

Part I: Exploratory Analysis of IoT Sensor Data

The developer's guide to Cognitive IoT

Module II:

Toolset Introduction

On ApacheSpark, NoSQL, ObjectStorage and the rest...

This module covers...

- Data Storage Options for real-time IoT sensor data
- Introduction to ApacheSpark
- The best data science programming language
- Functional programming for parallelizable code
- IBM Bluemix
- Node-RED

In this Video you will learn...

Different Data Storage Options

SQL

- Advantages
 - well known, well established
 - high integrity, high data normalisation
 - fast, indexed access
 - Open Standard

SQL

- Dis-advantages
 - change of schema needs DDL
 - hard to scale
 - high storage cost

NoSQL

- Advantages
 - Dynamic schema, change of schema no problem
 - Low storage cost
 - Linear scalable

NoSQL

- Dis-advantages
 - No data normalisation, no data integrity
 - Less well-established
 - Generally slower in access than SQL

ObjectStorage

- Advantages
 - Very low storage cost
 - Linear scalable
 - Seamless schema migration, schema-less
 - Open Standard

ObjectStorage

- Dis-advantages
 - not well established, but catching up
 - no integrity
 - no data normalisation
 - no indexed access

Decision Matrix

	SQL	NoSQL	ObjectStorage
Storage cost	high	low	very low
Scalability	low	high	very high
Query Speed	high	low	very low
Flexibility	low	high	very high

Summary

- Three main storage options exist
- Choice depends on
 - amount of data
 - variety of schema
 - query performance requirements
 - additional data types like images, audio or video

Quiz

- Which storage type has the most flexible schema?
 - A: NoSQL databases
False: Although NoSQL is very flexible, in most of the cases it is limited to JSON documents
 - B: ObjectStorage
Correct: ObjectStorage resembles in file storage, so you can basically store anything in files
 - C: SQL databases
False: SQL has a very strict schema

The next Video covers...

The storage cost calculator