release, and timing of any future features or functionality described for our products remains at our sole discretion.

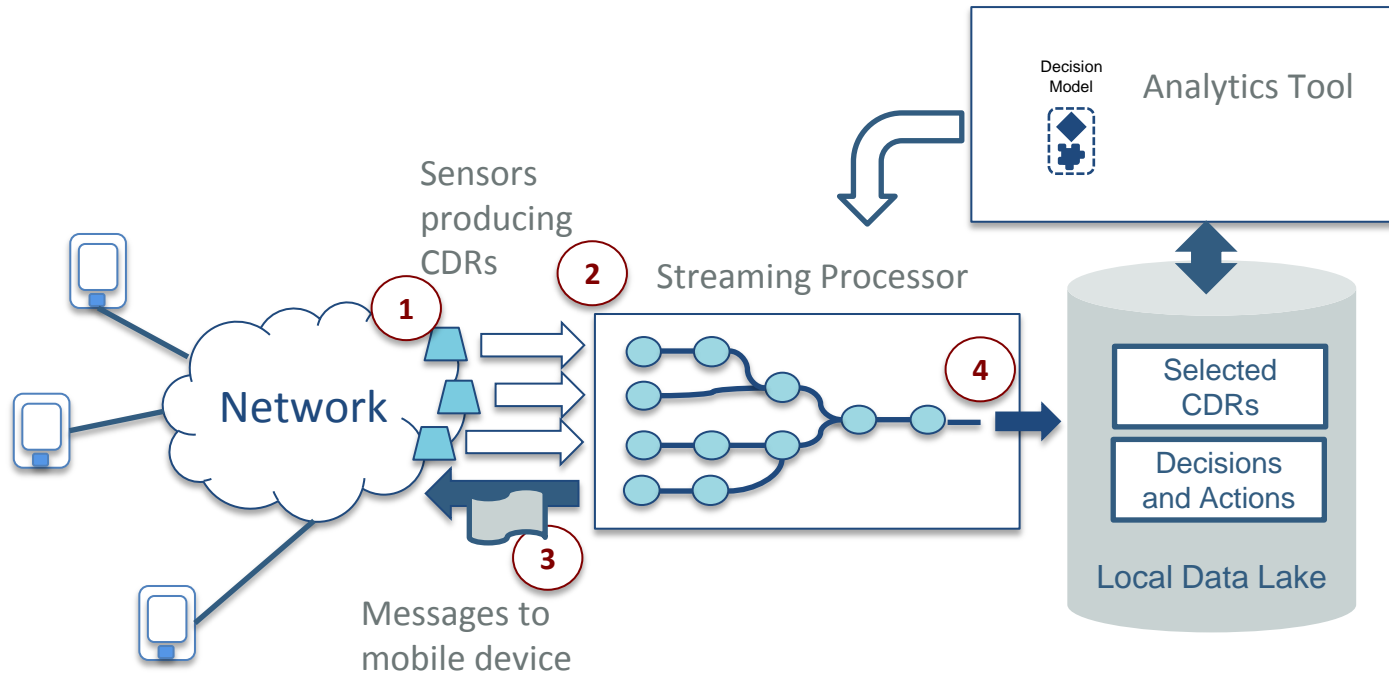
Performance is based on measurements and projections using standard IBM benchmarks in a controlled environment. The actual throughput or performance that any user will experience will vary depending upon many factors, including considerations such as the amount of multiprogramming in the user's job stream, the I/O configuration, the storage configuration, and the workload processed. Therefore, no assurance can be given that an individual user will achieve results similar to those stated here.



