# Leveraging SQL Spatial Analytics for Making Business Decisions

## Rolf Tesmer

**SQL & BI Solution Architect  --  Dimension Data**
**rolf.tesmer@dimensiondata.com**

SQL Saturday #296, Melbourne

5th April 2014

PASS

SQL saturday

#296 | MELBOURNE 2014

# Dimension Data **today**

2013 global revenues of **USD 5.8 billion**

Client-centric, services-focused business

**73%** of Global Fortune 100 and

**59%** of Global Fortune 500

are **Dimension Data clients**

Extensive experience in **emerging markets**

Over **15,000 employees** with operations in **52 countries** across **5 regions**

Over **6,000** clients across all **industry sectors**

# Agenda

- What are **Location Based Services** (LBS) solutions?

- **Technical overview** and **solution architecture**

- Leveraging **SQL Spatial** functions

- data, Data, **DATA**!

- **Reporting** and **analytics**

- Questions

# What are location based services (LBS)?

- A **value add** suite of **mobility services** that leverages Wi-Fi infrastructure

- Can **track device location, movements** and **browsing activity** in **real-time** over maps

- Can **push** interactive **services** and **content / data** back to devices

# What can they do?

1. **Capturing Traffic, Behaviours & Patterns**; opportunities to <u>optimise</u> or <u>monetise</u>
   - *Ex: Positioning long/short term Store location matched to traffic, time and demographics*
   - *Ex: Attach accurate sale value to a location based on traffic direction and time of day*
   - *Ex: Provide rules and alerting for "out of bounds" activities*
   - *Ex: Asset tracking (medical equipment, high risk patients, high value casino chips, etc.)*

2. **Provide "Consumer" Experiences**; personalised location based content & marketing
   - *Ex: Send personalised voucher based on location, demographic, dwell time, etc*
   - *Ex: Activate interactive media station on location, demographic, dwell time, etc*

3. **Provide "Consumer" Services**; "Where is it", "Where am I", "Take me to", eCommerce
   - *Ex: Locate store indicated in voucher, map path to store, track conversions (QR Code)*
   - *Ex: path finding based on capabilities and waypoints*

# Key metrics captured

1. **Primary Key (PK) – The MAC Address**
   - *The "unique key" of every device on the network*
   - *Can also be dissected to identify device manufacturer*

2. **Movement Measures**
   - *Date, Time, Map Location (X Y Z), Dwell Time, Direction of Travel, Speed of Travel, Distance*

3. **Registration Attributes**
   - ***Anything you want*** *– but recommended to build profile <u>slowly</u>*
   - ***Basics****; Name, DoB, Gender, Home Post Code, Email, Phone, Address, Marital Status*
   - ***Nice-to-Have****; Occupation, Education, Income, Associates (ie Partner Registration)*
   - ***Advanced****; Integration of purchase history, Integration of social media feeds*

4. **Aggregation Calculations and Forecasting Capability**
   - *Device Crowding, cumulative dwell times, common travel paths, likely travel paths, frequency*
   - *Calculated "Customer Value" based on business formula*

# But this is just the **Future**… *right*?

- Sydney Airport, **YESTERDAY** (2014-04-04)
  - Connected to free Wi-Fi (**Qantas Free Wi-Fi**) – *as I usually do*
  - Actually scanned the **T&C** this time around…

> 11.2 Everywhere Internet may collect and/or track certain information that is derived from your usage of the Qantas/Everywhere Internet service such as usage patterns, travel patterns and Web site page views and traffic patterns. Qantas/Everywhere Internet may use this tracking information for statistical purposes to improve our products and services and to manage our networks and systems more efficiently. Qantas/Everywhere Internet also records the MAC (Media Access Controller) address from the network card in or used with your computer or other device to enable Qantas/Everywhere Internet to accurately perform our billing functions.

  - *…may **collect and/or track** certain information that is derived from your usage of the service such as **usage patterns**, **travel patterns** and **Web site page views** and **traffic patterns**…*

  - *Qantas/Everywhere Internet also **records** the **MAC (Media Access Controller)**…*

# Agenda

- What are **Location Based Services** (LBS) solutions?

- **Technical overview** and **solution architecture**

- Leveraging **SQL Spatial** functions

- data, Data, **DATA**!

- **Reporting** and **analytics**

- Questions

# What are the solution components?

**Vendor OOTB Deployments**

- Wi-Fi Infrastructure / kit (**Cisco**)
- Mobility Service Appliances (**Cisco MSE**)
- Web Proxy Appliance (**Blue Coat**)
- Registration Appliance / Web Service

**Custom Developed Components**

- **Data Acquisition (ETL)** layer (.NET)
- **Data Warehouse (DW)** layer (SQL Server 2012/14)
- **Analytics & Reporting** layer (SQL Server 2012/14)
- **Device** layer (.NET / HTML5)

# Deployment architecture (macro)



**Networking**
**Switches / Routers / Controllers**

Movement Tracking

Outbound Traffic

Registration Feed

Push & Pull Services / Notifications

**Access Points**

**Cisco MSE**

**Blue Coat Proxy**

**Registration**

API (ETL)

API (ETL)

API (ETL)

Trigger Match Data

**SQL DW**
**SQL Reporting**
**SQL Analytics**

1. Media Content
2. Location Services
3. Notifications / Messages

PASS SQL saturday

#296 | MELBOURNE 2014

# Key SQL Server tables (simplified)

**FactBrowsing**

*Shopper Key*
Map Hierarchy
Browse Date
Browse Time
URL Browsed
Time Taken
Host Server
Port
Bytes Sent
Bytes Received
Referrer
Categories
Method
Content Type
...
...
...

A **row created** for **each** URL browsed while on the network *(anonymous cannot browse internet)*

**DimShopper**

*Shopper Key*
MAC Address
Name
Mobile Contacts
Email Contacts
Address Contacts
Device Details
Demographics
...
...
...

A **row** for **each** registered shopper / network user *(incl anonymous)*

**FactMovement**

Movement Date
Movement Time
*Shopper Key*
Map Hierarchy
Device Location
Dwell Time

A "**hot**" **row** for **each** device's **current** location that **is** on the network **RIGHT NOW** *(incl anonymous)*

**FactMovementHistory**

Movement Date
Movement Time
*Shopper Key*
Map Hierarchy
Device Location
Dwell Time
Depart Speed
Depart Direction
Depart Distance

A **row** for **each** device's **historical** location that **was** on the network *(incl anonymous)*

