



# Beyond Object Storage A Unified Storage Architecture for IoT

Sasikanth Eda, Sandeep Patil  
IBM



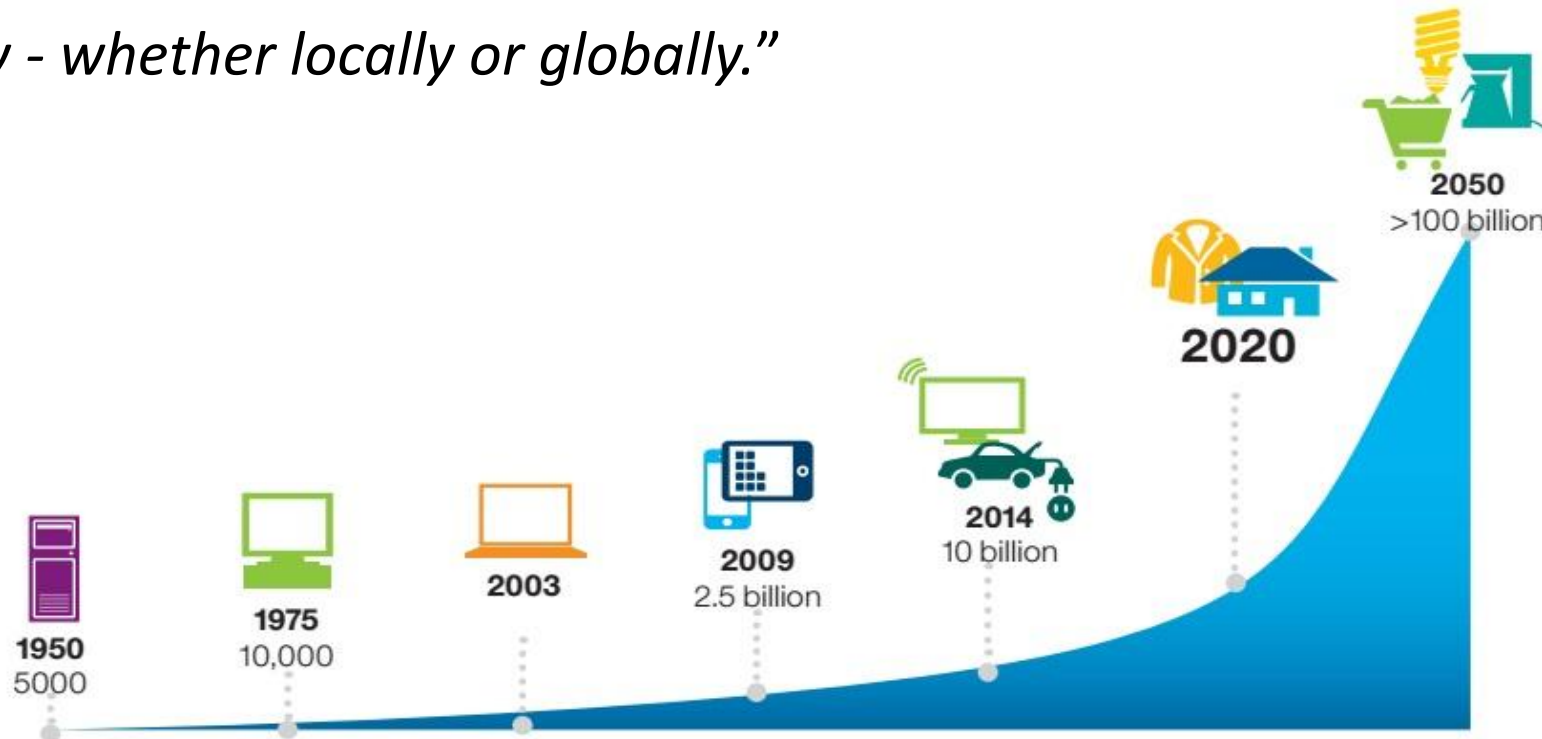
# Agenda

- Internet of Things (IoT): Market Estimates And Forecasts
- IoT: Challenges
- IoT Platform Architectures
- IoT Platform: Storage Requirements
- Enhancements to Object Storage to support IoT workloads Better



## IoT: Evolution of Devices

According to IDC - “The Internet of Things (IoT) is a *network of networks* of uniquely *identifiable endpoints* (or “things”) that *communicate* without human interaction using *IP connectivity* - whether locally or globally.”



