**Introduction**

- WAN bandwidth often limits replication throughput
- Deduplication and local compression increase replication throughput
- Delta compression with stream-informed caching adds 2X additional compression

**Remote Office Backup System**

**Central Office Backup System**

**Stream-Informed Cache**

- Full 1: A B C
- Store fingerprints and sketches in containers
- Full 2: A B C'
- A is duplicate
- B is duplicate
- C' is similar to C
- Container loaded to cache when A is seen again
- sk_A, fp_A
- sk_B, fp_B
- sk_C, fp_C

**Compression Results**

<table>
<thead>
<tr>
<th>Dataset</th>
<th>TB</th>
<th>Months</th>
<th>Dedup.</th>
<th>GZ</th>
<th>Delta w/ GZ</th>
<th>Delta Improv.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Source Code</td>
<td>4.6</td>
<td>6</td>
<td>24.9X</td>
<td>7.2X</td>
<td>14.9X</td>
<td>2.1X</td>
</tr>
<tr>
<td>Workstations</td>
<td>4.9</td>
<td>6</td>
<td>5.7X</td>
<td>2.8X</td>
<td>8.8X</td>
<td>3.1X</td>
</tr>
<tr>
<td>Email</td>
<td>5.2</td>
<td>7</td>
<td>6.9X</td>
<td>3.1X</td>
<td>5.8X</td>
<td>1.9X</td>
</tr>
<tr>
<td>System Logs</td>
<td>5.4</td>
<td>4</td>
<td>57.9X</td>
<td>4.6X</td>
<td>10.2X</td>
<td>2.2X</td>
</tr>
<tr>
<td>Home Dirs.</td>
<td>12.9</td>
<td>3</td>
<td>31.7X</td>
<td>3.1X</td>
<td>5.5X</td>
<td>1.8X</td>
</tr>
</tbody>
</table>

- Delta with GZ adds 2X more compression after deduplication than GZ alone. Compression factors are after an initial seeding period of one week.

**Customer Results**

- The median customer experienced 2X delta compression after deduplication.

**Conclusion**

- Delta locality closely matches deduplication locality for backup workloads
- Stream-informed delta compression is effective with a small cache
- Product allows customers to replicate and protect twice as much data across a WAN