PASS SQL saturday

# Managing large database tables (simplified)
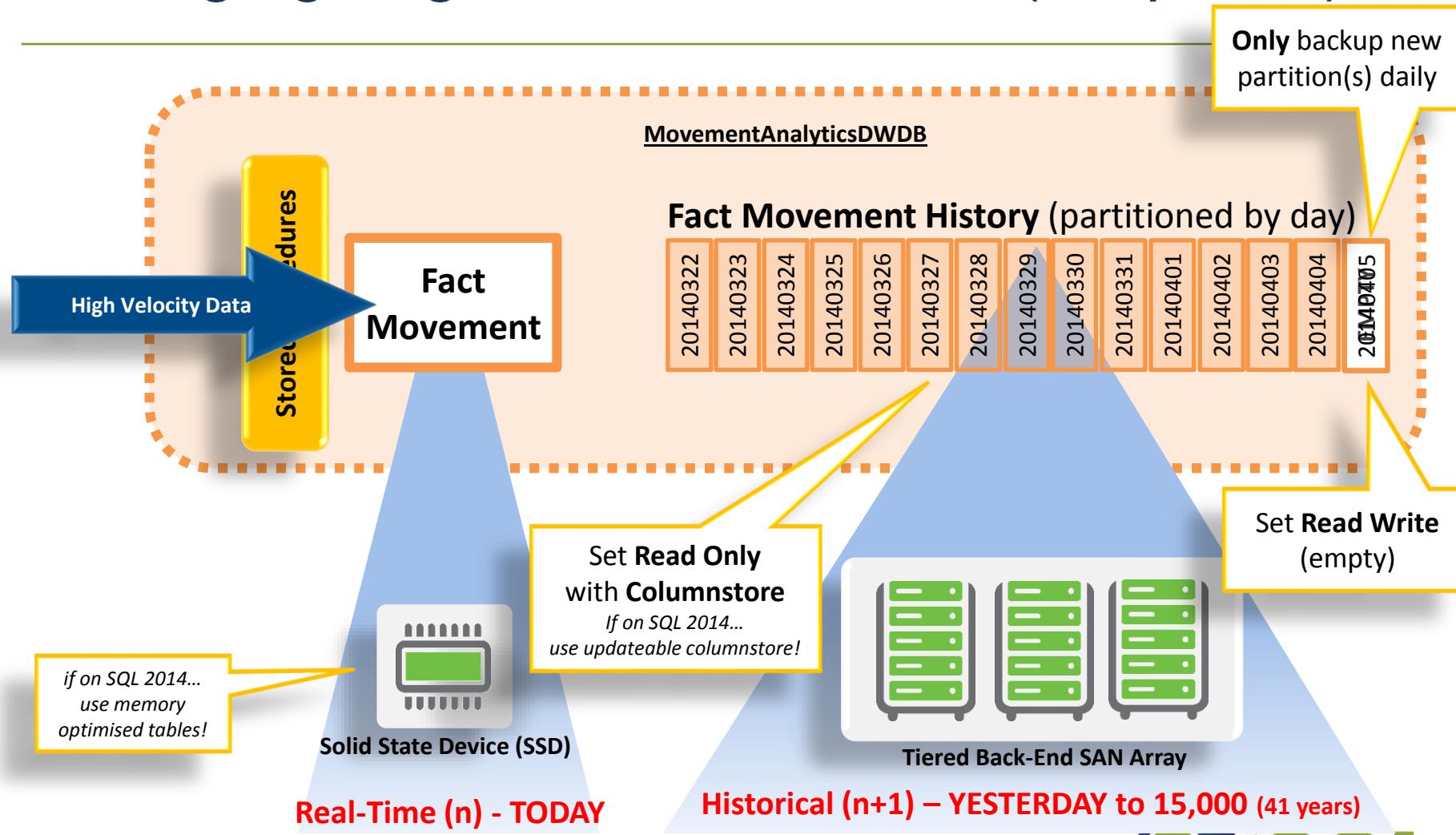
**MovementAnalyticsDWDB**

**Fact Movement History** (partitioned by day)

| 20140322 | 20140323 | 20140324 | 20140325 | 20140326 | 20140327 | 20140328 | 20140329 | 20140330 | 20140331 | 20140401 | 20140402 | 20140403 | 20140404 | 20140405 |

**Stored Procedures**

High Velocity Data →

**Fact Movement**

**Only** backup new partition(s) daily

Set **Read Write** (empty)

Set **Read Only** with **Columnstore**
*If on SQL 2014… use updateable columnstore!*

*if on SQL 2014… use memory optimised tables!*

**Solid State Device (SSD)**

**Tiered Back-End SAN Array**

**Real-Time (n) - TODAY**

**Historical (n+1) – YESTERDAY to 15,000 (41 years)**

# Agenda

- What are **Location Based Services** (LBS) solutions?

- **Technical overview** and **solution architecture**

- Leveraging **SQL Spatial** functions

- data, Data, **DATA**!

- **Reporting** and **analytics**

- Questions

- Calculating and tracking device **Movements**
  - Need **2 data points** – current location (**FactMovement**) & last location (**FactMovementHistory**)
  - Data points need to be related to the same device / movement pattern to make sense

- Calculating **Distance**
  - Can be done via geometry **STDistance** function or **mathematically** (*Pythagoras* **Theorem**)
  - $A^2 + B^2 = C^2$

  > **Not** the Geography datatype
  > Not on a **global scale!**

- Calculating **Speed**
  - **No geometry functions** to do this, has to be done **mathematically**
  - Speed (m/sec) = Distance (m) / Time (sec)

- Calculating **Direction**
  - **No geometry functions** to do this, has to be done **mathematically**
  - Need to use *Trigonometric* **functions**

- **DEMO: Calculating Movement Measures**
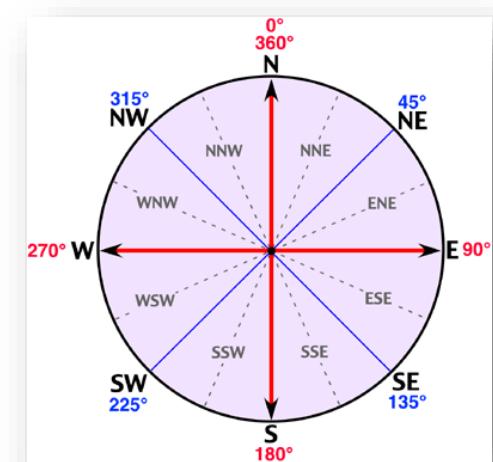
- ## Calculating device **Proximity**
    - Need **1 data point** – current or historical location (**FactMovement**) / (**FactMovementHistory**)
    - Need something to search against – such as a list of stores, other devices, etc

- ## Drawing a **Radius**
    - Can be done via <u>geometry</u> using **STBuffer** and **STContains** or **STIntersects** functions
    - Can probably be done **mathematically** – *but why bother*?!

- ## Calculating **Nearest** and **Furthest Elements**
    - Can be done via <u>geometry</u> using **STDistance** function
    - Can be done **mathematically** – *but again why bother*?!