Source: <http://www.gartner.com/newsroom/id/2636073>



## IoT: Market Estimates And Forecasts (Industry Specific)

- ✓ More than **50%** of **manufacturers** identify **lower operational costs** as the most significant driver of their organization's Internet of Things (IoT) initiatives over the next 12–24 months (05/06/2015, IBM Institute for Business Value).
- ✓ The value of Internet of Things in **healthcare** is forecasted to reach **\$163.24** billion by 2020, with a CAGR of 38.1% for 2015-2020 (01/05/2016, eMarketer).
- ✓ By **2018**, it is expected to drive development of over **200,000 new IoT apps and services** (11/03/2015, IDC).



## IoT: Market Estimates And Forecasts (Cloud Specific)

✓ **36%** of **hybrid leaders** are using **hybrid cloud** for Internet of Things.

(02/09/2016, IBM Center for Applied Insights)

✓ **52%** of IoT **developers** implement **cognitive computing and artificial intelligence** in their development work. (11/09/2015, Evans Data)

✓ By 2019, **80%** of new applications using the Internet of Things (IoT) will **analyze data in motion** as well as collect this information for **analysis of data at rest**.

(12/03/2015, Gartner)



## IoT: Challenges (Technology)

- **Cost of connectivity**

Solutions are **expensive** (high infrastructure, maintenance costs, service costs of middlemen).

- **Internet after trust**

In IoT networks trust (**privacy, anonymity**) is either **expensive** or difficult to guarantee.

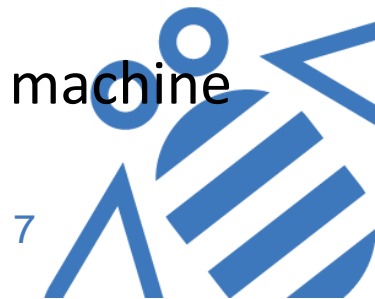
- **Lack of functional value**

- **Simple connectivity** enablement may or **may not** make a **device smarter** or better.
- **Connectivity** with **intelligence** makes a better product and experience.



## IoT: Challenges (Business)

- **Not future-proof**
  - **Refresh cycle** for basic pieces of IoT infrastructure (door locks, bulbs etc.) is **quite long** (may last for years, even decades).
- **Broken business models**
  - Unlike PC or smart phones, **IoT devices** (locks, bulbs etc.) worked **without apps and service** contracts, which make revenue expectations unrealistic.
  - **End to End IoT solutions** (Smart TV to speak to the toaster, Smart washing machine to speak to detergent vendor) get **cumbersome** quickly.



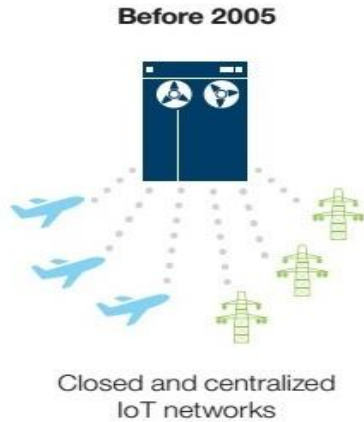
## IoT: Current State – Needs a Reboot

- ✓ **IoT landscape** is been evolving at a **rapid pace** and there has been **little innovation** to address the IoT market from a **storage infrastructure** standpoint.
- ✓ Currently **no storage vendor** is offering **specific storage solutions** for IoT deployments.
- ✓ There is **no single “best”** approach to data management in the context of IoT.
- ✓ Need for a **common framework**, enhancements in performance and security.





