

# THE POWER OF SEARCH

# JOHN STOVER

[johns@susqtech.com](mailto:johns@susqtech.com)

Lead Solution Architect, SusQtech  
PMP, MCT, MCTS, MCDBA, MCSE, MCSD

# Agenda

- **Expectations**
- **Website Search**
- **Enterprise Search**
- **SharePoint Topology & Architecture**
- **Search Relevancy**
- **Search Syntax**
- **Questions**

# Today's challenges need a unified solution

eam “”

roject “”

# The Search Challenge

We **EXPECT** to get it all from one simple box



# Internet Search

Blogs and Wikis

Published Documents

Extensibility

People & Knowledge

Search for the Website



Security

Websites

Multiple Indexes

Products

Events



# Enterprise Search is Different

Manageability

Team Workspaces,  
Blogs, Wikis

Business  
Relevance

Enterprise  
Documents

Extensibility

People &  
Knowledge

Search for the Enterprise



Security

Websites, Intranet,  
Extranet, Internet

Multiple  
Indexes

Line of Business  
Data

Desktop  
Information

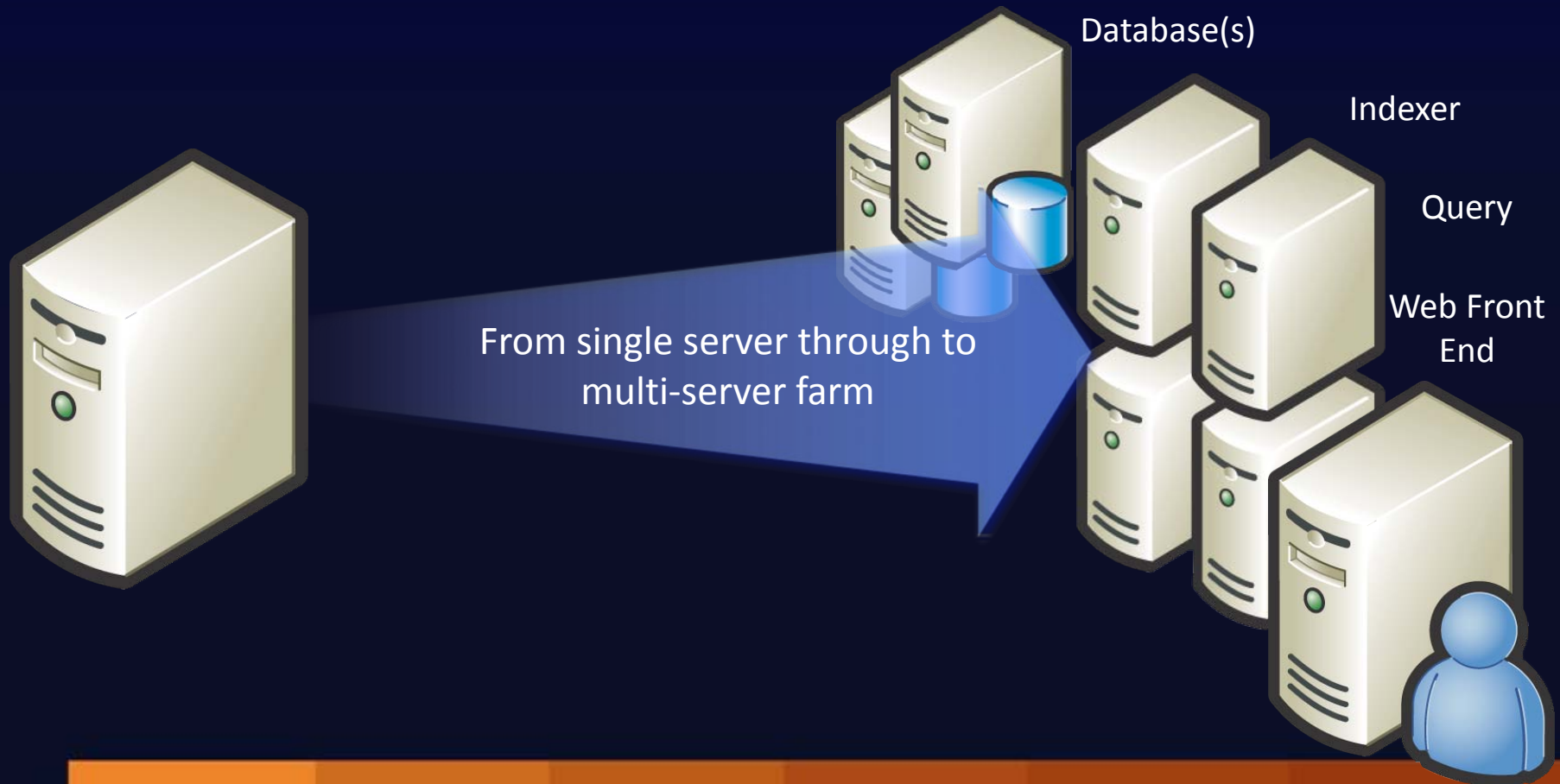
Business  
Intelligence  
Information

Expertise  
Social Networking

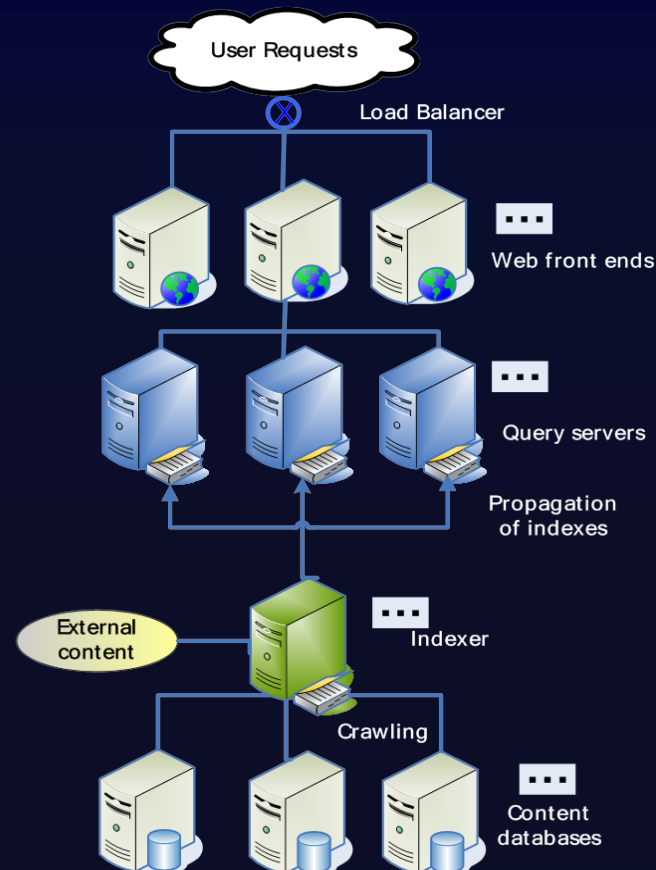


