Fast SAP HANA Fail Over Architecture with a SUSE High Availability Cluster in the AWS Cloud

Dr. Stefan Schneider
Partner Solutions Architect @ Amazon Webservices
stsch@amazon.com

Markus Gürtler
Senior Architect SAP @ SUSE
mguertler@suse.com
Agenda

- HANA Scenarios
- Implementing SUSE HA Scenarios on AWS
- Security, HA and DR in the Cloud
- A demonstration of a SAP HANA failover on AWS
- Q&A
SAP HANA Business Continuity

**Business Continuity**

- **HA per Datacenter**
  - SAP HANA Host Auto Failover (scale out with standby)
    - SAP
    - HW
  - SAP HANA System Replication
    - SAP

- **Disaster recovery between Datacenter**
  - SAP HANA Storage Replication
    - HW
  - SAP HANA System Replication
    - SAP
Automate SAP HANA System Replication

SAP HANA System Replication

“sr_takeover” is a Manual process
Automate SAP HANA System Replication

SAP HANA System Replication + SUSE High Availability Solution

Automates the “sr_takeover”
Automate SAP HANA System Replication

SAP HANA System Replication + SUSE High Availability Solution

improves

Service Level Agreement
Simplify Linux for SAP Workloads
SUSE Linux Enterprise Server for SAP Applications 11

- Page Cache Management
- Antivirus ClamSAP
- SAP HANA Security
- Simplified Operations Management
- High Availability SAP NetWeaver & SAP HANA
- SAP HANA HA Resource Agent
- Installation Wizard Faster Installation

- Reliable, Scalable and Secure Operating System
  SUSE Linux Enterprise Server

- 24x7 Priority Support for SAP
- Extended Service Pack Support 18 Month Grace Period
- Installation Wizard Faster Installation
- 24x7 Priority Support for SAP
SAP HANA System Replication
Powered by SUSE High Availability Solution

Performance optimized
- Secondary system completely used for the preparation of a possible take-over
- Resources used for data pre-load on Secondary
- Take-overs and Performance Ramp shortened maximally
From Concept to Implementation
SUSE High Availability Solution for SAP HANA

- **SUSE HANA Primary**
- **SUSE HANA Secondary**
- **suse01**
- **suse02**
- **vIP**
- **Cluster Communication**
- **Master**
- **Slave**
- **Clone**
- **SAPHana Master/Slave Resource**
- **SAPHana Topology Clone Resource**
- **Fencing**
Four Steps to Install and Configure

1. Install SAP HANA
2. Configure SAP HANA System Replication
3. Install and initialize SUSE Cluster
4. Configure SR Automation using HAWK wizard
SAPHanaSR HAWK Wizard
What is the Delivery?
SUSE Linux Enterprise Server for SAP Applications

The package **SAPHanaSR**
- the two resource agents
  - **SAPHanaTopology**
  - **SAPHana**
- HAWK setup Wizard (as technical preview)

The package **SAPHanaSR-doc**
- the important **SetupGuide**
Allowed Scenarios

- Scale-Up *performance-optimized* (synchronous ==>)
  A => B

- Scale-Up in a chain or *multi tier* (asynchronous ->)
  A => B -> C

- Scale-Up in a *cost-optimized* scenario (+)
  A => B + Q

- Scale Up in a mixed scenario
  A => B -> C + Q

- Now all with *multi tenancy* (%) - here cost optimized
  %A => %B + %Q
Performance-Optimized
Single-tier System Replication and memory preload

Performance optimized (A → B)

- Secondary system completely used for the preparation of a possible take-over
- Resources used for data pre-load on Secondary
- Take-over performance much faster than a cold start
**Cost-Optimized**

**Single-tier System Replication and DEV / QAS**

Cost optimized (A → B + Q)

- Operating non-prod systems on Secondary
- During take-over the non-prod operation has to be ended
- Take-over performance similar to cold start-up
- Needs another disk stack for non-prod usage load
starting with version 0.149

Multi Tier System Replication – Cascading Systems

Available since SAP HANA SPS7
(Three cascading systems)
beginning with version 0.151

Multi Tenancy (MCD)
Synchronizing multiple Databases within one System Replication

Performance optimized %A => %B
Cost optimized %A => %B -> %C
Multi tier %A => %B + %Q

Tenants are databases within the SAP HANA database system
System replication only replicate the complete database

node 1

SAP HANA PR1 primary

node 2

SAP HANA PR1 secondary

Pacemaker

System Replication

System PR1
SUSE SAPHanaSR in 3 Facts

✅ Reduces complexity
- provides a wizard for easy configuration with just SID, instance number and IP address
- automates the sr-takeover and IP failover ("bind")

✅ Reduces risk
- includes always a consistent picture of the SAP HANA topology
- provides a choice for automatic registrations and site takeover preference

✅ Increases reliability
- provides short takeover times in special for table preload scenarios
- includes the monitoring of the system replication status to increase data consistency
Our Community

Developed jointly in the SAP Linux Lab in Walldorf
Integration of the solution in partner products
Upstream open-source project

You are invited to join our community :-) 

Visit our booth or contact us via sapalliance@suse.com or saphana@suse.com
HANA System Replication on AWS

- Default Net (A)
  - Primary node: node1 10.1.0.11
  - Subnet: 10.1.0.0/8
  - AZ: az-A
  - HANA System Replication
- HA net (B)
  - Secondary node: node2 10.1.1.12
  - Subnet: 10.1.1.0/8
  - AZ: az-B
  - vPC: vpc-AB 10.1.0.0/16

Related connections:
- User Intranet
- VPN-A
- VGW-A
Cloud HA and Disaster Recovery Options

High Availability
Same Availability Zone (Data Center)
HANA synchronous replication
IP address switch in sub second intervals

