

Big Data Spatial Analytics at Ordnance Survey with IBM Netezza Spatial



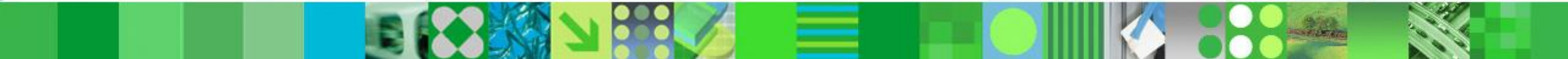
Matt Kadillak – Spatial Architect, IBM Netezza

Neil Taylor – Head of Commercial Markets, Ordnance Survey

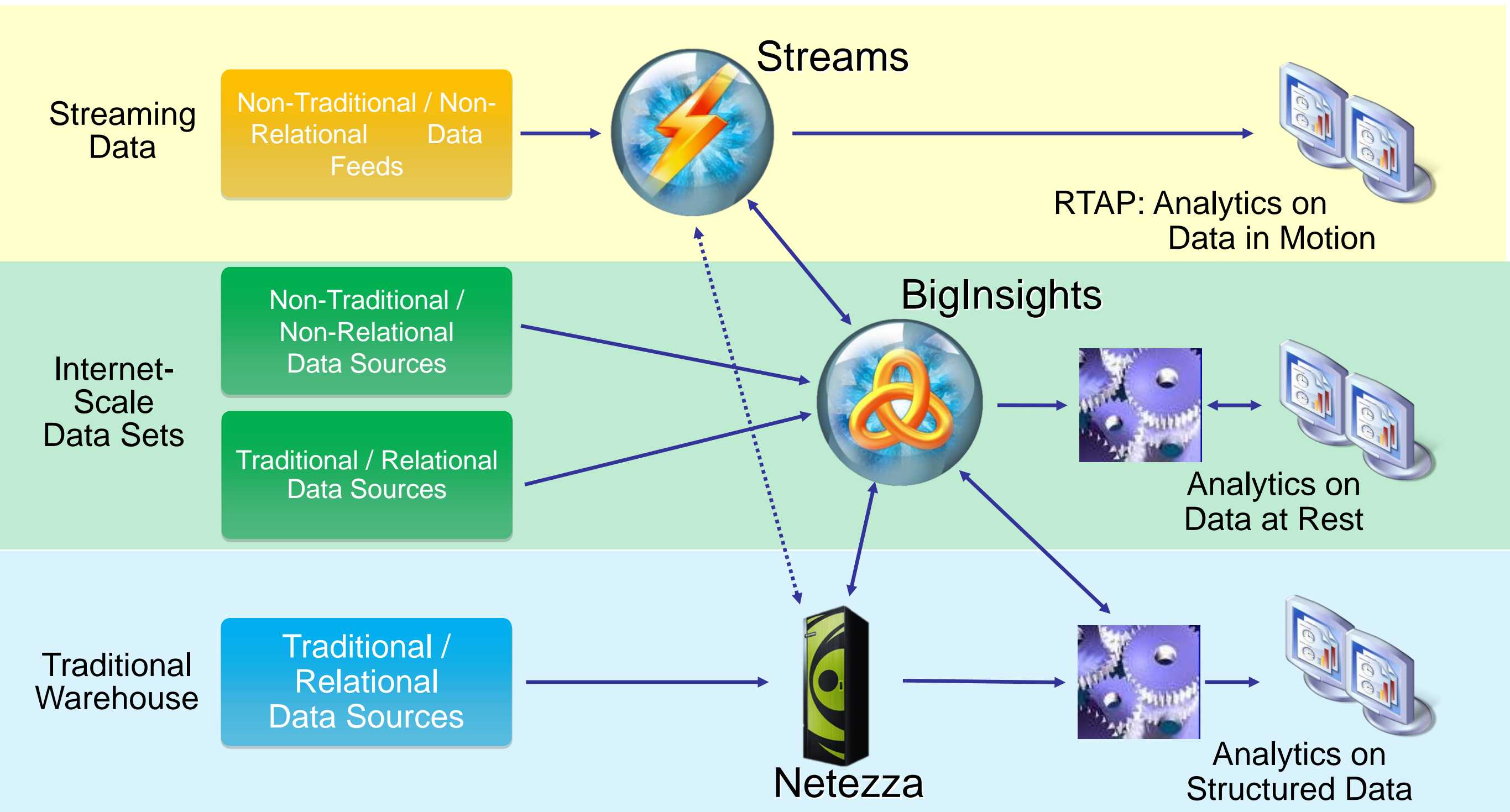
Agenda

- Introductions
- IBM Netezza & Netezza Spatial Overview
- What is Big Spatial Data?
- IBM Netezza and Esri
- Introduction to Ordnance Survey
- Spatial Analytics at Ordnance Survey
- Business Benefits
- Conclusion

IBM Netezza Overview



“Big Data Ecosystem”: Interoperability is Key



The Netezza Performance Server™

Core Value Proposition

- A high-performance, peta-scale data appliance that combines database, server and storage, and integrates easily into existing environments
- Delivers 10-100X the performance at half the cost of existing solutions
- Simplicity → installs in 1.5 days, load n' go architecture, no tuning or configuration



IBM Netezza
1000

Speed – Simplicity – Time to Value

Managing The Netezza Appliance

No indexes and tuning

No storage administration

- No dbspace/tablespace sizing and configuration
- No redo/physical/Logical log sizing and configuration
- No page/block sizing and configuration for tables
- No extent sizing and configuration for tables
- No Temp space allocation and monitoring
- No RAID level decisions for dbspaces
- No logical volume creations of files
- No integration of OS kernel recommendations
- No maintenance of OS recommended patch levels
- No JAD sessions to configure host/network/storage