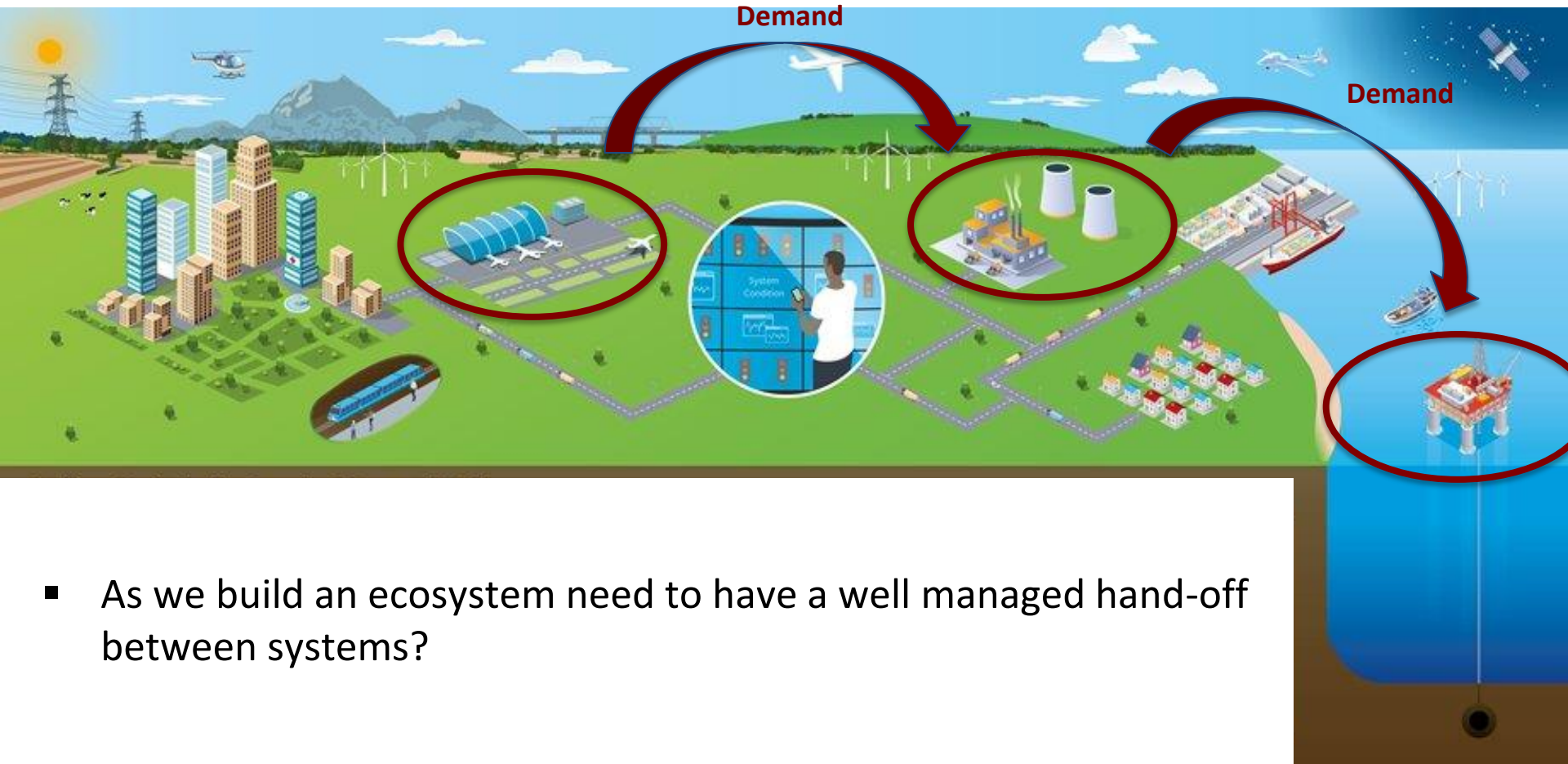
Sensors gather data but have little understanding of their context



Figure 7 - Streaming Analytics



The value of the ecosystem



- As we build an ecosystem need to have a well managed hand-off between systems?