- ## **DEMO: Calculating Proximity**
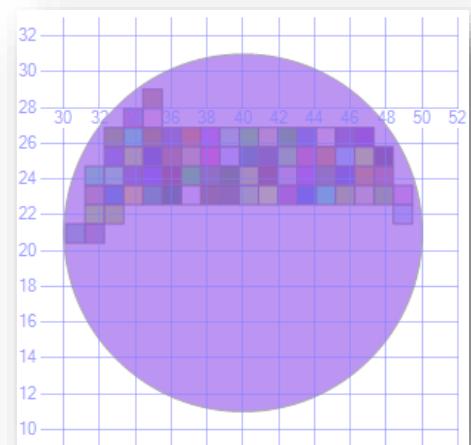
- Calculating device **Aggregations**
  - Need *LOTS* of **data points** – current or historical (**FactMovement**) / (**FactMovementHistory**)
  - Need something to reference against – such as a list of stores, other devices, etc

- Calculating a **Dwell Time Aggregations**
  - Can be done via **simple** SQL aggregation functions (avg, max, min, stdev, etc)
  - Overlay with maps to **visualise** where these aggregations **occur**

- Calculating a **Crowding Index Aggregations**
  - Can be done via simple SQL aggregation functions (**count** & **distinct count**)
  - Overlay with maps to **visualise** where these aggregations **occur**

- **DEMO: Calculating Aggregations**
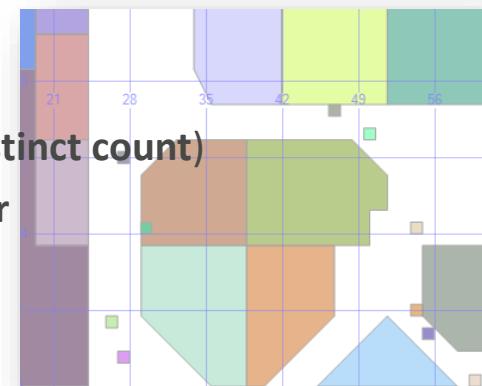
- **Route Finding Algorithm – A***
  - **Widely used** algorithm for **graph traversal**
  - The process of **plotting** an **efficiently traversable** path **between points**

> **NOT NEW**!
> First described in **1968**

- Finds a **least-cost path** from <u>an</u> **initial node** to <u>one</u> **goal node**
- Follows a path of the **lowest expected total cost** or **distance**
- Takes **distance already travelled** into account

- "**Complete**" algorithm – it **will** find a solution **if one exists!**

- **DEMO: A* Route Finding .Net Application**
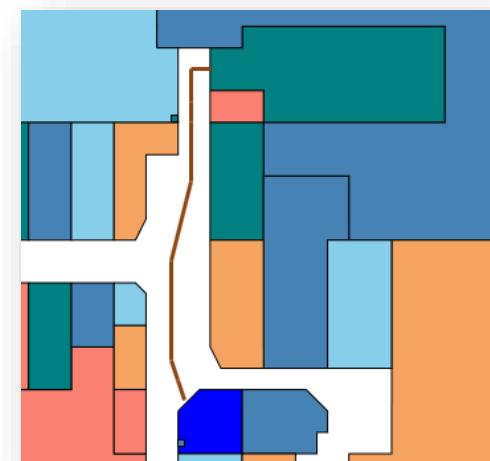
# Agenda

- What are **Location Based Services** (LBS) solutions?

- **Technical overview** and **solution architecture**

- Leveraging **SQL Spatial** functions

- data, Data, **DATA**!

- **Reporting** and **analytics**

- Questions

# So just how **BIG** could these solutions get?

Let's size up this **EXAMPLE** scenario…

- **Very Large Unnamed Victorian Retail Shopping Centre**
- Open **9 hours/day**, **7 days/week**
- **68,000** shoppers (@ peak hour & day)

- *Assume…* Peak of **3 hours/day** (6 hours @ **33% of peak load**)
- *Assume…* **40%** have a "trackable" device (**27,200**)
- *Assume…* **10** URL's browsed an **hour/device**
- *Assume…* **5%** YOY shopper growth rate

**(Q)** So what happens next?

# Data volumes and growth over time

| Year 1 | Year 2 | Year 3 | Year 4 | Year 5 | Year 6 | Year 7 | ... |

**39,058,737,600 rows**

or

**1.54 TB**

**289,311,991,785 rows**

or

**11.39 TB**

EXPANSION

**2,603,807,926,068 rows**

or

**103 TB**

**May need to consider specialised SQL solutions**
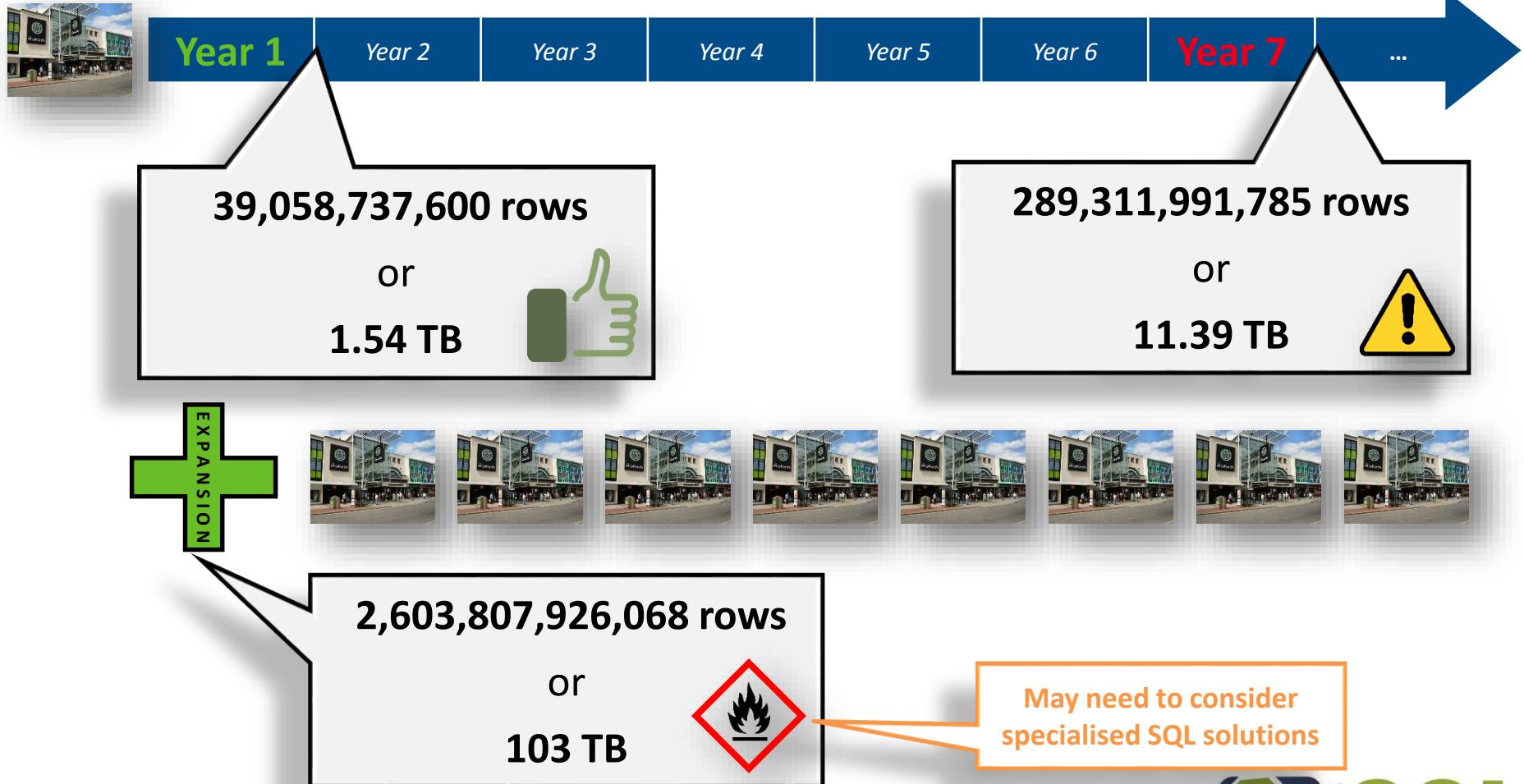
# So then, is this a **BIG DATA** problem?