No software installation

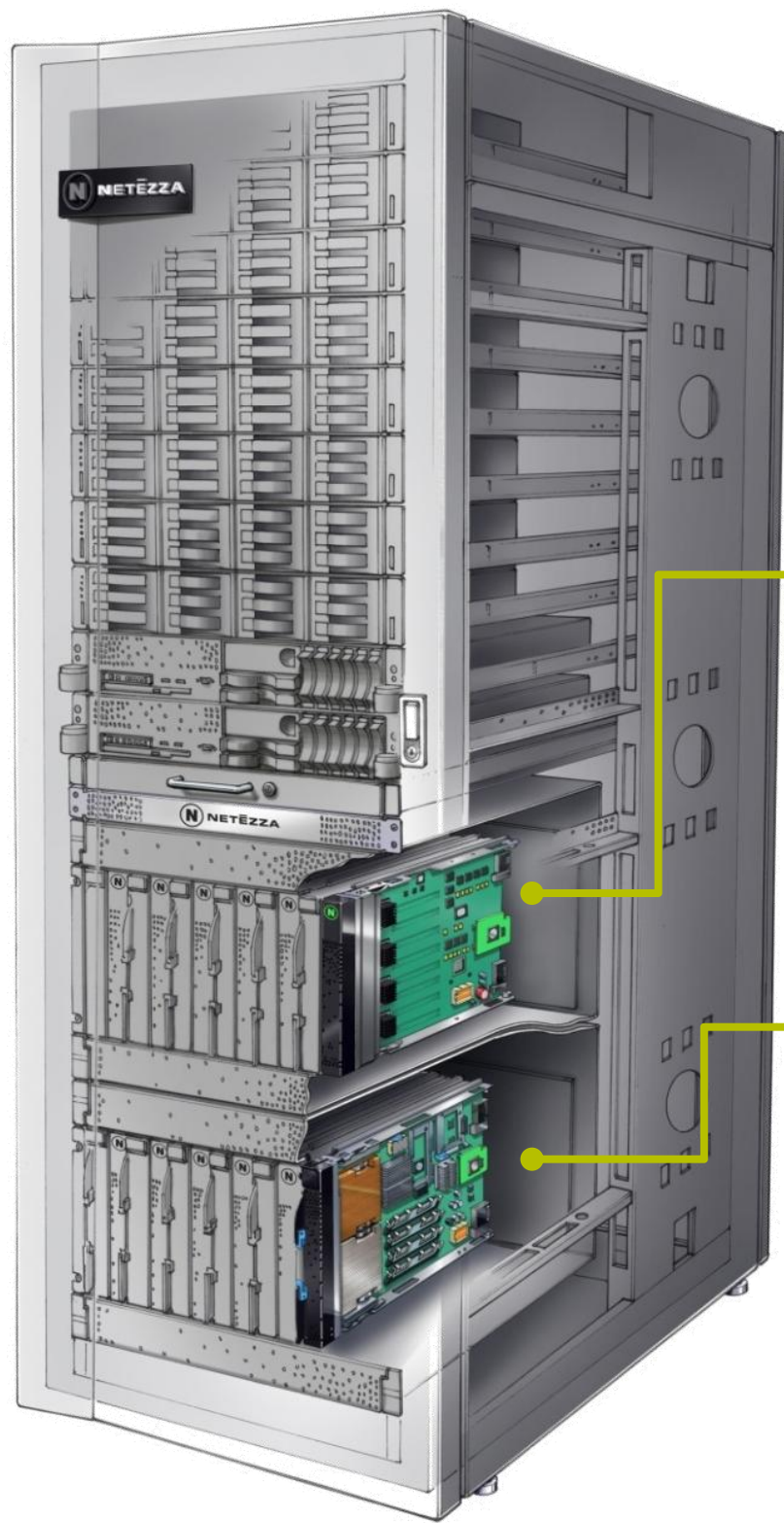
Resources become Data
Managers instead of
Database Administrators

IBM Netezza TwinFin™ Architecture

Inside the IBM Netezza 1000

Optimized Hardware + Software
Purpose-built for high performance analytics; requires no tuning

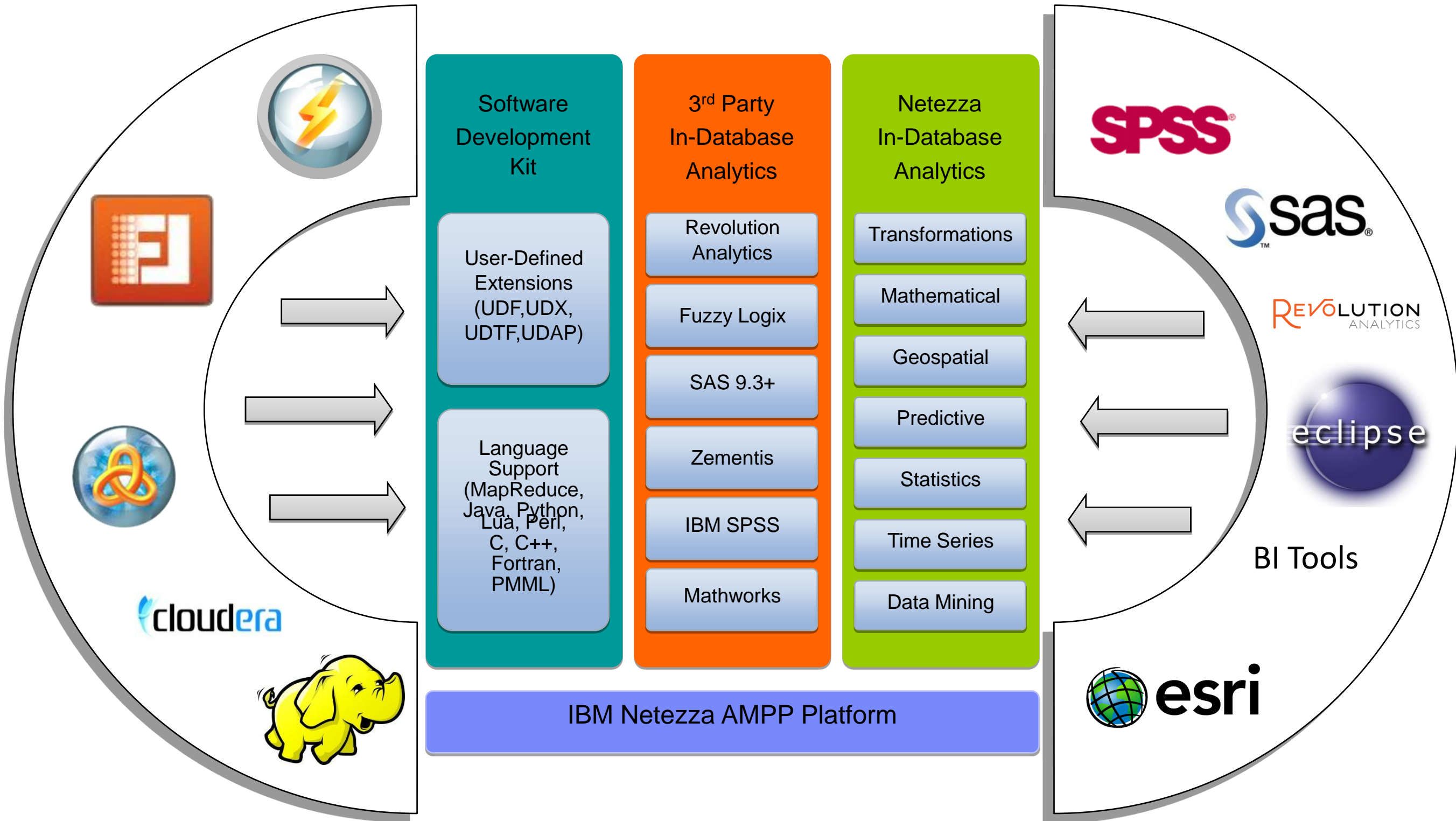
True MPP
All processors fully utilized for maximum speed and efficiency



