AWS

5 U M M I T

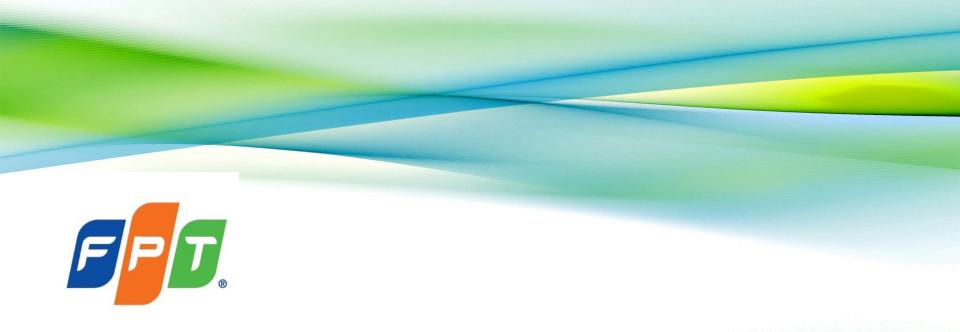
Singapore

Best Practices for Re-architecting & Optimizing Large-Scaled Data Systems on AWS: Oracle Migration and Big Data Streaming Solution

Cuong Bui

Chief Solution Architect, AWS Taskforce FPT Software





Best Practices for Re-architecting & Optimizing Large-Scaled Data Systems on AWS

From Oracle Migration to Big Data Streaming Solution







FPT- The Market Leader





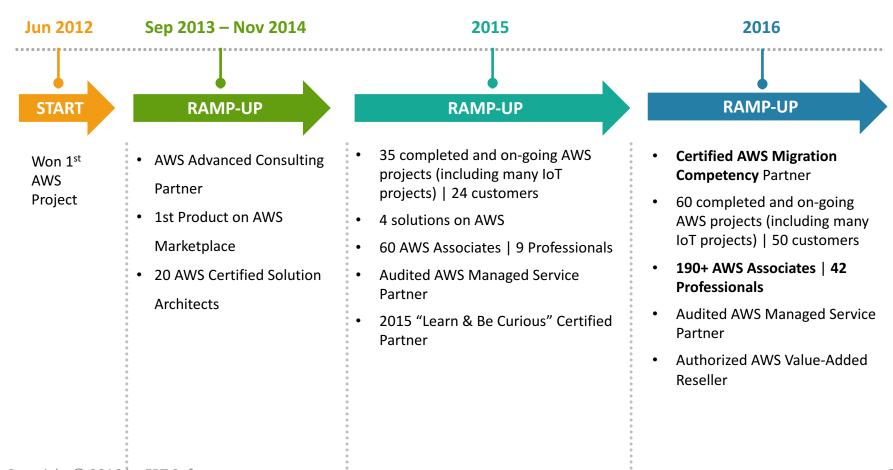
SEI CMMI Level 5 | ISO 27001:2013 (BS 7799 -2:2002) | ISO/TS 16949 | ISO 9001:2008 | ISO 20000-1:2011





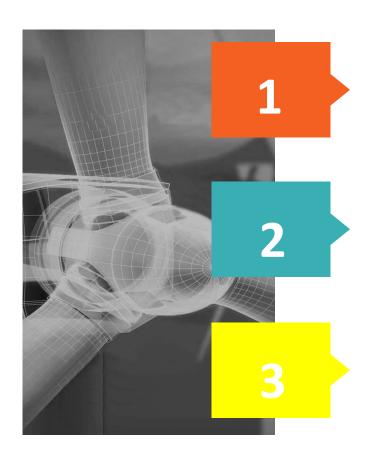
AWS – FPT Partnership History





What do we do?





Database and SAP Migration

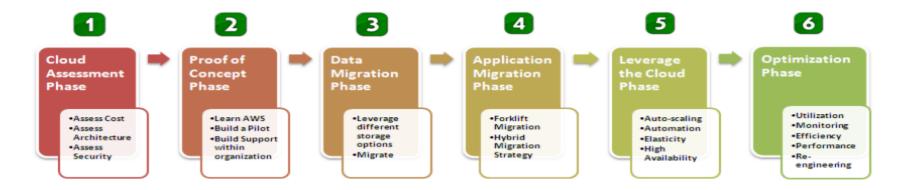
Big Data System Design

Digital Transformation: AI, IoT, Fintech, etc.