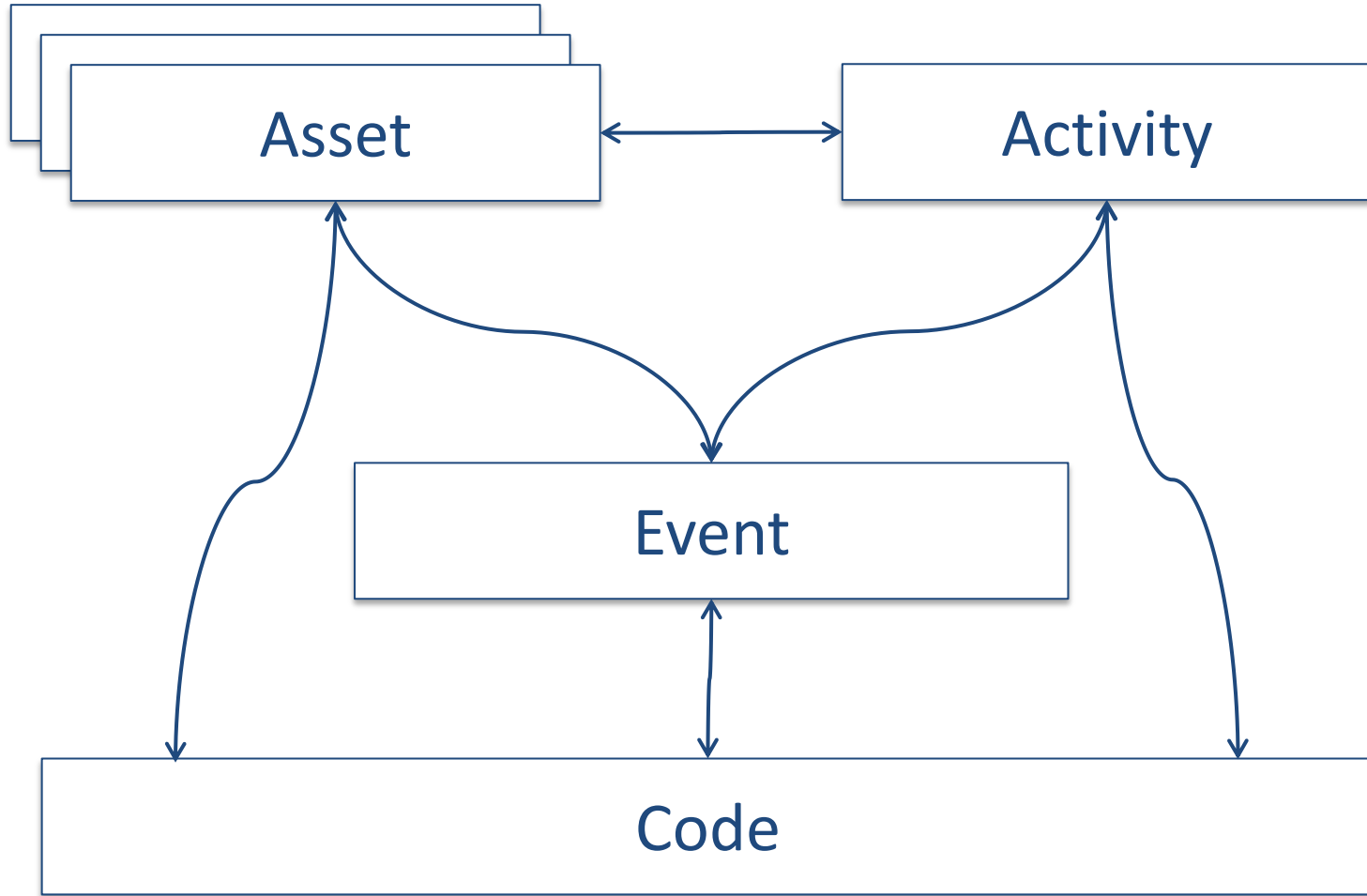
Data linkage

KWWD 2 US WILDWOOD US1 NJ Wildwood 39.02000000 -74.92000000 NJC009 NJZ023 34009 Y 504 281351 23