# IoT Platform: Centralized Architecture



- ✓ IoT centralized architecture uses a **hub** (typically powered by the cloud) that controls the execution of nodes (smart devices).
- ✓ Uses **Platform-as-a-Service** model.
- ✓ Current **database architecture** and **data management** strategy **may not** be sufficient to **handle large scale** IoT networks.
- ✓ **Big data** architecture gets **complex**, as IoT devices produce hundreds of thousands of data points per second.
- ✓ Offered Services: Event processing, Device discovery, Device management, Event Notifications, Real Time Analytics.



# IoT Platform: Decentralized Architecture



**ADEPT** platform – Autonomous Decentralized Peer-To-Peer Telemetry  
An effort to prove the foundational concepts around decentralized approach.

Key Objectives:

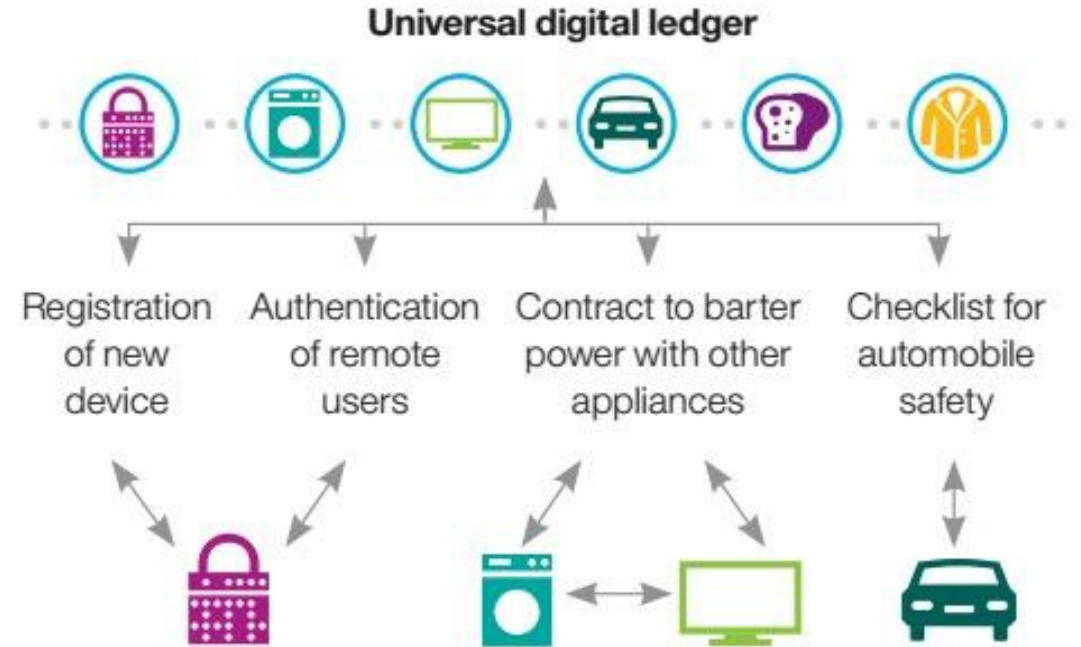
- ✓ **Distributed Transaction Processing & Applications:**
- ✓ **Robust Security**
- ✓ **Privacy** by design and default
- ✓ **Designed for commerce and market places**



# IoT Platform: Decentralized Architecture

## Key Solution Components:

- ✓ Trustless **peer-to-peer** messaging
- ✓ Secure distributed **data sharing** / file transfers
- ✓ **Autonomous** Device coordination.



# IoT Platform: Storage Requirements

1. **Scalable:** **Cost-effective scalability**, and the ability to decouple capacity and performance.
2. **Self-healing:** Traditional RAID mechanisms may be ineffective in large scale environments. Having a self-healing architecture with **configurable data availability** classes.
3. **Data awareness:** Ability to create a data-aware storage platform, which provides both **context** to the data and policy management, as well as visualization tools to harness the value of that data for better **business insights**.