How to migrate to AWS Cloud



A Phased Strategy for Migration: Step By Step Guide







Application Migration Phase: Lift and Shift

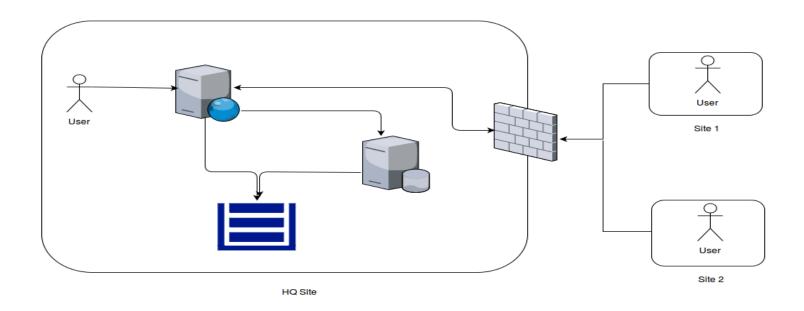
First case



- 1. Migrate the whole data center to AWS,
- Have some Oracle Applications like Oracle EBS, Siebel,
- 3. Have about **30K+ tables** to migrate.

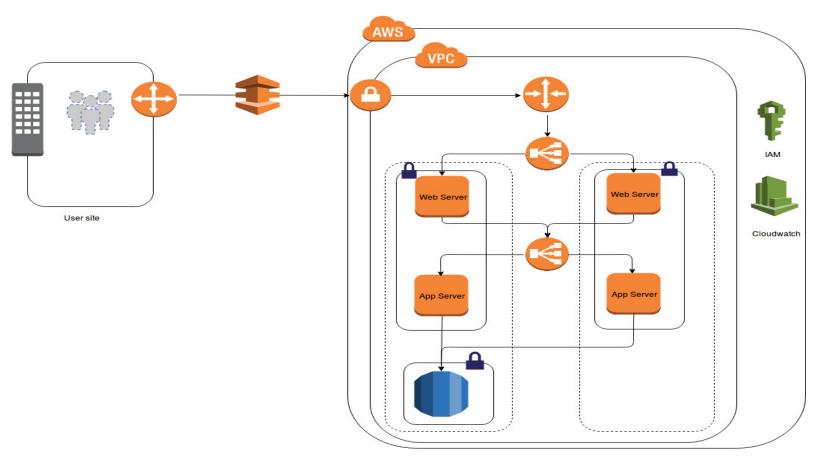
Existing System





Reference Architecture





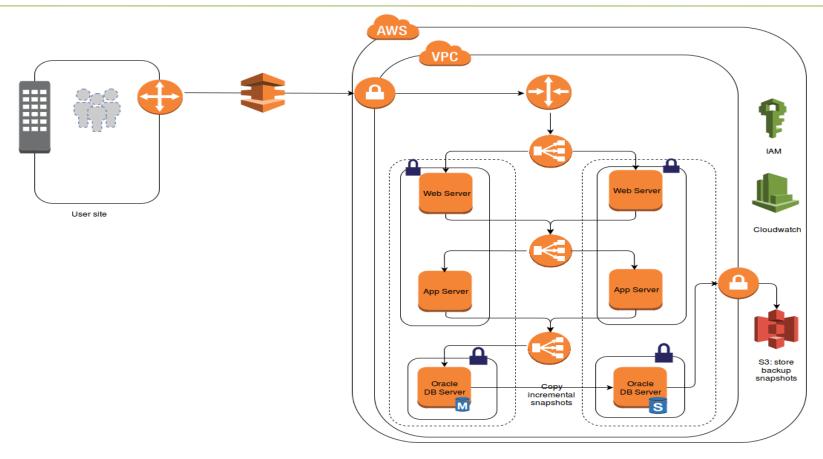
Real scenario



- Cannot use RDS due to special requirements from Customer Applications → Must use EC2 instances.
- 2. RTO of 2 hours, RPO of 15 mins.
- 3. No disk array on AWS → Have to deal with bit rot, disk failures.