Streaming Data
Hardware-based query acceleration for blistering fast results

Deep Analytics
Complex analytics executed in-database for deeper insights

IBM Netezza Analytics



Appliance Family for Data Lifecycle Management



Skimmer

TwinFin

High Capacity Appliance

(Netezza 100)

(Netezza 1000)

(Netezza 1000C)

Development & Test System

Data Warehouse Analytics

Queryable Archiving Back-up/DR

1 TB to 10 TB

1 TB to 1.5 PB

100 TB to 10 PB

IBM Netezza Evaluation Approach

- Prove the value at no risk to us, our business, or our timelines.
- In about two weeks, they will:
 - Install a machine in your data center
 - Load all of your data.
 - Integrate with existing applications
 - Train / advise our staff
 - Work with us to scope and deliver analysis to meet our business requirements (no smoke & mirrors)
 - Improve performance by 10-100x
 - Provide complete and full disclosure.
- No pre-tuning of systems or applications for the purposes of the evaluation.
- Open access to the machine throughout the evaluation.
- New queries welcomed
 - An analytical environment should be flexible to meet new business requirements



IBM Netezza Spatial™

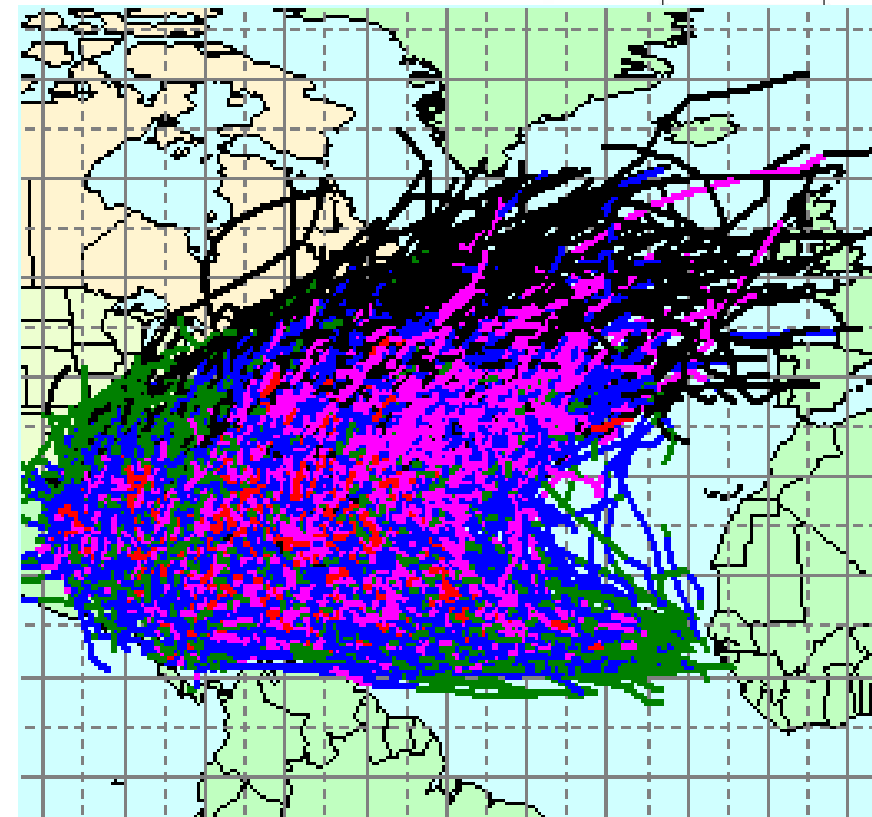
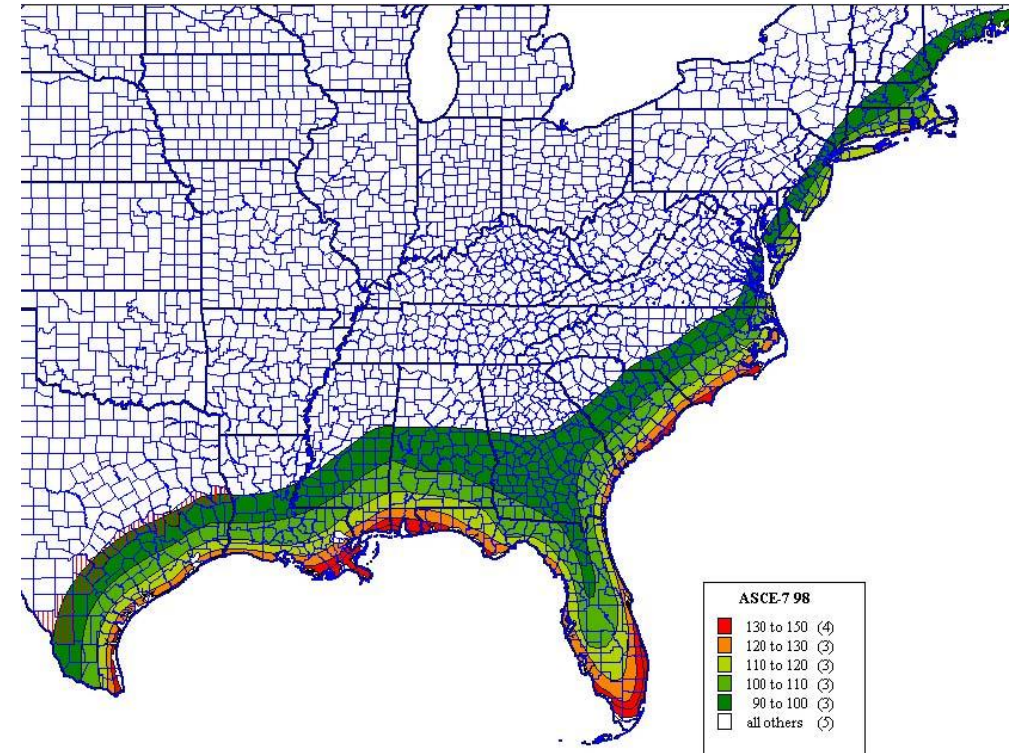
- Standard extension to the market leading NPS® DWA
- Native understanding of location and shape (vector)
- Orders of magnitude performance via parallelism
- Scales linearly
- Open, standards-based interface and data model
 - OGC Simple Features Specification 1.2.x
 - POSTGIS Libraries (GEOS/GGL/Custom)
 - ESRI Libraries – Shape/PE/Custom
- ETL via Plug-In for Safe Software FME (Desktop/Server)
 - High Speed LAT/LONG import via NZ Loader



