Streaming Reasoning
will it pay?

David Bowden – Dell EMC Research Scientist
Dell Technologies

Apps

PaaS

Virtualisation

Hardware

UTILITIES & SERVICES

FINANCE

PHARMA

TELCOS

ENERGY
Big Data

Volume
- Large quantity of data the system has to handle

Velocity
- The rate at which the data is hitting the system

Variety
- The variability of data types or forms handled by the system

Veracity
- The accuracy and completeness of data ingested into the system
Climbing the Semantic Pyramid

- Data
- Information
- Knowledge
- Wisdom

Semantic Level

Business Value
Climbing the Semantic Pyramid (of tomatoes)

Wisdom
Knowledge
Information
Data
Real-time Video Analytics

Man with bag of money, holding gun, outside bank. Car door open, with registration plate %$^&*!

Bank Robbery!
The Value of Now
Data Value over Time

Value

Real-time

Time

DSS

(decision support systems)
The Fire Hose Problem

Too much data hitting the system

Increased volume and velocity of:

- Social media
- Mobile phone traffic
- Internet of Things

Problem today has become more challenging

Somewhere in fire hose of data is valuable business information

If we can extract and process in real-time
Mobile Providers

Massive streams of data
- Voice
- Text
- Tweets
- Images

Plus customer metadata:
- Who
- What
- When
- Where

ask before you use it
Real-time Intelligence (RTi)
Business Value

Why?

Reduce the spam

Customer retention:

- Reduce service failures, **tell the customer before they tell you**
- Reducing OpEx costs, **better value for money**
- Providing real-time **sticky services** customers will stay for

Location Based Services (LBS) can provide customer with service relevant to **where they are now**

The customer is entrusting the provider with their personal data, they must **want to sign-up** for the services
How Long is Real-time?
How long is Real-time?

It depends on the application

- 1-5 seconds
- 5-15 seconds
- 15-180 seconds
Smart City

- Smart City PoC portal & framework
- Provides citizen portal
- Ingested data into data lake
- Supports for in-house and third-party app development
- 1-2 second response time
  need a push not pull design
Data Sources

- Hard & Soft data sources
- Structured & Unstructured data
- Veracity of data source
- Combining data sources
The Value of Prediction

Combining current and past data to predict future events has great business value

Predictions don’t have to be precise to be useful

- SmartCity PoC predicts how long a traffic obstruction is likely to last
- Dell Gateway used to monitor vibration of engineering machinery and predict possible failure
- Pivotal’s connected car PoC monitors current route and predicts likely destination
Security
Speed v Security?

To make things go fast we often strip the system down to the bare minimum, including security

- Security should not be an optional extra
- Retrofitting security into a system can be expensive
- Streaming security posses additional design challenges

Some common streaming data security challenges

- IoT devices are autonomous and often poorly hardened
- Streaming data gives little time to analyse security threats
- Extended IoT communication paths provide new attack surfaces
Some Simple Data Protection Precautions

- If you don’t need PII, better not to collect it
- Personally Identifiable Information (PII) should be encrypted, even when in memory
- Anonymise or aggregate data where possible
- Devices and communications channels should be hardened, particularly autonomous IoT ones
- Unstructured data may contain Personally Identifiable Information
GDPR

EU Regulation GDPR

- General Data Protection Regulation
- Made 2016, comes into force 2018
- Intended to give EU citizens control of personal data
- Up to €20M or 4% annual worldwide turnover per infringement
- Both the controlling organisation and any third-party they contract out to are liable
SPARKS is a collaborative research project funded by the European Commission. Smart meters represent a significant evolution in electrical grids, but are vulnerable to cyber-attack. The objective is to dynamically protect Smart Grid infrastructure against cyber-attacks. This work was carried out as part of the SPARKS project, sponsored by the EU FP7 Programme Contract No. 608224.
ESCUDO-CLOUD

ESCUDO-CLOUD is a collaborative research project funded by the European Commission.

The objective is to secure a user’s data in the Cloud, specifically on Cloud Service Provider platforms.

Introduces tools and techniques for creating “trust boundaries” for data in transit and at rest.

This work was carried out as part of the ESCUDO-CLOUD project, sponsored by the EU H2020 Programme Contract No. 644579.