More realistic architecture





Benefits





30% of cost reduction

Lower RTO and RPO

Enjoy AWS Data warehouse solution like Redshift and EMR

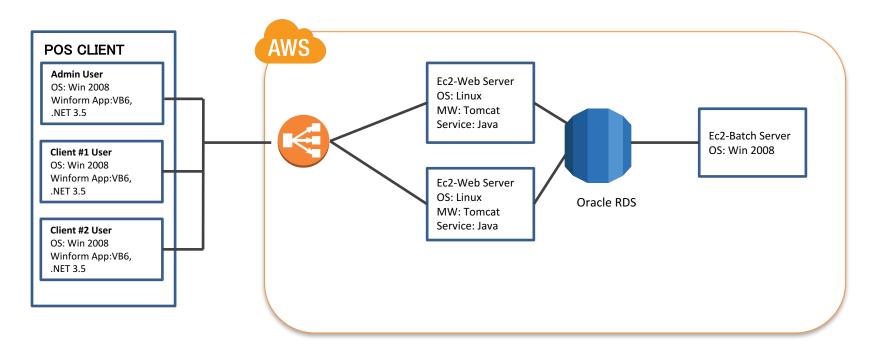




Leverage the Cloud: Database Migration to Aurora

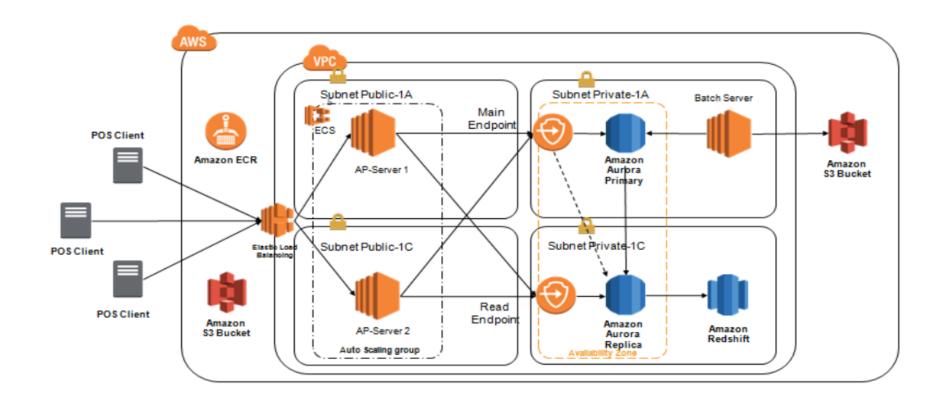
Existing Architecture for a Retailer





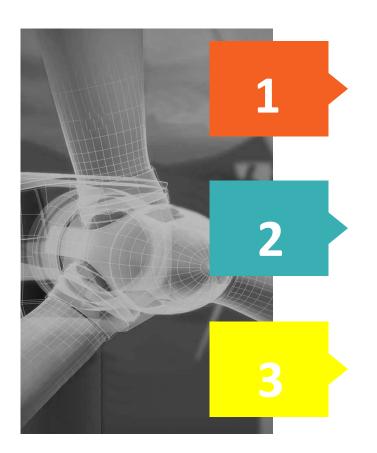
A reference architecture on Aurora





Challenges: Oracle to Aurora





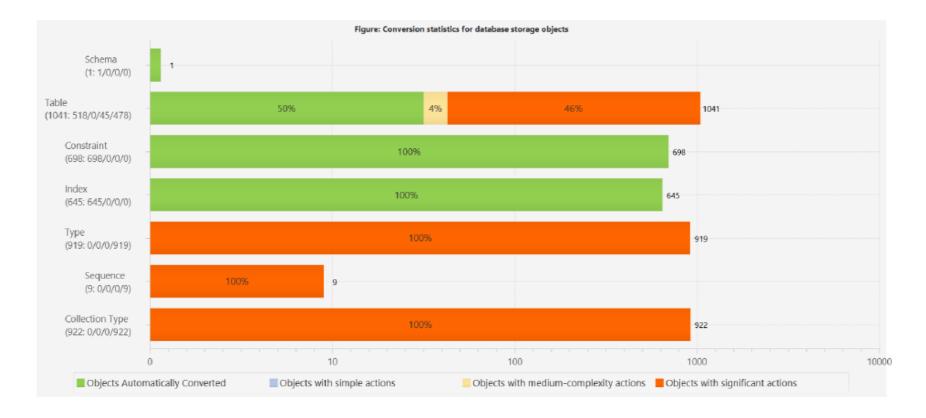
Rewrite PL/SQL stored procedures

Re-factor the application code

Enterprise grade support

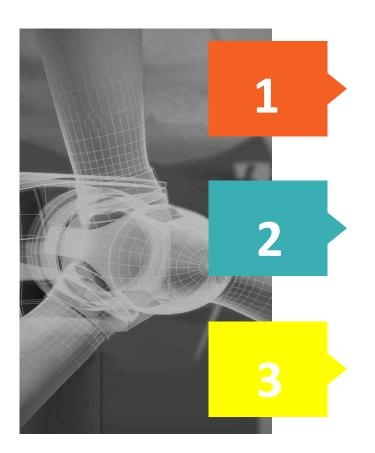
A report from Schema Conversion Tool (SCT)





Solutions





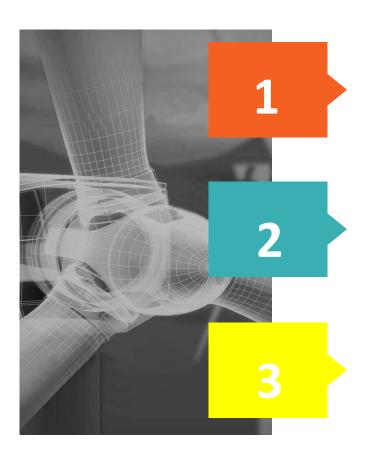
SCT is the key to migrate PL/SQL SP

Only have to update DAO

AWS support with 15 minute response time.