Analyze all your data all the time

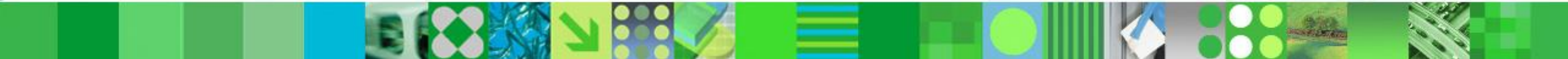
Spatial Analytic Applications - Big Spatial Data

- Insurance Demographics Data/ Risk Analysis
 - 120 million points (+/-) from Demographics providers
 - Coastal distance, elevation extraction (DTED), distance to Firestations, risk accumulation
- UK Government Mapping/Data Provider
 - 460 million features
- Government Intel
 - 400+ Million Points/Polygons
 - Spheroidal Buffer/Distance Calculations/Intersects
- Telco Location Data
 - 187 Billion Call Locations (CDR)
- Telematics/GPS in Rail, Trucking, Troop Movement, etc.
 - 50 Billion Lat/Long Locations
- Agriculture
 - -40-60 millions polygons (soil and land parcel)
- Cable Provider (Marketing)
 - Demographic assignment to market zones every 6-8 weeks
 - 116 million points
- Utilities – Smart Grid



Netezza Spatial scales to these data volumes and more.....

Esri & IBM Netezza Spatial





It's About Place, not Space

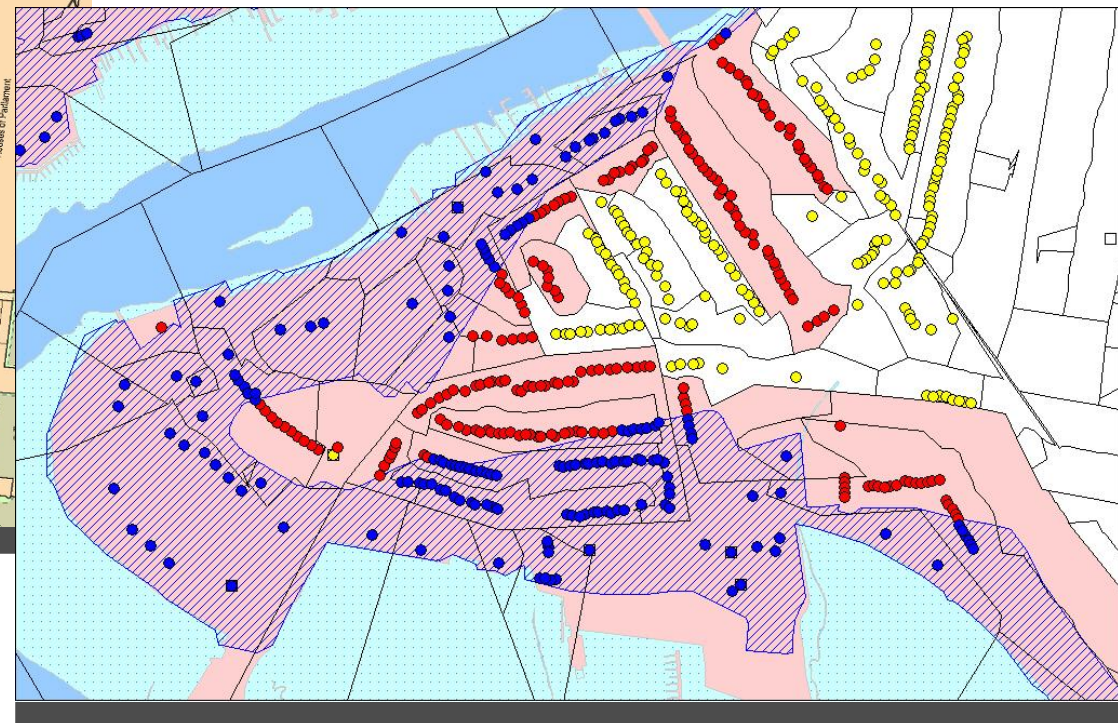
Neil Taylor

Head of Commercial Markets

July 2012

Location

- Everything happens somewhere
- Location is a key driver of decision-making for the public and private sectors
- Over 80% of data has a spatial element