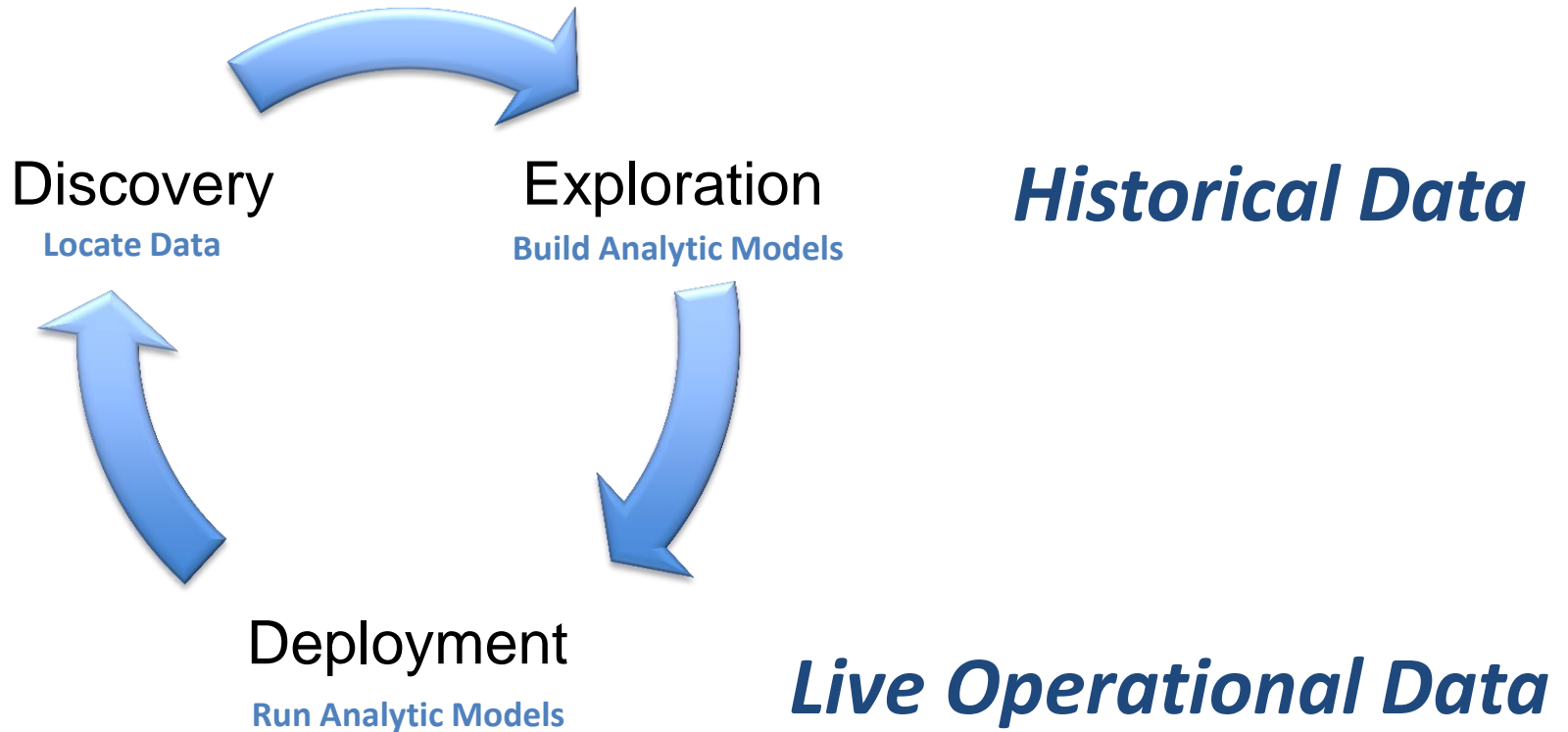
48982500 10.400000 (null) 10.400000 3300 (null) (null) 1005.757000 KWVL Waterville (null) (null) (null) 993.8 (null) (null) 2000-01-27 12:15:00
00 (null) (null) 15.800000 3400 (null) (null) 1006.096000 KWVL Waterville (null) (null) (null) 994.2 (null) (null) 2000-01-27 20:15:00 (null) (null) P
9282500 28.400000 (null) 28.400000 3300 (null) (null) 1019.980000 KWVL Waterville (null) (null) (null) 1007.9 (null) (null) 2000-01-30 23:35:00
9289700 26.600000 (null) 26.600000 3300 (null) (null) 1019.980000 KWVL Waterville (null) (null) (null) 1007.9 (null) (null) 2000-01-31 01:35:00
269300 41.384900 (null) 44.600000 3400 (null) (null) 1021.335000 KWVL Waterville (null) (null) (null) 1009.2 (null) (null) 2000-01-30 19:55:00 (I
146798500 50.000000 (null) 50.000000 2700 (null) (null) 1024.383000 KWWD Wildwood (null) (null) (null) 1023.5 (null) (null) 2000-01-02 05:35:
946869300 53.600000 18.412500 53.600000 3300 (null) (null) 1019.642000 KWWD Wildwood (null) (null) (null) 1018.7 (null) (null) 2000-01-03
946875300 55.400000 (null) 55.400000 3300 (null) (null) 1019.303000 KWWD Wildwood (null) (null) -0.700000 1018.4 Falling 2 2000-01-03 02:
946853700 57.200000 (null) 57.200000 3400 (null) (null) 1019.642000 KWWD Wildwood (null) (null) -1.400000 1018.7 Falling 2 2000-01-02 20:
946847700 62.600000 18.412500 62.600000 3400 (null) (null) 1019.980000 KWWD Wildwood (null) (null) (null) 1019.1 (null) (null) 2000-01-02
946842900 64.400000 (null) 64.400000 3400 64.400000 48.200000 1020.006000 KWWD Wildwood (null) (null) (null) 1020.1 (null) (null) 2000-C
946928100 66.200000 (null) 66.200000 3400 (null) (null) 1019.980000 KWWD Wildwood (null) (null) (null) 1019.1 (null) (null) 2000-01-03 17:35
47166900 24.800000 (null) 24.800000 3300 30.200000 23.000000 1036.574000 KWWD Wildwood (null) (null) 1.000000 1035.7 Rising 1 2000-C
7189700 40.163960 (null) 44.600000 3400 (null) (null) 1033.526000 KWWD Wildwood (null) (null) (null) 1032.6 (null) (null) 2000-01-06 18:15:0
7304900 36.397550 (null) 41.000000 3300 (null) (null) 1027.092000 KWWD Wildwood (null) (null) (null) 1026.2 (null) (null) 2000-01-08 02:15:0