Benefits





Performance gain of 20%

Reducing **50%** of the cost

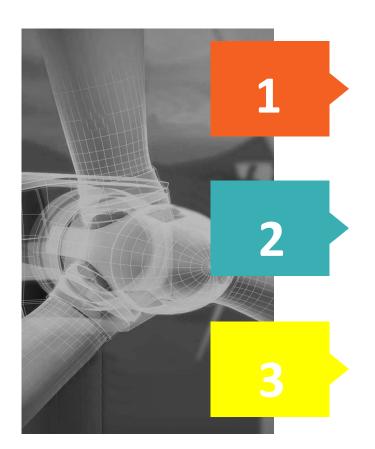
With built-in High Availability and DR





A Root Cause Analysis System





A data set of petabytes (PB) scale

Use machine learning to find the Root Cause

Data growth rate of multiple TB+ per day

Batch Processing Approach















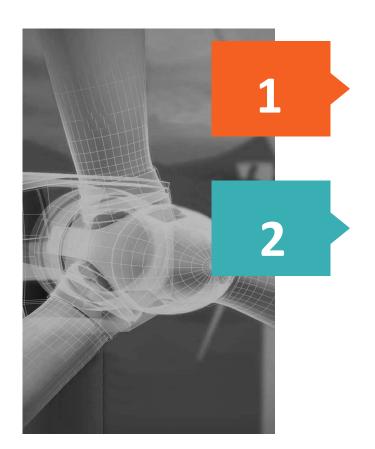






Some Problems with Batch Approach





It takes several **hours** if not days to complete

What if it costs **millions dollars** for each **hour** of RCA delay



Can we do it better?

Business Challenges





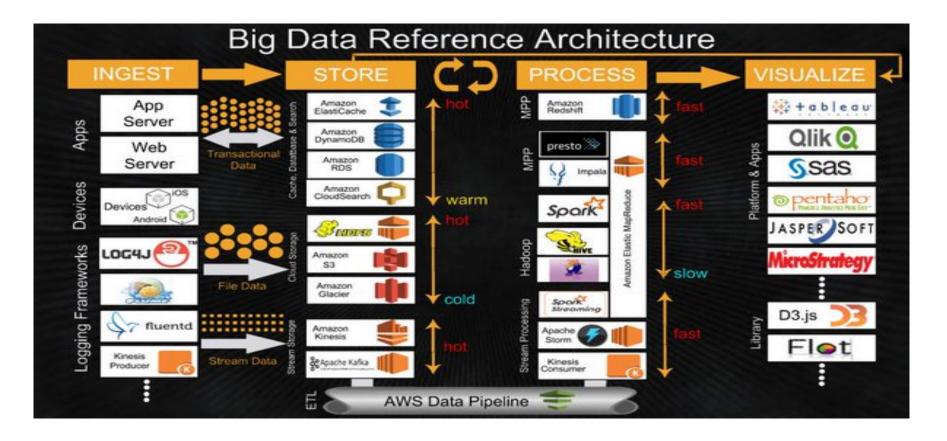
Reducing the RCA processing time from hours to minutes (100x performance)

Reduce the cost of the system

Retain as much data as possible

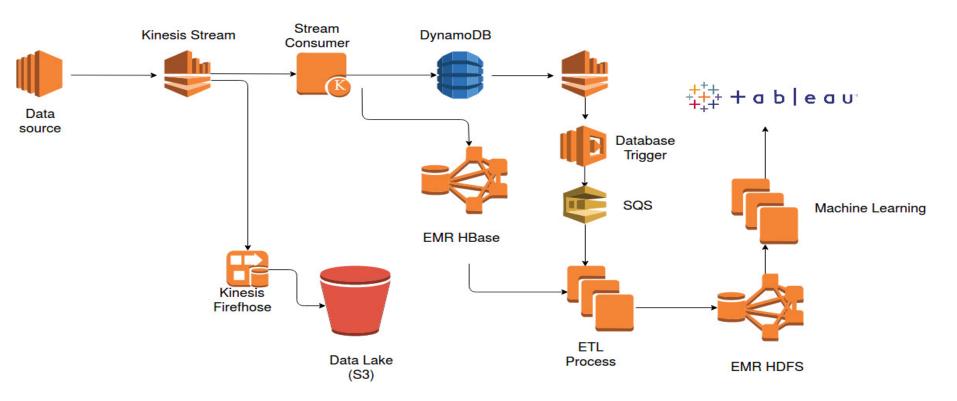
Reference Architecture for Big Data on AWS





Streaming Approach





Results





100 times faster for RCA

Reducing the cost of ownership

Much shorter time to implement their PoC and new algorithm experiments





Lessons Learnt