# Why Object Storage For IoT

Early predictions (why object storage suits best for IoT workload)

- ✓ Designed to support **large volumes** and velocity of data.
- ✓ Native **HTTP / REST** support.
- ✓ Linear **Scale out** architecture (Scale in all directions including IOPS, latency)
- ✓ **Distributed** Architecture (Performance, load distribution)
- ✓ **No** single point of failure.
- ✓ Support for **heterogeneous data types**, different data protection policies.



## Why Object Storage is falling short ?

Micro services of IoT	Type of Storage	Appropriate Storage Platforms
Batch Processing	Primary Storage	Distributed Filesystems
Real-Time Processing	Primary Storage	Memory-centric Filesystems, NoSQL Data Stores
Active Archiving	Secondary Storage	Distributed Filesystems, Object Storage
Cold Storage	Primary Storage	Object Storage with Spin down capabilities, Cloud Storage, Tape Libraries

\* Source: Gartner 2015

Based on a recent survey on IoT implementations, it is observed that **object storage** is **not** considered for **real time** processing of **IoT data** (However success of IoT implementation hugely depends on analytics platform).



# Factors that help Object Storage to hold its promise

**1: In-Place Analytics**

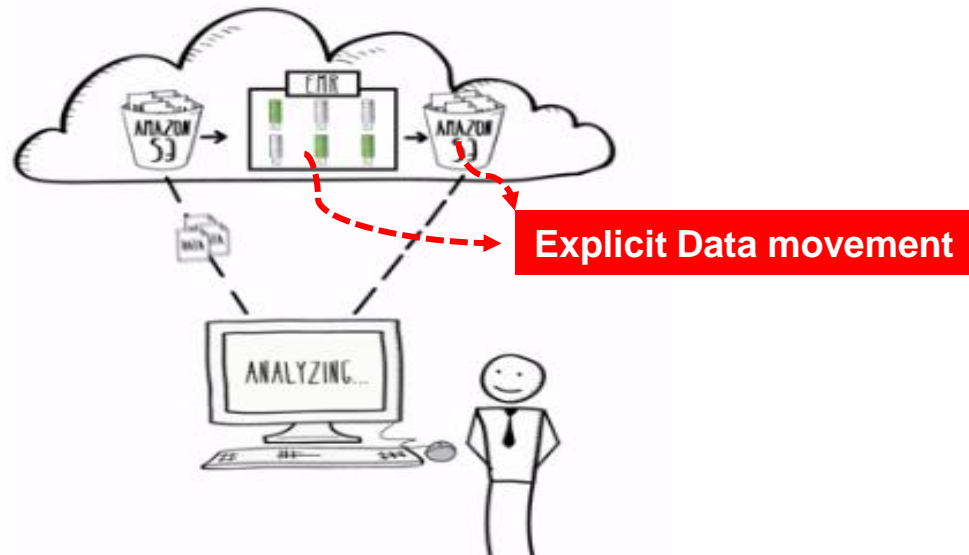
**2: IoT data collection and Privacy filters**