The industry **marketing definition** of **BIG DATA** is reasonably specific…

1. [**V**]elocity? **YES**! @ **245,000 data rows/second**

2. [**V**]olume? **YES**? @ **100+ TB of data**

3. [**V**]ariety? **NO**! **Data is well defined & structured!**

**(Q)** So how do we manage all this data?

# SQL Server can do it of course!
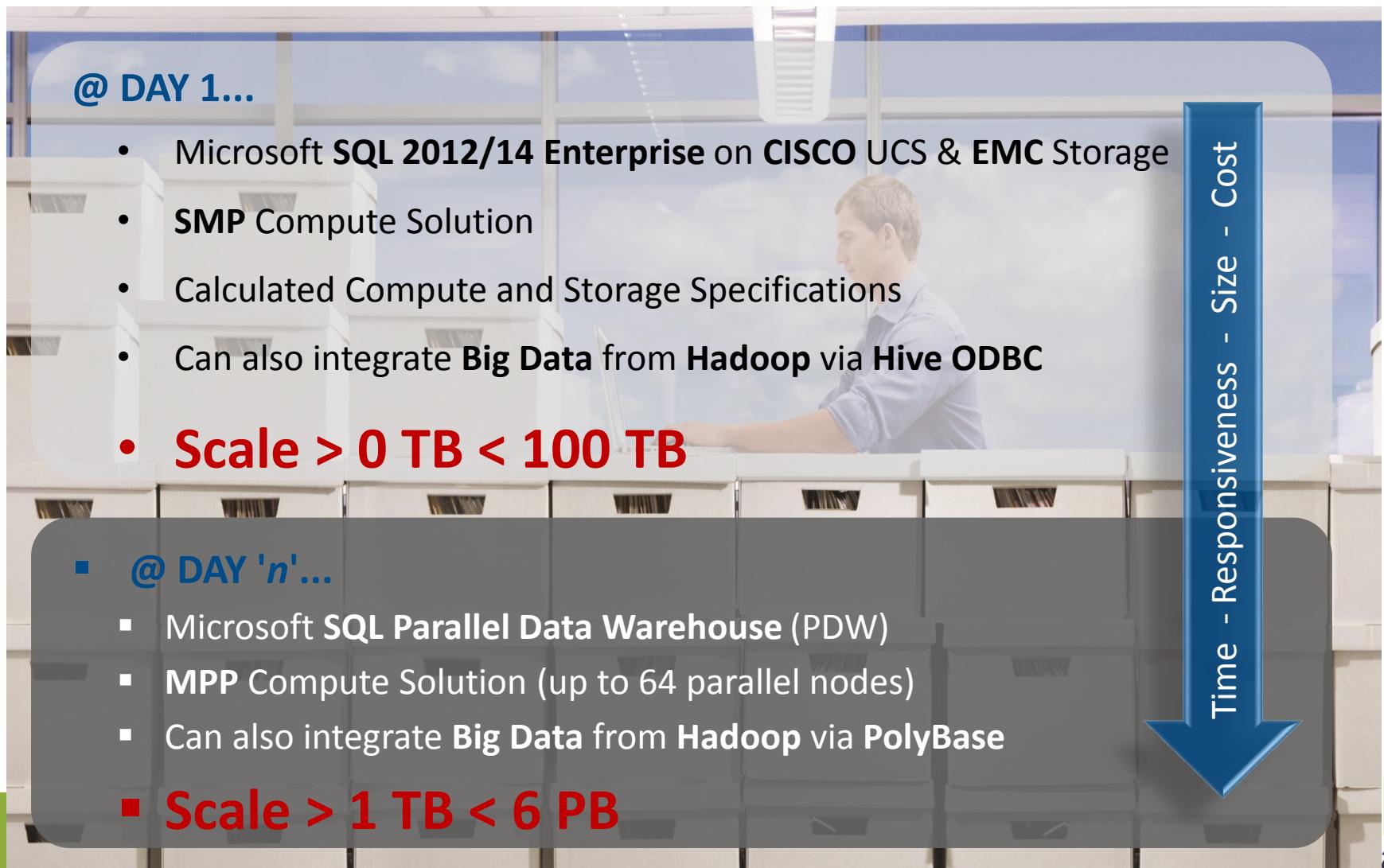
**@ DAY 1...**

- Microsoft **SQL 2012/14 Enterprise** on **CISCO** UCS & **EMC** Storage

- **SMP** Compute Solution

- Calculated Compute and Storage Specifications

- Can also integrate **Big Data** from **Hadoop** via **Hive ODBC**

- ## Scale > 0 TB < 100 TB

**@ DAY '*n*'...**

- Microsoft **SQL Parallel Data Warehouse** (PDW)
- **MPP** Compute Solution (up to 64 parallel nodes)
- Can also integrate **Big Data** from **Hadoop** via **PolyBase**