Disaster Recovery
Different Availability Zone (Data Center)
HANA synchronous or asynchronous replication
IP address switch in sub second intervals
Improved Security in the Cloud

Security

Policies to grant permission to stop and start systems by defined AIM users or systems

Policies to grant permissions to change network routing for defined AIM users and or systems

Auditing

AWS tracks when failover happened

AWS tracks tracks who started and shutdown systems
SUSE HanaSR Architecture on AWS

- Default Net (A)
  - primary node: node1 10.1.0.11
  - subnet-A 10.1.0.0/8
  - az-A

- HA net (B)
  - secondary node: node2 10.1.1.12
  - subnet-B 10.1.1.0/8
  - az-B

- SUSE HA CLUSTER
- Service IP 10.2.0.1

- vpn-A
- vgw-A
- vpc-AB 10.1.0.0/16
- VPC
- User Intranet

- System Replication
HanaSR in EC2

HA Resource Agents communicate to the Cloud via EC2 API

EC2

controls

API

controls

SAP HANA
PR1 primary

node 1

vIP

Pacemaker

System Replication

SAP HANA
PR1 secondary

node 2

System PR1

System PR1
STONITH fencing in HA clusters

- Loss of network connectivity results in split cluster partitions (split brain)
- STONITH fencing...
  
  ... solves split-brain situations in Pacemaker clusters ...
  
  ... by remotely shutting off or rebooting one or more nodes ...
  
  ... ensuring that just one cluster partition survives.
STONITH fencing in EC2

1. Cluster detects split-brain
   - network communication broken
   - node 1
   - node 2

2. Send STONITH request to EC2 API
   - node 1 requests force shut-off for node 2 via EC2 API
   - API request
   - EC2
   - node 1
   - node 2

3. EC2 API shuts-off node 2
   - EC2 instance shut-off on the hypervisor
   - node 1
   - node 2
   - shut-off
EC2 STONITH agent fence_ec2_sap

STONITH fencing agent for Pacemaker clusters running in AWS EC2

Agent uses EC2 API to hard-shutoff or reboot a cluster node (ec2-stop-instances <Instance ID> --force)

Allows to dynamically add or remove nodes without cluster re-configuration

Uses EC2 instance tags to Identify nodes belonging to a cluster
Floating IP address within VPC

Challenge
Move IP address (floating IP) between two EC2 instances in a VPC among different AV's

Research
Standard Pacemaker cluster IP failover mechanism not possible (→ EC2 instances / cluster nodes are not in the same Layer-2 LAN segment)
EC2 standard IP failover (EC2 Elastic IP) not available in VPCs
DDNS updates might not work with all SAP frontends (SAP GUI, HANA Studio, etc.)

Solution
Remotely changes routing table entries of a virtual router in the VPC (Setup of a /32 host-route pointing to an instance / cluster node)
Developed resource agent, that uses that mechanism to fail-over IP's
Resource Agent “aws-vpc-move-ip”

- Provides floating IP addresses for EC2 instances in VPC's among different AV's
- Locally adds & removes the “floating IP address”
- Changes routing table entry to route traffic to correct destination instance using EC2 API commands
Demonstration

The final presentation will have a 250MB video in this place. It has been omitted to limit the size of the presentation.
More information
http://www.suse.com/products/sles-for-sap
Visit us online to learn more about the SUSE® and SAP partnership at

http://www.suse.com/sap
http://www.suse.com/products/sles-for-sap/

Thank you.
Unpublished Work of SUSE. All Rights Reserved.
This work is an unpublished work and contains confidential, proprietary and trade secret information of SUSE. Access to this work is restricted to SUSE employees who have a need to know to perform tasks within the scope of their assignments. No part of this work may be practiced, performed, copied, distributed, revised, modified, translated, abridged, condensed, expanded, collected, or adapted without the prior written consent of SUSE. Any use or exploitation of this work without authorization could subject the perpetrator to criminal and civil liability.

General Disclaimer
This document is not to be construed as a promise by any participating company to develop, deliver, or market a product. It is not a commitment to deliver any material, code, or functionality, and should not be relied upon in making purchasing decisions. SUSE makes no representations or warranties with respect to the contents of this document, and specifically disclaims any express or implied warranties of merchantability or fitness for any particular purpose. The development, release, and timing of features or functionality described for SUSE products remains at the sole discretion of SUSE. Further, SUSE reserves the right to revise this document and to make changes to its content, at any time, without obligation to notify any person or entity of such revisions or changes. All SUSE marks referenced in this presentation are trademarks or registered trademarks of Novell, Inc. in the United States and other countries. All third-party trademarks are the property of their respective owners.
Multi-tier System Replication

Chain Topology (A → B → C)

Cluster

P
A

sync
B

async
C

Default Setup - Chain

only async now

Cluster

P
A

async
B

async
C
Multi-tier System Replication

Chain Topology ( A → B → C )

Default Setup - **Chain**

only async now
Multi-tier System Replication

Chain Topology (A → B → C)

Cluster

Default Setup - Chain

Not allowed from SAP

This would be a star
Multi-tier System Replication

Chain Topology ( A → B → C )

Clustering

Async

Only async
Multi-tier System Replication

Chain Topology (A → B → C)

Cluster

A

B

P

async

C

Only async

Cluster

A

B

P

async

C

ADMIN:
Break Replication complete
starting with version 0.149

Multi-tier System Replication

Chain Topology (A → B → C)

Cluster

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
</tr>
</thead>
</table>

async

C

Only async

Cluster

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
</tr>
</thead>
</table>

async

C

ADMIN:
Break Replication complete

Cluster

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
</tr>
</thead>
</table>

sync

sync

async

A B

C

ADMIN: Create new SystemReplication
Again a chain