**3: Native Event, Cross Device Notification, Life Cycle Management**



# Issue-1: In-Place Analytics

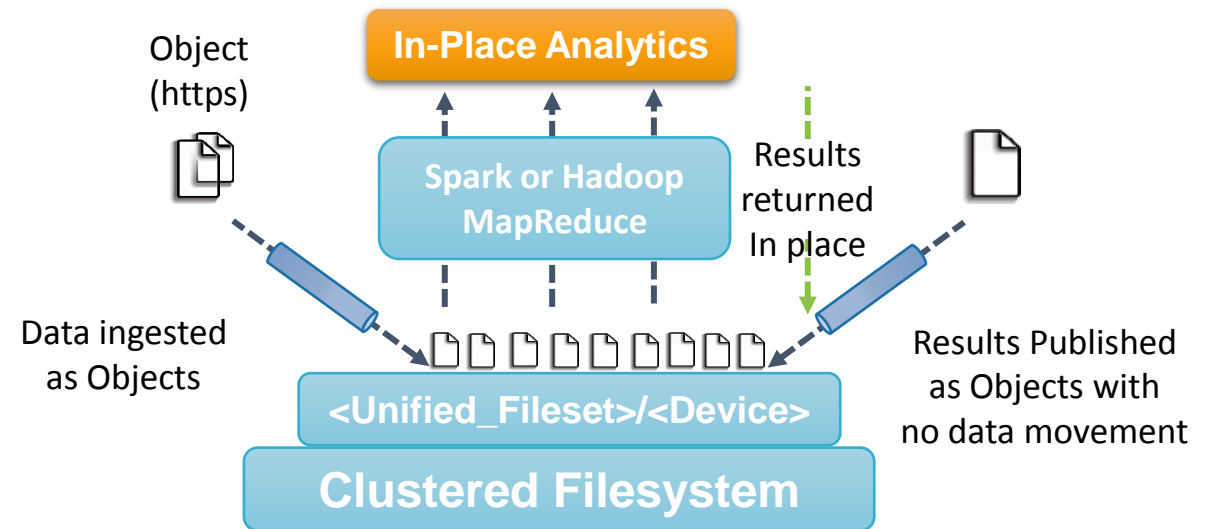
## Analytics on Traditional Object Store



1. Data to be **migrated** from object store to dedicated analytic cluster.
2. Perform the analysis and copy results back to object store for publishing.

Reference: <https://aws.amazon.com/elasticmapreduce/>

## Analytics With Unified File and Object Access



**Object data available as "Files"** on the same fileset.

Analytics systems (Hadoop, Spark) can **directly** leverage this data analytics.

**No data movement / In-Place immediate data analytics**





# Swift-on-File for In-Place Analytics

This object:

<http://swift.example.com/v1/acct/cont/obj>

was stored with Swift here:

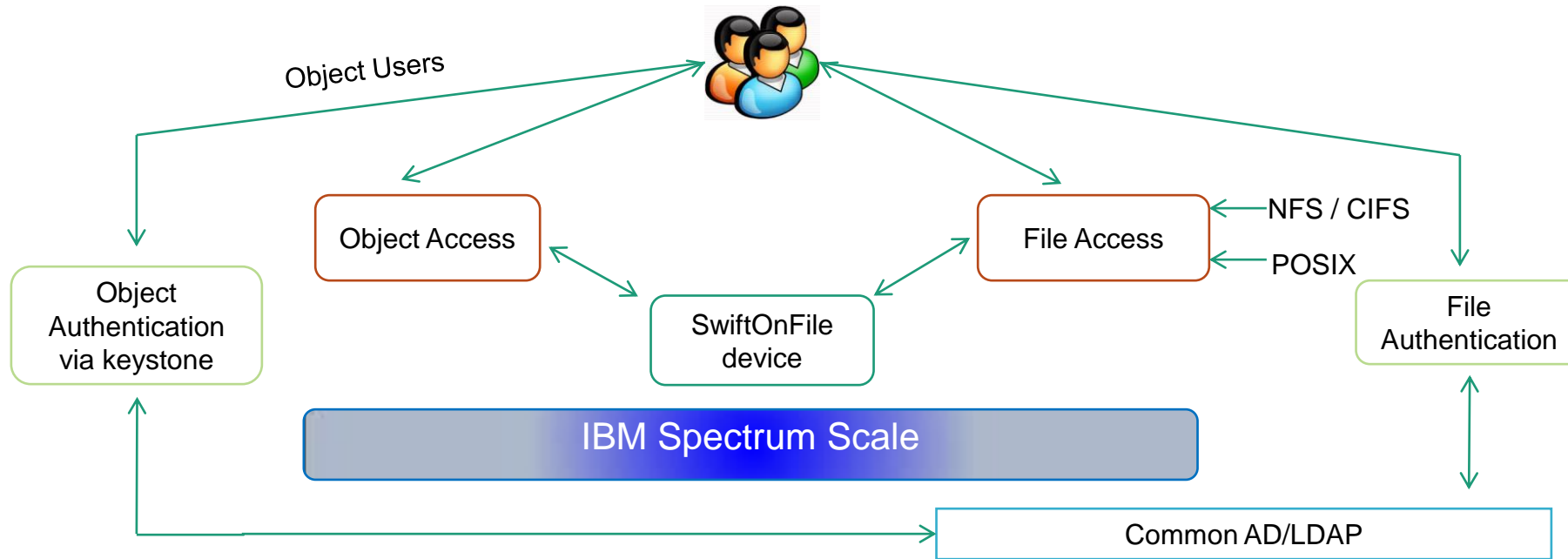
`/mnt/sdb1/2/node/sdb2/objects/981/f79/f566bd022b9285b05e665f  
d7b843bf79/1401254393.89313.data`