- ## Scale > 1 TB < 6 PB

Time - Responsiveness - Size - Cost

# DEMO – Data Sizing Model



## SQL DW DATA SIZING MODEL PARAMETERS

| Parameter | Value | Scale | Cals 1 | Desc 1 | Cals 2 | Desc 2 | Comments and Description |
|---|---|---|---|---|---|---|---|
| # locations | 9 | centres | | | | | Australia only |
| Peak # devices/location | 27,200 | peak devices | 244,800 | burst per/sec | 11.44 | burst MB | number of devices actively on a single centre wifi network (peak load) |
| Yearly Growth % on Devices | 5% | YoY percent | 28,560 | yr 1 Size | | | the % growth year on year of the number of devices on the WiFi network |
| MSE polling Interval (Sec) | 5 | sec | 12.00 | polls/min | | | number of seconds interval which the MSE API is called to get device location |
| NCS polling Interval (Sec) | 60 | sec | 1.00 | polls/min | | | number of seconds interval which the NCS API is called to get device network stats |
| # Movements/poll/device | 1 | movements/poll | | | | | number of device movements for each on every MSE poll |
| Total Hours/Day (Centres Open) | 9 | hours per day | | | | | total number of hours the location is available (avg) |
| Total Hours/Day (Peak Usage) | 3 | hours per day | 6 | hrs off-peak | | | total number of hours the location is at peak # of Devices |
| % Devices Off-Peak | 33% | percentage | 8,976 | off-peak devices | | | the % of total # Location Devices when centre is in off-peak |
| Days/Year (Location is available) | 365 | days | | | | | number of days per year the centre is open (avg) |
| Data retention period (Years) | 7 | years | | | | | number of years the data is retained in the DW |
| URL browsed/device/hour | 10 | per hour | 680.00 | burst per/sec | 332.03 | burst KB | number of URL's browser per device per session time at the location (estimated) |
| DB index % allowance | 60% | percentage | | | | | the % data size increase due to addition of indexes for each row |
| DB data compression % allowance | 50% | percentage | | | | | the % data size reduction due to page compression |
| DB row size - Movement (Bytes) | 49 | bytes | 39.20 | bytes | 0.03828 | kb | DB Movement Size (KB)+60% index allowance - 50% compression allowance |
| DB row size - Browsing (Bytes) | 500 | bytes | 400.00 | bytes | 0.48828 | kb | DB Browsing Size (KB)+60% index allowance - 50% compression allowance |
| DB row size - Network (Bytes) | 16 | bytes | 12.80 | bytes | 0.01563 | kb | DB Net Stats Size (KB)+60% index allowance - 50% compression allowance |
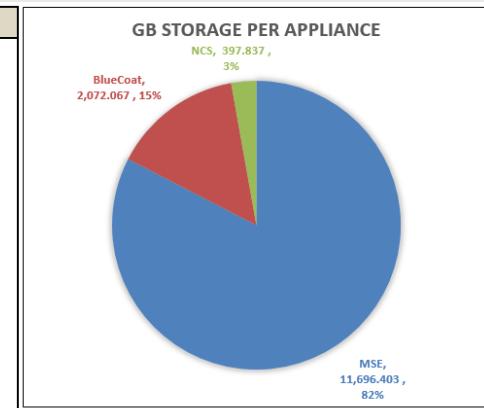
## TOTAL: ALL DATA & ALL LOCATIONS

| Measure | SINGLE LOCATION Rows | KB | GB | TB | Measure | 9 LOCATIONS Rows | KB | GB | TB |
|---|---|---|---|---|---|---|---|---|---|
| FACT /Sec | 3,303 | 140 | 0.000 | 0.000 | FACT /Sec | 29,725 | 1,256 | 0.001 | 0.000 |
| FACT /Min | 198,167 | 8,374 | 0.008 | 0.000 | FACT /Min | 1,783,504 | 75,365 | 0.072 | 0.000 |
| FACT /Hour | 11,890,027 | 502,434 | 0.479 | 0.000 | FACT /Hour | 107,010,240 | 4,521,902 | 4.312 | 0.004 |
| FACT /Day | 107,010,240 | 4,521,902 | 4.312 | 0.004 | FACT /Day | 963,092,160 | 40,697,120 | 38.812 | 0.038 |
| FACT /Week | 749,071,680 | 31,653,316 | 30.187 | 0.029 | FACT /Week | 6,741,645,120 | 284,879,842 | 271.683 | 0.265 |
| FACT /Month | 3,254,894,800 | 137,541,193 | 131.170 | 0.128 | FACT /Month | 29,294,053,200 | 1,237,870,741 | 1,180.526 | 1.153 |
| FACT /Year | 39,058,737,600 | 1,650,494,321 | 1,574.034 | 1.537 | FACT /Year | 351,528,638,400 | 14,854,448,891 | 14,166.306 | 13.834 |
| FACT /7 Years | 289,311,991,785 | 12,225,377,185 | 11,659.028 | 11.386 | FACT /7 Years | 2,603,807,926,068 | 110,028,394,661 | 104,931.254 | 102.472 |

### GB STORAGE PER APPLIANCE

NCS, 397.837 , 3%

BlueCoat, 2,072.067 , 15%

MSE, 11,696.403 , 82%

# Agenda

- What are **Location Based Services** (LBS) solutions?

- **Technical overview** and **solution architecture**

- Leveraging **SQL Spatial** functions

- data, Data, **DATA**!

- **Reporting** and **analytics**

- Questions

# DEMO – Analytics and Reporting
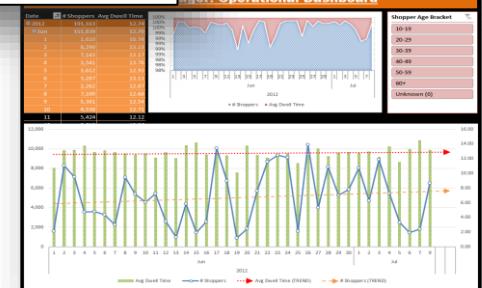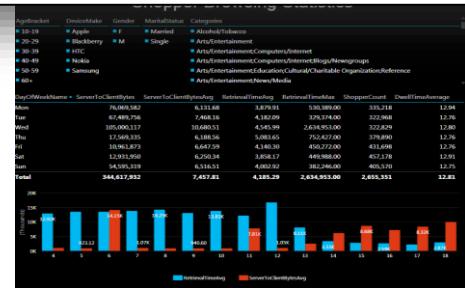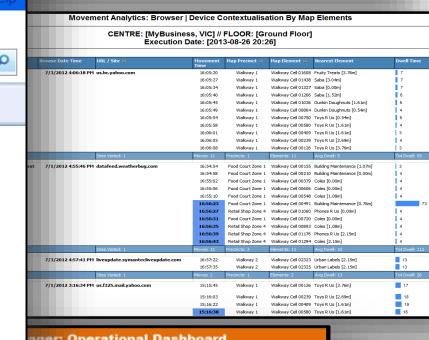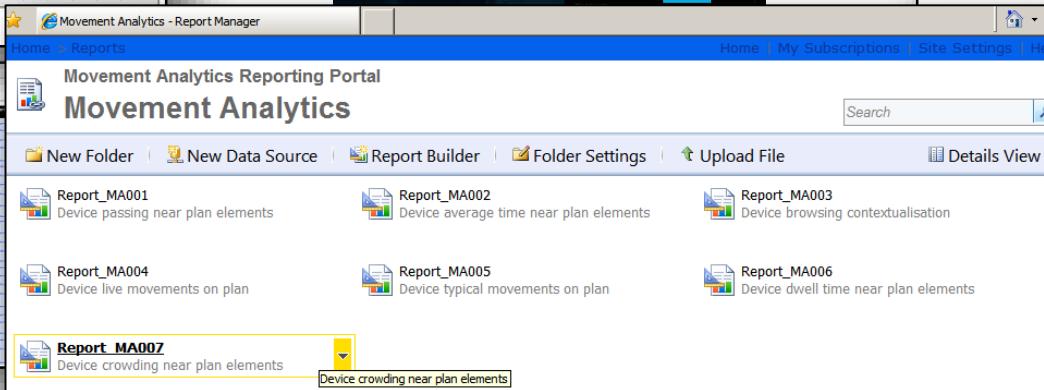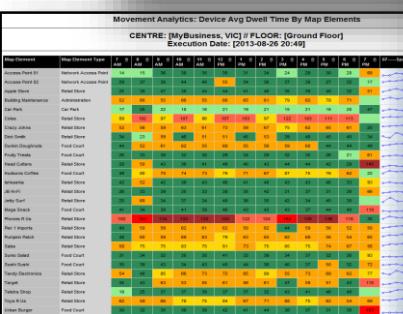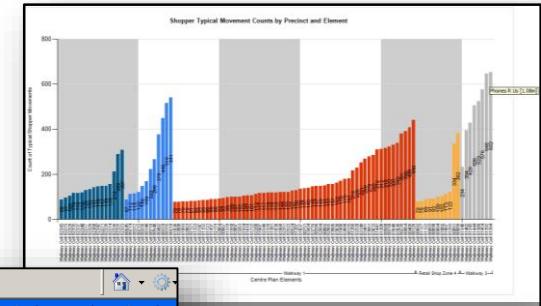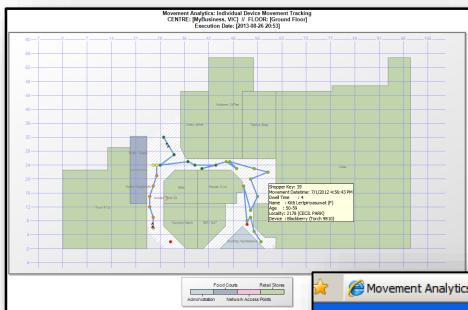
# SSRS – Browser Contextualisation