Product Features click a feature to learn more ▼	Microsoft Search Server 2008 Express	Microsoft Search Server 2008	Microsoft Office SharePoint Server 2007*	FAST ESP
<a href="#">Streamlined Installation</a>	●	●		
<a href="#">Rich Out-of-the-Box UI</a>	●	●	●	
<a href="#">No Pre-set Document Limits</a>	●	●	●	●
<a href="#">Extensible Search Experience</a>	●	●	●	●
<a href="#">Relevance Tuning</a>	●	●	●	★
<a href="#">Continuous Propagation Indexing</a>	●	●	●	●
<a href="#">Federated Search</a>	●	●	●	★
<a href="#">Security-trimmed Results</a>	●	●	●	●
<a href="#">Query and Results Reporting</a>	●	●	●	●
<a href="#">High Availability and Load Balancing</a>		●	●	●
<a href="#">Search Structured Data</a>			●	●
<a href="#">People and Expertise Searching</a>			●	
<a href="#">SharePoint Productivity Infrastructure</a>			●	
<a href="#">Extreme Scalability in Content Volume and Queries</a>				★
<a href="#">Extensive Query and Index Processing</a>				★
<a href="#">Out-of-the-Box Navigation and Clustering</a>				★

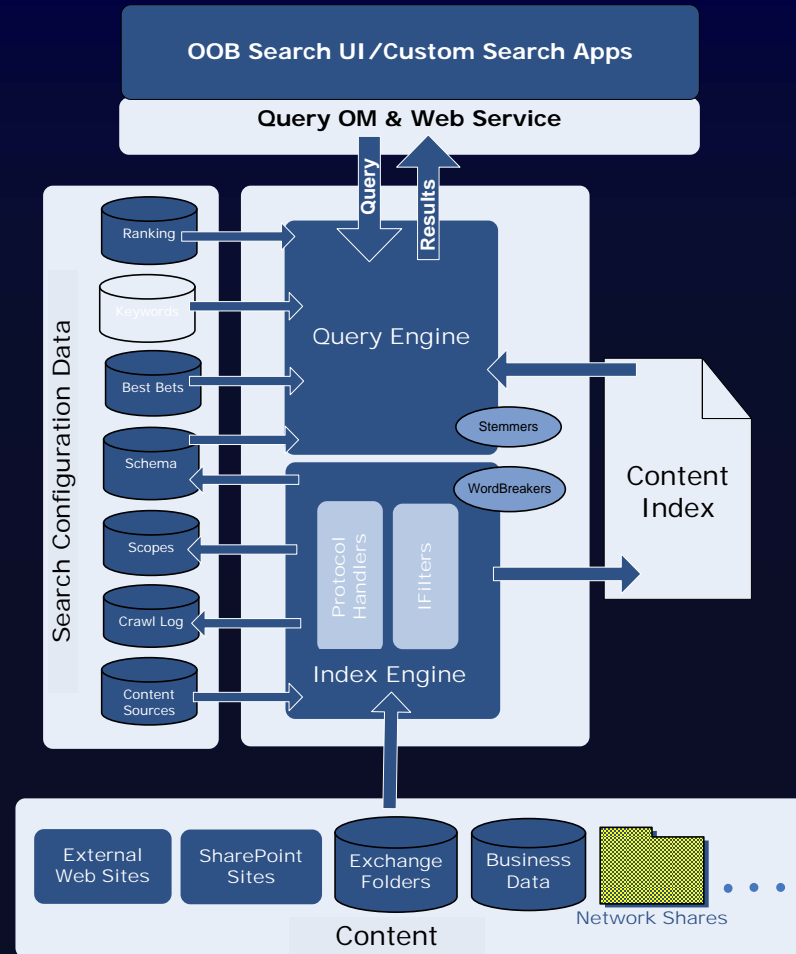
# High Level Architecture for SharePoint