7292900 38.795900 19.563280 44.600000 3300 (null) (null) 1024.383000 KWWD Wildwood (null) (null) (null) 1023.5 (null) (null) 2000-01-07 22
947555700 53.600000 24.166410 53.600000 3300 57.200000 53.600000 1001.694000 KWWD Wildwood (null) (null) 4.100000 1000.8 Rising 1 :
17529300 53.600000 (null) 53.600000 2600 (null) (null) 1005.757000 KWWD Wildwood 0.03 (null) (null) 1004.9 (null) (null) 2000-01-10 16:35:0
947549700 55.400000 27.618750 55.400000 3400 (null) (null) 998.307400 KWWD Wildwood (null) (null) (null) 997.4 (null) (null) 2000-01-10 22
1490900 41.384900 (null) 44.600000 2600 46.400000 42.800000 1015.917000 KWWD Wildwood (null) (null) -0.300000 1015.0 Falling 2 2000-0:
7644500 36.363880 33.372660 44.600000 3390 (null) (null) 1011.176000 KWWD Wildwood (null) (null) (null) 1010.3 (null) (null) 2000-01-12 00
7584500 43.521730 (null) 46.400000 3300 (null) (null) 1008.128000 KWWD Wildwood (null) (null) (null) 1007.2 (null) (null) 2000-01-11 07:55:0
100 38.967580 29.920320 46.400000 3000 59.000000 44.600000 1005.419000 KWWD Wildwood (null) (null) 1.000000 1004.5 Rising 1 2000-01
47730900 25.350150 (null) 30.200000 3300 (null) (null) 1023.028000 KWWD Wildwood (null) (null) (null) 1022.1 (null) (null) 2000-01-13 00:35:
7714100 36.970950 16.110940 42.800000 3400 (null) (null) 1022.012000 KWWD Wildwood (null) (null) (null) 1021.1 (null) (null) 2000-01-12 19
7716500 37.970820 (null) 42.800000 3400 (null) (null) 1022.351000 KWWD Wildwood (null) (null) (null) 1021.5 (null) (null) 2000-01-12 20:35:0
7718900 37.444560 (null) 42.800000 3400 (null) (null) 1022.351000 KWWD Wildwood (null) (null) (null) 1021.5 (null) (null) 2000-01-12 21:15:0
7782500 44.260220 (null) 46.400000 3000 (null) (null) 1006.773000 KWWD Wildwood (null) (null) -1.400000 1005.9 Falling 2 2000-01-13 14:55
7795700 44.550240 (null) 48.200000 3000 (null) (null) 1004.064000 KWWD Wildwood (null) (null) (null) 1003.2 (null) (null) 2000-01-13 18:35:0
1948177300 -2.510077 19.563280 12.200000 3300 (null) (null) 1027.092000 KWWD Wildwood (null) (null) (null) 1026.2 (null) (null) 2000-01-18