**Movement Analytics: Browser | Device Contextualisation By Map Elements**

**CENTRE: [MyBusiness, VIC] // FLOOR: [Ground Floor]**
**Execution Date: [2013-08-26 20:26]**

| Shopper | Browse Date Time | URL / Site | Movement Time | Map Precinct | Map Element | Nearest Element | Dwell Time |
|---|---|---|---|---|---|---|---|
| **Brenda Diaz** | 7/3/2012 4:06:18 PM | us.bc.yahoo.com | 16:05:20 | Walkway 1 | Walkway Cell 01608 | Fruity Treats [3.76m] | 7 |
| | | | 16:05:27 | Walkway 1 | Walkway Cell 01438 | Saba [3.04m] | 7 |
| | | | 16:05:34 | Walkway 1 | Walkway Cell 01327 | Saba [0.00m] | 7 |
| | | | 16:05:40 | Walkway 1 | Walkway Cell 01266 | Saba [1.52m] | 6 |
| | | | 16:05:45 | Walkway 1 | Walkway Cell 01036 | Dunkin Doughnuts [1.61m] | 5 |
| | | | 16:05:49 | Walkway 1 | Walkway Cell 00864 | Dunkin Doughnuts [0.54m] | 4 |
| | | | 16:05:54 | Walkway 1 | Walkway Cell 00750 | Toys R Us [0.54m] | 5 |
| | | | 16:05:58 | Walkway 1 | Walkway Cell 00580 | Toys R Us [1.61m] | 4 |
| | | | 16:06:01 | Walkway 1 | Walkway Cell 00409 | Toys R Us [1.61m] | 3 |
| | | | 16:06:05 | Walkway 1 | Walkway Cell 00239 | Toys R Us [2.69m] | 4 |
| | | | 16:06:08 | Walkway 1 | Walkway Cell 00126 | Toys R Us [3.76m] | 3 |
| Totals (Shopper: 16) | | Sites Visited: 1 | Moves: 11 | Precincts: 1 | Elements: 11 | Avg Dwell: 5 | Tot Dwell: 55 |
| **Kitti Lertpiriyasuwat** | 7/1/2012 4:55:46 PM | datafeed.weatherbug.com | 16:54:54 | Food Court Zone 1 | Walkway Cell 00155 | Building Maintenance [1.07m] | 3 |
| | | | 16:54:58 | Food Court Zone 1 | Walkway Cell 00210 | Building Maintenance [0.00m] | 4 |
| | | | 16:55:02 | Food Court Zone 1 | Walkway Cell 00379 | Coles [0.00m] | 4 |
| | | | 16:55:06 | Food Court Zone 1 | Walkway Cell 00606 | Coles [0.00m] | 4 |
| | | | 16:55:10 | Food Court Zone 1 | Walkway Cell 00548 | Coles [1.08m] | 4 |
| | | | 16:56:23 | Food Court Zone 1 | Walkway Cell 00491 | Building Maintenance [0.76m] | 73 |
| | | | 16:56:27 | Retail Shop Zone 4 | Walkway Cell 01060 | Phones R Us [0.00m] | 4 |
| | | | 16:56:31 | Food Court Zone 1 | Walkway Cell 00720 | Coles [0.00m] | 4 |
| | | | 16:56:35 | Retail Shop Zone 4 | Walkway Cell 00893 | Coles [1.08m] | 4 |
| | | | 16:56:39 | Retail Shop Zone 4 | Walkway Cell 01176 | Phones R Us [2.15m] | 4 |
| | | | 16:56:43 | Retail Shop Zone 4 | Walkway Cell 01294 | Coles [2.15m] | 4 |
| Totals (Shopper: 39) | | Sites Visited: 1 | Moves: 11 | Precincts: 2 | Elements: 11 | Avg Dwell: 10 | Tot Dwell: 112 |
| **Marc Ingle** | 7/3/2012 4:57:41 PM | liveupdate.symantecliveupdate.com | 16:57:22 | Walkway 2 | Walkway Cell 02323 | Urban Labels [2.15m] | 13 |
| | | | 16:57:35 | Walkway 2 | Walkway Cell 02325 | Urban Labels [2.15m] | 13 |
| Totals (Shopper: 71) | | Sites Visited: 1 | Moves: 2 | Precincts: 1 | Elements: 2 | Avg Dwell: 13 | Tot Dwell: 26 |
| **Ranjit Varkey Chudukatil** | 7/1/2012 3:16:34 PM | us.f325.mail.yahoo.com | 15:15:45 | Walkway 1 | Walkway Cell 00126 | Toys R Us [3.76m] | 17 |
| | | | 15:16:03 | Walkway 1 | Walkway Cell 00239 | Toys R Us [2.69m] | 18 |
| | | | 15:16:22 | Walkway 1 | Walkway Cell 00409 | Toys R Us [1.61m] | 19 |
| | | | 15:16:38 | Walkway 1 | Walkway Cell 00580 | Toys R Us [1.61m] | 16 |

# SSRS – Device Movements



Movement Analytics: Individual Device Movement Tracking
CENTRE: [MyBusiness, VIC]  //  FLOOR: [Ground Floor]
Execution Date: [2013-08-26 20:53]

