Challenges in starting an SRE team from scratch in an Enterprise

SRECON20 Americas
Purpose, goal and strategy

Purpose
Why we exist
We connect for good

2030 Ambition
Who we must become
To be the world’s most trusted
connector of people, devices
and machines

Values
What will guide us
Personal, Simple, Brilliant

Strategy
How we’ll grow value for all our stakeholders

1
Looking in
Build the strongest foundations

2
Looking out
Create standout customer experiences

3
Looking to the future
Lead the way to a bright, sustainable future
Let there be...
Where to start?

Observability
Cost Control
Security
Automation
Performance
Reliability
TOIL
The Pillars
Tools

WAY TO M.T. EVEREST BC.

Mount Everest Base Camp
We’ve migrated to the cloud... Now what?
Key challenges...
OUR KEY CHALLENGES & SOLUTIONS
The challenge: no SREs!
Forming the team

Advocate

Environments & Governance
Level 2 Support
Graduate programme

Capacity Management
Platform Services

Education
Tools & Automation

Objectives & Measures
What do we need to succeed??

SUCCESS

What people think it looks like

SUCCESS

What it really looks like
The challenge: starting from scratch
I'M ROOTING FOR YOU
Our goals

- Faster rollout with HA & reliability
- Best practice & automation
- Reduce cloud sprawl
- Shared goals across teams
Our principles

1. Automation
   • If an existing process cannot be automated, it must be redesigned and replaced.
   • If a proposed process cannot be automated, it will be rejected.

2. Disposable Infrastructure
   • Servers are disposable and treated like cattle.
   • Servers live in auto-scaling groups that self-heal.
   • Servers are provisioned from images that are fully equipped and operational; no post provisioning tinkering allowed.
   • Application servers are stateless.

3. Continuous Integration
   • All code changes are made in Git via pull requests, verified, and approved.
   • Functional tests run on every deploy, preventing (or rolling back) deploys when the build fails.

4. Continuous Deployment
   • Deploys do not require any human interaction other than to instigate them.
   • Deploy time matters and engineers should strive to make it faster.
   • Rollbacks happen automatically when a failed deploy is detected.
   • It is easy to tell which commit is deployed.

5. Monitoring & Alerting
   • All systems are monitored for the critical “four golden signals” metrics - latency, traffic, error rate and saturation.
   • Metrics are easily available and consumable in a single interface.

6. Security
   • Security is automated and baked into everything.
   • Security checks are run as part of CI/CD.
   • Intrusion detection systems are in place.
   • As few infrastructure components as possible are publicly accessible, ideally zero.

7. Cloud Computing Standards
Doing some serious self-reflection…
SRE Maturity Report

Level 5: Elite
Level 4: Advanced
Level 3: Defined
Level 2: Developing
Level 1: Initiation

BT Specific SRE Activities

<table>
<thead>
<tr>
<th>Ref</th>
<th>Activity</th>
<th>Rating*</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Disaster recovery is periodically (at least annually) tested against non-production environments.</td>
<td>x</td>
</tr>
<tr>
<td>2</td>
<td>SRE standards are documented.</td>
<td>x</td>
</tr>
<tr>
<td>3</td>
<td>SRE standards are continuously audited.</td>
<td>x</td>
</tr>
<tr>
<td>4</td>
<td>Manual operations tasks are regularly reviewed and automated.</td>
<td>x</td>
</tr>
<tr>
<td>5</td>
<td>Game days are run regularly to check the efficiency of change processes.</td>
<td>x</td>
</tr>
<tr>
<td>6</td>
<td>Development teams are provided with observability of production performance in real time.</td>
<td>x</td>
</tr>
<tr>
<td>7</td>
<td>Development teams are provided with observability of production incidents.</td>
<td>x</td>
</tr>
<tr>
<td>8</td>
<td>There is a clear SRE roadmap for the next 12 months.</td>
<td>x</td>
</tr>
<tr>
<td>9</td>
<td>The SRE team supports the Digital Security Manager with Security Incidents.</td>
<td>x</td>
</tr>
</tbody>
</table>

BT SRE Score: 16

Traditional SRE Activities

<table>
<thead>
<tr>
<th>Ref</th>
<th>Activity</th>
<th>Rating*</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>SRE team members participate in incident response procedures on a regular (i.e. weekly) basis.</td>
<td>x</td>
</tr>
<tr>
<td>11</td>
<td>There is a culture of authoring blameless post-mortems for production incidents.</td>
<td>x</td>
</tr>
<tr>
<td>12</td>
<td>There is a process to manage production incidents.</td>
<td>x</td>
</tr>
<tr>
<td>13</td>
<td>There is an SLI (Service Level Indicators) policy in place.</td>
<td>x</td>
</tr>
<tr>
<td>14</td>
<td>There is an SLO (Service Level Objectives) policy in place.</td>
<td>x</td>
</tr>
<tr>
<td>15</td>
<td>There are periodic reviews of SRE project work and impact with business leaders.</td>
<td>x</td>
</tr>
<tr>
<td>16</td>
<td>There are periodic reviews of SLIs and SLOs with business leaders.</td>
<td>x</td>
</tr>
<tr>
<td>17</td>
<td>TOL is monitored and work patterns adjusted accordingly.</td>
<td>x</td>
</tr>
<tr>
<td>18</td>
<td>There is a resilience testing program in place.</td>
<td>x</td>
</tr>
</tbody>
</table>

Traditional SRE Score: 11

Overall Score: 27
“If you fail to plan, you are planning to fail.”
The challenge: cloud sprawl
SRE Cloud Standards

1) Discovery method
2) Resolution method
non-compliant resource

Config rule detects non-compliance

EventBridge triggers SNS

If Lambda set up

Lambda auto-remediates non-compliance

SNS - sends a message to Slack channel

Compliant resource

SRE picks up notification on channel and investigates
The challenge: lack of the "traditional" SRE model
The “traditional” SRE activities

**SLI/SLOs**
- There is an SLI (Service Level Indicators) policy in place.
- There is an SLO (Service Level Objectives) policy in place.
- There are periodic reviews of SLIs and SLOs with business leaders.

**Disaster Recovery**
- There is a resilience testing program in place.

**Incident Management**
- SRE team members participate in incident response procedures on a regular (i.e. weekly) basis.
- There is a process to manage production incidents.

**Blameless Post-mortems**
- There is a culture of authoring blameless post-mortems for production incidents.

**Automation**
- TOIL is monitored and work patterns adjusted accordingly.

**Leadership buy-in**
- There are periodic reviews of SRE project work and impact with business leaders.

The diagram illustrates the timeline and activities related to the “traditional” SRE activities, with specific milestones and tasks for each phase.
Chaos Engineering: the art of breaking things purposefully
SRE team wins 🏆

⭐ Saved ££ in our AWS accounts
⭐ Focused on security in our AWS accounts
⭐ Improved collaboration with developers and operations
⭐ Visibility of production to developers
⭐ Resiliency at the forefront
⭐ Positive change!
Leadership support

What is the problem we’re trying to solve?

Plan, plan, plan!

Use SRE principles
Everyone is just trying to make it