Information elements



Analytics lifecycle



Metadata should bring as much information about the data sets to the data scientist as is known collectively by the organization.

Data Set Name: Employee Directory X

Description:
Core attributes describing all employees of OCO pharmaceuticals created from a daily extract from Kenexa.
Owner: [Penny Payer](#)

Classification Ranges:
Confidentiality: Public, Confidential, Sensitive
Confidence: Authoritative
Retention: Indefinitely

Status:
Last accessed: 6th May 2016
Records: 3488
Last Update: 1st May 2016

Contents:
[Structure ...](#)
[Contents ...](#)
[Lineage ...](#)

Column: X

Description **Characteristics** **Lineage**

CS

Position reference number for non-exempt employees. The value ranges from 01 to 06 where 01 is the most senior and 06 is the most junior.

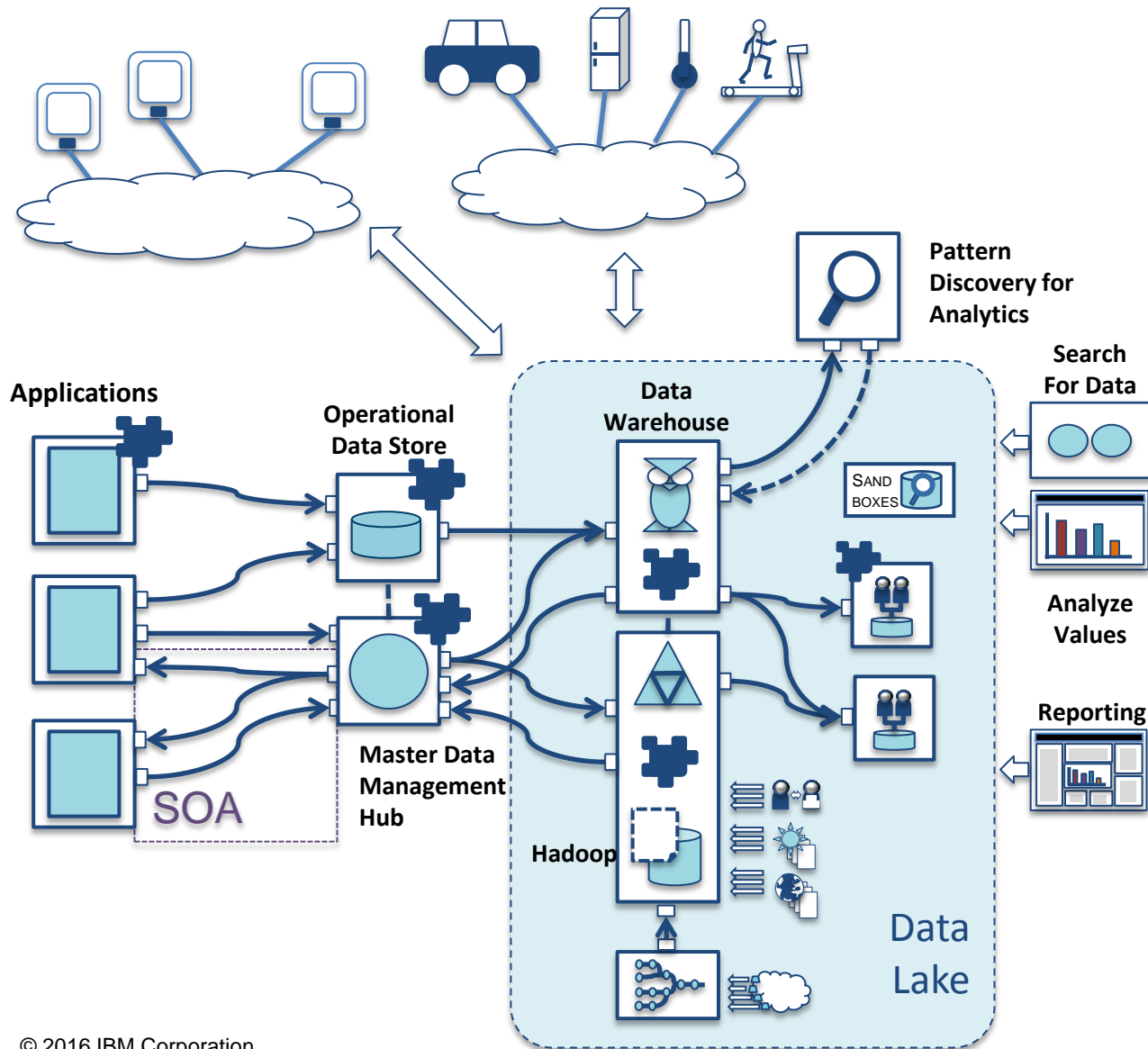
Type: String
Classification: Public



Name	Band	Job Title