Shopper Key: 39
Movement Datetime: 7/1/2012 4:56:43 PM
Dwell Time      : 4
Name    : Kitti Lertpiriyasuwat (F)
Age     : 50-59
Locality: 2178 (CECIL PARK)
Device  : Blackberry (Torch 9810)
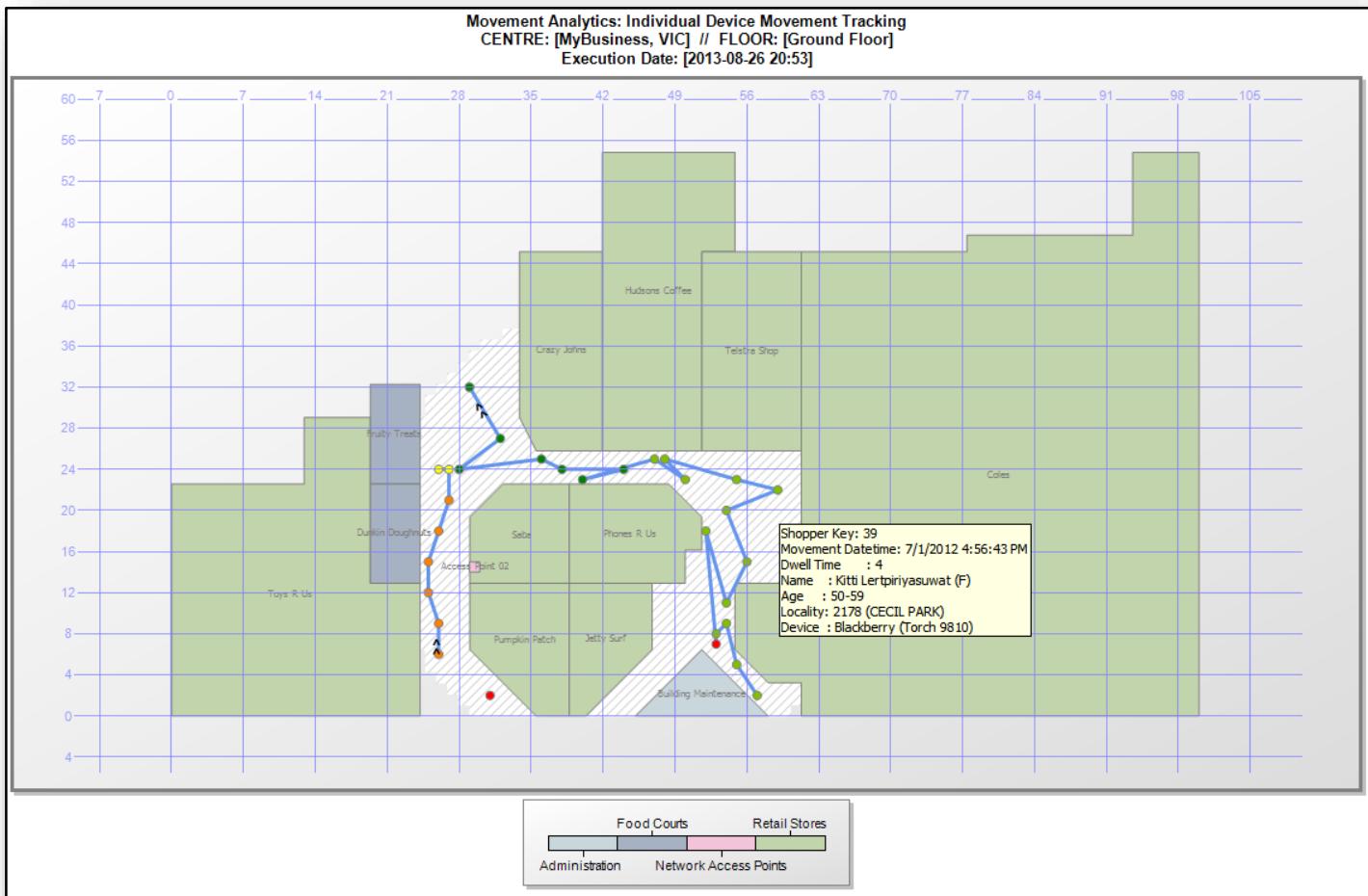
# PowerView – Locality & Demographics
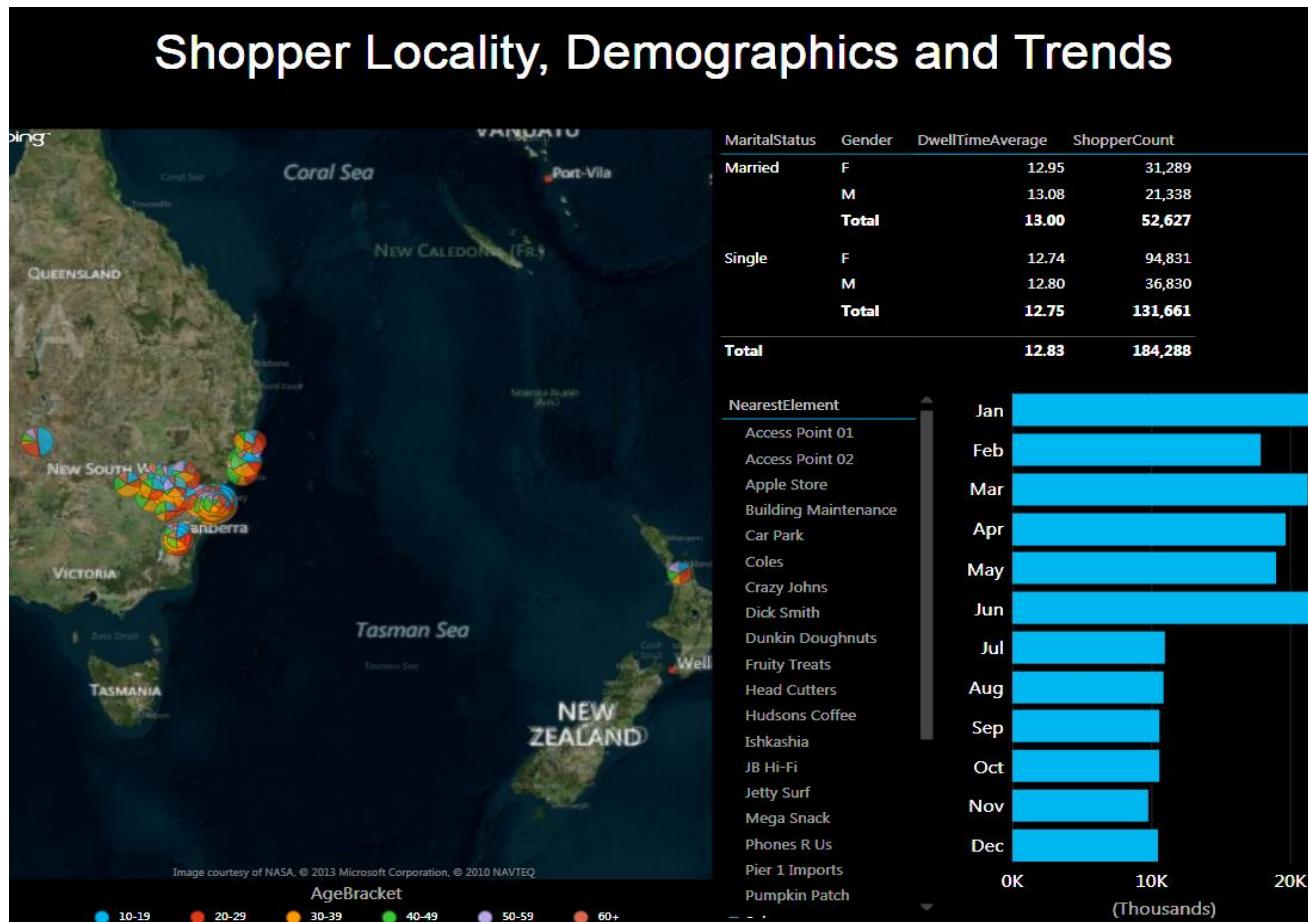