# Microsoft Office SharePoint Server Topology Model



# Microsoft Office SharePoint Server Architecture



# Syntax & Relevancy

# Search Syntax #1

Uses the implicit Boolean function

# AND

# Search Syntax #2

**The search service is *not* case-sensitive.**

# Search Syntax #3

The search service automatically ignores common words (called noise words) such as **the**, **it**, and **by**, as well as **single-digit numbers** and **single-digit letters**.

# Search Syntax #4

**Include or exclude keywords by using the plus (+) or minus (-) sign in front of the keyword.**

**For example, the query **art + picasso** would return a list of artwork only by Picasso.**

# Relevance & Ranking Ingredients

(partial)

- **Metadata Extraction**
  - Automatically extract properties like titles and authors from documents
- **Automatic Language Detection**
  - Helps bias toward results in your language
- **File Type Biasing**
  - For example, PPT docs tend to be more relevant than XLS
- **Text Analysis**
  - Text ranking based on matching terms, frequencies, word variants
- **Click Distance**
  - Browsing distance from authoritative sites
- **Anchor Text**
  - Hyperlinks act as annotations for their target
- **URL Depth**
  - URLs higher in the hierarchy tend to be more relevant
- **URL Matching**
  - Direct matches on text in URLs

# Protocol Handlers, iFilters, & the BDC

# Protocol Handlers

- **SharePoint sites (MOSS or WSS)**
- **Web sites (HTTP / HTTPS)**
- **OLEDB & ODBC**
- **NTFS**
- **Exchange Public Folders**
- **Lotus Notes**
- **People**
  - Built from AD, LDAP, BDC, & Manually (privacy)
  - Social relevancy
- **Partners**
  - Documentum, FileNet, Hummingbird, Interwoven, etc.

# iFilters

- **Reads a document and extracts the plain text content and properties (metadata) for the indexer to index**
- **Ships with support for over 200 file types**
- **Partners/Vendors providing support for many more, e.g. PDF**

# Business Data Search

- **Today**
  - Vertical applications lack full-text search
  - Most users can't locate or access vertical apps
  - Hard to crawl business data with search tools
- **Office 2007: Search any database/web service**
  - No need to write custom IFilters or protocol handlers
  - No need to create HTML representations of data
  - Highly customizable results
  - Integrated with scopes and search center

# Administrative Features

- **Consistent search system across WSS and SPS**
  - Indexing Management
  - Simpler UI
  - Better indexing logs
- **Flexible Scoping**
  - Scopes decoupled from content sources
  - Simple and compound scopes
- **Monitoring**
  - MOM Pack available
- **Query and Result Reporting**
  - Get visibility into user search behavior
  - Logging for more detailed custom reports
- **Keywords and Best Bets**
  - Editorial results for specific queries
  - Streamlined UI
- **Index Propagation**
  - Continuous; indexed content is immediately searchable
- **Object Model and APIs**
  - Improved; for custom tools and services
  - SDK available

# MOSS Search Definitions

- **Corpus**
  - The full collection of all indexed information
- **Content Sources**
  - OOB support for Web Content, File Shares, Exchange Public Folders, Lotus Notes Databases, SharePoint Content, SharePoint user profiles, Business Data Catalog, and much more!
- **Scopes**
  - Used to create logical slices of the corpus
  - Scopes can span multiple content sources
  - Scopes can be defined by Property, Site, Content Source, URL

# Q&A

# JOHN STOVER

[johns@susqtech.com](mailto:johns@susqtech.com)

Lead Solution Architect, SusQtech  
PMP, MCT, MCTS, MCDBA, MCSE, MCSD