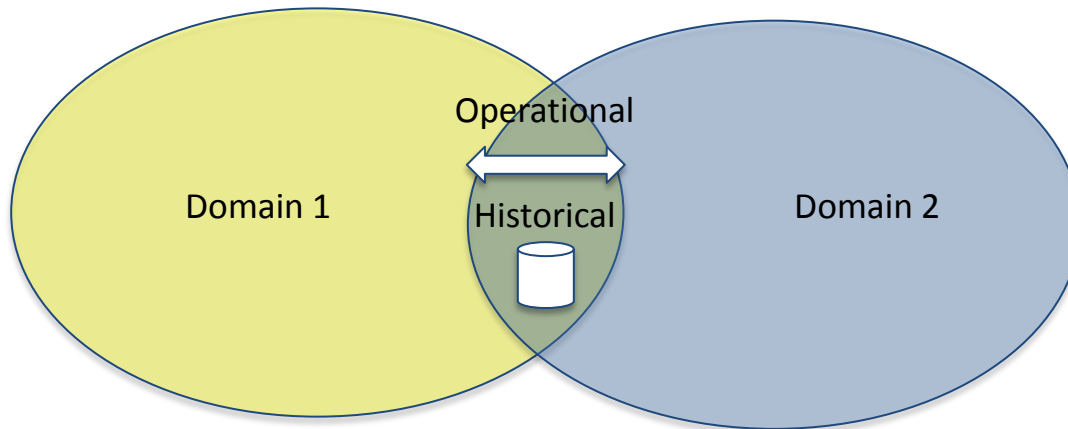


The system landscape of a digital enterprise



Boundary objects and data sets

- Act as the exchange point between different domains.
- Use terminology, reference data and values that both domains understand.
- For data services need an operational API and a historical data set/feed.



https://en.wikipedia.org/wiki/Boundary_object



Summary

- Nothing I have said today is new or earth shattering.
- Information architecture requires clarity of purpose, but beyond that it is straight forward.
 - However, it takes focus and I see so many organizations struggling to get hold of their data because they have skipped this step.
- IOT is a system of systems.
 - We need to think about the standards and mechanisms that allow appropriate abstractions and sharing of the reference data between systems so that both the operational and historical data is available in an understandable and actionable form.