Shopper Locality, Demographics and Trends

# Agenda

- What are **Location Based Services** (LBS) solutions?

- **Technical overview** and **solution architecture**

- Leveraging **SQL Spatial** functions

- data, Data, **DATA**!

- **Reporting** and **analytics**

- Questions

# Some SQL Spatial Geometry Fun…

- SQL Spatial can do **more** than just **boring** *business* **stuff**… *tsk*!
- We can create our own **Patch Work Animals**!
- 
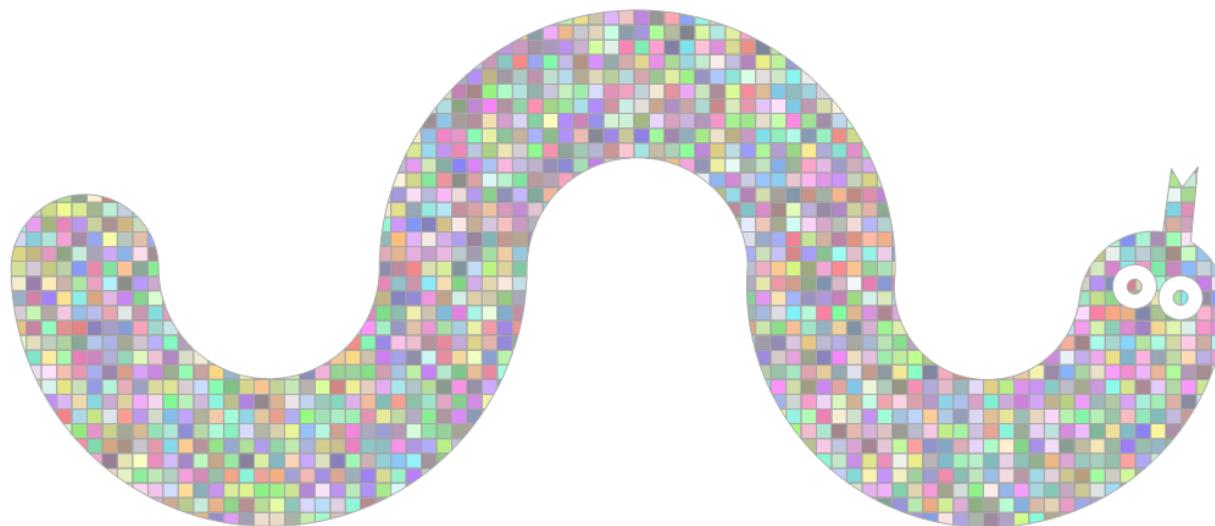
- A **BIG** kudos to **Alastair Aitchison (MVP & MCTS)**
  - http://alastaira.wordpress.com/2012/04/13/gridding-geometries-or-creating-patchwork-animals-in-sql-server/

# Questions?

## Please complete an Evaluation Form and return it to the Room Monitor

Please make sure you visit our fantastic sponsors:

PASS SQL saturday

#296 | MELBOURNE 2014

# And a *final parting note* to leave you with…



...and in other news...

According to this sign
one of the **great benefits**
of a pregnant woman
on a bus is the...
**Free Wi-Fi !**

# References

- Location Based Services – A WIKI Definition
  - http://en.wikipedia.org/wiki/Location-based_service

- Cisco Mobility Services Engine (MSE)
  - http://www.cisco.com/c/en/us/products/wireless/mobility-services-engine/index.html
  - http://www.cisco.com/c/en/us/products/collateral/wireless/mobility-services-engine/white_paper_c11-728970.html

- Gridding Geometries - Alastair Aitchison Blog
  - http://alastaira.wordpress.com/2012/04/13/gridding-geometries-or-creating-patchwork-animals-in-sql-server/

- A* Search Algorithm
  - http://en.wikipedia.org/wiki/A*_search_algorithm

# Appendix
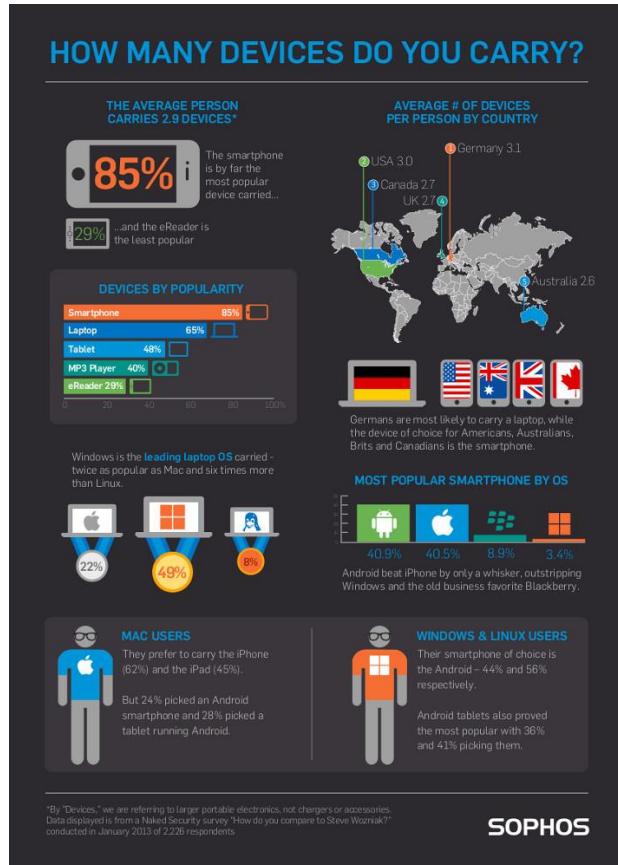


- **Australia** has **2.6 devices per person**

- **Germany** has the highest at **3.1 per person**

- Majority of Devices are **Smart Phones**

- **Android** is the **most popular** Smart Phone OS

- **Windows** is the **least popular** Smart Phone OS

- **Windows** is the **most popular** Laptop OS