Ordnance Survey Great Britain

- Ordnance Survey is 220 years old
- Civilian organisation since 1983; 1150 staff
- Independent Government Department and Executive Agency reporting directly to a Government Minister
- Turnover of £140m - £30m profit



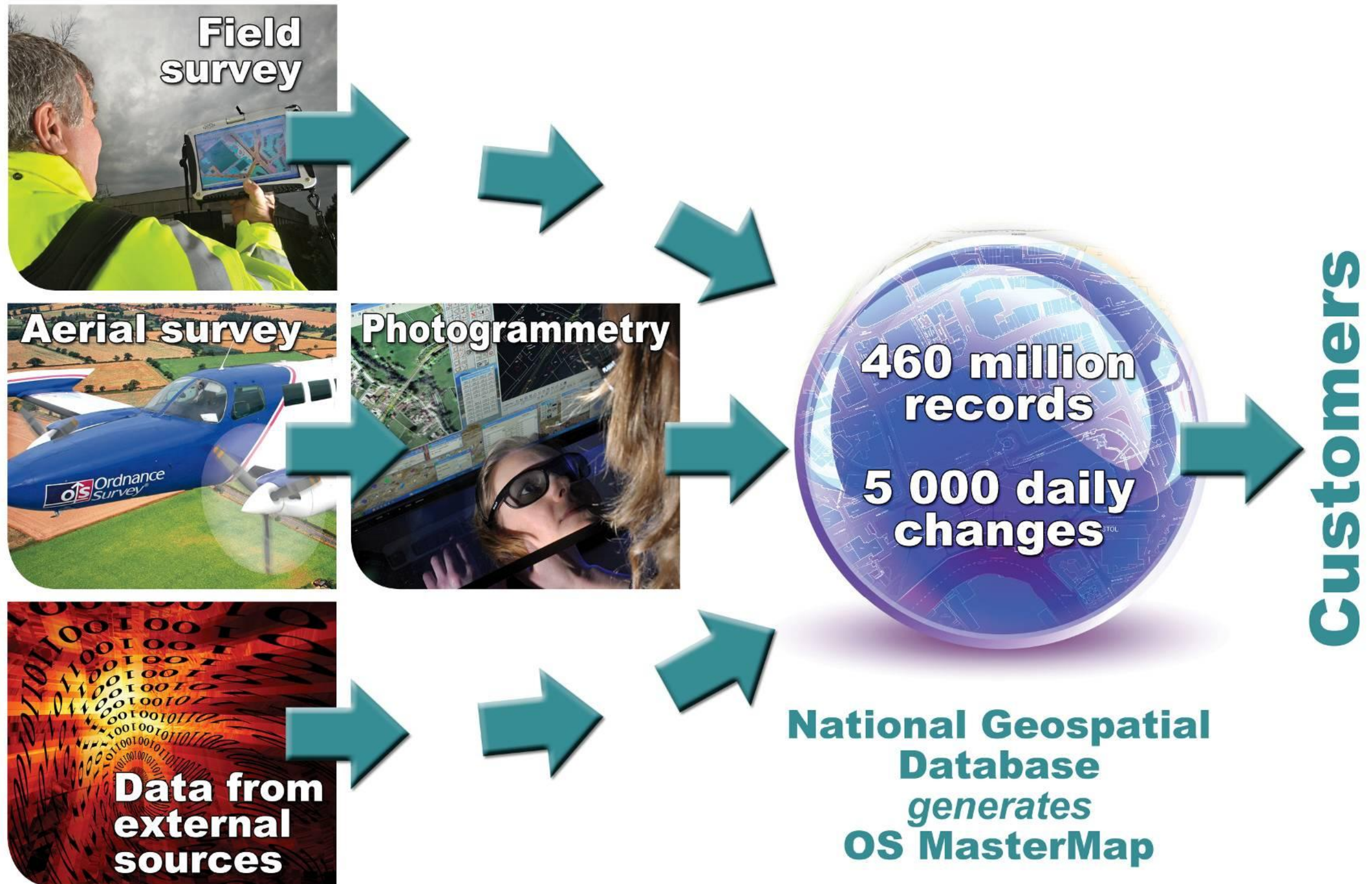
Ordnance Survey today

- Creates and maintains the 'master spatial database' of Great Britain from which others derive benefit
- Manages complete national large scale digital data down to building level detail
- Maintains a database of 460 million features with approximately 5,000 changes made daily
- In 2010/11, 99.9% of real world features were represented in the database within six months of completion on the ground
- From the database, Ordnance Survey produces a range of digital data and paper maps for business, leisure, educational and administrative use

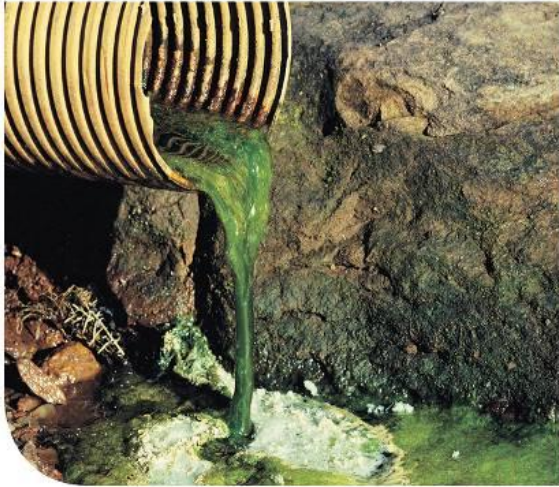
Provides the underpinning geographic framework for Great Britain