But is now stored with SwiftOnFile here:

`/mnt/scaleoutFS/acct/cont/obj`



# Unified Identity Between Object And File

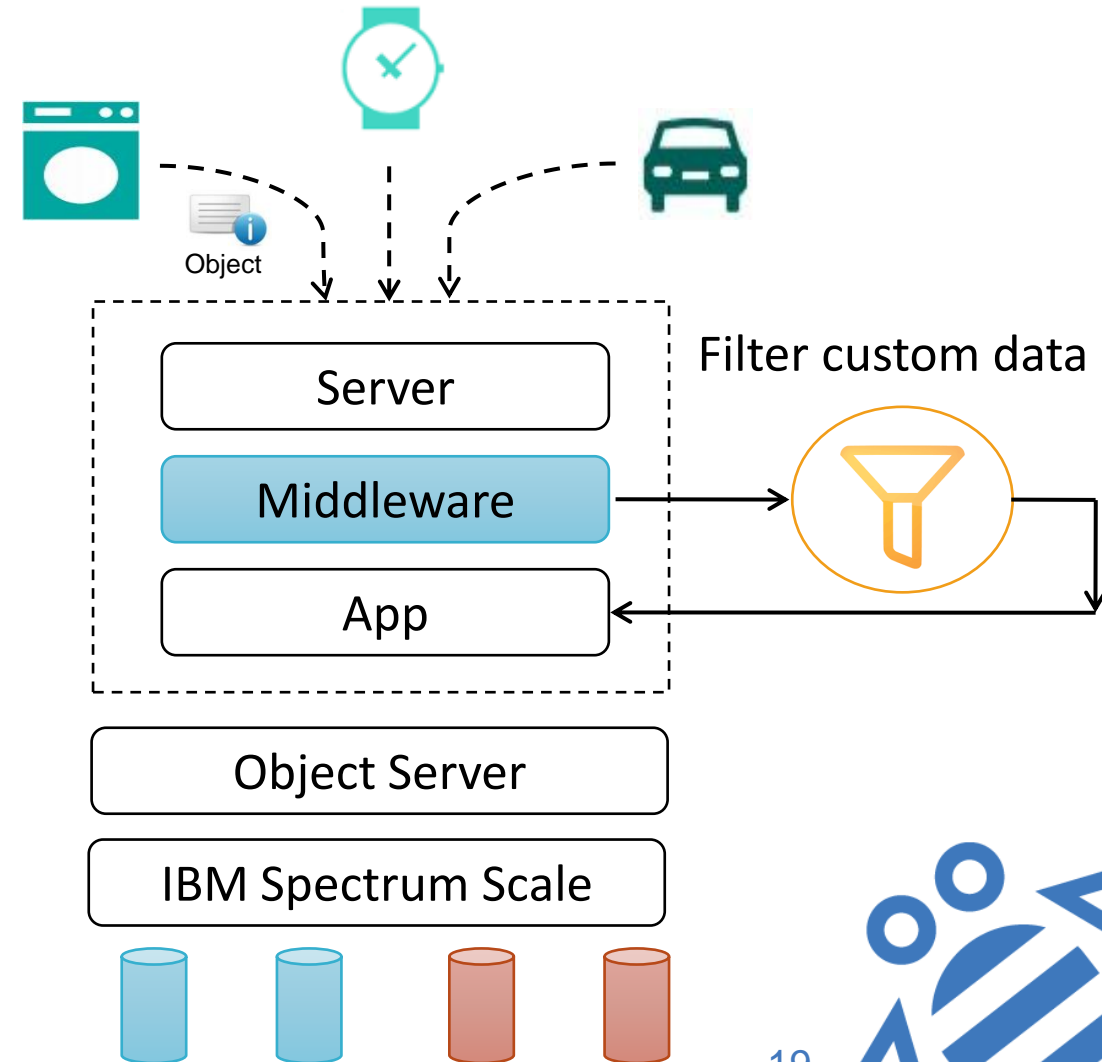


- ✓ Common set of Object and File users using same directory service (AD+RFC 2307 or LDAP)
- ✓ Objects created using Swift API will be owned by the user performing the Object operation (PUT)
  - Note that if object already exists, existing ownership of object will be retained

