MapReduce

- Developed as a research project out of Google.
- OSDI’04: “MapReduce: Simplified Data Processing on Large Clusters”
- **Big Idea:** Create a framework for processing data based on functions that can be “automatically parallelized”.
  - Allows many nodes to contribute to processing the data without human design/programming.


**Example #1: Word Count**

<table>
<thead>
<tr>
<th>The</th>
<th>quick</th>
<th>brown</th>
<th>fox</th>
<th>jumps</th>
<th>over</th>
<th>the</th>
<th>lazy</th>
<th>dog</th>
</tr>
</thead>
<tbody>
<tr>
<td>[0]</td>
<td>[1]</td>
<td>[2]</td>
<td>[3]</td>
<td>[4]</td>
<td>[5]</td>
<td>[6]</td>
<td>[7]</td>
<td>[8]</td>
</tr>
</tbody>
</table>

**Map:**

**Reduce:**

**Example #2: Mutual Friends**

Through asking about your friends about their friends, you have identified who are friends of whom (→ means “is friends with”):

A → B, C
B → A, C, D
C → A, B, D
D → B, C

You want to identify all **mutual friends** to any set of two people. For example: {A, B} → C, D.

**Map:**

**Reduce:**
**Software as a Service (SaaS)**

Q: What makes SaaS different from just software?

1:

2:

3:

Much like all of the other things talked about in the cloud, SaaS is a model of software that is independent of programming language and independent of architectural design.

Q: What features differentiate SaaS from PaaS?

1:

2:

3:

4:

Some software tools have characteristics of both SaaS and PaaS – these are general terms to understand the level of abstraction that you, as a developer, are working with or providing.

---

**Large Categories of SaaS Software**

- Human Resource Management (HRM)
- Customer Relationship Management (CRM)
- Enterprise Resource Planning (ERP)
- Content Management Systems (CMS)
- Billing Platforms
- Project Management
- Bookkeeping/Accounting
- Ecommerce
- Communication

---

**Examples of SaaS at Illinois**

---

**Advantages of SaaS**

1:

2:

3:

4:

**Legend:**

<table>
<thead>
<tr>
<th>IaaS Infrastructure as a Service</th>
<th>CaaS Containers as a Service</th>
<th>PaaS Platform as a Service</th>
<th>FaaS Functions as a Service</th>
<th>SaaS Software as a Service</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data</td>
<td>Data</td>
<td>Data</td>
<td>Data</td>
<td>Data</td>
</tr>
<tr>
<td>Functions</td>
<td>Functions</td>
<td>Functions</td>
<td>Functions</td>
<td>Functions</td>
</tr>
<tr>
<td>Applications</td>
<td>Applications</td>
<td>Applications</td>
<td>Applications</td>
<td>Applications</td>
</tr>
<tr>
<td>Runtime</td>
<td>Runtime</td>
<td>Runtime</td>
<td>Runtime</td>
<td>Runtime</td>
</tr>
<tr>
<td>Containers*</td>
<td>Containers</td>
<td>Containers*</td>
<td>Containers*</td>
<td>Containers*</td>
</tr>
<tr>
<td>OS</td>
<td>OS</td>
<td>OS</td>
<td>OS</td>
<td>OS</td>
</tr>
<tr>
<td>Virtualization</td>
<td>Virtualization</td>
<td>Virtualization</td>
<td>Virtualization</td>
<td>Virtualization</td>
</tr>
<tr>
<td>Hardware</td>
<td>Hardware</td>
<td>Hardware</td>
<td>Hardware</td>
<td>Hardware</td>
</tr>
</tbody>
</table>