Updating the Ordnance Survey database



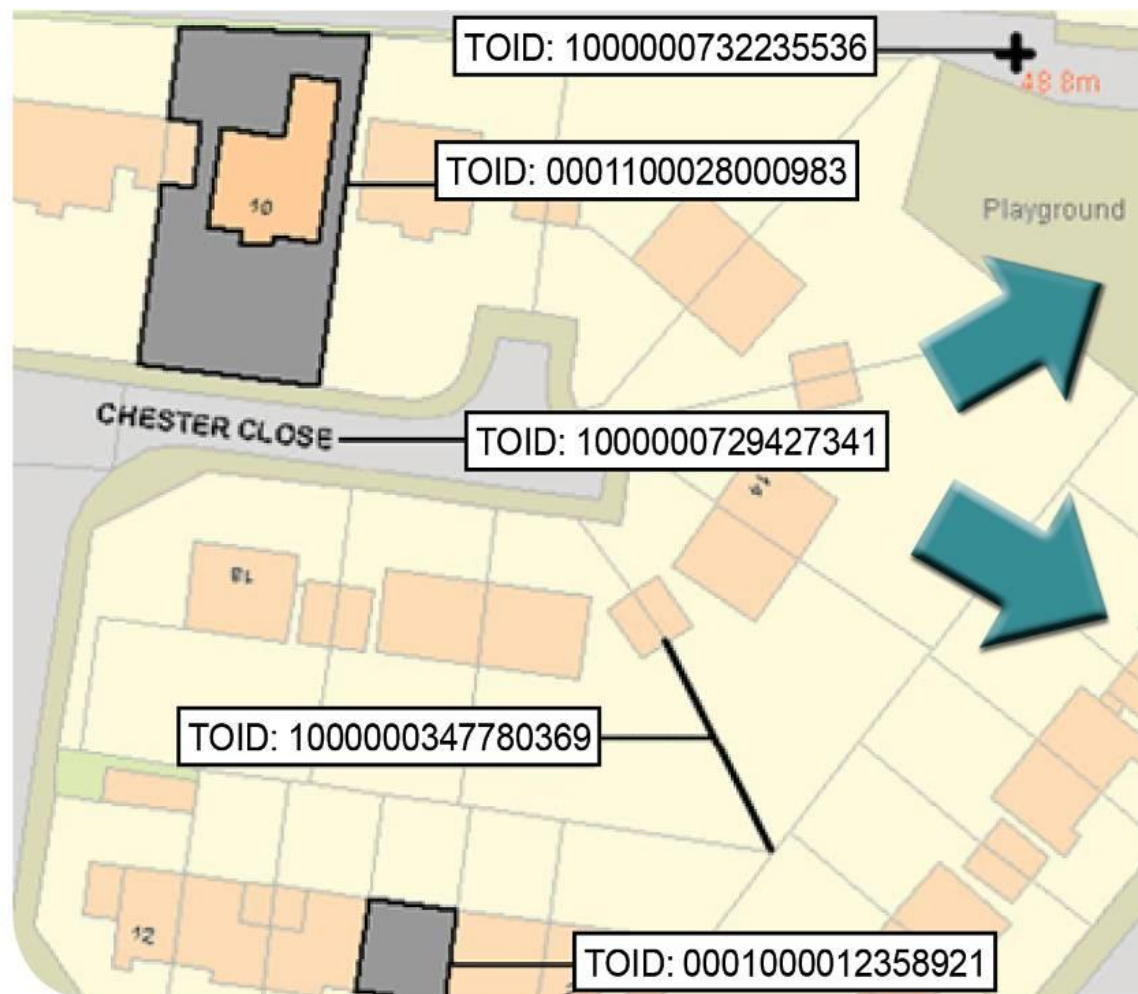
Wide Range of Customers and Markets





A database to connect via real world information

- Every object represented in OS MasterMap has a unique Reference identifier called a TOID. These TOIDs can be used to connect other information and are linked to other core references



Highways database

road_name:

Chester Close

road_type: **cul-de-sac**

lighting: **none**

TOID:
osgb1000000729427341

Crime report database

crime_type:

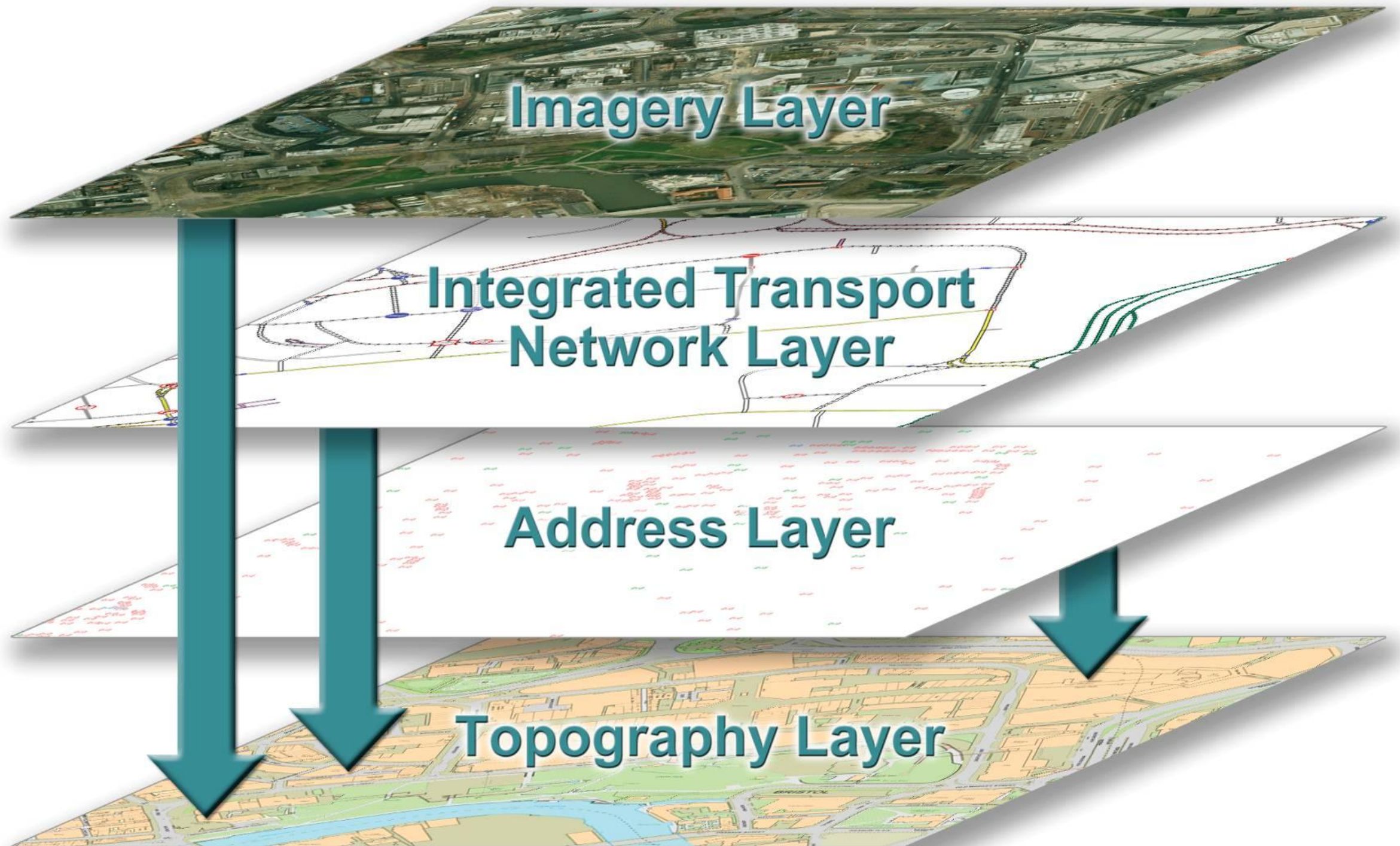
vehicle break-in

date: **09/05/2003**

TOID:
osgb1000000729427341

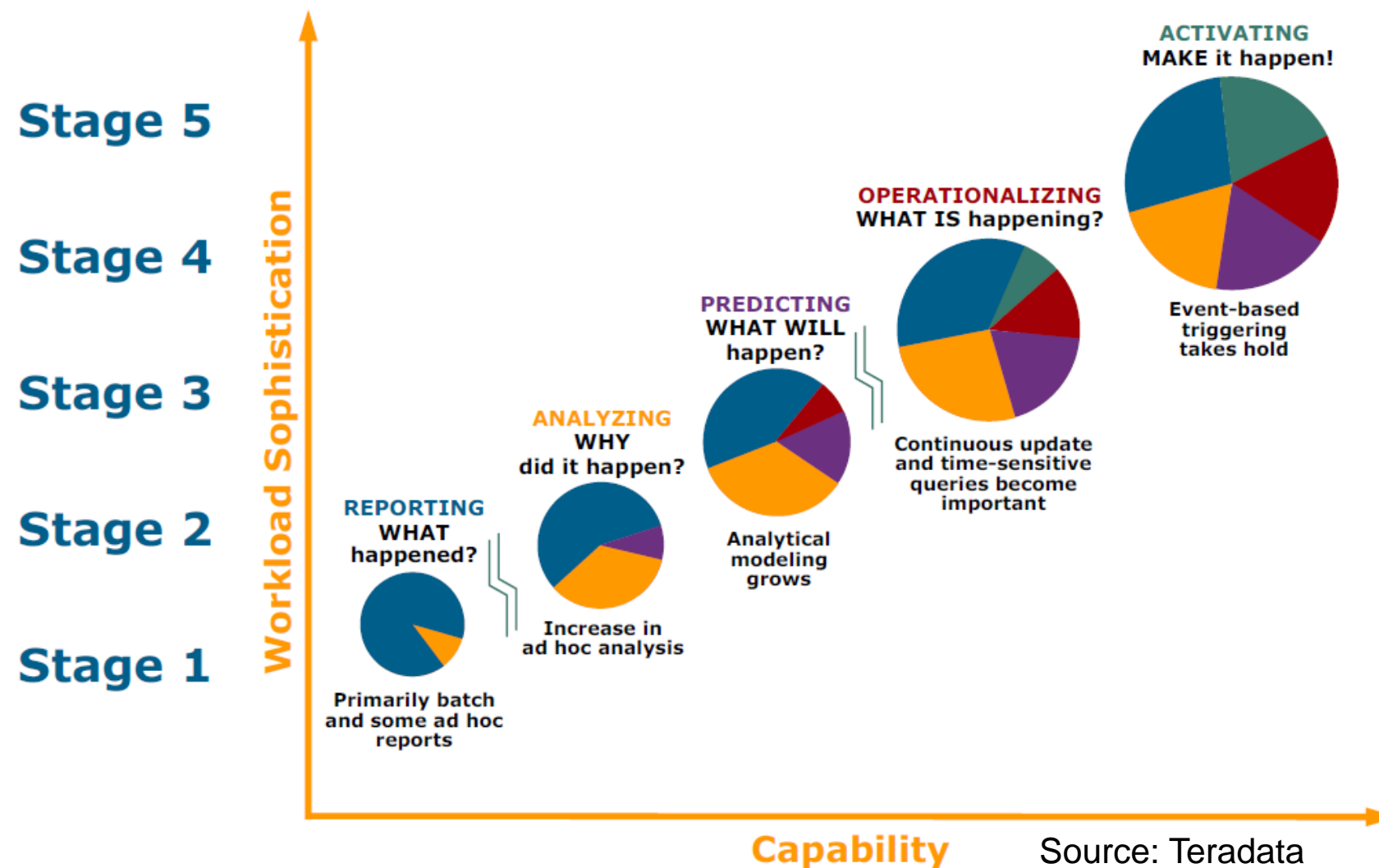
**Highways
and Police
can share
information**

OS MasterMap current layers



Big data and location analytics is changing the world

- Big data gives us the ability to:
 - harness structured and unstructured data
 - ask provocative questions of it





Ordnance Survey and IBM Netezza

What does TAFKAT do for us?