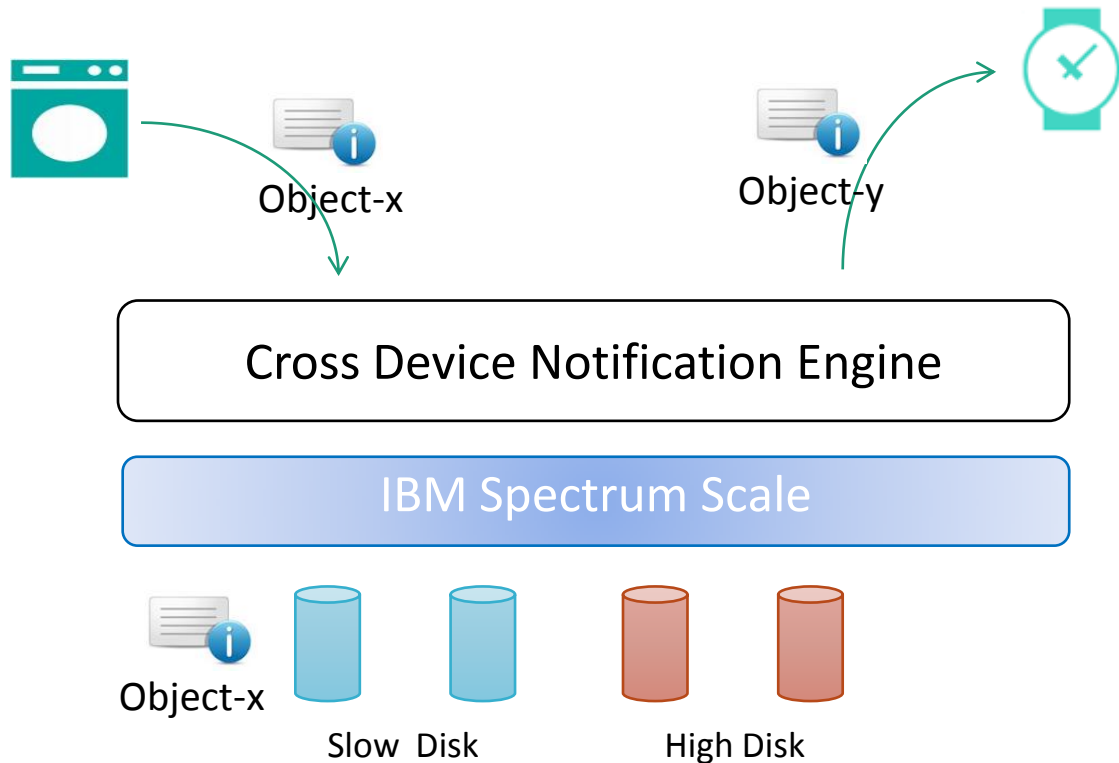


## Issue-2: IoT Data Collection and Privacy Filters

- ✓ Data **collection** and maintaining **privacy** on IoT devices is a big deal.
- ✓ **Lack of trust** on details collected by commodity IoT devices. **Storing** these details on **cloud** makes situation more **uncomfortable**.
- ✓ End user **configurable Middleware** to **filter** out the **privacy** parameters per IoT device.
- ✓ End user **configurable Expiry** of objects, metadata per IoT device.
- ✓ **Automated placement** of data per IoT device.



## Issue-3: Event, Cross Device Notification, Life Cycle Management



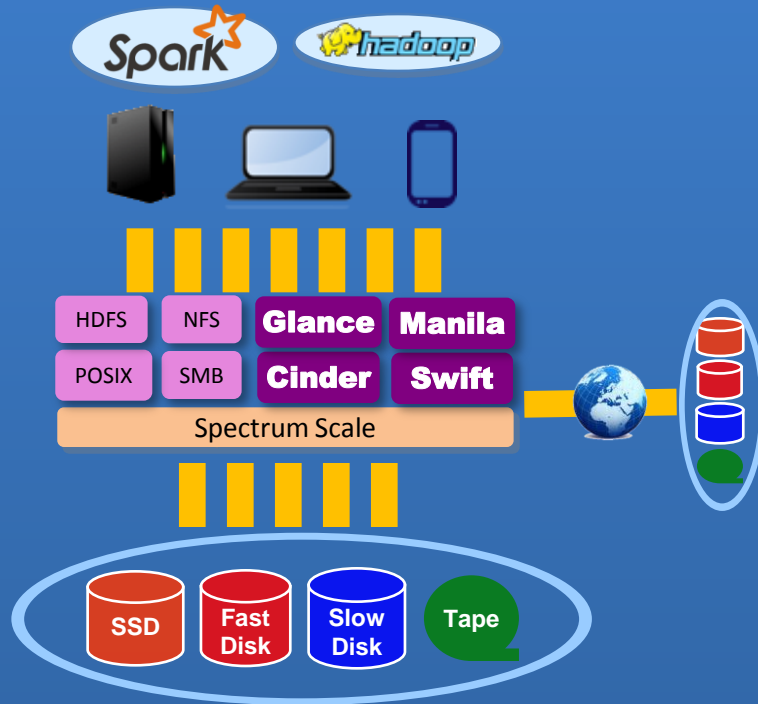
- ✓ **Avoids** need of **database triggers** or performance hampering analytic engine to act on a particular set of received IoT data.
- ✓ **Policies, rules** for cross device notification can be applied based on received object content, type, device type, metadata details and generated event type.
- ✓ **Auto-migration / placement** of data based on its dependency for cross device notification.



# IBM Spectrum Scale

Open by design™

## Data management at scale



- Avoid vendor lock-in with true Software Defined Storage and Open Standards
- Seamless performance & capacity scaling
- Automate data management at scale
- Enable global collaboration

## OpenStack and Spectrum Scale helps clients manage data at scale



**Business:** I need virtually unlimited storage



An open & scalable cloud platform



**Operations:** I need a flexible infrastructure that supports both object and file based storage



A single data plane that supports Cinder, Glance, Swift, Manila as well as NFS, et. al.



**Operations:** I need to minimize the time it takes to perform common storage management tasks



A fully automated policy based data placement and migration tool



**Collaboration:** I need to share data between people, departments and sites with low latency.



Sharing with a variety of WAN caching modes

### Results

- Converge File and Object based storage under one roof
- Employ enterprise features to protect data, e.g. Snapshots, Backup, and Disaster Recovery
- Support native file, block and object sharing to data.



# THANK YOU



## Acknowledgements

Dean Hildebrand

Bill Owen

Shyama VenuGopal