Stress Testing

Evaluation Results – Product Testing

Internal Analytics	Status	Current	Netezza
The Address Layer 2 queries listed in AI_M2_INTERFACE_DATA_SQL.sql (which test for any invalid combinations of attributes in the dataset).	Improve	~3 hours	17 mins
Topology check of OS MasterMap Topography Layer Topo Lines and Topo Areas.	New	N/A	3 mins 3 secs
Address Layer 2 features that are matched to Topography Layer buildings (TOID reference) but do not sit spatially within them.	New	N/A	25 mins
Features within each of the OS MasterMap layers which reference a feature that no longer exists in the product (e.g. Address Layer 2 "Reference to Cartographic Text TOID" to Topography Layer "TOID").	New	N/A	6x queries at ~1 min each
Count of OS MasterMap Topography Layer Topo Lines below 5cm (by length split at 1mm intervals).	New	N/A	37 secs
Number of OS MasterMap Topography Layer Topo Lines with a Descriptive Term of "Inferred Property Closing Links" which don't have a Topo Area with a Descriptive Term of "Multi Surface" on either side of them.	New	N/A	36 hrs

?? ??

Data Queries

?? ??



Evaluation Results – Stakeholder Questions

External Analytics	Status	Current	Netezza
What length of the GB coastline is made up of "beach"?	Improve	~2 weeks	4 mins 26 secs
What is the remotest point (point farthest from a metalled road) in the country?	Improve	~1 day	27 mins 22 secs
What is the farthest point from the coastline?	Improve	~1 day	9 mins 12 secs
What is the number of each Address Layer 2 base function falling within each of the GOR boundaries?	Improve	~2 days	1 min 9 secs
Locate potential high rise blocks of flats (using OS MasterMap Address Layer 2 and Topography Layer).	Improve	~3 days	49 secs

$f(x) = x^2$
 Find the derivative

$\frac{y_1 - y_0}{x_1 - x_0} = \frac{g(x+h) - g(x)}{(x+h) - x} = \frac{g(x+h) - g(x)}{h}$

$f'(x) = \lim_{h \rightarrow 0} \frac{f(x+h) - f(x)}{h}$

$f(x) = \lim_{h \rightarrow 0} \frac{(x+h)^2 - x^2}{h}$

$= \lim_{h \rightarrow 0} \frac{x^2 + 2xh + h^2 - x^2}{h}$

$= \lim_{h \rightarrow 0} \frac{2xh + h^2}{h}$

$= \lim_{h \rightarrow 0} (2x + h)$

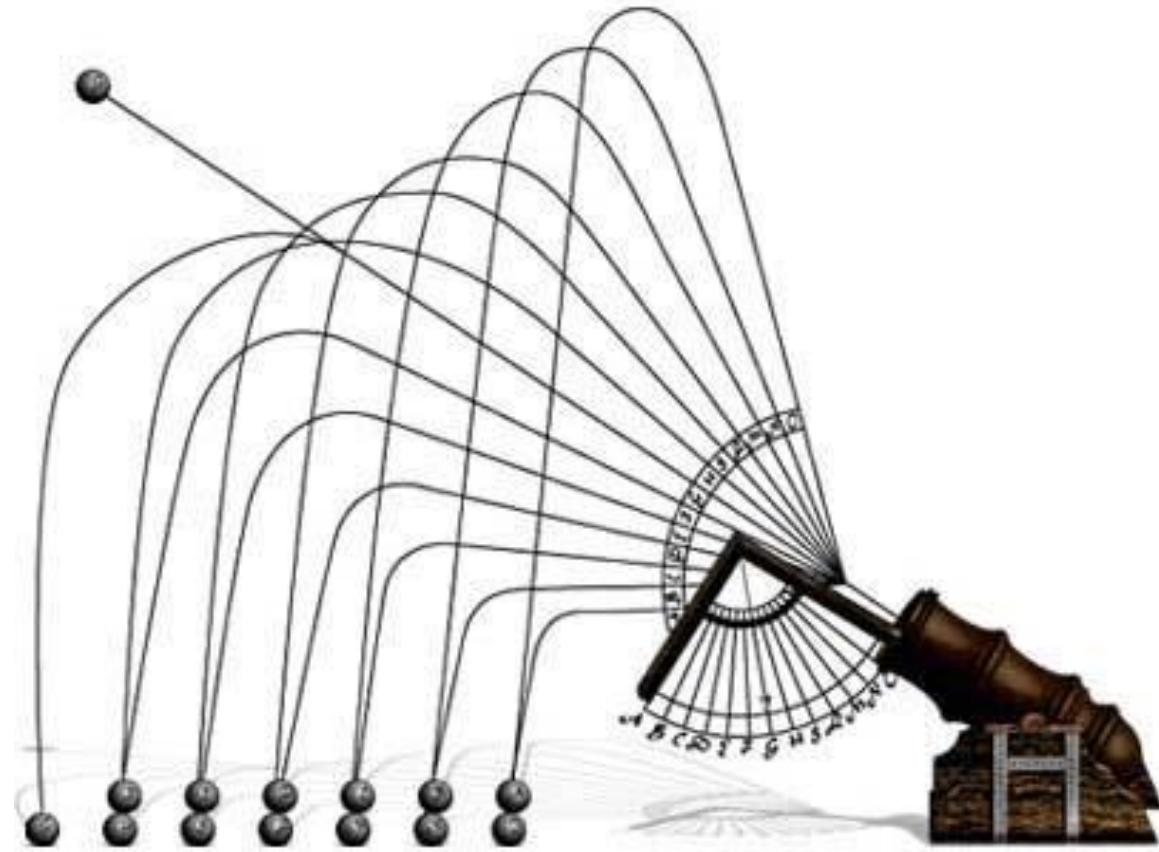
$= 2x$

$f'(x) = 2\sqrt{x}$

Develop Understanding

Evaluation Results – External Analytics

External Analytics	Status	Current	Netezza
Counts the numbers of postcodes [Code-Point]; addresses [Address Layer 2]; and addressable building polygons [Topography Layer] that fall with the Environment Agency flood polygons.	New	N/A	2 hrs 28 mins 24 secs



Customer Prototyping

Adventures with Customer Data



Smarter
Customers



Conclusions

Stress Testing Data Queries

**Develop
Understanding**

**Customer
Prototyping**

Thank you!

